PhD Thesis

Essays in Bounded Rationality and Strategic Interaction

by Frederik Roose Øvlisen

University of Copenhagen

May 2009

Contents

Preface

This PhD thesis was written during my enrollment as a PhD student at Department of Economics, University of Copenhagen from June 2006 through May 2009. During this time, I have benefitted from the good research and travel conditions provided by the Department of Economics and by Center for Experimental Economics (CEE). I am particularly grateful to my supervisor Jean-Robert Tyran who, as my advisor and co-author, has shaped me, my thinking and my research. His encouragement, enthusiasm, curious comments and inspiration had a positive influence on my motivation and I have learned so much in the process due to him.

My papers have gained from valuable comments and remarks from my fellow PhD students, participants at workshops and summer schools, and from guests at CEE and at the Department of Economics. I would also like to express my thanks to my other co-authors, Julie Beugnot, Zeynep Gürgüç and Michael W.M. Roos, as well as I would like to thank Alexander Sebald for valuable comments and suggestions. A special thanks goes to Carsten Søren Nielsen for our daily interesting conversations that have helped me sharpen my thoughts and papers.

During my PhD, I spent one year at Ludwig-Maximilians-Universität (LMU) in Munich. I thank professor Klaus Schmidt for his generous hospitality, and the people at his chair and at LMU for welcoming me in their research environment. For financial support of my stay in Munich, I would like to thank *The Ryoichi Sasakawa Young Leaders Fellowship Fund*, *Augustinusfonden*, *Oticonfonden*, and "*Christian and Ottilia Brorsons Rejselegat for Yngre Videnskabsmænd og Kvinder*".

The three years have been great fun, full of discoveries and it has been very insightful. However, at times it has been tough and demanding. I would like to thank my parents, Annette and Bjarni, for their unconditional and endless support.

Frederik Roose Øvlisen

Copenhagen, May 2009.

I am very grateful to the members of my committee, Lise Vesterlund, Charles Noussair and David Dreyer Lassen, for their valuable comments and suggestions made both in their report and at my defense September 11, 2009.

Frederik Roose Øvlisen

Copenhagen, October 2009.

Introduction and Summary

This PhD thesis presents five essays on bounded rationality and strategic interaction. The common theme of the five papers is that real people, i.e. people with limited ability to compute and reason, do not behave like game theorists suggest in some strategic situations. The thesis presents investigations into the extent and the determinants of deviations from standard game-theoretic predictions in controlled experiments and provides potential explanations for these deviations by referring to notions of bounded rationality.

The specific questions addressed in the essays cover a broad range from how education relates to strategic uncertainty and bounded rationality, how increasing the number of options available for players affects behavior, and how incentives induced by the strategic properties – i.e. how the action of other players affect your optimal decision – shape decisions. In the essays, I also study how disequilibrium and failure to coordinate actions can be induced by irrelevant information, so-called sunspots, and how it can be mitigated by collective action.

All essays in this thesis use experimental methods to investigate the multi-faceted role of bounded rationality in strategic interaction. The experiments are guided and aided by theoretical considerations, and the data used has been collected in the laboratory, in the field (over the internet) and is in places complemented by registry data collected by the Danish Bureau of Statistics. The main advantage of the experimental method is its unrivalled ability to control other factors. This is a key requirement in studying bounded rationality because behavior that looks as if it was boundedly rational can often be "rationalized" or explained away by invoking auxiliary assumptions when the data generating process is not known – as is often the case with field data.

The thesis studies the relevance of bounded rationality in strategic interaction, i.e. situations in which what is optimal to do for one player depends on what other players do. The assumption routinely made in standard game-theoretic reasoning is that players are unboundedly rational, and consists of two components (Camerer and Fehr, *Science* 2006: p. 47). "First, individuals are assumed to form, on average, correct beliefs about events in their environment and about other people's behavior; second, given their beliefs, individuals choose those actions that best satisfy their preferences. If individuals exhibit, however, systematically biased beliefs about external events or other people's behavior or if they systematically deviate from the action that best satisfies their preferences, we speak of bounded rationality." The second aspect is particularly relevant for the five essays discussed here because it highlights the importance of expectations about the behavior of other, possibly boundedly rational, players.

The first three essays are closely related since the data was collected in the same experiment using variations of the "guessing game". A total of 19,196 subjects participated in these guessing games over the internet. The experiment had two periods. In the first period, all participants played a standard guessing game in which all participants choose a number between 0 and 100. The player being closest to 2/3 of the average number chosen by all players wins a prize of DKK 5,000. Game theory predicts that all players pick 0, which is the unique pure strategy Nash equilibrium in this game. Yet, observed behavior is grossly consistently with this prediction. Only about 2 percent of the players behave according to this prediction, and most players choose numbers around 30. In the second period, participants of the first period were reinvited. In total, 7,635 subjects accepted the invitation and were randomly allocated to one of five treatments which were variations over the guessing game in the first period. The first treatment was simply a repetition of the first period, serving to test convergence to equilibrium with repetition and learning. The second treatment transformed the game into an individual optimization problem in which strategic uncertainty is totally absent, and it therefore is a dominant strategy to choose 0. The purpose of this treatment is to test for the role of strategic uncertainty by eliminating it in a controlled manner. In the third treatment, rather than allowing for guesses between 0 and 100, guesses from 0 to 180 could be submitted. The purpose of this extension was on the one hand to provide a control for treatments 4 and 5, but also to test for how participants respond to a change in the environment that ought to be irrelevant according to standard theory. Treatments four and five serve to investigate how incentives arising from the strategic environment of the game shape behavior. In treatment four, the game exhibits strategic complementarity (inducing an incentive to "follow the crowd") and in treatment five the game exhibits strategic substitutability (inducing "contrarian" behavior, i.e. incentives to choose a "high" number when others choose "low" numbers).

The first paper entitled *Making an Educated Guess* (joint work with Jean-Robert Tyran) explores the role of education on strategic thinking in the guessing game. In addition to using the data from the first period of the experiment, this paper compares the first two treatments. Subjects taking part in these treatments were matched with Danish registry data to obtain detailed individual level socio-economic variables. The experiment is the largest guessing game ever performed, with one out of every 280 Danes participating. It is the first paper to match large-scale experimental data with detailed registry data, and the paper is methodologically innovative in that it developed routines for running multi-period large-scale experiments. A main result of the paper is that “the educated”, i.e. people with high school education and better grades in high school and tertiary education perform better in the guessing game. Specifically, we show that the educated are more able players in the first round since their choices are closer to the target number. Interestingly, we show that “the educated” are also more likely to solve the individual optimization task correctly. Thus, education is favorable for both cognitive – solving an individual optimization task – and social – predicting others’ behavior – skills in a strategic game.

The second paper entitled *Step-level Thinking and Changes in the Action Set* (singleauthored) exploits the data from the first and third treatment. The only difference between these two treatments is the range of possible guesses, i.e. the action set. In one treatment, the highest possible guess was 100 while it was 180 in the other treatment. Although not predicted by any theory, we find very different behavior in the two treatments. Interestingly, we can explain the behavior if we normalize the guesses proportional to the action set. With appropriate adaptations, the well-known step-level theory is consistent with the observed behavior.

The third paper entitled *Strategic Properties in Guessing Games* (single-authored) has two purposes. First, I introduce a new general formulation of the guessing game, prove existence of equilibrium, and provide the conditions for its uniqueness. Second, the paper investigates the causal effect of the strategic environment on guesses through controlled *ceteris paribus* variation. Treatments exclusively differ by whether actions are strategic substitutes or strategic complements. The strategic environment is defined by the slope of the best reply function – if it is optimal for an agent to choose a low guess when others also choose low, we say that actions are strategic complements (positive slope). If, on the other hand, it is optimal for an agent to choose a high guess when others guess low, we say that actions are strategic substitutes (negative slope). After subjects play the first round of the guessing game, they are randomly assigned to the two treatments. We find that subjects under strategic substitutes guess higher and closer to the equilibrium value than under strategic complements. The results suggest that strategic substitutes induces subjects reason more thoroughly (have more advanced depth of reasoning) than strategic complements. The intuition behind the results is straightforward – under strategic complements it is optimal to partly imitate others whereas it is optimal to act opposite of others under strategic substitutes.

Assuming that the first period guessing game anchored subjects on low values, the intuition provides an explanation of the observed difference in behavior and depths of reasoning.

Whereas the first three papers are based on a large-scale internet experiment, the fourth and fifth paper report results from coordination games run in a standard laboratory environment.

The fourth paper entitled *Collective Action and Coordination* (joint with Jean-Robert Tyran) presents two studies on the effectiveness of collective action in improving coordination in games with multiple, pareto-ranked equilibria. The first study shows that majority voting does not select a pareto-dominant equilibrium, does not prevent lock-in in a inferior equilibrium, and that, once coordination failed, it does not break lock-in despite the possibility to collectively "jump out of the trap". The second study uses a setting in which coordination failure again looms large in decentralized play, but collective action per se is quite effective in improving coordination. Yet, we show that collective action has a limited coordinating effect by "setting a precedent", i.e. in providing assurance by coordinating expectations on a superior equilibrium in subsequent decentralized play.

The fifth paper entitled *Coordination Failure Caused by Sunspots* (joint with Julie Beugnot, Zeynep Gürgüç and Michael W.M. Roos) also studies coordination failure. While most research on coordination games investigate which mechanisms that can prevent coordination failure, the fifth paper asks how coordination failure can arise. Specifically, we ask if coordination failure can be caused by "sunspots" – i.e. a random exogenous signal, not related to the payoffs. In a game with pareto-ranked equilibria – that absent of sunspots is characterized by full coordination on the pareto superior equilibrium – we show that the introduction of sunspots leads to coordination failure and thus to significant inefficiency. Hence, the strategic uncertainty caused by the sunspots causes coordination failure.

Making an Educated Guess

Making an Educated Guess

Jean-Robert Tyran and Frederik Roose Øvlisen*

May 2009

We run a large-scale two-period guessing game over the internet and match participants' choices to their socio-economic background information from registry data. The first period is a standard guessing game. In the second period, some participants simply repeat the game. Others play a version in which the game is transformed into an individual optimization task. Here, it is dominant to choose 0. We show that "the educated", i.e. people with high school education, better grades in high school and tertiary education, perform better in the guessing game. Specifically, we show that the educated are more able players in the first period since their choices are closer to the target number, and are more likely to solve the individual optimization task. Thus, education is favorable for both cognitive – solving an individual optimization task – and social – predicting others' behavior – skills in a strategic game.

Keywords: Bounded rationality, strategic uncertainty, disequilibrium, guessing game, beauty contest, newspaper and internet experiment, step-level reasoning

JEL-codes: D50, D83, D84, C93

---

* University of Copenhagen, Department of Economics, Studiestræde 6, DK-1455 Copenhagen.
Jean-Robert.Tyran@econ.ku.dk and Frederik.Oevlisen@econ.ku.dk.
We gratefully acknowledge financial support from the Carlsberg Foundation and University of Copenhagen. We are grateful for comments by Ralph Bayer, Dirk Engelmann, Nagore Iriberri, Wieland Müller, Charles Noussair, and seminar participants at the LEaF Conference at LSE, and at M-BEES 2008 at Maastricht University. We thank Guan Yang for valuable research assistance.

1 Introduction

Education and training may matter for how people play games. It seems obvious that academic training will influence the ability to grasp strategic incentives and thus the ability to accurately predict how other players will behave. Yet, little is known about the relation. One reason is simply that most experimental studies have used students as participants with relatively high cognitive abilities, a high level of education and with little variation of these variables in the subject pool. Some studies (e.g. Güth, Schmidt, and Sutter 2007, Thaler 1997, Drehmann, Oechssler, and Roider 2005, Bosch-Domènech and Nagel 1997) did use participants with lower and more varied educational background, but normally little is known about that background.

We present a repeated large-scale standard guessing game (also known as a beauty contest) with more than 19,000 participants to shed new light on several aspects of education and game play. We develop and present new treatments in the perhaps largest controlled experiment ever performed. Furthermore, we link experimental data to detailed, individual socio-economic registry data on the participants to provide new correlates of bounded rationality. This detailed , that allow for statistical controls at an unprecedented level.

Our study shows that the distribution usually observed in guessing games, including the prominent spikes, can be explained by assuming behaviorally unsophisticated players that make individual errors rather than highly sophisticated players with strategic uncertainty. More specifically, we refer to *k*-level "depth of reasoning" theories to account for the typical decision pattern observed in these games. We show that about one tenth of the players play strategically and that they guess both lower and closer to the target number than the non-strategic players. Furthermore we show that high school attendance, a higher grade point average in high school and tertiary education improves guessing in the game – both by overcoming individual errors and by better estimation of the others' actions. However, other socio-economic and personal characteristics such as income, gender, and age have no effect. Thus, attending high school, getting better grades in high school and completing a higher education not only mitigate individual errors, but also improve behavior in the first period of the strategic game.

The role of cognitive ability and behavior in games has received little attention. Noteworthy exceptions are Frederick (2005), Benjamin, Brown and Shapiro (2006) and Dohmen, Falk, Huffmann and Sunde (2007), who find a correlation between cognitive ability and impatience and risk preferences. Millet and Dewitte (2006) study cognitive ability and bargaining games and show that unconditional altruistic behavior is related to general intelligence. In a meta-study looking at prisoners dilemma experiments, Jones (2007) reports that students in schools with higher SAT scores cooperate more. Rydval, Ortmann and Ostatnický (2007) find that subjects with lower working memory are more likely to make errors in reasoning. In contrast Brandstätter and Güth (2002) study ultimatum and dictator games and find that intelligence has no significant effect on bargaining behavior.

Previous results on cognitive ability and behavior in the guessing game is scarce. However, the guessing game is used by Dickinson and McElroy (2009) to study the effect of sleep deprivation and circadian effects. They find that when players are deprived of sleep or when they make choices at non-optimal times-of-day, choices are consistent with lower levels of iterative reasoning. Yet, they fail to show whether this is due to strategic effects or to bounded rationality.

Camerer (2003) performed a guessing game with Caltech undergraduates (who have high SAT-scores), and he found that "the fact that the Caltech students do not choose numbers that are much closer to the Nash equilibrium than average folks refutes the hypothesis that simply being good at math will automatically lead players to a Nash equilibrium" (p. 216). He also finds that "unusual analytical skill and training in game theory move choices about one iteration closer to the equilibrium" (p. 218). These observations were made with homogeneous groups (Caltech students playing with Caltech students). Finally, Camerer concludes that when looking at heterogeneous groups, "samples of self-selected newspaper readers also make choices closer to equilibrium" (p. 218).

Palacios-Huerta (2003) is contrasting this, when he asks subjects in a Monty Hall experiment to report their SAT score and Grade Point Average. He classifies students with higher scores as "more able", and finds that more able subjects are more likely to find the optimal solution. In another paper, Palacios-Huerta and Volij (2006) perform the centipede game with professional chess players, and find that 69% of the players stop immediately (100% when only chess Grandmasters are considered), as theory suggests. In these two papers it seems that logical training mitigates boundedly rational behavior.

The most closely related to ours, is Burnham, Cesarini, Wallace, Johannesson, and Lichtenstein (2007). These authors report results from a guessing game experiment performed with Swedish twins. While their study differs in a number of ways from ours, some of the conclusions are strikingly similar. Differences concern a slightly different parameterization, the game was not played over the internet (but face to face) and data about education, sex and age are self-reported rather than drawn from a registry. Another important difference is that they administer an IQ test to directly measure cognitive ability. They find that subjects with higher IQ and some college education tend to guess lower in the guessing game. In addition, they find that older subjects guess higher, whereas gender has no significant effect. As such, their results on cognitive ability an education are similar to ours. However, Burnham et al. (2007) focus on explaining the submitted guess rather than the “difference to the winning number” in their regressions. We argue that what counts in the strategic game is to be close to the winning number and simply “guessing lower” is not equivalent to “guessing better”.

The guessing game is an ideal workhorse to study disequilibrium behavior, since it is very simple, has a clear theoretical prediction and risk preferences or social preferences do not plausibly affect choices. Many experiments in both newspapers and laboratories have been performed with varying group composition (both with respect to size, education and information). This has led to some stylized facts which are robust with respect to varying designs and parameter values. The experimental data has led to new theories developed to explain behavior in theses games.

The guessing game has been studied extensively because of its apparent simplicity and the striking deviation from the standard economic prediction it reliably produces (see Bosch-Domènech, Montalvo, Nagel and Satorra (2002) and Camerer (2003), for an overview). The guessing game was first suggested by Hervé Moulin (1986) as a textbook example of iterated elimination of dominated strategies. Yet, when writing the textbook, Moulin was aware that the average person did not act according to this reasoning. Moulin came across the game in the French popular science journal *La Recherche* in the late 1970s, where it had been used as a tie-breaking question in a quiz of logical questions. When Moulin contacted *La Recherche* to see the results, he found the typical spike pattern that was confirmed in later research.[1] Thus, the guessing game started as a newspaper experiment, and was introduced in the economic literature as an illustration of a iterated elimination of dominated strategies. Ironically, the guessing game would later become a favorite tool in experimental economics to demonstrate the failure of this reasoning process in practice.

The structure of the guessing game is quite simple. N participants simultaneously choose any number x_i ($i = 1,\ldots,N$) from the closed interval l to h. The guess with the minimum distance to the "target number", defined as p times the average number ($X = N^{-1} \sum x_i$), wins a prize. If several people are equidistant to the target number, the prize is equally shared among these people. The most common parameterization is $l = 0$, $h = 100$, $p = 2/3$ and $N > 2$, where standard economic theory predicts that all participants choose 0, the unique Nash equilibrium. For a more general description of the game and target function, see Øvlisen(2009a).

Numerous guessing games have been run both in the laboratory and in newspapers contests (e.g. *Financial Times*, *Spektrum der Wissenschaft* and *Expansión* – see Duffy and Nagel (1997) and Bosch-Domènech et al. (2002) for an overview) but few have focused on the role of education. Most laboratory studies use students as subjects. This is a subject pool with relatively high levels of education (and cognitive ability) and little heterogeneity. Both the experiments performed in *Financial Times* and in *Spektrum der Wissenschaft* suffer from a similar limitation, since the readership of these papers is relatively homogenous. Few newspaper contests had heterogeneous participants, but those who do (e.g. *Expansión*) cannot say much about the role of education because observed game play cannot be matched to educational variables. In contrast, we have high quality register data and considerable variation in the sample. For example the average annual net income in our sample around 187,547 DKK (25,000€) with a standard deviation of around 114,855 DKK (15,000 €) and ranges from 0 to more than 15 million DKK (2 million €).

[1] We are grateful to professor Hervé Moulin for his help in clarifying the origins of the guessing game in personal communication.

We conduct an internet experiment with two periods. For the first period, we recruit 19,196 subjects through announcements in the Danish newspaper *Politiken*. Politiken is a popular, “middle of the road” newspaper, and it is known to have a diverse readership, and is neither a tabloid nor an elitist newspaper. We can thus expect a very heterogeneous group of participants to take part in the experiment. The number of subjects corresponds to one out of every 280 Dane participating in the standard guessing game in the first period. In the second period, participants of the first period are invited by e-mail to participate again. They are then assigned (randomly) to different treatments each with approximately 1,500 participants, using a web-interface. To the best of our knowledge, this experiment is the first internet experiment using randomized treatment allocation and it is the first repeated large-scale guessing game.

Internet experiments have recently become popular, since they have many advantages over traditional laboratory experiments. The internet allows the researcher to maintain many of the virtues of the laboratory experiments (e.g. induced preferences, randomized matching and randomized treatment allocation). In addition, collecting a large number of observations is easy and it can be cheaply done. Furthermore you have a group of “real” people participating that will exhibit large heterogeneity in age, income, educational background etc. In this study, we not only perform an internet experiment with a heterogeneous subject pool. We also learn about their personal characteristics, such that we can control for e.g. educational background and income in our analysis. However, internet experiments have a drawback, in that some control is lost. Obviously, selection effects cannot be excluded – but it is not clear that they should be a larger problem in internet experiments than they are in standard laboratory experiments.

For example, only subjects with access to the internet can participate. This gives rise to a potential selection bias, if e.g. only rich or well-educated persons have access to the internet. In Denmark, 89 percent of the population had at the time of the experiment, access to the internet – either at home, at work or both places[2]. Making use of the internet is in Denmark a daily activity and it is common among all age and income groups to use the internet for e-banking, shopping etc. We will later see that participants are very heterogeneous with respect to e.g. income, age, and education.

---

[2] Statistics Denmark – www.statistikbanken.dk

In addition, Denmark has the advantage that all contact with the government is registered on an individual level in a central registry. Scientists can (subject to rules of anonymity) gain access to these data, that on an individual level includes a variety of detailed socio-economic characteristics, including grade point average (GPA) from high school, income variables, country of origin, and various education variables. We have linked our participants in the experiment with the Danish registry data, and obtain socio-economic correlates of bounded rationality that explains disequilibrium behavior.

We proceed as follows. Section 2 describes the experimental design, section 3 reports the results of the experiment, section 4 reports regression results from matching socio-economic data and behavior in the experiment, and section 5 concludes.

2 Experiment

Section 2.1 provides a broad description of the experiment, and section 2.2 explains the predictions of the experiment.

2.1 Design

The experiment has two periods and two treatments.[3] The first period was announced in the hard-copy version of the Danish daily newspaper *Politiken* and on its website. Participation was only possible through their website. Participation was not limited to subscribers of the newspaper, but open to all visitors at the newspapers website.

All participants played the standard guessing game with l = 0, h = 100 and p = 2/3, and could guess any real number (allowing for up to nine digits) between 0 and 100 (both included). In total 19,196 participants took part to win the prize of DKK 5000 (at the time approx. $1,000 or € 670), that was split equally among the four winners that were closest to the target number.

In the first period, subjects provided their guess, name, address, e-mail address, phone number and an optional comment. The e-mail address served as the unique identifier, and only one guess per e-mail was allowed. The experiment was presented to

[3] A total of five treatments were run in the second period. The other three treatments address a different question and are discussed in companion papers, see Øvlisen(2009a,b).

the readers as a competition and the scientific aim of the competition was not revealed. However, the readers were told that the aim and the result of the exercise would be discussed in the paper the following week. The experiment was open to participants for one week. Instructions (translated from Danish) can be found in the appendix.

In the second period, we used the e-mail address provided in the first period to invite all previous participants to participate again. Thus, only participants of the first period could participate in the second period. Each participant received a personalized e-mail that contained a unique link. By clicking this link, they visited a personalized homepage and were randomly allocated to one of five different treatments, each with a prize of DKK 5000 (at the time approx. $1000 or € 670). In addition to a short description of the experiment and the instructions (see appendix), they were informed about their own guess, the average guess and the target number of the first period. The information page also included two links providing additional information (in a pop-up window). One link showed the distribution of guesses in the first period (similar to Figure 1) and the other link displayed a newspaper article about the guessing game, with detailed comments of how to think of the game (see Schou 2005). This description included an intuitive explanation of the Nash equilibrium in plain words. As can be seen from the appendix, the instructions of the game included a line indicating the slope of the best response function. Of the 19,196 invited participants, 7,635 participated in the second period.

In this paper, we analyze two treatments in the second period. Treatment REPEAT with 1,625 participants was simply a repetition of the first period, where the rules and parameters were not changed.

The other treatment, ROBOT, with 1,520 participants transformed the game into an individual optimization task. Subjects were informed that each subject was matched with a computer, and that the computer would randomly draw a number between 0 and 100 from a uniform distribution. Whoever was closest to 2/3 of the average of the participant's and the computer's number would win the contest. They were told that among all the "pairs" in this treatment, one would be randomly drawn and it would be determined if the participant or the computer won the DKK 5,000.

Chou et al. (2007) run two-person guessing games with various protocols and instructions and claim that the participants' ability to solve the optimization problem is not due to cognitive bias but results from the "unnatural" way the game is presented and explained. We agree that instructions and framing matter in this game, and we believe we would have found much a higher share of participants choosing 0 had the game been explained as, for example, "if the person choose a lower number than the computer, the person wins". However, that is not our point. What we show is that people who are able to solve the optimization problem are better educated and, most importantly, are more sophisticated in the game in the first period. Thus, we chose to keep the description of the ROBOT game as similar as possible to the game in the first period and to REPEAT.

In both ROBOT and REPEAT, participants submitted their guess, birth date, gender, and postal code. They could also update their e-mail address and leave an (optional) comment.

2.2 Theory

Applying standard economic game theory using successive elimination of dominated strategies to the guessing game with the parameters presented above, results in one pure strategy Nash equilibrium, namely that everyone picks 0. In equilibrium, the average is 0, two thirds of that is 0 and all the participants share the prize and walk away with a strictly positive gain. This game theoretic prediction thus holds for both the first period and for the REPEAT treatment. However, the N-person game exhibits strategic complementarity in the sense that it is optimal to choose high numbers if others are expected to choose high numbers.

In ROBOT, the standard economic prediction is different. Here, there are only two players: the computer and the participant. Multiplying a factor less than one and the average of two numbers, yields a number that is closer to the smaller of the two numbers. Therefore, in a guessing game with 2 players, the lower number always wins. Choosing 0 is a weakly dominant strategy, since it is always the lower number. Thus, we would predict that all subjects choose 0. The subjects know that the computer picks a random number (with equal probability) and thus the strategy of the computer is known. In other words, there is no strategic uncertainty present in this treatment. So, if we observe that some subjects do not choose 0, we can attribute the choice to an individual error (i.e. bounded rationality). The ROBOT treatment is thus similar in its design to the two-person guessing game (Costa-Gomes and Crawford (2006) and Nagel and Grosskopf (2007, 2008)). Yet, in the two-person guessing games beliefs are not properly controlled for. Given some belief about the other's action, it can still be rational to pick a number greater than 0. In the computerized counterpart, where the strategy of the computer is explained, beliefs are eliminated and thus there is no strategic uncertainty.

Guessing games have been much studied and different theories have been developed to explain the robust patterns of non-Nash behavior that is found. Nagel (1995) is the first to suggest a theory, directly inspired from the guessing game, and also Stahl and Wilson (1994, 1995), Stahl (2004) and Camerer, Holt and Ho (2004) provide theories applicable to the guessing game.

All these theories apply step level cognition, i.e. characterize participants into different categories according to their cognitive depth of reasoning. In these models, subjects perceive that they themselves are the most sophisticated and thus they best respond to their belief of the other players applying less steps than themselves.

The classic game theoretic concept of iterative elimination of dominated strategies use a similar logic. By applying indefinitely many steps, the equilibrium will be reached. Thus, a step-1 player will do one round of elimination of dominated strategies (and thus choose 66.7), a step-2 player will do two rounds (and choose 44.5) whereas a step-infinity player will choose 0. This approach typically finds little support from laboratory data.

Nagel (1995) puts forward another and yet more simple approach that has had great success in explaining the spikes that guessing game results provide. Here, it is assumed that a step-0 player in first period will randomize over the interval (i.e. on average choose 50). A step-1 player best respond to this, believing that everyone else is a step-0 player, and thus choose 33.33. A step-2 player will best respond to everyone else being a step-1 player and choose 22.22 and so forth. For subsequent periods, a step-0 player is assumed to choose the average number of the previous period.

Camerer, Holt and Ho (2004) provide a theory that is based upon the same principle, although players here are assumed to be more sophisticated. Rather than assuming that everyone else applies one step of reasoning less, they assume that participants choose best replies to a distribution of types having less cognitive depth than them. Thus, a step-2 player will best respond to a mixture of step-1 and step-0 players. Types are assumed to be Poisson distributed.

Finally, Stahl (1996) suggests to augment the Nagel-type step-model with a learning module. The idea is that as the game is repeated, players learn more than just the average (and target) of prior periods. They also learn which models outperform others, and Stahl (1996) argues that they over time will shift to better performing rules.

3 Results

3.1 One-Shot Newspaper Guessing Games

While our experiment is the largest newspaper guessing game, it is not the first. Thaler (1997) did an experiment in collaboration with *Financial Times*[4], Bosch-Domènech and Nagel (1997) with *Expansión*, Fehr and Renninger (2000) with *Die Zeit* and Selten and Nagel (1997) with *Spektrum der Wissenschaft*. Although the characteristics of the readership of the various newspapers differ, the three studies show a similar pattern (see Bosch-Domènech et al. (2002) for a survey).

All previous studies were one-shot games while ours has two periods. Thus, our first period is comparable to these previous newspaper guessing game studies. Broadly speaking, the previous newspaper experiments show similar results to ours, shown in Figure 1. In the first period, we find the prominent spikes at numbers around 33, and 22 that can be explained with steps 1 and 2 using step-level reasoning (see Nagel 1995).

In contrast to the previous newspaper experiments, we also find a spike at 67. Also, our average guess of 32.41 is 30-50% higher than any of the three previous guessing games (with averages between 19 and 25). Furthermore, we have only around 2 percent equilibrium choices compared to between 6 and 12 percent in the previous newspaper experiments. Among the 19,196 guesses, four participants split the prize of DKK 5000 (at the time approx. $1000 or €670) equally with a winning guess of 21.6 that got closest to the target number of 21.61.

---

[4] Due to legal restrictions the FT experiment allowed only for guesses in integers – this changes the game slightly, since also 1 is a Nash Equilibrium, which is also reflected in the guesses provided. The guessing game is easy to implement from a practical perspective, and only rarely have other games have been run in collaboration with newspapers (e.g. Güth et al. 2007).

Our results differ remarkably from experiments in *Financial Times* and *Spektrum der Wissenschaft*. These outlets have a relatively homogenous and highly educated readership, and it is likely that readers know that other readers have a similar background. *Politiken* on the other hand is a popular newspaper with no particularly distinguished readership. We believe that it is the heterogeneous readership that makes education an important determinant of disequilibrium play.

3.2 REPEAT treatment

Subjects who responded to our invitation to participate in period 2 were randomly assigned to treatments. There were 5 treatments in total. The other treatments are discussed in companion papers (Øvlisen 2009a,b). In REPEAT the 1,625 participants chose an average number of 21.662 and thus a target number of 14.441. One participant won the prize of DKK 5000 (at the time $ 1,000 or € 670) with a guess of 14.451.

Figure 1 and Figure 2 show the distributions of guesses in the first period and in REPEAT treatment, respectively. Comparing the two figures, we clearly find that subjects converge towards equilibrium as the game is repeated. We see that the entire distribution shifts closer to 0 – the average falls from about 32 to 21. Second we observe coordination of expectations – the mass around the target number about doubles, from 23 percent to 46 percent of the choices within +/– 5 from the target number.[5]

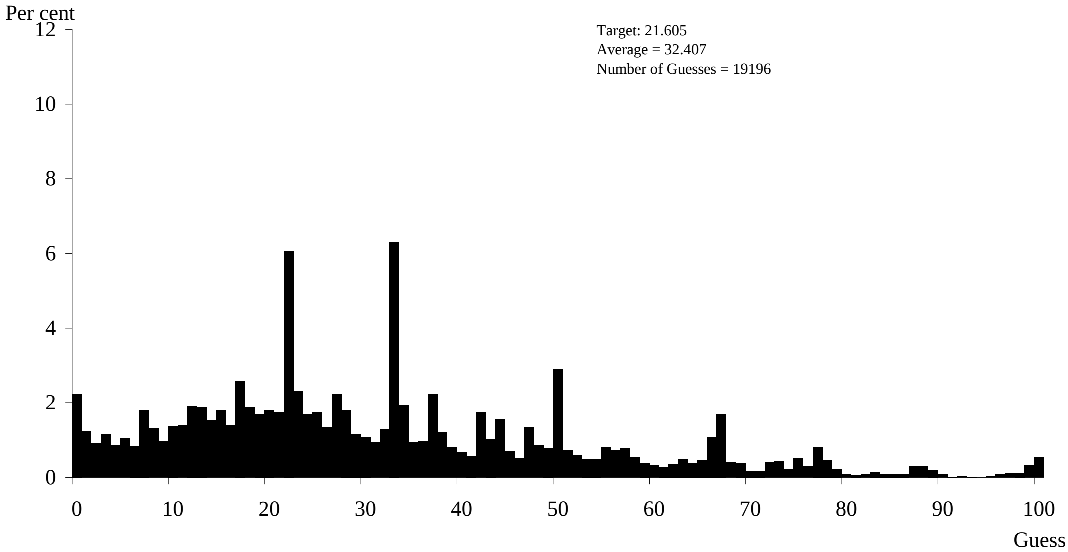
Another indication of convergence is that extreme choices are less common with repetition. In the first period 8.4 percent of the choices were in the weakly dominated interval between [$66^2/_3$,100], whereas only 3.6 percent of choices in REPEAT were within this interval. Similarly, we find fewer equilibrium choices (0) in REPEAT than in the first period.

When we look for the spikes we found in the first period, we see that the spikes at 67 and 50 have disappeared. In the first period, we found spikes at 33 and 22 that corresponded to 1 and 2 steps of reasoning in the step-level theory. In REPEAT, we find that the 33 spike is almost gone, whereas the 22 spike is still there. Most prominent are two new spikes at 15 and 10 arise. Using step-level theory, and assuming step-0 players

[5] Similar results holds for other intervals, e.g 12 percent in the first period vs. 21 per cent in REPEAT of choices +/– 2 from target number

to choose last period's average number, the spike at 22 corresponds to 1 step of reasoning, whereas the spikes at 15 and 10 correspond to 2 and 3 steps, respectively.

Figure 1: Distribution of guesses in the first period



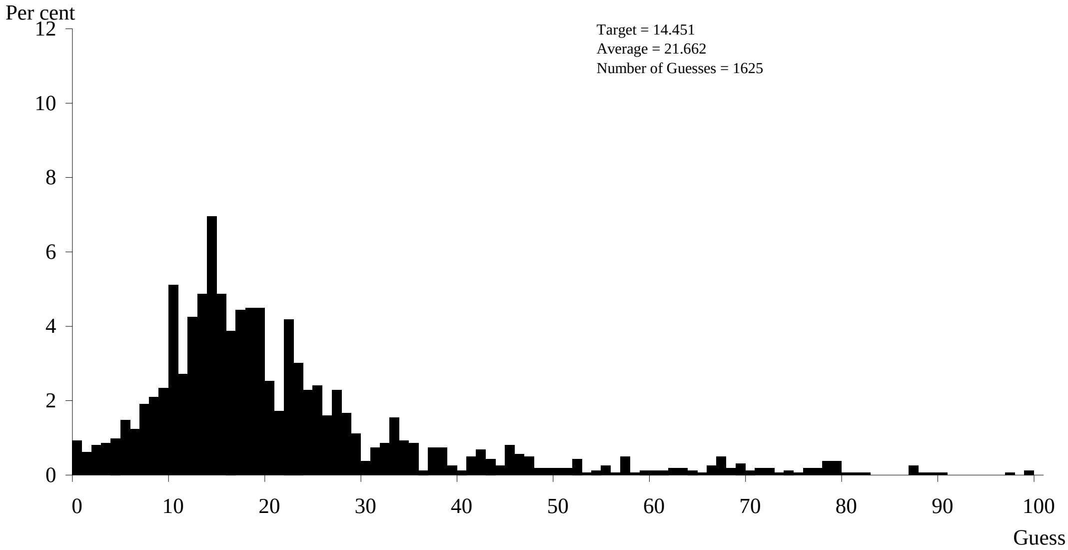
Since no newspaper (or any large-scale) guessing game has been performed for multiple periods, we turn to the lab for comparison of our results. The convergence we observe looks very similar to convergence of similar experiments performed in the laboratory.

Kocher and Sutter (2005) find that using more than 3 steps is rare and they find no evidence of increased depth of reasoning with repetition. In Figure A1 in the appendix we compare our distribution of steps to Kocher and Sutter (2005). We find that the distribution we obtain (with many more observations) looks strikingly similar in the first period, and is almost identical in the second period. Thus it seems that the results obtained with respect to depths of reasoning are robust.

One could hypothesize that a participant would use the same number of steps in all periods (and adjusting the step-0 starting point to the average of the previous period). However, we see from Table A1 in the appendix that this does not seem to be the case overall. Only about 25 percent of the participants apply the same level of reasoning in both periods. We find that 72% of the players in the first period and 78% of players in

REPEAT, use 2 or fewer steps of reasoning. A Wilcoxon Signed-Rank rejects that the distribution of steps are the same for subjects in both periods (p =.0000). This finding is similar to Stahl (1996), who also reject the hypothesis that step-players use the same step-*k* rule in every period.

Figure 2: Distribution of guesses in REPEAT (second period)



However, when we limit our attention only to players who are classified as level-1 or level-2 in the first period we find that a large share of the subjects classified as level-1 players in the first round switch to level-2 play in the second round. We classify 39 subjects as level-1, since they choose numbers between 32 and 34, and 27 subjects as level-2, since they choose numbers between 21 and 23. If these players would maintain the same behavioral sophistication they would again be level-1 and level-2 players. Level-1 play in REPEAT corresponds to choosing around 21 (assuming step-0 players choose the average of the first period), and level-2 play corresponds to choosing around 14.

Figure 3: First period Level-1 players' decision in REPEAT

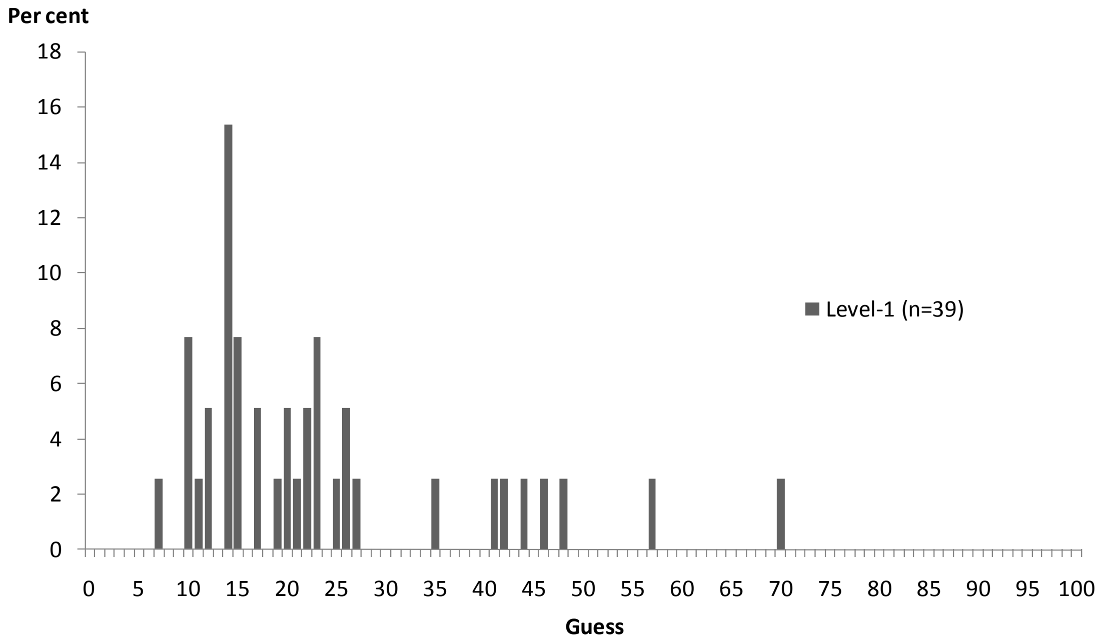
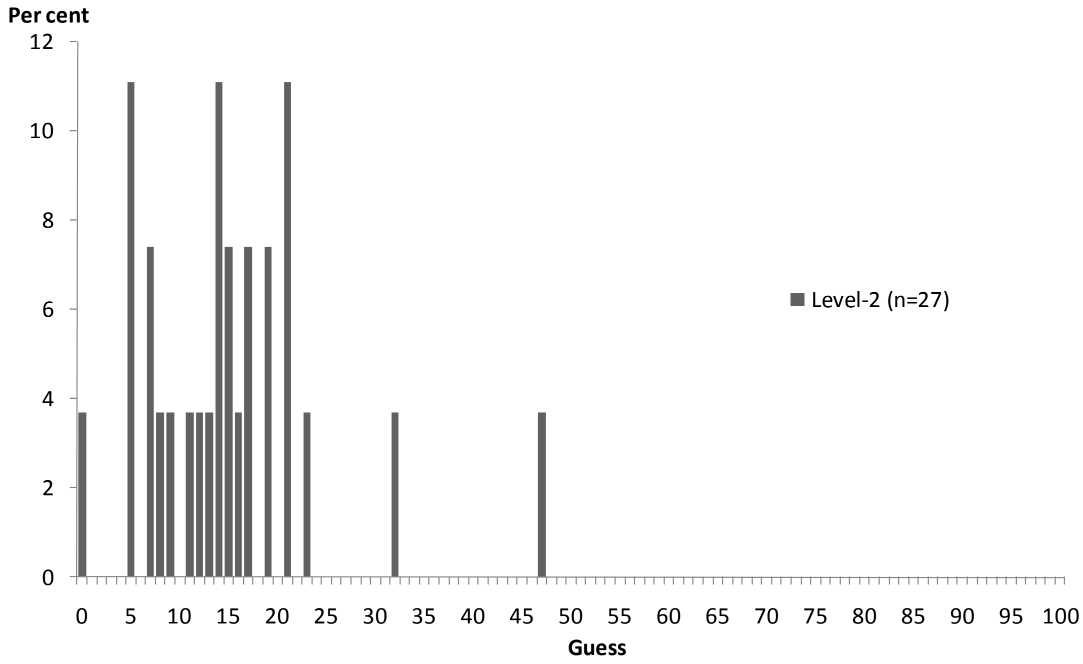


Figure 3 shows the distribution of choices in REPEAT for the 39 subjects classified as level-1 players in period 1.The average choice of 23.4 is close to the prediction of 21. However, Figure 3 shows that we do not find a spike at 21. Instead, we find the largest spike at 14, corresponding to level-2 play. In fact, 20.5 percent of the "level-1" players chose between 20 and 22 in the second period. This is also where we find the largest spike among all subjects (see Figure 2). Thus, it would be alarming if the step-1 players would exhibit a very different pattern.

Figure 4 shows the distribution of choices for subjects who were classified as level-2 players in the first period. The results from looking at "level-2" players are inconclusive. We do indeed find a large spike at the predicted value of 21, corresponding to level-2 play in REPEAT. However, equally many did not adjust their guess and changed to level-1 play in REPEAT. Finally, a large spike at 5 remains unexplained by step-level thinking. We will later return to this topic, when we look at how socio-economic variables correlate with steps of reasoning.

Figure 4: First period Level-2 players' decision in REPEAT



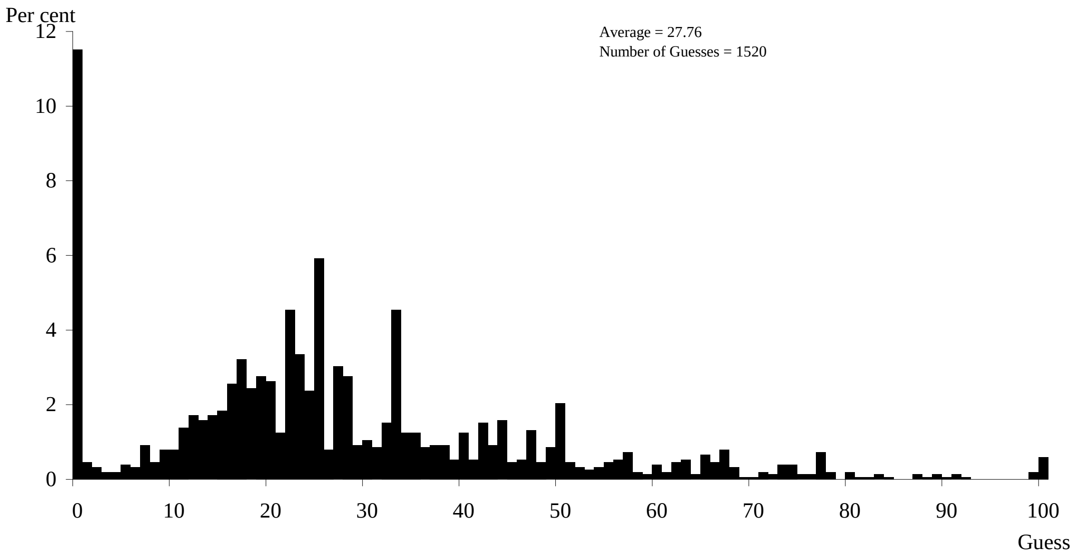
3.3 ROBOT treatment

We conclude that the typical spike-pattern found in guessing games can be explained using step-level theory, assuming non-sophisticated players. However, we can expect a small fraction of the participants to correctly solve the game, and in expectation of violation of rationality, they themselves best-respond to this belief and choose a number greater than 0.

In ROBOT, subjects faced a much simpler game than in REPEAT since they knew the strategy of the computer. Thus they did not face strategic uncertainty, and the game was reduced to an individual optimization task, where it is a dominant strategy to choose 0. This is very interesting, since it controls for expectations and strategic uncertainty, and allows us to attribute any non-0 guesses to bounded rationality.

Figure 5 displays the distribution of choices in the ROBOT treatment. Surprisingly, less than 12 percent correctly chose 0. Thus, the vast majority of subjects failed to optimize in this simplified setting. One could argue that since a deviation from 0 is not very costly, we should not expect all submitted numbers to be exactly 0, but just expect them to be "low". Guessing one unit higher reduces the probability of winning by 1 percent, so we should then expect to observe a triangular shaped histogram with decreasing observations in guesses. We clearly do not observe such a pattern, and we find that guesses between 1 and 10 are rare.

Figure 5: Distribution of guesses in ROBOT (second period)



The structure of our game is similar to the two-person guessing game in Grosskopf and Nagel (2007, 2008) [6] where the weakly dominant strategy is 0. However, their design is different in a crucial way. In the two-person guessing game, the strategy of the opponent is not known and subjects thus still face strategic uncertainty. Any behavior different from 0 could be caused by disequilibrium expectations. Thus, the theoretic prediction rests on both the assumption of rationality and of common knowledge of rationality. In contrast in ROBOT, the strategy of the computer is known and strategic uncertainty eliminated. Here, 0 is a weakly dominant strategy, controlling for expectations and the theoretic prediction solely relies on the assumption of rationality. Despite the differences, the two-person guessing game is an interesting comparison for our ROBOT treatment. Grosskopf and Nagel (2007, 2008) report similar findings in that only about 10 percent of their student participants choose 0 when they play the 2-person game for the first time.[7]

---

[6] Costa-Gomes and Crawford (2004) study the relation of information search and responses in two-person guessing games.

[7] Grosskopf and Nagel (2008) report that 21.2% of the guesses are above 50 whereas we find only 12.0% above 50. Grosskopf and Nagel (2008) perform different sequences of the game and find that behaviour in the first round of a 2-person guessing game is not different for subjects who have previously been exposed to a n > 2 guessing game.

We find that about 12 percent of participants correctly optimize when they face a version of the guessing game without strategic uncertainty. Consequently, about 88 percent of the subjects failed to optimize in this setting. In the following, we study these two groups of subjects into more detail.

In Figure 6, we look at how the 88 percent (1,345 subjects) “non-optimizers” in ROBOT played the game in the first period. When comparing Figure 1 and 6, we find that the distribution of choices in the first period of the guessing game (with 19,196 subjects) is the same as for non-optimizers in ROBOT. Thus, subjects who failed to optimize in ROBOT have a typical behavioral decision pattern in the first period. A Mann Whitney test fails to reject that guesses from non-optimizers in ROBOT come from the same distribution as all non-ROBOT subjects decisions (p = .0598). This result suggests that the typical outcome in the first period of the guessing game can be explained by non-sophisticated players exhibiting bounded rationality and not by sophisticated players who are subject to strategic uncertainty. We conclude that subjects failing to optimize in ROBOT act as the “typical participant” in a normal guessing game.

How do optimizers in ROBOT play the normal guessing game? Figure 7 shows the choices of the 12 percent (175 subjects) that optimized in the simplified ROBOT in the standard game in the first period. When comparing to Figures 1 and 6, we clearly see that these subjects played totally different to the overall distribution of guesses in the first period and a Mann Whitney U-test clearly rejects that the two distributions are the same (p = .000). We do not find the typical spikes, we find only 1 percent (two observations) dominated choices, and we find lower numbers (average of 18 and a median of 17.93).

The 12 percent can be regarded as the upper bound for the number of subjects that understood the game in the first period, because of two reasons. First, we know from Weber (2003) that subjects converge towards 0 even without feedback. Second, some subjects might have read about the game in the newspaper article that followed the first period. In fact, only 1 percent (2 observations) of the 175 optimizers chose 0 in the first period. In the first period, about 24 percent of their choices are +/- 5 from the winning number, about 45 percent are below the winning number -5 and the remaining 21 percent are above the winning number +5. So, although optimizers guess closer to the winning number than non-optimizers, it seems that they overestimate their fellow participants in that they tend to guess below the winning number.

Figure 6: Distribution of choices in period 1 for subjects not guessing 0 in ROBOT

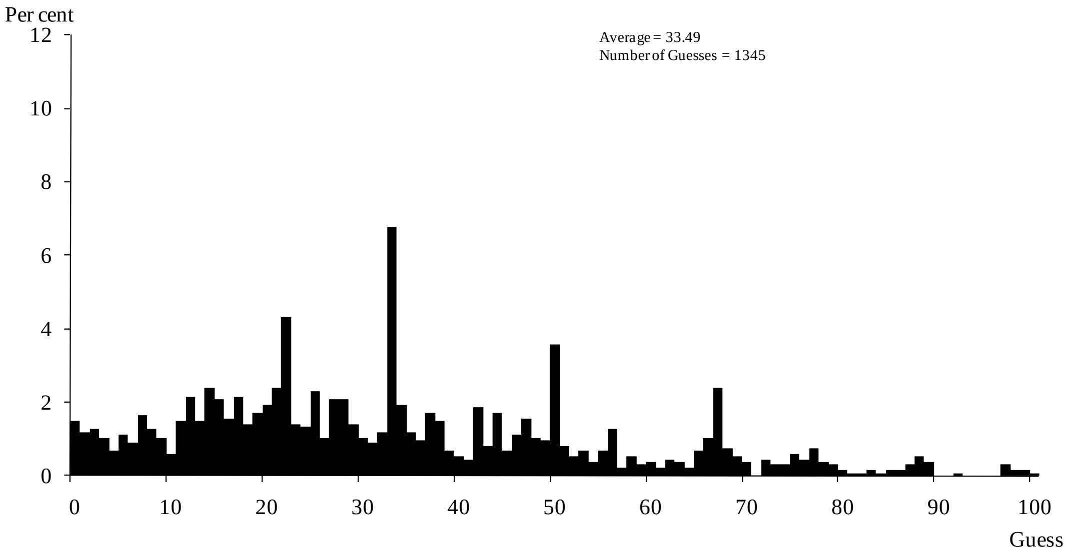
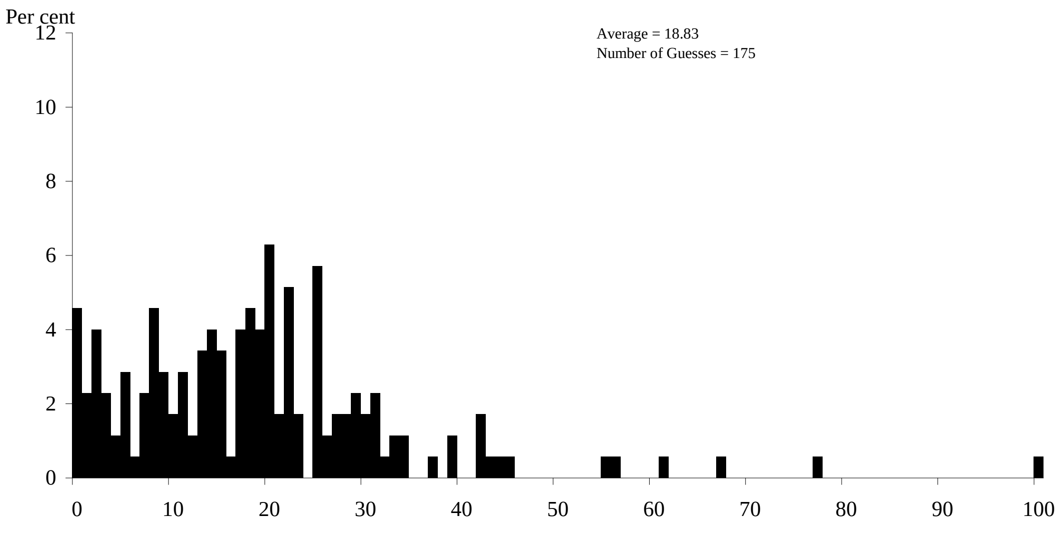


Figure 7: Distribution of choices in period 1 for subjects guessing 0 in ROBOT



4 Socio-economic correlates

Statistics Denmark collects registry data on the Danish population. They keep records of all public databases, such that all data collected by any governmental agency (health care, tax authorities, education, transfers etc.) is stored on an individual level, using a common personal identifier. Under strict anonymity rules, researchers can gain access to this data such that individual-level analysis can be carried out.

We have gained access to the data, and have uniquely identified about 500 subjects per treatment, corresponding to about one third of our participants of the second period. We use the participants' zip-code, gender and birthdates to identify unique individuals in the Statistics Denmark database. This identification allows us to for the first time to match information like income, grades, education, country of origin, type of employment and other socio-economic variables with the experimental data.

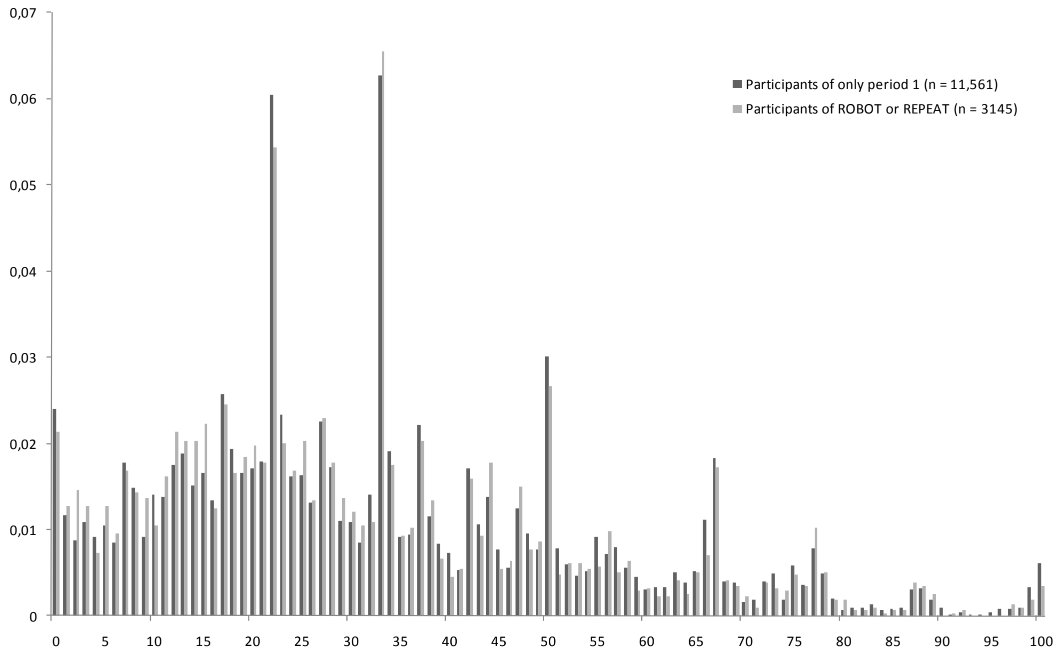
4.1 The data

We conclude that our sample looks different from the population when it comes to completion of high school and gender. The other variables look similar to the population as a whole. We do not claim that we have a representative sample, but we do claim that we have high quality socio-economic data and a large heterogeneous sample that allows for a control for all relevant variables. We argue that the identification process does not bias the data, and we argue that there is not systematic attrition from period 1 to period 2 of the experiment. While there clearly is selection into participating in the experiment, we believe that the rich socio-economic data allows us to control for relevant variables.

With 19,196 participants in period one, but only 7,635 participants in the second period (randomly assigned to 5 treatments) attrition is an obvious issue. We believe the following factors account for this attrition. As the second period of the experiment took part about a year later than the first period, some participants never received the invitation to participate in the second period (e.g. due to changed e-mail addresses, or not reading the e-mail on time). This source of attrition seems to induce no selection bias. Yet, one could fear that selection would be an issue in the second part of the experiment – e.g. only subjects close to the target in the first period would participate again. However, this seems not to be the case. In fact, the distribution of guesses made in the first period by participants in the second period ($n = 7,635$) is not different from the distribution of those who only took part in period one ($n = 11,561$) (Kolmogorov-Smirnov test, $p = .787$). Figure 8 shows the striking similarity of choices of these two groups in period 1. We also find no difference when comparing period 1 choices of participants in REPEAT and ROBOT (n = 3145) (Kolmogorov-Smirnov test, p = .763).

Of the total 7,635 participants[8] in both periods, a unique match was found for 2,599 of them. These participants are the only persons in Denmark with the exact combination of gender, zip code and birth-date the participant indicated. So, about one third of our subjects have been uniquely identified. We only include uniquely identified subjects in the regression analysis.

Figure 8: Irrelevance of attrition for choices in period 1



To test if our identification strategy causes systematic bias, we compare the variables obtained from the experiment for our uniquely identified sample and our full sample. Table 1 shows that the variables in question do not differ between the two samples, except for the variable age. While both the average and the standard deviation exhibit no large numerical difference we find that we cannot reject that slightly older participants were selected into period 2.

We find that the distributions of submitted guesses for all participants (Figure 1, 2 and 5) and for the identified subjects (Figure A2, A3 and A4, respectively) are similar.

[8] Note that there were 5 treatments in total, of which 2 are discussed in this paper

Thus, the key variables of interest (guesses, clicked distribution) are no different between the full sample and the matched sample.

In total 7,635 subjects participated in the second period, of which 3,145 were randomly allocated to take part in either the ROBOT or the REPEAT treatment. The random allocation ensured no selection effect into treatments. Of all the participants, 2,599 were uniquely matched participants, having 1,100 of them in either ROBOT or REPEAT (among these, few observations might be dropped from the regressions due to missing values – this will be clear from the regression tables).

Table 1: Summary Statistics

Variable	ROBOT			REPEAT		
	Non-Identified	Identified	Mann-Whitney (p-values)	Non-Identified	Identified	Mann-Whitney (p-values)
Guess in period 1	31.36	32.64	0.15	30.92	31.28	0.96
Guess in period 2	28.23	26.87	0.13	21.17	22.56	0.39
Clicked distribution	0.21	0.23	0.46	0.26	0.28	0.60
Age[1]	34.13	36.53	0.00	35.09	36.64	0.02
Pct. male	0.72	0.71	0.80	0.71	0.71	0.89
n	996	524		1049	576	

[1] Using self-reported age-data

The variables obtained in the experiment are the guess in period 1, the guess in period 2, gender, age, and whether they in period 2 clicked the link to see the distribution of period 1. Table 1 shows that around 25 percent of the subjects clicked the distribution in REPEAT, whereas slightly less clicked it in ROBOT – in the latter treatment, the information provided by the distribution has no value, which explains why fewer subjects chose to acquire the information.

The variables obtained from Statistics Denmark are type of employment (employed, retired, self-employed etc.), type of highest education completed, income (both net and gross), and grade point average (GPA) in high school. Table 2 provides summary statistics for the socio-economic variables used in the regression analysis. With an average age close to 37 and a standard deviation of 12-13 years, our participants are two years younger on average than the population of Denmark (average 39.2 in 2006)[9]. The standard deviation of the entire Danish population in 2006 was 22.95, and our sample has thus a smaller standard deviation. About 71 percent of participants in the second period are male. This is a much higher share than in the Danish population as a whole (49 percent). Our average income of 187,547 DKK (25,000 €) is very close to the average net income in Denmark of 172,754 DKK (23,000 €) in 2006. The large standard deviation of 114,855 DKK (15,000 €) in our sample shows that subjects in our experiment are rather heterogenous. About half the identified sample completed high school (or equivalent), which is significantly higher than the 32 percent in the Danish population. Table 2 shows that the random allocation to treatments was effective, since all relevant variables are tested similar in REPEAT and ROBOT.

Table 2: Uniquely Identified Sample

	ROBOT		REPEAT		Mann-Whitney
Variable	Mean	Std.dev.	Mean	Std.dev	(p-values)
Guess in period 1	32.64	20.78	31.23	21.14	0.17
Guess in period 2	26.87	20.00	22.56	17.21	0.00
Clicked distribution (dummy)	0.23	0.38	0.28	0.45	0.05
Age	36.62	12.43	36.81	12.98	0.99
Pct. male	0.71	0.45	0.71	0.45	0.95
Net Income (DKK)	193,201	134,353	185,600	121,187	0.33
Pct. High School Completion(dummy)	0.52	0.50	0.52	0.50	0.96
GPA	8.66	1.02	8.71	0.99	0.55
n	524		576		

[9] All country data averages come from Statistics Denmark online statistics on the Danish population: http://www.statistikbanken.dk. Data from year 2006.

4.2 The Educational Variables

In this section we will briefly describe the educational variables in our dataset. The grade point average (GPA) variable is available for all persons completing high school (or equivalent) from year 1978 onwards. Danish grades are {00, 03, 5, 6, 7, 8, 9, 10, 11, 13}, where 6 is the passing grade. A higher GPA indicates better performance in high school. Since the GPA is an average of all grades, it can take any value from [6,13][10]. Note, that missing values of GPA can come from two sources: either the person did not complete high school and has thus no GPA recorded, or the person completed high school before 1978.[11]

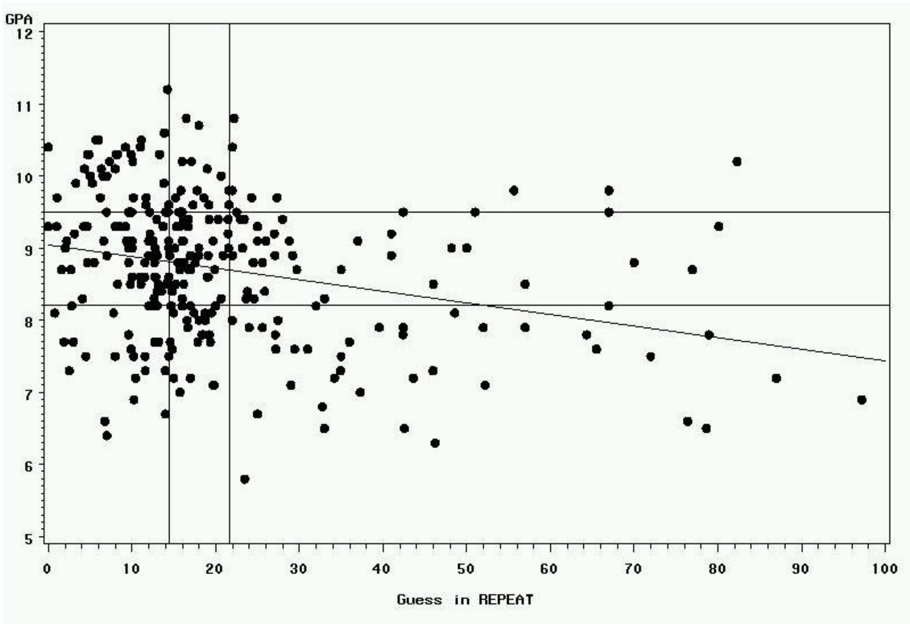
Figure 9 plots the submitted guesses in REPEAT against the GPA. The horizontal lines depicts the beginning of the top and bottom quartile of grades (GPA above 9.5 and below 8.2, respectively). The two vertical lines mark the target value (at 14.451) and the average (at 21.662). The downward sloping line is the fitted regression line. With a highly statistical significant negative slope of -0.016, the regression line shows the negative relationship between grades and guesses without controlling for any additional variables. This is finding is similar to Burnham et al. (2007), who find that subjects in the lowest quartile of cognitive ability make higher guesses, while subjects in the top quartile make almost no choices above 50 and one-third make choices below 12.5.

We also have very detailed information of the highest completed education. A total of 2,318 education types are available. We use this information to create a (dummy) variable “tertiary education”, that takes the value one if a bachelor, a master or a PhD-degree has been obtained. As expected, we find a significant positive correlation of 0.17 between GPA and whether a PhD has been obtained.

[10] The GPA can take values slightly lower than the passing grade of 6, since exceptional situations can arise where high school will be completed even with a grade point averages below the passing grade.

[11] Adding a dummy to control for this yield no changes in the regressions – coefficients remain similar, and the significance of the coefficients does not change.

Figure 9: Grades and Guesses



4.3 Regression Specifications

In the period 1 and REPEAT we measure the success of a participant by the absolute difference to the winning number. We regress the absolute difference to the winning number using a TOBIT model with censoring at 0.

In ROBOT, the dependent variable is a dummy that takes the value 1 if the participant solved the optimization task correctly (chose 0) and the dummy is 0 otherwise.

Since only about half the sample has completed high school, we include a GPA dummy (that takes the value 1 if high school is completed) and interact this GPA dummy with the GPA. This allows us to use all the observations, and interpret the results both with respect to high school completion and grade level in high school.

All regressions control for age, gender and income after tax and interests(in 10,000 DKK). The regressions run on second period data (REPEAT and ROBOT) include two additional independent variables. The first is the difference to the target in period 1. This variable measures if subjects are “good” at the game and implicitly measures whether subjects who guessed high in the first period also guess high in the second period (inertia). The second is a dummy that takes the value 1 if subjects in the second period clicked the link to the distribution of choices in the first period.

4.4 High School Completion

In the following tables the regression results for the first period, REPEAT, and ROBOT will be presented. We will discuss this triple three times, adding an education variable each time, also to check for robustness of results. First, we look at only upper secondary education (high school, ages 15-18, about 32% of a cohort), and then we look at tertiary education. Finally, we add the grade point average in high school.

We will now comment on the results of the first set of regressions, found in Table 3. The dummy capturing high school attendance is significant and of a magnitude of around 3 points in the first period. Thus, having attended high school implies a guess of 3 points closer to the winning number. In ROBOT we find, that subjects who attended high school are more likely to choose the dominant strategy of 0. In ROBOT, we see that this education variable is highly significant. We find this plausible because solving the optimization task is not trivial (only 12% could do it) and we know from other studies that cognition helps for optimization. Interestingly we find that in REPEAT the coefficient is insignificant. As seen from Figure 2, there is very little variation in the guesses provided. Hence, the regressions have less variation to pick up, and it seems plausible to expect less statistically significant variables in the regressions run on REPEAT data.

Table 3: Regression Results –High School Completion

Method	TOBIT	TOBIT	PROBIT
Dependent variable	Diff to target Period 1	Diff to target REPEAT	0 in ROBOT
Diff to target in FIRST round		0.28*** (0.04)	-0.01 (0.006)
Feedback 1=Clicked distribution link in PERIOD 2		-4.48*** (1.452)	0.26 (0.221)
Gender 1=Male	-3.76*** (0.700)	-2.02 (1.409)	0.40* (0.23)
Age Years	0.00 (0.033)	-0.06 (0.063)	-0.01 (0.011)
Income after tax and interests Gross income yields same results coefficients and significance	-0.83** (0.324)	-0.11 (0.606)	0.09 (0.07)
Education, secondary (Completed High School)	-3.76*** (0.711)	-2.09 (1.405)	0.60*** (0.223)
Constant	23.38*** (1.410)	13.39*** (2.872)	-1.85*** (0.49)
Sigma Standard deviation of errors	15.93*** (0.222)	14.28*** (0.437)	
Control for type of employment:	No	No	No
Observations	2568	535	461

Standard deviation in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

For the second period data, we have two additional independent variables, namely the difference to the winning number in the first period and whether the link to the distribution of guesses in the first period was clicked. For the REPEAT treatment, we find that both coefficients are statistically significant, and the magnitude high. The results imply, that for each unit a participant was away from target in the first period, she remained 0.28 points away from the target in REPEAT. Thus, the adaption is incomplete. Table A2 in the appendix shows the regression results with the difference in differences in the first period and REPEAT. We find that age is the only significant variable and thus that the degree of inertia is uncorrelated with educational variables.

The distribution dummy yields a very large coefficient in REPEAT, implying a distance of 4.5 less to the winning number if the link was clicked. So, the additional information provided in the distribution graph helped subjects guess closer to the winning number. One could argue that the coefficient cannot be interpreted causally, since the “smart” or “interested” subjects – who guess lower and closer to the winning number – could be the subjects who click the link to the distribution. In ROBOT however, the coefficient is insignificant. Getting feedback about the distribution of choices in the first period does not help figuring out the solution to the optimization problem. ROBOT is an optimization task in which strategic considerations or experience should play no role. Yet, you would still expect the “smart” and “interested” subjects to be interested in the game and click the distribution link. Since the variable is insignificant in ROBOT, it suggests that we can interpret the effect causally in REPEAT.

In the literature dealing with overconfidence, many studies find that men more overconfident than women (e.g. Lundeberg, Fox, and Punccohar 1994 and Barber and Odean 2001). Since roughly two thirds of our participants are male, systematic overconfidence by males might be self-fulfilling. Seen in this light, the statistically significant gender coefficient is interesting. In the first period, we find males guessing closer to the winning number. In the second period, males’ guesses are not statistical significantly closer to the winning number. Thus, in the second period systematic male overconfidence might be self-fulfilling. However, we find very little gender effect in ROBOT, ruling out systematic overconfidence as an explanation for the behaviour in REPEAT.

Income is only statistically significant in the first period. Income is highly correlated with education, and the significance of income might therefore be due to an omitted variable bias. However, we control for this in later regressions.

Finally we see that the constant is much lower in REPEAT than in the first period, showing that there is convergence, in that subjects in REPEAT are – on average – much closer to target than in the first period.

The interesting part here is that education was significant in the first period. However, in REPEAT education is now not significant and the two controls, gender and income, lose significance. In the first period, players have to form expectations “out of the blue”. This is cognitively difficult and good cognitive abilities are therefore helpful. But in REPEAT, it is more or less an issue of using the anchor provided by the average in the first period and applying the same reasoning (e.g. *k*-level) as before. This is merely an adaptation task that does not require much cognitive ability, yet information about the distribution of guesses in the previous period is helpful. Simply from looking at the distribution of choices in REPEAT (see Figure 2), the very condensed distribution does not leave much heterogeneity for the regressions to pick up.

4.5 Tertiary Education

Table 4 below provides the results from the next set of regressions. Similar to the previous table, we show the regression results for the first period, REPEAT and ROBOT.

First, we see that the coefficient for tertiary education is, as high school was in the first set, highly significant for the first period and ROBOT guesses. We again see that education in the REPEAT is insignificant. As a consequence of including tertiary education, the effect secondary education is getting, as expected, weaker in both the first period and ROBOT. It is remarkable that the coefficients on the controls are very stable, compared to Table 3.

4.6 Grade Point Average (GPA)

Table 5 below reports the results from the last set of regressions. Here, grade point average from high school is included (interacted with a dummy for high school completion).

Table 4: Regression Results –Tertiary Education

Method	TOBIT	TOBIT	PROBIT
Dependent variable	Diff to target Period 1	Diff to target REPEAT	0 in ROBOT
Diff to target in FIRST round		0.28*** (0.04)	-0.01 (0.007)
Feedback 1=Clicked distribution link in PERIOD 2		-4.46*** (1.46)	0.21 (0.224)
Gender 1=Male	-3.67*** (0.696)	-2.02 (1.409)	0.41* (0.236)
Age Years	0.01 (0.033)	-0.06 (0.063)	-0.01 (0.011)
Income after tax and interests Gross income yields same results coefficients and significance	-0.43 (0.331)	-0.10 (0.623)	0.06 (0.074)
Education, secondary (Completed High School)	-2.46*** (0.752)	-2.05 (1.489)	0.42* (0.242)
Education, tertiary (Completed BA, MA or PhD)	-3.80*** (0.74)	-0.13 (1.486)	0.47** (0.208)
Constant	22.79*** (1.407)	13.37*** (2.874)	-1.78*** (0.503)
Sigma Standard deviation of errors	15.85*** (0.221)	14.28*** (0.437)	
Control for type of employment:	No	No	No
Observations	2568	535	461

Standard deviation in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Regression Results – GPA

Method	TOBIT	TOBIT	PROBIT
Dependent variable	Diff to target Period 1	Diff to target REPEAT	0 in ROBOT
Diff to target in FIRST round		0.27*** (0.04)	-0.01 (0.007)
Feedback 1=Clicked distribution link in PERIOD 2		-3.75*** (1.475)	0.17 (0.229)
Gender 1=Male	-3.70*** (0.695)	-2.18 (1.401)	0.46* (0.242)
Age Years	0.00 (0.033)	-0.08 (0.063)	-0.01 (0.011)
Income after tax and interests Gross income yields same results coefficients and significance	-0.41 (0.331)	-0.10 (0.619)	0.04 (0.075)
Education, secondary (Completed High School)	9.47** (3.869)	18.49** (7.79)	-2.16* (1.197)
Education, tertiary (Completed BA, MA or PhD)	-3.14*** (0.768)	0.87 (1.523)	0.33 (0.221)
High School * GPA in High School	-1.42*** (0.451)	-2.43*** (0.903)	0.30** (0.136)
Constant	23.23*** (1.411)	14.10*** (2.868)	-2.02*** (0.518)
Sigma Standard deviation of errors	15.82*** (0.221)	14.19*** (0.434)	
Control for type of employment:	No	No	No
Observations	2568	535	461

Standard deviation in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

The inclusion of GPA changes the coefficients for secondary education dramatically - both in sign and size. However, when interpreting the numbers we must take the level of GPA into account. The average GPA in our sample 8.7, which means that the net effect of high school for an "average high school student" is:

Period 1: 9.47+(8.7*-1.42)=-2.88 (close to -2.46 in Table 4)

REPAT: 18.49+(8.7*-2.43)=-2.65 (close to -2.05, but insignificant in Table 4)

ROBOT: -2.16+(8.7*0.3)=0.71 (close to 0.42 in Table 4)

Note that all coefficients have the expected sign, i.e. inclusion of GPA reduces the positive effect of high school on the target, as expected. Again note that all the controls are strikingly similar to the previous regressions.

The above analysis shows that education is a good predictor of behavior. Better education – having completed high school or tertiary education – and higher grades in high school lead to guesses closer to the target number and to a higher likelihood of correct optimization. In the second period, guesses improve if the distribution of the previous period is seen. The coefficients are not only statistically significant, but also of an economically large magnitude. Thus, education matters for game play.

4.6 Step-Level Reasoning and Socio-Economic Variables

We previously saw, that subjects who played step-2 in the first period did not exhibit a distinct pattern in REPEAT. Now, we will see if the socio-economic variables can explain the steps of reasoning applied. More specific, we will focus on the predictions of step-1 and step-2 play, and let the absolute difference to this number be the dependent variable.

In the first period, step-1 and step-2 corresponds to 33 and 22, respectively. In REPEAT, step-1 corresponds to 21.605 (assuming step-0 to pick the average of the first period) and step-2 corresponds to 14.403. Note, that step-2 is (almost) the same as the actual target number for both periods.

Table 6 below shows the regressions for both the first and second period. The first two columns report step-1 and step-2, respectively, for the first period, whereas column 3 and 4 reports step-1 and step-2 for REPEAT.

For the first period, we find that the regression have similar sized coefficients and similar statistical significance as the difference to the target number, we saw in Table 5. For step-1 (i.e. the difference to 33), we find very few statistical significant coefficients. In particular, we see that GPA and high school completion are insignificant. Tertiary education is however significant, but the coefficient is much smaller than it is in the difference to step-2 play. We can thus conclude that in the first period, education as a whole – and in particular high school completion and better grades in high school – are more a characteristic of step-2 players than of step-1 players.

Table 6: Regression Results – Difference to step-1 and step-2 play

Method	TOBIT	TOBIT	TOBIT	TOBIT
Dependent variable	Period 1 Diff. to Step-1	Period 1 Diff. to Step-2	REPEAT Diff. to Step-1	REPEAT Diff. to Step-2
Feedback 1=Clicked distribution link in PERIOD 2			-1.29 (1.273)	-3.83*** (1.481)
Gender 1=Male	-1.99*** (0.578)	-3.62*** (0.698)	0.18 (1.209)	-2.17 (1.407)
Age Years	-0.04 (0.027)	0.00 (0.033)	-0.10* (0.054)	-0.07 (0.063)
Income after tax and interests Gross income yields same results coefficients and significance	-0.15 (0.275)	-0.43 (0.332)	-0.11 (0.534)	-0.12 (0.621)
Education, secondary (Completed High School)	3.47 (3.217)	9.05** (3.889)	13.39** (6.724)	18.09** (7.826)
Education, tertiary (Completed BA, MA or PhD)	-1.39** (0.639)	-3.04*** (0.771)	2.05 (1.314)	0.85 (1.529)
High School * GPA in High School	-0.55 (0.375)	-1.37*** (0.453)	-1.64** (0.78)	-2.38*** (0.907)
Constant	20.81*** (1.174)	23.00*** (1.418)	12.80*** (2.476)	14.14*** (2.879)
Sigma Standard deviation of errors	13.14*** (0.186)	15.88*** (0.224)	12.25 (0.375)	14.24*** (0.437)
Control for type of employment:	No	No	No	No
Observations	2568	2568	535	535

Standard deviation in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

For REPEAT, we similarly see that step-2 play is very similar to the results provided in Table 5. The main difference between step-1 and step-2 in REPEAT is, that the distribution of choices in period 1 had no explanatory power for step-1 players. Thus, subjects who saw the distribution are more step-2 players than they are step-1 players. However, the causality is not evident. We find the educational variables to be of similar statistical significance, but the coefficients are smaller for step-1 players than for step-2 players. Hence, the relationship between step-play and education is qualitatively the same, but the impact of education on step-2 players is larger than on step-1 players. For

REPEAT we can conclude that education characterizes step-2 players more than step-1 players, whereas feedback view only characterizes step-2 players.

4.5 Robustness Checks

To test the robustness of the above reported results, different specifications have been run. For instance, all regressions have been run with additional independent variables. The type of employment, using four dummy variables (self-employed, leading position, unemployed, and retired) resulted in no significant employment coefficients. Similarly, we got insignificant coefficients on a dummy for a university degree (a bachelor, a master or a PhD) and “years of schooling”. We also used gross income rather than income after tax and interests, but this led to similar coefficients that were also insignificant. Instead of using age, we ran regressions using age squared and found similarly that the coefficients were insignificant. However, in restricted regressions where we dropped independent variables we found that the age coefficient became statistically significant. Burnham et al. (2007) find that age and guesses are significantly correlated when using a comparable regression setup. However, their controls are fewer than ours and self-reported. We believe this correlation to be spurious, since it does not remain significant when we introduce additional controls from our high-quality data.

As an independent variable we also included the time subjects spent in the second period from loading the website to submitting their guess. Most subjects guessed within a short time after loading the page for the first time. However, it was possible to load the page and close the browser only to return at a later point in time to finish the experiment. As some subjects did this, we only used the time measurement if it was less than 15 minutes not to give too much weight to these relatively few observations with extreme values, compared to the average 3 minutes and 20 seconds it took subjects to submit their guess. Including the time spent was in some specifications significant, and led to a less significant distribution dummy. A likely reason is that subjects who clicked the distribution link also spent more time before submitting their guess. Consequently, we dropped the time usage in the main specifications.

Instead of using the difference to the winning number in the first period as independent variable, the same regressions have been run with the guess in the first period. This yields coefficients of similar sign, magnitude and significance.

Finally, all regressions were conducted with those independent variables included that were known in both the matched and the unmatched data set. Here, the coefficients were similar in sign, magnitude and significance, supporting the view that no systematic bias was introduced during the matching with Statistics Denmark data.

5 Conclusion

We have presented a large-scale newspaper and internet guessing game experiment with 19,196 participants. This is the largest guessing game ever, and the first two-period newspaper experiment. Furthermore, it is the first newspaper experiment with randomized treatment allocation. We have as such, carried many of the usual advantages of laboratory experiment over to newspaper and internet experiments. A two-period newspaper/internet experiment is a methodological novelty, that can be applied to many different types of experiments.

Moreover, we have matched our participants in the experiment with registry data and obtained socio-economic background variables that we use in analyzing subjects behavior. We find that subjects that attended high school, and subjects with a tertiary education make better guesses in the first period of game. Also, subjects with better high school grades make better guesses, whereas age and income have no effect. Finally we find evidence that supports a causal relationship between getting detailed feedback about what others did in the prior period and guessing lower and closer to the winning number.

We find that step-level theory explains behavior well in the first period. Repeating the experiment leads to convergence towards the equilibrium, and also here step-level reasoning explains behavior. Reducing the game to an individual optimization problem reveals that slightly more than one tenth of the subjects are able to solve the optimization problem, and that these “optimizers” outperform other subjects in the strategic guessing game. In fact, subjects who fail to optimize exhibit the typical choice pattern observed in guessing games. Disequilibrium behavior can thus to a large extent be explained by bounded rational behavior, and to a minor extent by strategic uncertainty. Thus, the typical distribution observed in these guessing games, can be explained by non-sophisticated players using step-level theory. However, about one tenth of the players act strategically and choose non-equilibrium guesses because of strategic uncertainty.

As such we have in a rich environment that combines a large-scale internet experiment with socio-economic registry data shown, that bounded rationality plays a large role for disequilibrium behavior, and that subjects with the right education help to mitigate disequilibrium and that they outperform subjects without education.

References

Barber, M.B. and Odean, T. (2001): Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment. *Quarterly Journal of Economics* 116, 261-292.

Benjamin, D.J., Brown, S.A. and Shapiro, J.M. (2006): Who is "Behavioral"? – Cognitive Ability and Anomalous Preferences. Levine's Working Paper Archive, UCLA Department of Economics.

Bosch-Domènech, A. and Nagel, R. (1997): El Juego de Adivinar el Número X: Una Explicación y la Proclamación del Vencedor. *Expansión*, June 16th, 42-43.

Bosch-Domènech, A., Montalvo, J.G., Nagel, R. and Satorra, A. (2002): One, Two, (Three), Infinity, ...: Newspaper and Lab Beauty-Contest Experiments. *American Economic Review* 92, 1687-1701.

Brandstätter, H., and Güth, W. (2002): Personality in Dictator and Ultimatum Games. *Central European Journal of Operations Research* 10, 191-215.

Burnham, T.C., Cesarini, D., Wallace, B., Johannesson, M. and Lichtenstein, P. (2007): Billiards and Brains – Cognitive Ability and Behavior in a *p*-Beauty Contest. SSE/EFI Working Paper No. 684.

Camerer, C.F. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.

Camerer, C.F., Ho, T.-H. and Chong, J.-K. (2004): A Cognitive Hierarchy Model of Games. *Quarterly Journal of Economics* 119, 861-98.

Chou, E., McConnell, M., Nagel, R. and Plott, C.R. (2007); The Control of Game Form Recognition in Experiments: Understanding Dominant Strategy Failures in a Simple Two Person "Guessing" Game. Social Science Working Paper 1274, Caltech.

Costa-Gomes, M.A. and Crawford, V.P. (2006): Cognition and Behavior in Two-Person Guessing Games: An Experimental Study. *American Economic Review* 96, 1737-1768.

Dickinson, D.L and McElroy, T. (2009): Naturally-occurring Sleep Choice and Time of Day Effects on p-Beauty Contest Outcomes. Working Paper 09-03, Appalachian State University.

Duffy, J. and Nagel, R. (1997): On the Robustness of Behaviour in Experimental “Beauty Contest” Games. *Economic Journal* 107, 1684-700.

Dohmen, T., Falk, A., Huffmann, D., Sunde, U. (2007): Are Risk Aversion and Impatience Related to Cognitive Ability? IZA Discussion Papers number 2735, Institute for the Study of Labor (IZA).

Drehmann, M., Oechssler, J. and Roider, A. (2005): Herding and Contrarian Behavior in Financial Markets: An Internet Experiment. *American Economic Review* 95, 1403-1426.

Fernández, V.A., Branas-Garza, P., Jiménez Jiménez, F. and Cosano, J.R. (2004): Communication, Coordination and Competition in the Beauty Contest Game: Eleven Classroom Experiments. Working paper 04-01, University of Grenada.

Fehr, E. and Renninger, S-V. (2002): Gefangen in der Gedankenspirale.” *Die Zeit* 48, 31.

Frederick, S. (2005): Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19, 25-42.

Grosskopf, B. and Nagel, R. (2007): Rational Reasoning or Adaptive Behavior? Evidence from Two-Person Beauty Contest Games. Working Paper – Universitat Pompeu Fabra.

Grosskopf, B. and Nagel, R. (2008): The Two Person Guessing Game. *Games and Economic Behavior* 62, 93-99.

Güth, W., Schmidt, C. and Sutter, M. (2007): Bargaining Outside the Lab – a Newspaper Experiment of a Three-Person Ultimatum Game. *Economic Journal* 117, 449-69.

Güth, W., Kocher, M.G. and Sutter, M. (2002): Experimental “beauty-contests” with Homogeneous and Heterogeneous Players and with Interior and Boundary Equilibria. *Economics Letters* 74, 219-28.

Ho, T.H., Camerer, C. and Weigelt, K. (1998): Iterated Dominance and Iterated Best Response in Experimental *p*-Beauty-Contests. *American Economic Review* 88, 947-69.

Holt, C. and Roth, A.E. (2004): The Nash Equilibrium: A Perspective. *PNAS* 101, 3999-4002.

Jones, G. (2006): Are Smarter Groups More Cooperative? Evidence from Prisoner's Dilemma Experiments, 1959-2003, Working Paper – George Mason University.

Kocher, M, Sutter, M. and Wakobinger, F. (2007): The Impact of Naïve Advice and Observational Learning in Beauty-contest Games. Working paper 2007-015, Tinbergen Institute.

Kocher, M. and Sutter, M. (2005): The Decision Maker Matters: Individual versus Group Behavior in Experimental Beauty-Contest Games. *Economic Journal* 115, 200-23.

Kovac, E., Ortmann, A. and Vojtek, M. (2007): Comparing Guessing Games with Homogeneous and Heterogeneous Players: Experimental Results and a CHM Explanation. *Economics Bulletin* 9, 1-9.

Lundeberg, M.A., Fox, P.W. and Punccohar, J. (1994): Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments. *Journal of Educational Psychology* 86, 114–121.

Millet, K. and Dewitte, S. (2006): Altrusitic Behavior as a Costly Signal of General Intelligence. *Journal of Research in Personality* 41, 316-26.

Morone, A., Sandri, S. and Uske, T. (2006): On the Absorbability of the Guessing Game Theory. Working paper December 2006.

Moulin, H. (1986): Game Theory for the Social Sciences. New York University Press, 2nd edition.

Nagel, R. (1995): Unraveling in Guessing Games: An Experimental Study. *American Economic Review* 85, 1313-26.

Øvlisen, F. (2009a): Strategic Properties in Guessing Games. Working paper.

Øvlisen, F. (2009b): Step-level Thinking and Changes in the Action Set. Working paper.

Palacios-Huerta, I. (2003): Learning to Open Monty Hall's Doors. *Experimental Economics* 6, 235-251.

Palacios-Huerta, I. and Volij, O. (2006): Field Centipides. Working paper.

Rydval, O., Ortmann, A., and Ostatnický, M. (2007): Three Very Simple Games and What it Takes to Solve Them, Jena Economic Research Paper no. 2007-092, Max-Planck-Institute of Economics, Jena.

Schou, A. (2005): Konkurrence afslører at danskere er irrationelle. *Politiken*, 22. September 2005, 1st Section, p. 13.

Sbriglia, P. (2004): Revealing the Depth of Reasoning in *p*-Beauty-Contest Games. University of Naples II, Working Paper.

Slonim, R.L. (2005): Competing Against Experienced and Inexperienced Players. *Experimental Economics* 8(1): 55-75.

Stahl, D.O. (1996): Boundedly Rational Rule Learning in a Guessing Game. *Games and Economic Behavior*, 16(2), 303-330.

Stahl, D., and Wilson, P. (1994): Experimental Evidence of Players' Models of Other Players. *Journal of Economic Behavior and Organization*, 25, 309–327.

Stahl, D., and Wilson, P. (1995): On Players Models of Other Players: Theory and Experimental Evidence. *Games and Econonomic Behavior*, 10, 218–254.

Sutter, M. (2005): Are Four Heads Better than Two? An Experimental Beauty-Contest Game with Teams of Different Size. *Economics Letters* 88: 41-6.

Thaler, R.H. (1997): Giving Markets a Human Dimension. *Financial Times*, June 16th, 6.

Weber, R. (2003): 'Learning' with No Feedback in a Competitive Guessing Game. *Games and Economic Behavior* 44: 134-44.

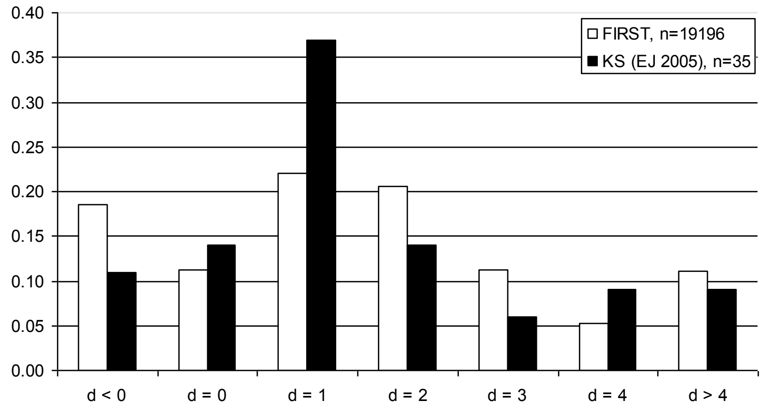
Appendix

Table A1: Distribution of step-levels (+/- 0.5) in the first period and in REPEAT treatment

REPEAT	*First round* 0 g_i>40.8	1 27<g_i<41	2 18<g_i<27	3 12<g_i<18	4 8<g_i<12	>4.5 g_i<8	Total
0	**0.12**	0.05	0.03	0.02	0.01	0.01	0.23
1	0.08	**0.07**	0.06	0.02	0.01	0.02	0.26
2	0.05	0.06	**0.06**	0.06	0.02	0.03	0.29
3	0.01	0.02	0.03	0.01	0.02	0.03	0.12
4	0.01	0.01	0.01	0.01	0.00	0.01	0.04
>4.5	0.01	0.01	0.01	0.00	0.00	0.02	0.05
Total	0.27	0.22	0.21	0.12	0.06	0.12	1.00

The corresponding step, *k*, of each individual guess, g_i, is calculated for the 1625 participants in the REPEAT treatment by solving for *k* in the equation: $g_i = (2/3)^k * R$, where the reference number *R* is 50 in the first period, and 32.407 in REPEAT.
In the table, we present the steps in grouped intervals, such that any guess with a step lower than 0.5 is represented as "0", a step btw. 0.5 and 1.5 as "1" and so forth.

Figure A1: Step-level in the laboratory and in the field



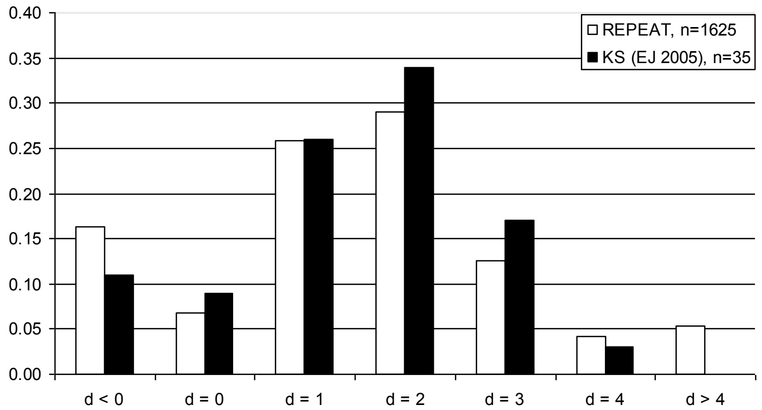


Figure A2: Distribution of choices in first period for matched participants only

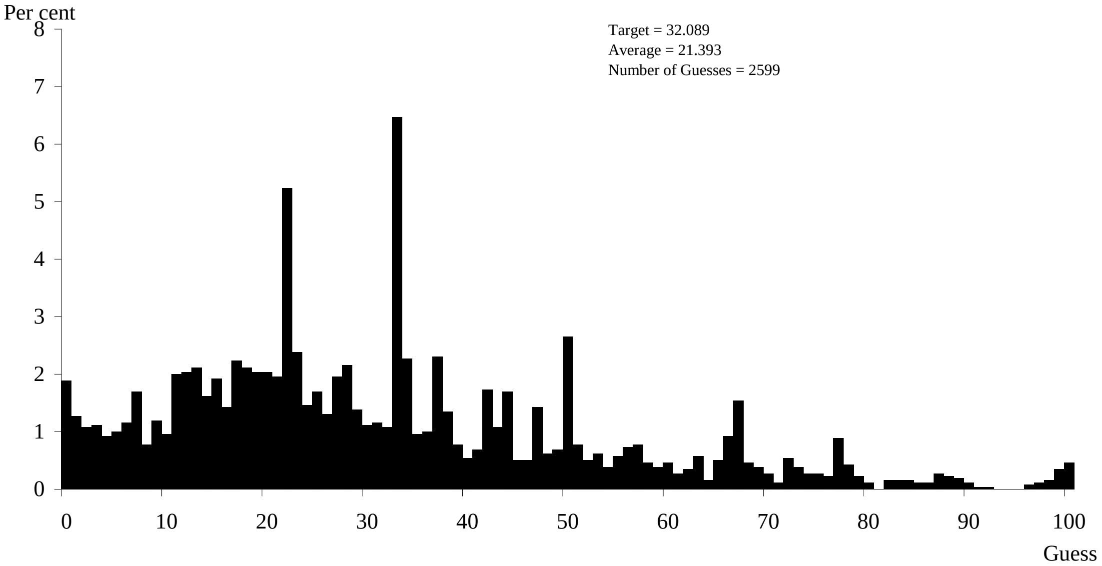


Figure A3: Distribution of choices in REPEAT for matched participants only

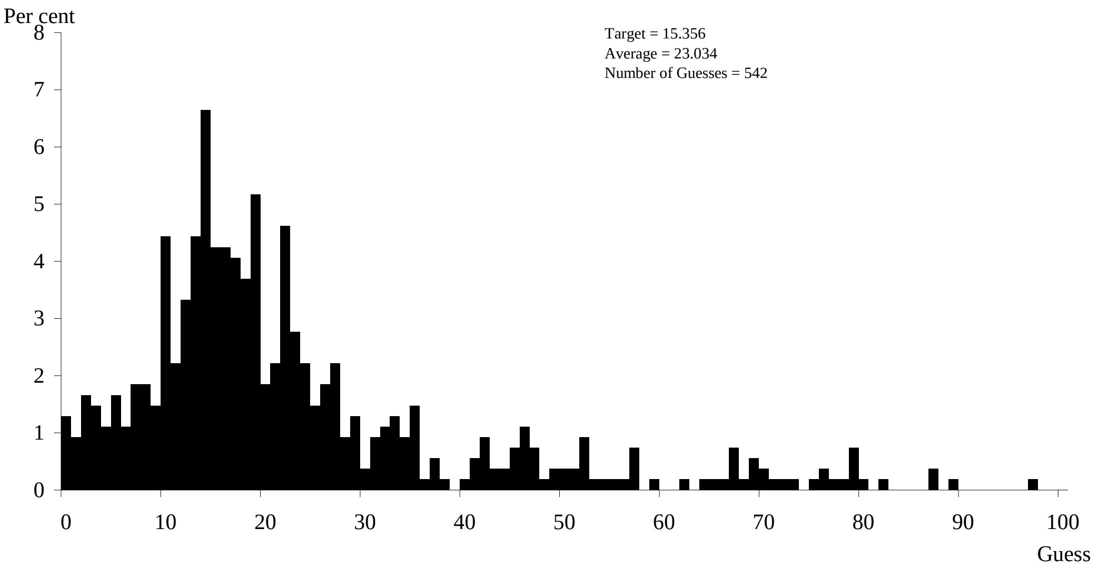


Figure A4: Distribution of choices in ROBOT for matched participants only

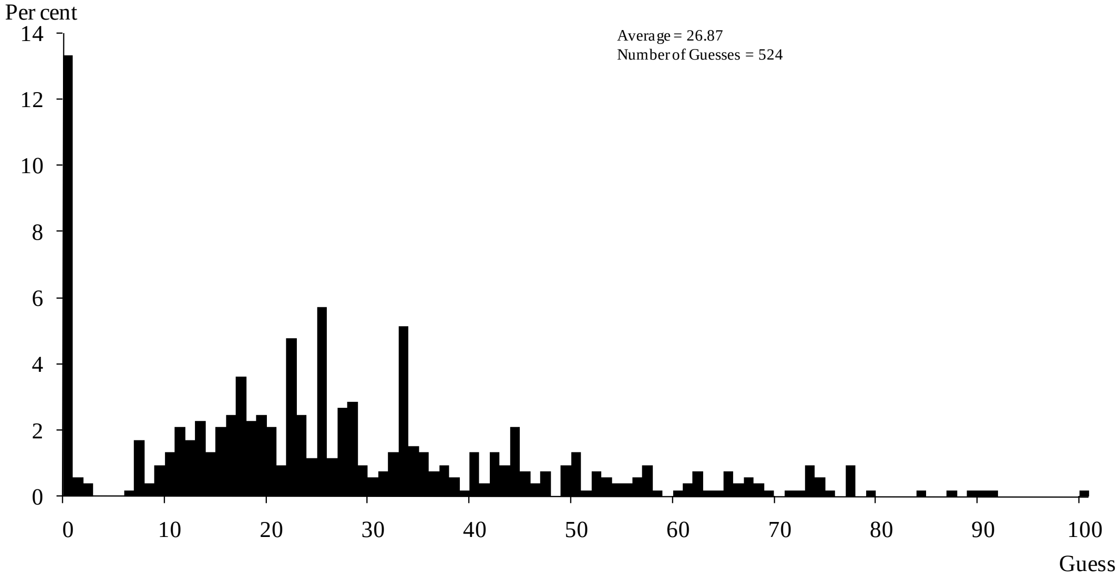


Table A2: Regression of difference to target in the first period – difference to target in REPEAT.

Method	TOBIT
Dependent variable	Difference in Difference First period - REPEAT
Feedback 1=Clicked distribution link in PERIOD 2	0.99 (1.868)
Gender 1=Male	-0.84 (1.772)
Age Years	0.16** (0.080)
Income after tax and interests Gross income yields same results coefficients and significance	-1.22 (0.783)
Education, secondary (Completed High School)	-13.52 (9.914)
Education, tertiary (Completed BA, MA or PhD)	-2.63 (1.935)
High School * GPA in High School	1.66 (1.149)
Constant	2.12 (3.472)
Sigma Standard deviation of errors	18.07*** (0.552)
Control for type of employment:	No
Observations	535

Standard deviation in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Instructions

The original instructions were in Danish. Below follows an English translation.

The first period:
The participants were told the rules, and that the purpose of the competition would be revealed four days after the competition closed for new entries, where the newspaper ran an article about the game and about experimental economics.
More specifically, the participants were told:
Headline: "Guess a number and win 5,000 kroner"
Politiken gives you the opportunity to win 5,000 kroner. All you have to do is to guess a number.

You must submit a guess according to these rules:

- Pick a number between 0 and 100 (both included), which you believe will be the closest to two thirds of the average of all submitted.
- Your guess does not have to be an integer
- You must submit your guess by Sonday the 18. September [2005] at 12:00.
- The winner is the reader closest to two thirds of the average off all the submitted numbers.
- If more than one participant are equally close, the prize will be split among them.
- The winner – and the purpose of the competition – will be revealed in Politiken and at politiken.dk on Thursday den 22. September [2005].
- Employees of JP/Politikens Hus can not participate in the competition.

Submit your guess (between 0 and 100 – both included): __________
Short reasoning for your guess (not mandatory): ___________________________
Name:_____________
Address: __________________
E-mail: ____________________
Phone: _____________________

< Submit your guess >

The second period:
The instructions consist of two parts. A common part, that was the same for all treatments and a treatment specific part, explaining the treatment specific rules.

Common part:
In September 2005, you participated in “Guess a Number” at the webpage of Politiken, with the possibility of winning DKK 5,000.
Now, we again offer you the opportunity to win DKK 5,000.
This time, only the 19,196 participants of the first period can participate. The participants of this round will be randomly divided into five different groups. In each group, a prize of DKK 5,000 will be awarded (such that the total prize pool is DKK 25,000).
Just as in the first round, the rules are that:
Your guess does not have to be an integer
If more than one participant are equally close, the prize will be split among them.
You must submit your guess by Wednesday the 21. March 2007 at 12:00.

In the first round the average of all guesses were 32.407.
The winning number (which was 2/3 of the average of all guesses) was 21.605
Your guess in the first round was [the participants guess in the first period was displayed here].
You can read more about the first competition from September 2005 in the *newspaper article from Politiken, 22. September 2005* [link opens a new window]. In addition, you can see a graphic distribution of guesses in the first round here [link opens a new window].

Treatment Specific Part:
REPEAT: In your group, the rules in this round are the same as in the first round.
Pick a number between 0 and 100 (both included), which you believe will be the closest to two thirds of the average of all submitted.
The winner is the participant closest to two thirds of the average off all the submitted numbers.
The lower you expect the other to choose on average, the lower you should choose.
So, you should choose 2/3 of your expectation of the average of the others’ guesses.

ROBOT: In your group, the rules in this round are slightly different from the first round.
You will be matched with a computer.
The computer chooses a random number between 0 and 100 (both included). All numbers appear with the same probability.
Pick a number between 0 and 100 (both included), which you believe will be the closer to two thirds of the average of the computers number and your number.
A participant in your group will be picked at random. Is this participants number closer than the computers number to two thirds of the average, the participant will win.
So, the lower you expect the computers number, the lower you should choose.

Common part (continued): Submit your guess: __ Comment (optional): __

Step-level Thinking and Changes in the Action Set

Step-level Thinking and Changes in the Action Set

Frederik Roose Øvlisen*

May 2009

Guessing games have been studied in many variations and parameterizations, yet few studies have analyzed the effect of a change in parameters from one period to another. Here, we study how a change in the action set between the first and the second period of a large-scale internet guessing game influences behavior. Although not predicted by existing theories, we find a dramatic change in behavior that corresponds exactly to the relative change in the action set. However, if the assumptions of the theories are normalized by the relative change in the action set, the predictions explain the observed behavior. Thus, with appropriate assumptions, step-level thinking is a good predictor of behavior when a change is the action set is introduced.

Keywords: Bounded rationality, guessing game, beauty contest, newspaper and internet experiment, step-level reasoning.

JEL-fcodes: D03, D50, D83, D84, C93.

* University of Copenhagen, Department of Economics, Studiestræde 6, DK-1455 Copenhagen. Frederik.Oevlisen@econ.ku.dk. I gratefully acknowledge financial support from the University of Copenhagen. I am grateful for comments by Jean-Robert Tyran, Ralph Bayer, Dirk Engelmann, and Charles Noussair. I thank Guan Yang for valuable research assistance.

Introduction

Since the introduction to the experimental literature in the seminal paper by Nagel (1995), guessing games have been much studied, and theories have been develop to explain the behavior we typically observe in this class of game. Although the game has been studied and compared across many variations and parameterizations (e.g. Güth, Kocher and Sutter (2002) and Morone and Morone (2008)), very few studies have devoted any attention to how such changes are anticipated by the participants.

In this paper we analyze how behavior is affected when subjects experience that the action set is extended from the first to the second period in a 2-period game. Although not predicted by existing theories, we show that behavior is very different when changing the action set. We then show how modifications to the assumptions of the existing theories can predict the observed pattern. More specifically we show that the model's reference point should be normalized by the change in the action set.

The guessing game is a very nice workhorse to study such a change. There exists vast experimental evidence on the game, there is a simple game theoretic prediction, the game is absent of social and risk preferences, and theories have been developed to explain the non-Nash behavior typically observed in the game. The general structure of the guessing game is quite simple. N participants simultaneously choose a number x_i ($i = 1,\ldots,N$) from the closed interval l to h. The guess with the minimum distance to the "target number", defined as p times the average number ($X = N^{-1}\sum x_i$), wins a prize. If several people are equidistant to the target number, the prize is equally shared among these people.

The distribution of choices normally observed in these guessing games are well explained by step-level (or level-k) theories, as proposed by e.g. Nagel (1995), Camerer, Holt and Ho (2004) and Stahl (1996). These theories assume that subjects apply step level cognition and characterize participants into different categories according to their cognitive depth of reasoning. In these models, subjects perceive that they themselves are the most sophisticated and they best respond to their belief of the other players applying less steps than themselves. The intuition of the models is as follows. A step-0 type is assumed to pick some reference number x_0 (an equivalent interpretation is that all step-0 players pick x_0 on average). A step-1 type believes everybody else to be step-0 players, and picks the best reply to step-0. Thus, a step-k type believes everyone else to be of type k-1, and best replies to this

(In the cognitive hierarchy model proposed by Camerer et al. (2004) a type k is assumed to best reply to a distribution of types of less cognitive depth than k).

Most of the step-level theories are developed for a one-shot purpose. However, Nagel (1995) apply the theory to a repeated guessing game. It is then assumed that step-0 play corresponds to playing the average number of the last round.

In the first round as well as in both treatments of the second round, standard game theory predicts a single pure strategy Nash equilibrium of 0. Yet, numerous studies of the game have shown that behavior differs from the Nash prediction[1]. In the laboratory, repeating the game has lead to convergence towards the Nash equilibrium. We here present the first repeated large-scale guessing game and provide novel insights as to the convergence in a large-scale guessing game.

Here, we present a large-scale experiment with a novel design to address how subjects in a large-scale experiment adjust guesses in a repeated setting when the environment is non-stable. We let subjects play one of the most common parameterizations, with p=2/3, l=0 and h=100, in the first round. In the second round, we introduce two treatment variations. The first treatment (REPEAT) is similar to previous guessing game experiments (see e.g. Nagel (1995) and Kocher & Sutter (2005)), in that it is a simple repetition of the first round. However, it is the first large-scale repeated guessing game. Here we find that subjects play similar to previous laboratory studies, in that repetition leads to convergence towards the Nash prediction. Just as first round behavior has been found similar in large-scale and laboratory studies, we conclude that this also holds for second period behavior.

The second treatment (EXTEND) takes a different approach, in that we extend the action set by setting h=180. We should expect no difference in behavior in REPEAT and in EXTEND. Standard game theory would predict the behavior in REPEAT and EXTEND to be the same, namely that everyone pick the equilibrium of 0. The prediction of step-level thinking rests on the assumption that step-0 types pick the average of the previous round, and would thus also not predict any difference in behavior. Yet, we find that behavior differs dramatically. However, if we normalize the submitted guesses by 1.8 (the factor with which the action set was extended), the provided guesses in the two treatments are the same. Thus we show that subjects facing a change in the range of possible guesses apply the same reasoning as subjects in the stable environment, but that they adjust for the change in parameters. By modifying the assumption on step-0 play in step-level theory, we can explain the

[1] In Tyran & Øvlisen (2009) we show that bounded rationality and to some extent strategic uncertainty causes this non-Nash behavior.

behavior observed. Rather than assuming step-0 types to play the average of the last period, we assume that step-0 types pick the *normalized* average of the last period.

In the next section the game and experimental design will be introduced in more detail, just as our design and procedure will be explained. Then a section on the results and analysis follows, preceding a short conclusion.

Design

The experiment consists of two rounds and two treatments.[2] The first round was announced in the hard-copy version of the Danish daily newspaper *Politiken* and on its website. Participation was only possible through their website, and not limited to subscribers of the newspaper but open to all visitors at the newspaper's website.

In the first round, all participants played the standard guessing game with l = 0, h = 100 and p = 2/3, and could guess any real number (allowing for up to nine digits) between 0 and 100 (both included). In total 19,196 participants took part to win the prize of DKK 5000 (at the time approx. $1,000 or € 670), that was split equally among the four winners that came closest to the target number.

In the second round, all previous participants were invited to participate again. The participants in the second round were randomly allocated to one of five different treatments, each with a prize of DKK 5000 (at the time approx. $1000 or € 670). In addition to a short description of the experiment (see appendix), they were informed about their own guess, the average guess and the target number of the first round. The information page also included two links providing additional information (in a pop-up window). One link showed the distribution of guesses in the first round (similar to Figure 1) and the other link displayed a newspaper article about the guessing game, with detailed comments of how to think of the game (Schou 2005). 7,635 of the 19,196 invited participants took part in the second round.

In this paper, we analyze two treatments in the second round. Treatment REPEAT with 1,625 participants was simply a repetition of the first round, where the rules and parameters remained unchanged with l = 0, h = 100 and p = 2/3. The other treatment, EXTEND, had 1,560 participants. Here, the rules were slightly different in that we changed the upper bound of guesses to h = 180,

[2] A total of five treatments were run in the second period. The other three treatments address a different question and are discussed in the companion papers Tyran and Øvlisen (2009) and Øvlisen(2009). A more detailed description of the game and the experiment can be found in Tyran & Øvlisen (2009).

keeping $p = 2/3$ and $l = 0$ unchanged. Thus, rather than choosing numbers between 0 and 100, as in the first period and in REPEAT, the action set was extended such that subjects in EXTEND could choose numbers between 0 and 180.

All present theories that have been applied to guessing games suggest the same: namely that we should observe the same behavior in REPEAT and in EXTEND. In EXTEND, we simply add some high numbers that should not affect choices since they are in a region not considered by the players.

In the next section we will see, that the actual behavior is much different from what our theories suggests.

Results

The results of the first period with 19,196 participants are similar to previous large-scale guessing games (e.g. Bosch-Domènech, Montalvo, Nagel and Satorra 2002). Tyran & Øvlisen (2009) provides an in-depth analysis of the experiments of the first round, where the average was 32.4 and the winning number 21.6. The results exhibit the typical distribution of choices, with the spikes predicted by the step-level model.

Figure 1 show the guesses in the REPEAT treatment. Compared to the winning number of the first period, the winning number of 14.4 is now closer to the Nash prediction and is 2/3 lower than the first period. We also see the distribution of guesses to be much more condensed, and that most choices are around the winning number of 14.4. This is very similar to previous experiments performed in the laboratory (Nagel 1995). We find similar convergence towards the Nash prediction in a large-scale experiment with heterogeneous subjects as we see in the laboratory. The result is well explained by level-*k* thinking, where step-0 players are assumed to on average pick 32.4 (the average of the first period). The model would predict that step-1 players choose 21.6, step-2 players choose 14.4, and step-3 players choose 9.6. When we look at the distribution, these values are where we find the largest spikes. The largest spike at 14 corresponds to step-2 play, and the second-largest spike at 10 corresponds to step-3 play. Thus, we find the same behavior in a large-scale repeated internet guessing game with 1,625 subjects as we find in the typical laboratory studies with much smaller group sizes.

Figure 2 show the guesses in the EXTEND treatment. When comparing to Figure 1, it is immediately clear that the two distributions are very different (t-test, p=0.00, Mann-Whitney p=0.00, Kolmogorov-Smirnov p=0.00). Guesses in EXTEND are much more dispersed than guesses in REPEAT, and the average number of 40.91 is higher than the average in the first period. Whereas we observe convergence towards the Nash-prediction in REPEAT, we observe the opposite in EXTEND. From both the step-level models and standard game theory, we predicted behavior in the two treatments to be the same. Yet we observe a strikingly different behavior in the two treatments.

We find approximately the same fraction of strictly dominated choices in both treatments (3 per cent choosing above 67 in REPEAT and 2 per cent choosing above 120 in EXTEND). If the observations in REPEAT were bulked at 100, it could be because subjects preferred numbers outside the action set. However, we can rule this out since only two subjects choose 100 in REPEAT.

Figure 1: Guesses in REPEAT

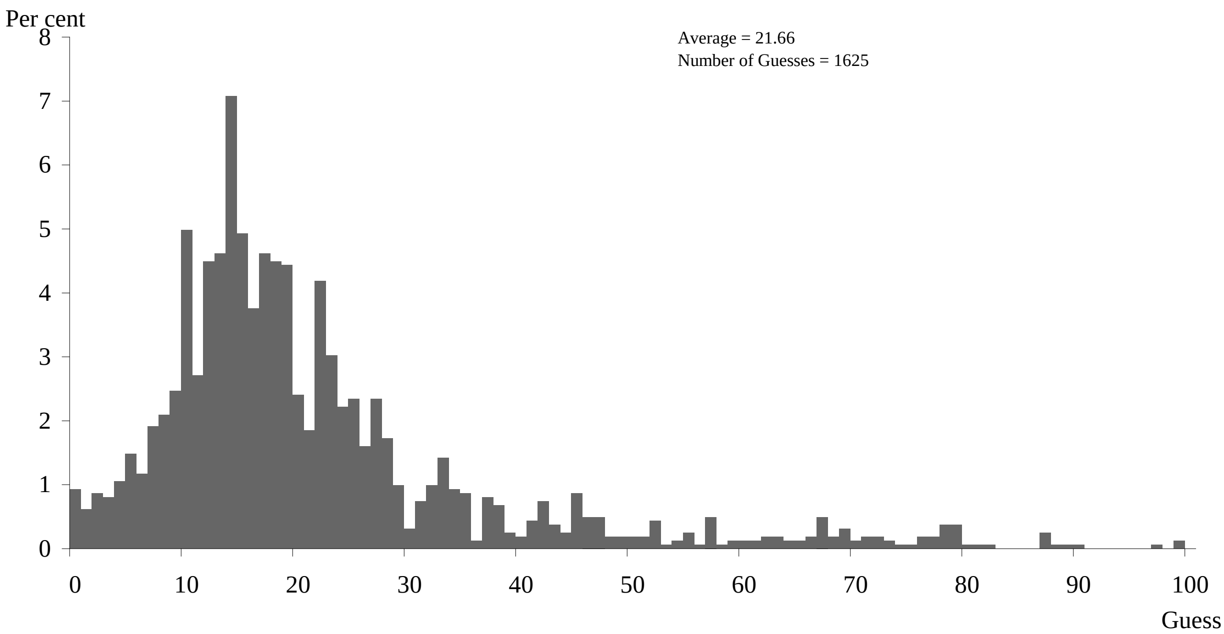
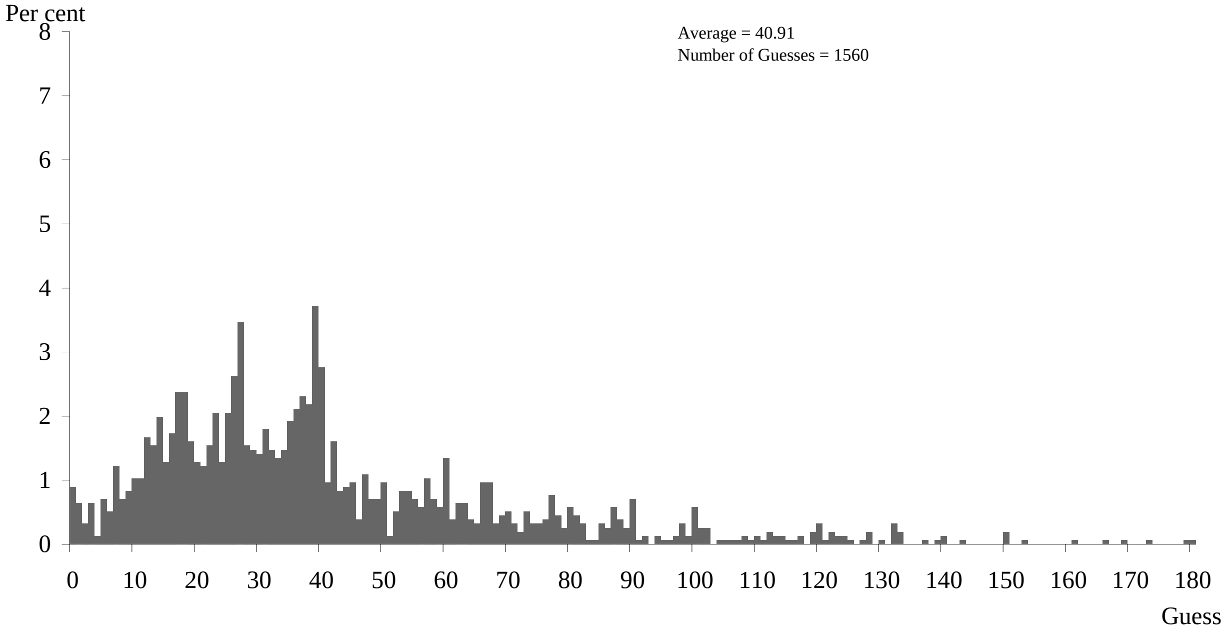


Figure 2: Guesses in EXTEND



This seems like a striking and peculiar finding. Yet, we might gain some useful insight when we look for the largest spike in Figure 2. We find that the largest spike at 39, since just below 4 per cent of the participants (57 subjects) guessed higher than 38.5 and below 39.5 (both included). Of these 57 guesses, 18 subjects guessed exactly 38.889. In REPEAT, less than one per cent (4 subjects) chose guesses in this interval, an no one guessed at 38.889.

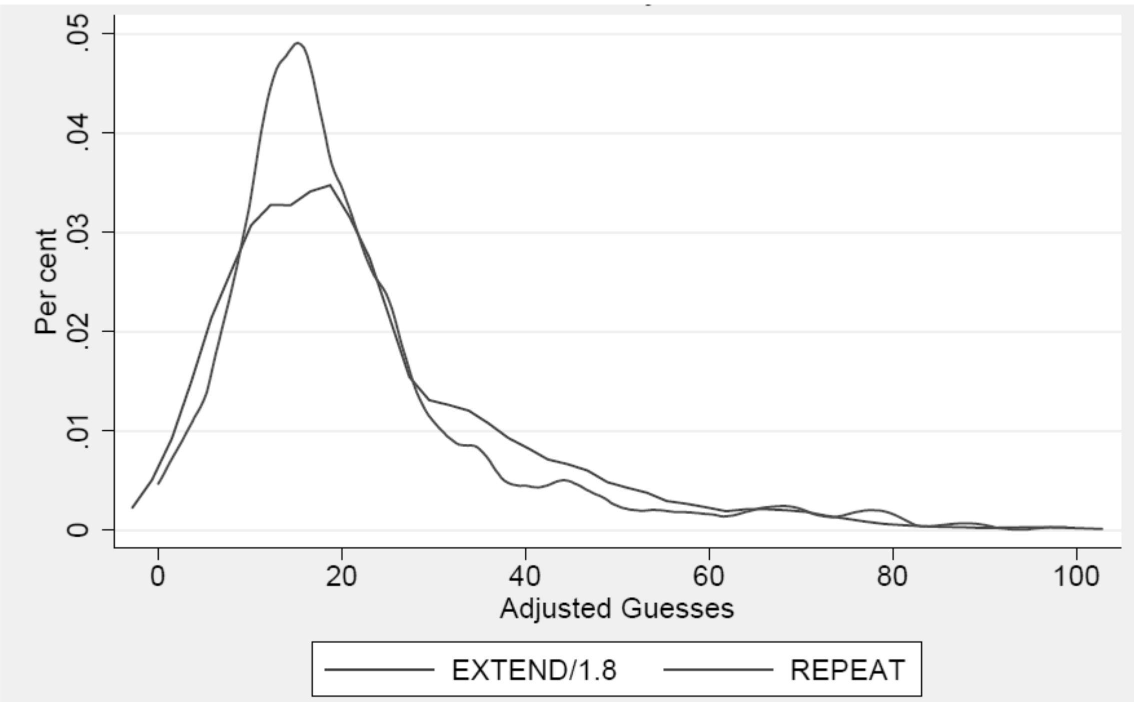
At first glance, this is indeed a peculiar number to choose for so many people. However, a closer look reveals what logic these subjects followed. They have taken the winning number of the first period, and adjusted it by the relative change in the action set. The action set was extended by a factor 1.8, and since the winning number in the first period was 21.605, we get: 21.605*1.8=38.889. If we relax the precision of the adjustment a little, we can even explain the spike we find at 40.

Now we incorporate this logic into the step-level model. Taking the average number of the previous period and adjusting it with 1.8 thus gives what we will call the “normalized step-0 play”. Normalized step-0 types are thus assumed to guess 32.5*1.8=58.5, normalized step-1 players will consequently choose 39, and normalized step-2 types will choose 26.

We have already seen how the largest spike in Figure 2 can be explained by normalized step-1 players. Similarly, we can explain the large spikes at 26 by normalized step-2 types. The spike at 27 can be explained using the relaxed precision of the adjustment on step-0 players.

Now, if this logic was to hold true we could simply scale all guesses in EXTEND downwards by 1.8, and we should get similar distribution of guesses EXTEND and REPEAT. Figure 3 shows the kernel density of the guesses in REPEAT and the normalized guesses in EXTEND (i.e. the guesses divided by 1.8). We see that the distributions are strikingly similar, although testing does not confirm that the two distributions are the same (t-test, p=0.06, Mann Whitney p=0.03, Kolmogorov-Smirnov p=0.00). The distribution of REPEAT still exhibits lower guesses and more guesses around the winning number than observed in the normalized EXTEND.

Figure 3: Kernel Density of guesses in REPEAT and adjusted guesses in EXTEND.



Conclusion

In this paper, a two-period large-scale guessing game is presented. In the first period our results are similar to other comparable large-scale and laboratory studies. In the second period, we present two different treatments.

The first treatment is simply a repetition of the first period. We find convergence towards the Nash equilibrium, similar to that of previous laboratory experiments. Convergence has not previously been studied in a large-scale experiment, and we conclude that the behavior in the repeated large-scale guessing game is similar to that of laboratory studies.

The second treatment introduces a change in the environment in that the action set is extended. Rather than allowing for numbers between 0 and 100 (as in the first period), subjects can choose numbers from 0 to 180.

Step-level models would predict the same behavior in the two treatments, applying the usual assumption that step-0 players pick the mean of the previous period (Nagel, 1995). Yet we observe a strikingly different behavior. Whereas there was convergence towards the Nash equilibrium in the repeated game, the treatment with the extended action space showed the opposite, namely higher numbers being chosen on average.

However, if we modify the usual assumption for step-0 play, and instead assume step-0 to pick the normalized mean of the previous period, the predictions of the two treatments would not be the same. In our example, step-0 would not choose the average of the previous period, but rather the average times 1.8 – the factor by which the action set was extended. Applying this assumption, we can explain the spikes we observe in the data. Normalizing the entire distribution of choices by 1.8 reveals a strikingly similar distribution in the repeated treatment and in the treatment with extended action set.

References

Bosch-Domènech, A., Montalvo, J.G., Nagel, R. and Satorra, A. (2002): One, Two, (Three), Infinity, ...: Newspaper and Lab Beauty-Contest Experiments. *American Economic Review* 92, 1687-1701.

Camerer, C.F., Ho, T.-H. and Chong, J.-K. (2004): A Cognitive Hierarchy Model of Games. *Quarterly Journal of Economics* 119, 861-98.

Duffy, J. and Nagel, R. (1997): On the Robustness of Behaviour in Experimental "Beauty Contest" Games. *Economic Journal* 107, 1684-700.

Güth, W., Kocher, M. G. and Sutter, M. (2002): Experimental "beauty-contests" with Homogeneous and Heterogeneous Players and with Interior and Boundary Equilibria. *Economics Letters* 74, 219-28.

Kocher, M. and Sutter, M. (2005): The Decision Maker Matters: Individual versus Group Behavior in Experimental Beauty-Contest Games. *Economic Journal* 115, 200-23.

Morone, A. and Morone, P. (2008): Boundary and interior equilibria: what drives convergence in a 'beauty contest'?, MPRA Working Paper.

Nagel, R. (1995): Unraveling in Guessing Games: An Experimental Study. *American Economic Review* 85, 1313-26.

Schou, A. (2005): Konkurrence afslører at danskere er irrationelle. *Politiken*, 22. September 2005, 1st Section, p. 13.

Stahl, D.O. (1996): Boundedly Rational Rule Learning in a Guessing Game. *Games and Economic Behavior*, 16(2), 303-330.

Tyran, J-R and Øvlisen, F.R.(2009): Making an Educated Guess. Working Paper.

Øvlisen, F.R.(2009): Strategic Properties in Guessing Games. Working Paper.

Appendix - Instructions

The original instructions were in Danish. Below follows an English translation.

The first period:
The participants were told the rules, and that the purpose of the competition would be revealed four days after the competition closed for new entries, where the newspaper ran an article about the game and about experimental economics.
More specifically, the participants were told:
Headline: "Guess a number and win 5,000 kroner"
Politiken gives you the opportunity to win 5,000 kroner. All you have to do is to guess a number.

You must submit a guess according to these rules:

- Pick a number between 0 and 100 (both included), which you believe will be the closest to two thirds of the average of all submitted.
- Your guess does not have to be an integer
- You must submit your guess by Sonday the 18. September [2005] at 12:00.
- The winner is the reader closest to two thirds of the average off all the submitted numbers.
- If more than one participant are equally close, the prize will be split among them.
- The winner – and the purpose of the competition – will be revealed in Politiken and at politiken.dk on Thursday den 22. September [2005].
- Employees of JP/Politikens Hus can not participate in the competition.

Submit your guess (between 0 and 100 – both included): __________
Short reasoning for your guess (not mandatory): __________________________
Name:_____________
Address: __________________
E-mail: ____________________
Phone: _____________________

< Submit your guess >

The second period:
The instructions consist of two parts. A common part, that was the same for all treatments and a treatment specific part, explaining the treatment specific rules.

Common part:
In September 2005, you participated in "Guess a Number" at the webpage of Politiken, with the possibility of winning DKK 5,000.
Now, we again offer you the opportunity to win DKK 5,000.
This time, only the 19,196 participants of the first period can participate. The participants of this round will be randomly divided into five different groups. In each group, a prize of DKK 5,000 will be awarded (such that the total prize pool is DKK 25,000).
Just as in the first round, the rules are that:
Your guess does not have to be an integer
If more than one participant are equally close, the prize will be split among them.
You must submit your guess by Wednesday the 21. March 2007 at 12:00.

In the first round the average of all guesses were 32.407.
The winning number (which was 2/3 of the average of all guesses) was 21.605
Your guess in the first round was [the participants guess in the first period was displayed here].
You can read more about the first competition from September 2005 in the *newspaper article from Politiken, 22. September 2005* [link opens a new window]. In addition, you can see a graphic distribution of guesses in the first round here [link opens a new window].

Treatment Specific Part:
REPEAT: In your group, the rules in this round are the same as in the first round.
Pick a number between 0 and 100 (both included), which you believe will be the closest to two thirds of the average of all submitted.
The winner is the participant closest to two thirds of the average off all the submitted numbers.
The lower you expect the other to choose on average, the lower you should choose.
So, you should choose 2/3 of your expectation of the average of the others' guesses.

EXTEND: In this round, the rules are in your group slightly different than in the first round.
Pick a number between 0 and 180 (both included), which you believe will be the closest to two thirds of the average of all submitted.
The winner is the participant who gets closest to two thirds of the average of all the submitted guesses.
The lower you expect the other to choose on average, the lower you should choose.
So, you should choose 2/3 of your expectation of the average of the others' guesses.

Common part (continued): Submit your guess: __ Comment (optional): __

Strategic Properties in Guessing Games

Strategic Properties in Guessing Games

Frederik Roose Øvlisen*

May 2009

In this paper we present a two period large-scale internet experiment. First we develop a general formulation of target function in the guessing game. We show existence of equilibrium and provide conditions for uniqueness. Then we let subjects play an initial round of the guessing game to anchor them on low values. In the second round we present two treatments that differ only in their strategic property. We ask if behavior under strategic complements and strategic substitutes differs, and find that guesses under strategic substitutes are closer to the equilibrium prediction and exhibit higher depth of reasoning than guesses under strategic complements.

Keywords: Guessing game, beauty contest, step-level reasoning, strategic complements, strategic substitutes.

JEL-codes: D03, D50, D83, D84, C93.

*

* University of Copenhagen, Department of Economics, Studiestræde 6, DK-1455 Copenhagen. Frederik.Oevlisen@econ.ku.dk. I gratefully acknowledge financial support from the University of Copenhagen. I am grateful for comments by Jean-Robert Tyran, Ralph Bayer, Dirk Engelmann, Wieland Müller, and Charles Noussair. I thank Guan Yang for valuable research assistance.

Introduction

In this paper, we ask if behavior in the guessing game is different under strategic complements than under strategic substitutes, i.e. if it matters for behavior if the slope of the best reply function is positive or negative. We report experimental results from a two-period large-scale guessing game. In the first period, the 19,196 subjects play the standard guessing games (Nagel 1995). The first period anchors subjects on “lower numbers”. In the second period, we adjust the action space and change the equilibrium value from the boundary to an interior equilibrium. We then assign subjects to two treatments that differ only in their strategic environment. We find that guesses under strategic substitutes are closer to the equilibrium value and exhibit larger depth of reasoning, than guesses under strategic complements.

Since Moulin (1986) and Nagel (1995) introduced the guessing game in the literature, experiments with various parameterizations and variations have been performed using different types of target functions. In this paper, we propose a general model of the target function that allows for all possible variations of the guessing game. We show existence of equilibrium and provide conditions for uniqueness. The model captures all variations of the guessing game, and in particular it makes a change of the strategic environment easy. Thus, our formulation eases experimental design as well as the theoretical analysis of the game.

The strategic environment is essential for any game. If the slope of the best-reply function is positive, we say that actions are strategic complements, whereas strategic substitutes is defined as a negative slope. Fudenberg and Tirole (1984) and Haltiwanger and Waldman (1985) were among the first to study strategic complements and substitutes. Here, a simple intuition explains the difference in behavior that we observe when the strategic properties are changed. Under strategic complements, rational subjects have the incentive to mimic bounded rational subjects, whereas in strategic substitutes rational subjects have an incentive to act in the opposite way of the bounded rational subjects. Camerer and Fehr (2006) apply the same logic to explain how behavior in business entry games (with strategic substitutes) comes surprisingly close to the equilibrium prediction, whereas behavior in the (classic) guessing game (with strategic complements) systematically deviates from equilibrium prediction.

A recurring theme in dealing with strategic properties is the issue of Cournot and Bertrand competition. Quanties in Cournot competition are strategic substitutes, while prices are strategic complements in Bertrand. Being the two workhorse models of industrial organization, the theme has gained some attention (e.g. Davis 2008). Suetens and Potters (2007) show experimentally that there is more tacit collusion in a Bertrand than in Cournot setting. However, a *ceteris paribus* comparison between treatments is not possible.

Fehr and Tyran (2008) present one of the first direct comparisons of behavior under strategic complements and substitutes. Using a price-setting game, they show that after a nominal shock, subjects equilibrate quicker under strategic substitutes than under strategic complements. The explanation follows the same intuition as above, namely that expectations are flexible and forward-looking under strategic substitutes, but adaptive and sticky under strategic complements. Similarly, Suetens and Potters (2009) find that in two-player cooperation games, there is more cooperation when actions are strategic complements than when they are strategic substitutes.

Closest to our analysis here is a recent paper by Willinger and Sutan (2009). They also compare behavior in the guessing game under strategic substitutes and complements. With subjects in groups of 8 people, they let subjects submit guesses between 0 and 100. In a one shot setting with an interior equilibrium of 60, they find that guesses under strategic substitutes come closer to equilibrium than guesses under strategic complements. Furthermore they find, that subjects in both treatments apply 1.5 steps of reasoning on average. However, their treatment variation does not only change the strategic environment – the slope of the best-reply function and the range of dominated actions is different between the two treatments. Our formulation of the target function ensures that the slope of the best-reply function and the range of dominated strategies remain the same in the two treatments.

In this paper we confirm the finding of Willanger and Sutan (2009) that guesses are lower and further away from equilibrium under strategic complements than under strategic substitutes. However our results suggest that the depth of reasoning is larger under strategic substitutes than under strategic complements.

We here report results from a two-period large-scale guessing game. In the first period, the 19,196 subjects play the standard guessing games (Nagel 1995). This period anchor subjects on “lower numbers”. In the second period, subjects are assigned to either a treatment with strategic substitutes or strategic complements. We find that guesses under strategic complements are closer to the equilibrium value and exhibit larger depth of reasoning, than guesses under strategic complements.

The organization of the paper is as follows. First a general theory of the target function of the guessing game is presented. It is shown that there exist a Nash equilibrium and the conditions for uniqueness are derived. Then the experimental design is explained, before the hypotheses are presented. After the results a detailed discussion on how the results relate to step-level theory follows, before we conclude.

Theory

In the guessing game, N participants simultaneously guess a number between l and h. Let x_i denote the guess of participant $i = 1,\dots, N$. The guess with the minimum absolute distance to the target number, T, wins a prize[1]. The target number is defined as

$$T = Z + Ip(\overline{X} - Z)$$

, where $Z \in [l, h]$ is a constant, $\overline{X}$ is the average of all submitted numbers ($\overline{X} = N^{-1}\sum x_i$), p>0 is a parameter and I={-1,1} is an indicator that determines the strategic environment. For I=1 we have strategic complements and for I=-1 we have strategic substitutes. The parameter $p > 0$ is the (absolute) slope of the best reply function. Note that a negative p is equivalent to a change of I.
The following proposition gives the equilibria of the guessing game.

Proposition 1: *(i)* There exist a unique pure strategy Nash equilibrium $\bar{x} = x_i \,\forall\, i=1,\dots N$ with $\bar{x} = Z$ when I=-1 $\vee$ $0 < p < 1$, and
(ii) there exist multiple equilibria $\bar{x} = x_i \,\forall\, i=1,\dots N$ with $\bar{x} = \{Z, l, h\}$ when I=1 $\wedge$ $p > 1$, and
(iii) with $\bar{x} = [l, h]$ for I=1 $\wedge$ p=1.

Proof:

Since the game is about minimizing the absolute distance to T, the distance must in equilibrium be 0. This can only be ensured when all subjects pick the same number, so the equilibrium is symmetric where $x_i = \overline{X} \equiv x$. In order to win, subjects want to minimize the absolute distance to T.

$$\min_x (T - x)^2 = \min_x \left((1 - Ip)(Z - x)\right)^2 \qquad (1)$$

The first order condition of (1) is:

$$-2(1 - Ip)(Z - x)(1 - Ip) = 0 \qquad (2)$$

Equation (2) has two solutions, namely x=Z and I=1 $\wedge$ p=1. Since we are interested in the global minimum we must have that

[1] If several people are equidistant to the target number, different tie rules can be applied, e.g. that the prize is equally shared among these people. Alternatively, all subjects can be paid according to their distance to the target number.

$$2(Ip-1)^2 > 0 \qquad (3)$$

Thus, if I=1 $\wedge$ p=1 we do not have a unique solution. From (2) and (3) we get that $x=Z$ is always an equilibrium and any x is an equilibrium when I=1 $\wedge$ p=1.

The target T can take on any number, but participants are restricted to guesses $x \in [l,h]$. If $T<l$ then the minimization problem has the solution $x=l$. Similarly if $T>h$ the solution is $x=h$. T only take on numbers $T \notin [l,h]$ if I=1 $\wedge$ p>1. We thus have uniqueness of equilibrium if I=-1 $\vee$ 0 < p < 1. ■

The table below summarizes this and provides an overview of the possible equilibria for given parameter values.

Table 1: Equilibria	Strategic Complements (I=1)	Strategic Substitutes (I=-1)
0 < p < 1	$x = Z$	$x = Z$
p = 1	$x \in [l,h]$ (including $x=Z$)	$x = Z$
p > 1	$x = Z$ $x = h$ $x = l$	$x = Z$

From Table 1 we see that regardless of the strategic environment and p, we always have that $x=Z$ is a pure strategy equilibrium. We furthermore have, that this is the unique pure strategy Nash equilibrium under strategic substitutes. For strategic complements, $x=Z$ is the unique pure strategy Nash equilibrium when 0 < p < 1. For strategic complements there are multiple equilibria for $p \geq 1$.

This formulation of the target function is general, and all existing descriptions of the guessing game are captured as special cases. It follows straightforward that the standard representation of the guessing game from Nagel (1995) is a special case of our more general formulation (let $I=1$ and Z=0).[2]

Theories building on cognitive depths of reasoning have been applied to explain behavior in guessing games, e.g. step-k in Nagel 1995 or the Cognitive Hierachy model (CH) in Camerer, Ho, Chong (2004). Following the model of Nagel (1995), we assume step-0 players to choose x_0. Then we have that a k-level thinker will choose the number x_k given by:

$$x_k = Z + (Ip)^k (x_0 - Z)$$

[2] Another example is the target function from Sutan and Willanger(2009), that arises by setting $c=Zp-Z$, I=1 for complements (in their paper called BCG+) and $h=Z+pZ$, I=– 1 for substitutes (BCG–).

Again, this general formulation captures all existing guessing games, e.g. Nagel (1995) where $x_k = p^k x_0$ for I=1 and Z=0.

We have in this section proposed a general formulation of the target function applied in the guessing game. We then showed existence of an equilibrium at x=Z, and saw that existence does not depend on the strategic environment or the parameter value of p. Then we showed the requirements for uniqueness of the equilibrium. We believe that this formulation is helpful in both designing and analyzing experimental designs and results. It highlights the key properties of the game, and allows for an easy treatment variation of the parameters of interest.

Design

The experiment has two treatments. Common for the two treatments are the parameters p = 2/3, l=0, h=180, Z=90 and N>2. We choose the p value such that we have a unique equilibrium in both treatments. To keep the complexity of the two treatments the same, the equilibrium must be in the middle of the interval. We thus choose an interval such equilibrium and step-level predictions are not prominent numbers.

The two treatments differ in their strategic property, in that we consider an identical setup under strategic complements and strategic substitutes (I=1 and I=-1, respectively). We will denote the two treatments COMPLEMENTS and SUBSTITUTES. Applying the parameter values reveals the following target functions:

$$T^{COMPLEMENTS} = 30 + \frac{2}{3}\overline{X}$$

$$T^{SUBSTITUTES} = 150 - \frac{2}{3}\overline{X}$$

If we let $x^*_{i,C}$ and $x^*_{i,S}$ denote the Nash equilibrium for strategic complements and substitutes, respectively, we have (since 0<p<1) the same unique Nash equilibrium $x^*_{i,C} = x^*_{i,S}$ =Z=90 in both treatments. The slope of the best-reply function is 2/3, where the sign of the slope – negative in COMPLEMENTS and positive in SUBSTITUTES – is the only difference between the treatments. Sonnemans and Tuinstra (2008) show that convergence towards equilibrium in guessing games with strategic complements depend heavily on the p-value, yet we keep p constant across treatments.

Thus, the two treatments represent two different games. Yet, the game is fully symmetric in the sense that for any average choice of others, the distance to the Nash equilibrium is the same in both strategic environments. More specifically, finding the best reply given any expectation of the average choice of the others is equally difficult in the two strategic environments since both reaction functions have the same slope. The only difference is the sign. Thus, from a theoretic point of view, both environments offer the same degree of complexity.

The experiment consists of two rounds. The first round was announced in the hard-copy version of the Danish daily newspaper *Politiken* and on its website. Participation was only possible through their website. Participation was not limited to subscribers of the newspaper, but open to all visitors to the newspapers website.

In the first period, participants played the standard guessing game with l = 0, h = 100 and p = 2/3, and could guess at any real number (allowing for up to nine digits) between 0 and 100 (both included). In total 19,196 participants participated to win the prize of DKK 5000 (at the time approx. $1,000 or € 670). The prize was split equally among the four winners who all submitted the same guess closest to the target number.

In the first round, subjects provided their guess, name, address, e-mail address, phone number and an optional comment. The e-mail address served as the unique identifier, and only one guess per e-mail was allowed. The experiment was presented to the readers as a game and the scientific aim of the competition was not revealed. However, the readers were told that the aim and the result of the exercise would be discussed in the newspaper the following week. The experiment was open to participants for one week. Instructions can be found in the appendix.

In the second round, we used the e-mail address provided in the first round to invite all previous participants to participate again. Thus, only participants of the first round could participate in the second round. Each participant received a personalized e-mail that contained a unique link. By clicking this link, they visited a personalized homepage and were randomly allocated to one of five different treatments[3], each with a prize of DKK 5000 (at the time approx. $1000 or € 670). In addition to a short description of the experiment and the instructions (see appendix), they were informed about their own guess, the average guess and the target number of the first round. The information page also included two links providing additional information (in a pop-up window). One link showed the distribution of guesses in the first round and the other link displayed a newspaper article about the guessing game, with detailed comments of how to think of the game. This description included an intuitive explanation of the Nash equilibrium in plain words. As can be

[3] A total of five treatments were run in the second period. The other three treatments address a different question and are discussed in the companion papers – see Tyran and Øvlisen (2009) and Øvlisen (2009).

seen from the appendix, the instructions of the game included a line indicating the slope of the best response function. Of the 19,196 invited participants, 7,635 participated in the second round.

In the second round, treatment COMPLEMENTS had 1,477 participants and treatment SUBSTITUTES had 1,453 participants. Recall that both treatments differ from the first period in two ways. First, the action set was changed to h = 180. Second, we set Z=90 such that a new Nash equilibrium and target function arose. Subjects were informed that the rules had changed from the previous period in that they could guess a number between 0 and 180 (both included). In COMPLEMENTS, subjects were told that the winner(s) was determined as whoever got closest to 30 plus 2/3 times the average number in their group. Participants in SUBSTITUES were told that the winner(s) was determined as whoever got closest to 150 minus 2/3 times the average of all numbers on their group. The prize in both treatments was DKK 5000, and participants were informed that it would be split equally in case of a tie.

Hypotheses

In this section we derive hypotheses of expected behavior from different theories. We begin with the equilibrium prediction. As we saw in the previous section, in both treatments 90 is a unique pure-strategy Nash equilibrium. This gives us the first hypothesis.

Hypothesis 1: The target number is 90 in both SUBSTITUTES and COMPLEMENTS.

The next hypotheses are based on step-level theory (see Nagel 1995) that has proven a good predictor of behavior in this type of games. Step-level thinking assumes step-0 players to (on average) choose a reference number. In the first period of a guessing game, the reference number has previously been shown to be the middle of the interval. As the game is repeated, the reference number is normally assumed to be the average of the previous period. Since the parameters change from the first to the second period, where the participants face the two treatments, it is not clear what we should expect.

First, suppose subjects regard the changed rules as a new experiment. We would then expect them to assume step-0 players to choose 90. Any k-step-players best reply to a k-1-step belief would yield the equilibrium choice of 90. Thus, the first step-level hypothesis coincides with hypothesis 1.

Next, suppose that subjects consider the second period merely a second period of a repeated guessing game. Then they would assume step-0 players to choose an average of 32.4068, which was the average number in the first period. Step-1 players would expect everyone else to be step-0, and best-reply to that and choose 128 in SUBSTITUTES and 52 in COMPLEMENTS. Similarly, we can derive the predictions for step-2 and step-3 and we get arrive at the next two hypotheses.

Hypothesis 2.1 SUBSTITUTES

We expect spikes at 128, 64 and 107 corresponding to one, two, and three steps of iteration in SUBSTITUTES.

Hypothesis 2.2 COMPLEMENTS

We expect spikes at 52, 64 and 73 corresponding to one, two, and three steps of iteration in COMPLEMENTS.

Note, that the steps in COMPLEMENTS are all below the equilibrium value of 90, whereas every second step in SUBSTITUTES is above. Also, note that all even steps are the same in the two environments. This is a characteristic of the design, and persist regardless of the assumed step-0 play.

Now, in this second period subjects face a different environment. Not only was the target function changed, but also the interval of possible numbers was extended from [0,100] to [0,180]. Øvlisen 2009 show that subjects, when facing an extension of the action set, apply step-level thinking and adjust for the new action set. So let us suppose that subjects accounted for this nominal change in interval, such that they assumed step-0 players to choose the "adjusted average" of last period. The adjusted average is the average, adjusted for the change of interval, such that step-0 players on average would choose 32.4068*1.8=58.3. That gives us:

Hypothesis 3.1 SUBSTITUTES

We expect spikes at 111, 76 and 99 corresponding to one, two, and three steps of iteration in SUBSTITUTES.

Hypothesis 3.2 COMPLEMENTS

We expect spikes at 69, 76 and 81 corresponding to one, two, and three steps of iteration in COMPLEMENTS.

When we compare the predictions in the two strategic environments, we see that the steps in COMPLEMENTS are much closer to one another than the SUBSTITUTES. Consequently, we should expect the overall distribution of choices to be more clustered in COMPLEMENTS whereas guesses in SUBSTITUTES should be more dispersed.

As we saw in the theory section, the complexity of the two treatments is the same. Applying step-level thinking, we would therefore expect the depth of reasoning in the two treatments to be the same as also Willanger and Sutan (2009) finds it.

Hypothesis 4

Depth of reasoning is the same in both treatments

We now turn to the empirical findings.

Results

The distributions of guesses in the two treatments are shown in Figure 1 and Figure 2. In COMPLEMENTS, one participant came with a guess of 75.7 closest to the target of 75.55. In SUBSTITUTES, 4 participants came with a guess of 98 equally close to the target of 97.97 and split the prize. From Figure 1 and 2, it is immediately clear that the two distributions are not similar (Kolmogorov-Smirnov test (p=0.0000), Wilcoxon-Rank-Sum test (p=0.0000)). In both treatments, we see a large spike at the equilibrium value of 90. In COMPLEMENTS, around 6 percent of choices are 90 whereas just below 11 percent in SUBSTITUTES guess 90. The spike at 90 is by far the largest in both treatments. Despite the large spikes at 90, the vast majority of participants in both treatments chose something very different from 90. We can thus reject our Hypothesis 1, and further add that guesses in COMPLEMENTS and SUBSTITUTES do not come from the same underlying distribution.

From hypothesis 2.1 we should expect spikes at 128, 64 and 107 in SUBSTITUES. Looking at Figure 2, we see that this is not the case. We find no sign of a spike at any of the three values, and the hypothesis finds no support. Similarly, we should expect spikes at 52, 64 and 73 in COMPLEMENTS. Here, we do not find any spikes either. Thus it seems that a simple application of step-level theory applying the average of the previous round (with a different action set) as step-0 play is not applied by participants, and hypotheses 2.1 and 2.2 find no support.

The next two sets of hypotheses assume that the subjects apply step-level thinking, but adjust their expectations with the nominal re-scaling of the action set. The predictions from the two sets of hypothesis are very near to each other. In SUBSTITUTES, we find evidence supporting hypothesis 3.1 in that a large spike is present at 111 (and the second largest spike very close at 110), corresponding to level-1 play. At the remaining predicted values, we find an above-average representation of guesses, but nothing that we would classify as a spike.

For COMPLEMENTS we similarly find that the second largest spike with 3.5 per cent of the guesses is at 69, and the largest spike is very close at 70, as predicted by hypothesis 3.2. Thus, also in this treatment we that step-1 thinking is consistent with the spikes found. For the remaining step-level predictions we find an above-average percentage of guesses.

The average guess in COMPLEMENTS was 68.33 and the target number was 75.55. In SUBSTITUTES, the average was 78.05 and the target was 97.97. So, on average subjects guessed higher numbers in SUBSITUTES than in COMPLEMENTS. We also see that guesses – on average – were closer to the equilibrium in SUBSITUTES. If we compute the absolute differences to 90 in both treatments, we find – using a Wilcoxon Rank-Sum test – that not only are the differences not the same in the two treatments (z=4.992, p=0.0000), moreover the differences are larger in COMPLEMENTS than in SUBSTITUTES. So, participants' guesses are closer to the equilibrium prediction in SUBSTITUTES than in COMPLEMENTS.

Note that the second-largest spike is to the left of the equilibrium for COMPLEMENTS and to the right for SUBSITUTES, and that both correspond to adjusted level-1 players. This illustrates an effect that describes the overall distributions. In COMPLEMENTS we see that most the guesses are below the equilibrium value, and that there is strong inertia. This is implicitly expected from the hypothesis 3.1, since all steps are below 90. In SUBSTITUTES the average guess is also below 90. However, we find more observations above 90 in this treatment. This follows empirical finding is in line with the adjusted step-level models.

Figure 1: Distribution of Guesses - COMPLEMENTS

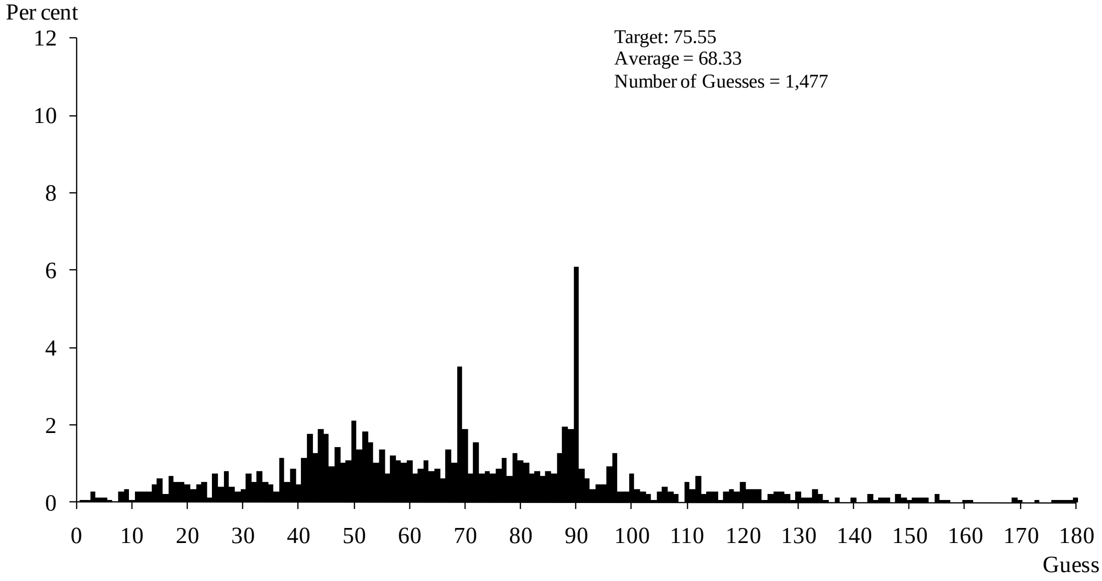
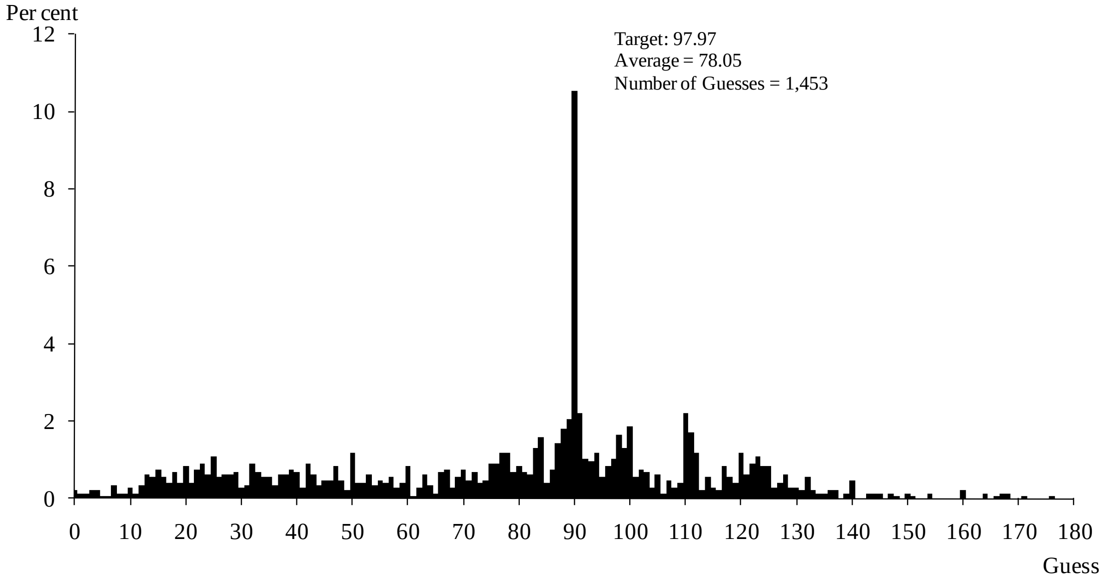


Figure 2: Distribution of Choices - SUBSTITUTES



Furthermore we see that 77 per cent of the choices in COMPLEMENTS are below the equilibrium, whereas only 55 per cent in SUBSTITUTES are below. Similarly, we find that only 19 per cent of choices are above equilibrium in COMPLEMENTS, whereas 39 per cent of choices are above equilibrium in SUBSTITUTES.

Since the steps are much closer to one another in COMPLEMENTS than in SUBSTITUTES, it is reasonable to expect the overall distribution to be more clustered in COMPLEMENTS than in SUBSTITUTES. This view is supported by the data, in that e.g. the standard deviation of guesses in COMPLEMENTS (31.2) is smaller than that of SUBSTITUTES (34.9).

Now, let's turn to look at the dominated choices. In COMPLEMENTS, 10 per cent of choices are in the range of dominated choices (<30), whereas only 1 per cent are within this range for SUBSTITUTES (>150). However, 13 per cent of choices in SUBSTITUTES are <30 and 1 per cent of choices in COMPLEMENTS are >150. So, the share of choices below 30 and above 150 in the two treatments is essentially the same. Thus, the choices in this interval are given by subjects not responding to the strategic environment of the treatment, and we can interpret these guesses as noise play (or equivalently step-0 players). The large number of guesses in the very low range of possible numbers (<30) is due to a very efficient anchoring on low numbers. The anchoring might work in multiple ways. Either, subjects play purely adaptive to the first round, and thus chooses a number below that of the average number of the first round, which essentially for almost all steps is below 30. Yet another explanation consistent with the data could be that subjects were confused about the new rules of the game. It lies in the nature of running an internet experiment that some control is lost. Thus, if subjects believe they are facing the old environment – although the instructions clearly state otherwise – we should expect to find low guesses. However, the important thing is that subjects do not hold false beliefs of how the others will adjust to the new strategic environment. If this was the case, we should not expect to find the same fraction of guesses below 30 in SUBSTITUTES and COMPLEMENTS.

In both treatments, we see many subjects choosing the equilibrium value of 90 in the second period. Now, we turn to see what they did in the first period. Figure A2 in the appendix show the distribution of guesses in the first period only for the subjects choosing 90 in SUBSTITUTES and COMPLEMENTS. The distribution of these choices are not the same for the two treatments (Mann Whitney, p=0.0038), with the main distinctions being more equilibrium guesses (0) and step-2 guesses (22) for subjects choosing 90 in SUBSTITUTES, and more step-1 (33) guesses for subjects choosing 90 in COMPLEMENTSS. We conclude that behavior in the first round for the equilibrium players of the second round, was not the same.

In this section we saw that the spikes we observe in the data are consistent with step-level thinking, using the average of the previous period, adjusted by the relative increase of the action set, as step-0 play. The two treatments exhibit equally many dominated choices, but in SUBSTITUTES guesses are on average higher, less clustered and closer to equilibrium. Yet, the difference to the target value seem not different in the two treatments.

Step-level thinking

In the previous section we found that step-level thinking, assuming that step-0 players adjust the average of the previous period with the relative extension of the action set, explained the spikes observed in the data. Under this assumption we find that the average number chosen in COMPLEMENTS (68.33) is closest to step-1 play (68.9), whereas the average in SUBSTITUTES (78.05) is closest to step-2 play (75.91). The average action of the two treatments is thus characterized by different step-levels.

In Table 2 below, the frequency of choices for the predicted steps using the above assumption is depicted. The table shows two levels of tolerance for the distance to the prediction.

TABLE 2	COMPLEMENTS	SUBSTITUTES
+/- 1 from prediction		
Step-1	0.051	0.029
Step-2	0.017	0.020
Step-3	0.022	0.034
+/- 5 from prediction		
Step-1	0.131	0.075
Step-2	0.097	0.077
Step-3	0.093	0.097

For both levels of tolerance, we see fewer step-1 players in SUBSTITUTES than in COMPLEMENTS, whereas the fraction of step-2 and step-3 players seem to be the same.

In Table 3 below, we classify subjects according to the minimum difference to the step-prediction for step-1 through step-4 (and above – note that this includes equilibrium play of 90 in both treatments).[4]

TABLE 3	COMPLEMENTS	SUBSTITUTES
Step-1	0.58	0.21
Step-2	0.05	0.42
Step-3	0.04	0.13
Step-4 or higher	0.33	0.24

[4] We cannot measure the exact step-level in each treatment, using e.g the CH-model. Whereas it is possible to calculate a non-integer step for a given guess in COMPLEMENTS, it is not possible in SUBSTITUTES. The reason is simply, that any guess between step-0 and step-1 can be explained by any level-*k*, with *k*>2. We therefore let subjects be classified by the step with the minimal absolute distance to their guess.

The distribution of steps are clearly different between the two treatments (Mann-Whitney, p=0.0000). We also see, that the main difference is that in SUBSTITUTES there are fewer step-1 players, and in SUBSTITUTES there are more step-2 players. This is not due to prominent numbers or low anchoring, since the value of Step-2 play is the same in both treatments, namely 76. We can thus reject hypothesis 4, in that the depth of reasoning is higher in SUBSTITUTES than in COMPLEMENTS.

Conclusion

In this paper, we have introduced a new general formulation of the target function. We have proved existence of equilibrium, and derived the conditions for uniqueness. We argue, that this target function has many advantages both in designing experiments and analyzing experimental results. All treatment variations can easily be done by changing just one parameter.

We report results from a two period large-scale internet experiment in which we do exactly that. We vary the strategic property of the game, and have one treatment with strategic complements and one with strategic substitutes. Using a rich data set with more than 1,400 subjects in each treatment, we find – similar to Willanger and Sutan (2009) – that the average guess in COMPLEMENTS is lower and further away from equilibrium than in SUBSTITUTES. Contrary to Willanger and Sutan (2009), our results suggests that subjects act consistent with higher steps of reasoning in SUBSTITUES. This is in line with previous experimental results in other games (e.g. Fehr and Tyran (2008)).

References

Camerer, C.F. and Fehr, E. (2006): When does "Economic Man" Dominate Social Behavior?. *Science* 311, 47-52.

Camerer, C.F., Ho, T-H. and Chong, J-K. (2004): A Cognitive Hierachy Model of Games. *Quarterly Journal of Economics* 119, 861-898.

Davis, D. (2008): Do Strategic Substitutes Make Better Markets? A Comparison of Bertrand and Cournot Markets. Working paper.

Fehr, E. and Tyran, J.R.(2008): Limited Rationality and Strategic Interaction. The Impact of the Strategic Environment on Nominal Inertia. *Econometrica*, 76(2): 353-94.

Fudenberg, D. and Tirole, J (1984): The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look. *American Economic Review* 74(2), 361-366.

Haltiwanger, J. C. and Waldman, M. (1985): Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity, *American Economic Review* 75(3), 326-340.

Moulin, H. (1986): Game Theory for the Social Sciences, 2nd and revised edition. New York University Press.

Nagel, R. (1995): Unraveling in Guessing Games: An Experimental Study. *American Economic Review* 85, 1313-26

Suetens, S. and Potters, J. (2007): Bertrand colludes more than Cournot. *Experimental Economics 10,* 71-77

Suetens, S. and Potters, J. (2009, *forthcoming*): Cooperation in experimental games of strategic complements and substitutes. *Review of Economic Studies*

Sonnemans, J. and Tuinstra, J. (2008): Positive Expectations Feedback Experiments and Number Guessing Games as Models of Financial Markets. Working Paper.

Tyran, J-R and Øvlisen, F.R.(2009): Making an Educated Guess. Working Paper.

Øvlisen, F.R.(2009): Step-level Thinking and Changes in the Action Set. Working Paper.

Willinger, M and Sutan, A.(2009): Guessing with negative feedback: An experiment. *Journal of Economic Dynamics & Control* 33, 1123-1133.

Appendix

Figure A1: Distribution of choices adjusted by 1.8.

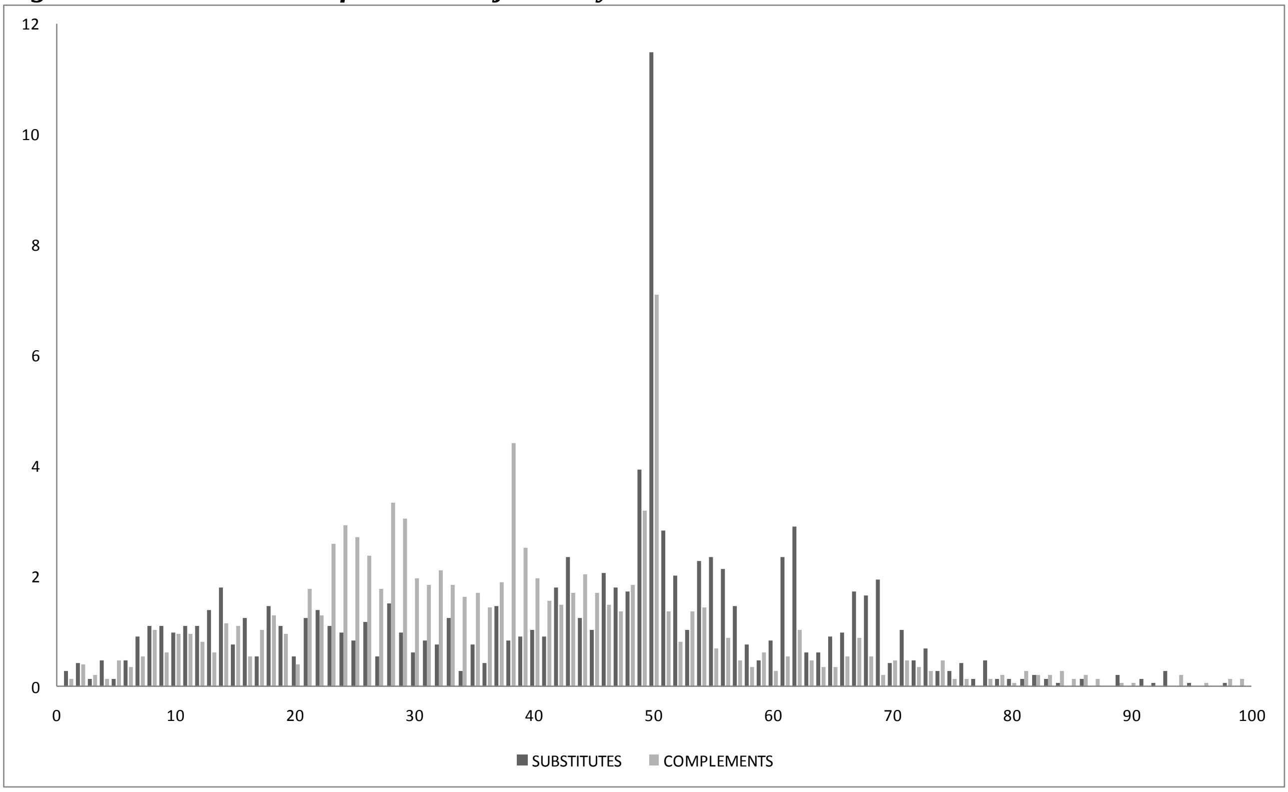
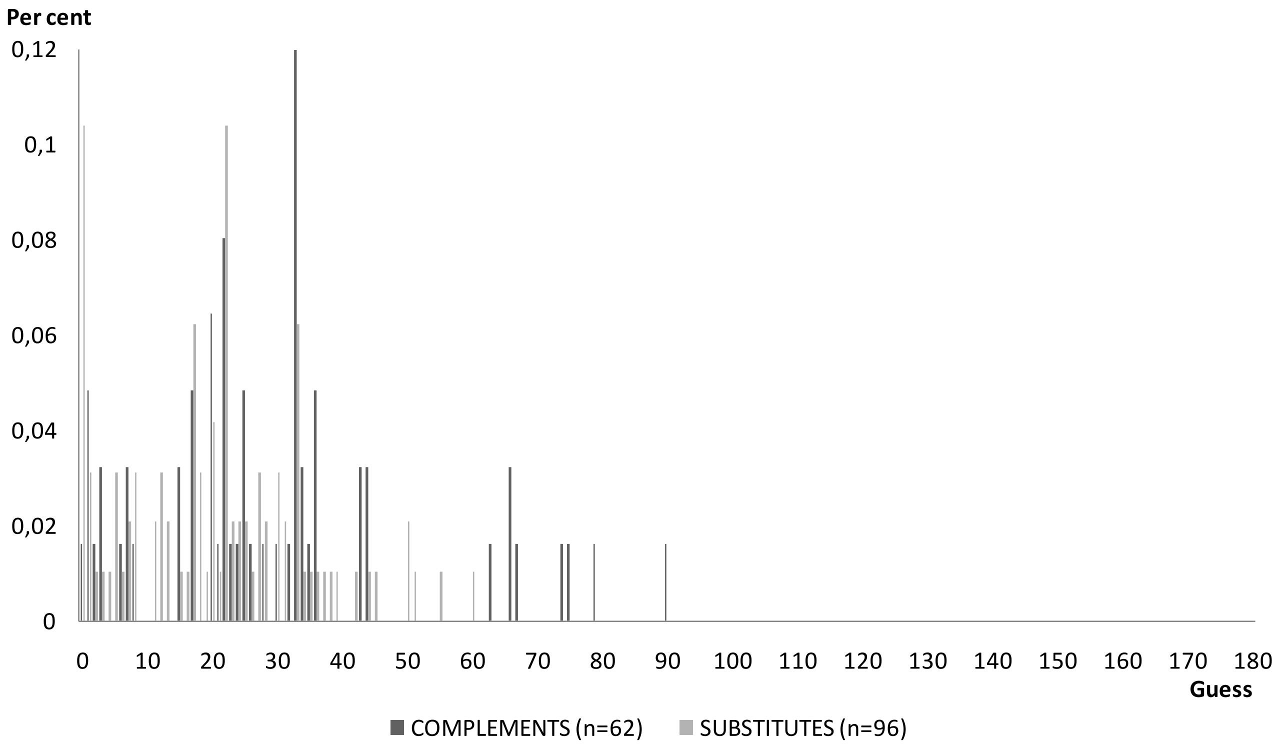


Figure A2: Distribution of choices in the first round for subjects choosing 90 in the second round.



Instructions

The original instructions were in Danish. Below follows an English translation.

The first period:

The participants were told the rules, and that the purpose of the competition would be revealed four days after the competition closed for new entries, where the newspaper ran an article about the game and about experimental economics.
More specifically, the participants were told:
Headline: "Guess a number and win 5,000 kroner"
Politiken gives you the opportunity to win 5,000 kroner. All you have to do is to guess a number.

You must submit a guess according to these rules:

- Pick a number between 0 and 100 (both included), which you believe will be the closest to two thirds of the average of all submitted.
- Your guess does not have to be an integer
- You must submit your guess by Sonday the 18. September [2005] at 12:00.
- The winner is the reader closest to two thirds of the average off all the submitted numbers.
- If more than one participant are equally close, the prize will be split among them.
- The winner – and the purpose of the competition – will be revealed in Politiken and at politiken.dk on Thursday den 22. September [2005].
- Employees of JP/Politikens Hus can not participate in the competition.

Submit your guess (between 0 and 100 – both included): __________
Short reasoning for your guess (not mandatory): ___________________________
Name:_____________
Address: __________________
E-mail: ____________________
Phone: ______________________

< Submit your guess >

The second period:

The instructions consist of two parts. A common part, that was the same for all treatments and a treatment specific part, explaining the treatment specific rules.

Common part:

In September 2005, you participated in “Guess a Number” at the webpage of Politiken, with the possibility of winning DKK 5,000.
Now, we again offer you the opportunity to win DKK 5,000.
This time, only the 19,196 participants of the first period can participate. The participants of this round will be randomly divided into five different groups. In each group, a prize of DKK 5,000 will be awarded (such that the total prize pool is DKK 25,000).
Just as in the first round, the rules are that:
Your guess does not have to be an integer
If more than one participant are equally close, the prize will be split among them.
You must submit your guess by Wednesday the 21. March 2007 at 12:00.

In the first round the average of all guesses were 32.407.
The winning number (which was 2/3 of the average of all guesses) was 21.605
Your guess in the first round was [the participants guess in the first period was displayed here].
You can read more about the first competition from September 2005 in the *newspaper article from Politiken, 22. September 2005* [link opens a new window]. In addition, you can see a graphic distribution of guesses in the first round here [link opens a new window].

Treatment Specific Part:

COMPLEMENTS: In this round, the rules are in your group slightly different than in the first round. Pick a number between 0 and 180 (both included) which you believe will be the closest possible to 30 plus two thirds of the average of all submitted guesses. The winner is the participant who gets closest to 30 plus two thirds of the average of all the submitted guesses.

The lower you expect the others to choose on average, the lower you should choose to win.
Thus, you should choose 30 + 2/3 of your expectation of the others’ average guess.

SUBSTITUTES
In this round, the rules are in your group slightly different than in the first round.
Pick a number between 0 and 180 (both included) which you believe will be the closest possible to 150 minus two thirds of the average of all submitted guesses. The winner is the participant who gets closest to 150 minus two thirds of the average of all the submitted guesses.

The lower you expect the others to choose on average, the higher you should choose to win.
Thus, you should choose 150 – 2/3 of your expectation of the others’ average guess.

Common part (continued): Submit your guess: __ Comment (optional): __

Collective Action and Coordination

Collective Action and Coordination

Jean-Robert Tyran and Frederik Roose Øvlisen*

May 2009

We experimentally study the effect of collective action on coordination in games with pareto-rankable equilibria. We use a game in which coordination failure looms large absent collective action, i.e. when players make decentralized choices, and show that collective action is surprisingly ineffective in improving coordination. Study 1 shows that majority voting does not select a pareto-dominant equilibrium, does not prevent lock-in in a inferior equilibrium, and that, once coordination failed, it does not break lock-in despite the possibility to collectively "jump out of the trap". Study 2 uses a setting in which coordination failure again looms large in decentralized play, but collective action *per se* is quite effective in improving coordination. Yet, we show that collective action has a limited coordinating effect by "setting a precedent", i.e. in providing assurance by coordinating expectations on a superior equilibrium in subsequent decentralized play.

Keywords: Collective action, voting, coordination failure, equilibrium selection, strategic uncertainty, money illusion, experiment

JEL-codes: C9, C72, E52

* Both authors are at the University of Copenhagen, Department of Economics, Studiestræde 6, DK-1455 Copenhagen. e-mails: Jean-Robert.Tyran@econ.ku.dk and Frederik.Oevlisen@econ.ku.dk.
We gratefully acknowledge financial support from the University of Copenhagen and FSE. We are grateful for helpful comments by Simon Gächter, Dirk Engelmann and Sigrid Suetens and we thank Christian Menck and Karen Winnie Larsen for effective research assistance.

Introduction

The fundamental problem of coordinating economic activities can in essence be captured in coordination games. In such games, players typically have incentives to make the same choices as others do. Examples are to drive on the same side of the street, use the same language, measurement unit or technical standard as others do. Such problems are in practice often solved by some form of collective choice, authoritarian decree, custom, or historical precedent. Once players are coordinated on the same action, outcomes tend to be stable if no single individual has an incentive to deviate from what has become the norm. Thus, players are in an equilibrium state. Yet, a defining feature of coordination games is the existence of multiple equilibria. In some cases, it is a matter of indifference which equilibrium players coordinate on. For example, it may not matter much whether cars drive on the left or the right side of the street, or whether engineers use centimeters or inches to measure distances. In other cases, equilibria may be pareto-rankable. In these cases, one equilibrium is better for everyone than another, yet players may be “trapped” in an inferior equilibrium. For example, a country may be stuck in poverty (a “development trap”) because investors are reluctant to invest simply because the level of other investments are low, or citizens may evade taxes because taxation is high as a result of common evasion (see Cooper 1999 or Crawford 1997 for overviews, Camerer 2003: Ch. 7 or Tyran 2007a for further examples of coordination failure).

In this paper, we study the effect of collective action in improving coordination in games with pareto-rankable equilibria. More specifically, we study if collective action prevents coordination failure and if it can break lock-in once coordination failure occurred.

Collective action may prevent coordination failure by selecting the pareto-dominant equilibrium before the game is played in a decentralized way. The prevention effect is particularly strong if decentralized play would lead to eventual coordination on a dominated equilibrium but collective action selects the pareto-dominant equilibrium right away. But collective action may also have efficiency-increasing effects if collective action and decentralized play select the same equilibrium, simply by avoiding disequilibrium play. In any case, efficiency increases more if collective action selects an equilibrium and if collective action sets an effective “precedent”. A precedent is effective if the (equilibrium) outcome of collective action coordinates expectations of players in subsequent decentralized play on the equilibrium selected by collective action.

Collective action can break lock-in once it occurs and enable players to collectively "jump" from an inferior to a superior equilibrium. This "curing effect" of collective action is particularly strong if (at least some of the) voters are aware that jumping to the pareto-superior equilibrium is indeed in their interest and if collective action is setting an effective precedent. In this case, setting an effective precedent means that the precedent set by collective action dominates the precedent set by previous decentralized play. In contrast, a "curing effect" of collective action is absent if players' experience of equilibrium play in the decentralized game sets a precedent for collective action. In this case, players would simply remain locked in the inferior equilibrium. The "curing effect" would also be of limited relevance if collective action succeeds in "jumping" to the superior equilibrium but players who are not assured quickly revert back to playing the inferior equilibrium.

Intuitively, collective action ought to be very effective if all players are fully aware that everyone is indeed better off in the pareto-dominant equilibrium and if switching equilibria is costless. However, collective action, for example in the guise of majority voting, can also be expected to be quite effective in games in which some players are in doubt or confused about payoffs. As long as the confused players are in minority, collective action effectively prevents coordination failure by selecting a superior equilibrium. If confused players adapt to the precedent set by collective action in subsequent decentralized play, perfect coordination can be achieved right away and coordination failure be prevented through collective action. In this sense, collective action is more robust to a small amount of confusion than decentralized play and can therefore be expected to improve coordination.

Another reason why collective action can be expected to improve coordination is that it eliminates or at least reduces strategic uncertainty. Strategic uncertainty may induce coordination failure even if most players are not confused about payoffs and the ranking of equilibria. The reason is that non-confused players may find it optimal not choose the action implied by the pareto-superior equilibrium in decentralized play for fear that others do not play it. However, non-confused players may find it optimal to vote for implementing the pareto-superior equilibrium because expectations matter less or are outright irrelevant in voting (We use two types of collective action in our experiment. In majority voting, the relevance of strategic uncertainty is reduced, while with a random dictator procedure it is totally eliminated).

We test the effect of collective action on coordination in games in which at least some players are known to be uncertain or outright mistaken about the true payoffs in the game, and therefore about the ranking of equilibria. We think that using such games provides a more demanding, and perhaps more realistic, test for the ability of collective action to improve coordination. We use a coordination game developed by Fehr and Tyran (2007, henceforth FT) who present payoffs to players in nominal terms but actual payments to participants are in real terms. If something is "lost in translation", to use an expression of Akerlof and Shiller (2009: 47), in going from nominal to real payoffs, i.e. if players are prone to money illusion, some players will be confused about the ranking of equilibria, and the non-confused players' beliefs are perturbed. While our paper does not argue that money illusion is a relevant aspect of economic choices, and does not focus on the role of money illusion in causing coordination failure (see Fehr and Tyran 2007 on that issue), we note that recent literature suggests money illusion affects how people think (Shafir, Diamond and Tversky 1997, Weber et al. 2009) and how people behave in various strategic situations and markets (e.g. Noussair, Richter and Tyran 2008 for financial markets, Brunnermeier and Juillard 2008 for housing markets, see Akerlof and Shiller 2009: 41-50, or Tyran 2007b for a survey).

Study 1 introduces majority voting for inexperienced (INEXP) and experienced (EXP) voters into the game by FT which is known to produce massive coordination failure in the sense that decentralized play eventually converges to the pareto-inferior equilibrium. In EXP, voting takes place after 30 periods of decentralized play, when most players have experienced coordination failure and lock-in, to test if voting helps to break the lock-in. In INEXP, voting takes place in period 1, i.e. before players have experienced the coordination failure, to test if it selects the superior equilibrium and prevents coordination failure in the 30 periods of subsequent decentralized play. The results are surprisingly clear. Neither do voters select the superior equilibrium in period 1, nor do they manage to jump out of the trap after having experienced lock-in. Since voters do not succeed in selecting the superior equilibrium in period 1, collective action also does not coordinate expectations in subsequent decentralized play by setting a "precedent".

To further investigate the "precedent" effect, we modified the game by FT such that collective action is likely to be more successful in selecting the dominant equilibrium while maintaining the property from study 1 that coordination failure looms large in decentralized play. Important modifications concern a reduced strategy space and the use of a random dictator voting procedure, which totally eliminates strategic uncertainty for voters, rather than majority voting.[1]

Study 2 has 3 treatments.[2] In treatment NoVote, the coordination game is played in the usual decentralized way for 30 periods. NoVote serves as a benchmark against which the efficiency-improving effect of voting is measured. Treatment RepeatVote combines voting of inexperienced and experienced voters and allows for learning effects in voting. In this treatment, players vote and then play 9 periods of the decentralized game, and this sequence is repeated 3 times. Treatment AllVote has collective action in all 30 periods. We find that coordination failure indeed looms large in decentralized play and is almost perfectly eliminated in AllVote, as expected. We find that voting in RepeatVote is quite effective, but surprisingly does not coordinate expectations by setting a precedent.

Two striking examples of regime switches may serve to illustrate how difficult it is in real life to solve coordination problems that seem straightforward from a theoretical perspective.[3] The Swiss government planned to introduce daylight saving time (DST) simultaneously, i.e. coordinated with, the neighboring states in 1980. However, the decision was subject to popular referendum in which the introduction of DST was rejected by a majority of voters. As a result, Switzerland was a "time island" in the summer 1980,

---

[1] Democratic collective action is also quite likely to induce cooperation if voting requires unanimity. The drawback here is that unanimity is unlikely to obtain (even in a common interest situation) in the absence of communication. Yet, as Feri, Irlenbusch and Sutter (2009) show in a series of coordination games, the combination of communication and unanimity (reached by informal agreement rather than through a formal voting rule) does improve coordination in most cases, but the success is far from perfect.

[2] Both studies were run using the program z-Tree (Fischbacher 2007). In total, 167 undergraduate students (55 for study 1, 112 for study 2) from all faculties were recruited using ORSEE (Greiner, 2004). Study 1 lasted about 90 minutes and average earnings were about €17, study 2 lasted about 30 minutes and earnings were about €12.

[3] Mueller (2003: 17) suggests that coordination games should be particularly easy to solve by collective action: "if all problems caused by social interaction were as simple as deciding on which side of the road everyone should drive, one might well imagine that it would be possible to do away with the state. But, alas, this is not the case …"

having a different time zone than its neighbors. Due to the high cost of such discoordination, the government overruled the referendum by authoritarian decree and introduced DST in 1981 anyway. A popular referendum launched in response to that decision with the purpose to abolish DST failed in 1981 and Switzerland has had DST ever since. A second illustrative example is the switch from left-hand to right-hand driving (RHD) in Sweden. In 1953, government proposed a switch to RHD mostly motivated by the wish to coordinate with neighboring countries with RHD, but the idea was controversial. A referendum was held in 1955 in which RHD was rejected by a vast majority (82.9%) of voters. The government decided to implement RHD anyway in 1963, and the switch was implemented in 1967, 14 years after the first government proposal. Sweden has had RHD ever since.

These examples can be thought of as textbook-cases of pure coordination, i.e. games in which equilibria are not ranked which makes popular resistance in both examples difficult to understand. A potential explanation for resistance is that voters' expectations have been so strongly coordinated by the historical precedent that they were not assured that collective action will set an effective precedent and disequilibrium may prevail after a regime switch. Alternatively, one may also argue that the equilibria are pareto-ranked – as the proponents of daylight saving time and right-hand driving argued in the referendum campaigns.[4] We are uncertain about which interpretation is correct, and we believe that this was also true of many voters at the time. That is, we believe that not all voters had a clear idea of the payoffs or the ranking of the two equilibria when they made their choices. Thus, the examples serve to highlight that collective action in the guise of majority voting may or may not lead to coordination when payoffs are fuzzy.

Given that collective action is an intuitive remedy to improve coordination and given the difficulties inherent in using field data to investigate the issue, it is surprising that only few experimental studies investigating collective action and coordination are available in the literature.[5] Examples are Capra et al. 2009 who study voting to escape poverty traps, Cabrales, Nagel and Rodriguez Mora (2006) who study voting on redistribution and

---

[4] Another interpretation is that equilibria are not ranked and indeed have the same equilibrium payoffs, but that switching itself is costly. Indeed the Swedish parliamentary service estimated the cost of switching of to RHD to several hundred million Swedish crowns in 1967.

[5] In contrast, collective action has been extensively studied in its ability to solve cooperation problems, e.g. Dal Bo, Foster and Putterman 2008, Noussair and Tan 2009, or Tyran and Feld 2005.

subsequent effort choices. Feri, Irlenbusch and Sutter (2009) do not find, among other things, in a replication of the weakest link game (van Huyck et al. 1990) that voting or the random dictator procedure improves coordination or increases payoffs, provided that players had opportunity to communicate before decentralized and collective action.

Collective action by decree seems to have mixed effects. Sefton and Yava (1996) find in a game with two pareto-rankable equilibria that the introduction of an outside authority (a third player) threatening to "regulate" play had little effect on outcomes. In contrast, Duffy and Kim (2005) find an external authority to be quite effective in a game where players chose to be peasants or predators and to produce or defend.

2. Does collective action prevent or break coordination failure? -- Study 1

Study 1 investigates if collective action improves coordination in a game in which coordination failure looms large absent collective action. Study 1 has two treatments which introduce collective action in the guise of majority voting either before participants experience coordination failure (INEXP) or after they have experienced coordination failure (EXP). Treatment INEXP serves to test if collective action prevents coordination failure by selecting the superior equilibrium from the start, treatment EXP serves to test if it can break lock-in if it occurred. In the periods without voting, both treatments use the same parameters and procedures as Fehr and Tyran (2007, henceforth FT).

FT investigate a coordination game in which n participants are in the role of firms who simultaneously choose integer prices from 1 to 30 and the payoff of player i depends on the price i chooses and the average price chosen by the other ($-i$) players. FT study if money illusion causes coordination failure. They induce money illusion by representing payoffs in either real or nominal terms. In the nominal representation (their treatment NH), the payoff table shows nominal payoffs Π which equal real payoffs π after "deflationing", i.e. after division by the average price of other firms P_{-i}. That is, $\pi_i = \Pi_i / P_{-i}$.

The coordination game has 3 pure-strategy equilibria. In equilibrium A, all players choose $p_A = 4$ and the payoff is $\pi_A = 28$ for each player, equilibrium B prevails at prices of $p_B = 10$ and the real payoff is $\pi_B = 5$ each, and equilibrium C, at prices $p_C = 27$, has a payoff of $\pi_C = 21$ for all players. The corresponding nominal payoffs are $\Pi_A = 112$, $\Pi_B =$

50, and Π_C = 567 (see the nominal payoff table C1).[6] That is, in the nominal representation, the coordination game exhibits a tension between real payoff dominance – equilibrium A pareto-dominates the other equilibria in real terms – and nominal payoff dominance – equilibrium C has higher nominal payoff than A. If players take nominal payoffs as a proxy for real payoffs, they may fail to coordinate on the efficient equilibrium A. FT find that money illusion causes coordination failure since coordination is close to perfect in the real representation but looms large in the nominal representation.[7] We are not concerned with the cause of coordination failure in this paper but with the effectiveness of collective action in improving coordination. We therefore use FT's game in the nominal representation as a workhorse since coordination failure looms large in this case.

2.1. Voting *ex post* to break lock-in (treatment EXP)

Treatment EXP has 3 phases. All phases use payoff table C2 (see appendix) showing nominal payoffs. Phase 1 is a replication of treatment NH in FT in which participants play the pricing game in groups of n = 5 for 30 periods. The game is played in a decentralized manner, i.e. each player independently and simultaneously picks a price. In decentralized play, the coordination problem has 2 aspects: players need to coordinate on the *same* action and players need to coordinate on the *best* action. To illustrate that the coordination problem potentially involves substantial strategic uncertainty[8], we note that the game has 900 potential outcomes for a player in total of which 27 symmetric outcomes in which players are coordinated on the same action, 3 symmetric equilibria but only 1 "best" equilibrium.

Phase 2 of EXP is a short-cut to majority voting in which all participants simultaneously vote for a price from 1 to 30 and the median choice is implemented for the entire group. Thus, majority voting by definition induces coordinated actions in the sense that all play the same action but the median voter's choices may fall on a disequilibrium outcome

---

[6] In treatment NH of FT, subjects received written instructions explaining how to calculate real payoffs from nominal payoffs. At the end of each period, players are informed about the implemented prices and their real payoffs π_i.

[7] FT devised the game as a test-bed to study the effects of money illusion on coordination failure. They do not claim that their game captures common features of oligopolistic pricing. For example, multiple equilibria do not necessarily occur in pricing games and, if they do, a low-price equilibrium does not necessarily pareto-dominate other equilibria. However, the authors note that such a situation is possible in principle.

[8] See Heinemann, Nagel and Ockenfels (2009) for an approach to measure strategic uncertainty.

or on an inferior equilibrium. The purpose of introducing collective action in phase 2 is twofold: first, voting provides an opportunity to escape lock-in on an inferior equilibrium if it occurred and second, voting serves as a coordination device if no coordination occurred because collective action induces all group members to set the same price by definition. Thus, majority voting potentially eliminates coordination failure if voters are aware that they are "trapped" in an inferior equilibrium. In addition, collective action allows a group to coordinate on a superior equilibrium even if some players (a minority) is not aware of its existence.

Phase 3 of EXP is the same as phase 1 except that the game is now only played for 9 rather than 30 periods. Phase 3 serves to evaluate if successful collective action fosters coordination when the game is again played in a decentralized manner by setting a precedent, i.e. by coordinating expectations. To test, we compare coordination before and after the vote.

Figure 1: Average group prices in decentralized play and in collective action for experienced voters (EXP, vote in period 31)

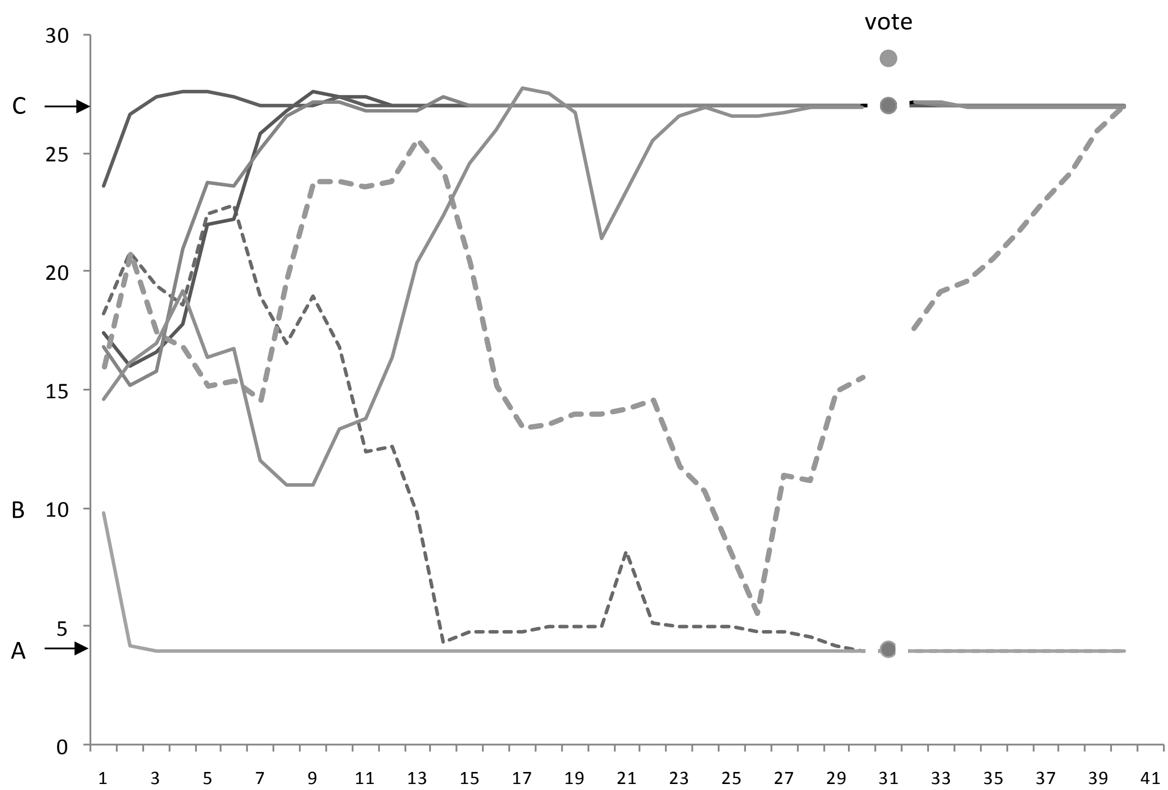


Figure 1 shows the main results for EXP. Six groups start out at average prices considerably above P_B = 10, 1 group just below. The group that starts below P_B almost instantaneously converges to the pareto-superior equilibrium A. Of the 3 groups starting above P_B, 3 groups converge rather quickly and 1 group rather slowly to the inferior equilibrium C. The thin dotted line shows average prices for a group that started above P_B and converged towards C for the first 6 periods but then converged to A and finally coordinated on A in the last period before the vote. The thick dotted line in figure 1 shows a group that started at prices above P_B, first converged to C up to period 13, then to A for the next 13 periods, then again to C for the next 4 periods, and has a disequilibrium average price of 15.6 in the last period before the vote. Thus, in the period before the vote, 4 out of 7 groups are coordinated on the inefficient equilibrium C with $\pi_C = 21$, 2 groups are coordinated on the dominant equilibrium A with $\pi_A = 28$, and 1 group is uncoordinated (average payoff over all groups is 20).

Best-reply dynamics or ficticious play induce convergence to the pareto-dominant equilibrium A if the average price happens to be below the unstable equilibrium B in the first period, but to the pareto-dominated equilibrium C if the group happens to start out at prices above B. The logic of the convergence process appears to be similar to what has been observed in the “continental divide game” (Van Huyck, Battalio and Cook 1997). In this game, the intermediate equilibrium can be thought of as a kind of “watershed” (or separatrix).[9]

In study 1, equilibria A and C are stable while B is unstable. Once players coordinate on equilibrium C, groups are “trapped” because small unilateral deviations from equilibrium action p_C do not provide incentives for others – even if they are aware of coordination failure – to move away from equilibrium C. The question is thus if the introduction of the majority vote in period 31 of EXP served to break lock-in. We find that collective action had essentially no impact.

---

[9] However, note that one group which eventually coordinated on equilibrium A in figure 1 started out at a price of above P_B. Thus, best-reply dynamics are not the full story behind the convergence process. See also Feri, Irlenbusch and Sutter (2009) for a replication in which either 5 individuals or 5 groups of 3 players play the game. In the teams treatment, individuals communicate in free form via electronic chat are requested to somehow unanimously agree on a joint action (no formal procedure is imposed). The authors find that groups who communicate in free form and agree earn significantly higher (22.5 percent) payoffs on average than individuals who do not communicate. However, coordination failure looms large both conditions. “Perfect coordination”, i.e. all players choosing the same equilibrium is only reached in 4 percent of the periods in either case. In a replication of the weakest link game with 7 pareto-rankable equilibria of van Huyck et al (1990), the authors let teams first communicate and then vote (median vote) or randomly assign one member in each group to be the dictator. They do not find that voting or the random dictator procedure improves coordination or increases payoffs.

All groups that coordinated on some equilibrium before the vote selected the same equilibrium through collective action. In particular, the 4 groups that coordinated on equilibrium C did not "jump out of the trap" but remained in the dominated equilibrium C. We do see some attempts to break lock-in in these groups. In fact, 5 voters who played p_C = 27 in period 30, vote for A in period 31. But since these voters are dispersed across different groups, they have no impact on the outcome of the vote. Thus, collective action was unsuccessful in breaking lock-in. On the other hand, it is comforting to note that collective action was not a disequilibrating force or otherwise counterproductive either. The 2 groups that successfully coordinated on A before the vote remained in A.

In the uncoordinated group (see thick dotted line in figure 1) one voter opts for A, one for C but 3 voters opt for a high disequilibrium price of 29. As a result, the collective choice (the median vote) was for a disequilibrium price of 29 in this group. Interestingly, the collective action does not serve as a "precedent" in this case. That is, the vote does not coordinate the group's expectations or actions in the period following the vote. Instead, the group just continues on the convergence path towards equilibrium C and eventually (by period 40) coordinates on C. Overall, behavior in the decentralized pricing game in the period before and after the vote is not significantly different (p = 0.481, Wilcoxon signed-rank test, henceforth WSR).

In EXP, we essentially replicate FT's finding of massive coordination failure and in particular the tendency of groups to coordinate on the equilibrium C which "looks" attractive judging by nominal payoffs ($\Pi_C > \Pi_A$) but which is dominated by A in terms of real payoffs ($\pi_A > \pi_C$). FT find that groups that coordinate on C tend to remain locked-in in decentralized play. FT speculate that strategic properties may inhibit learning to overcome coordination failure, perhaps because players, once they are locked-in, search the vast payoff table only locally for optimal solutions. We surprisingly find that majority voting does not induce any of the "trapped" groups to "jump" to the superior equilibrium A. We thus provide support for FT's conjecture that strategic properties may prevent learning. Interestingly, there seems to have been a reverse "precedent" effect in the sense that voting was for the equilibrium that players coordinated on in decentralized play.

2.2. Voting ex ante to select the superior equilibrium (treatment INEXP)

Treatment INEXP is the same as EXP but reverses the sequence of phases 1 and 2 of EXP. In phase 1 of INEXP, the median voter determines the price on which the group coordinates in period 1. Phase 2 consists of 29 periods of the price setting game in a decentralized manner as in phase 1 of EXP. The purpose of introducing majority voting in period 1 is to test if collective action prevents coordination failure by selecting the superior equilibrium A from the start. Thus, if collective action in period 1 succeeds in selecting equilibrium A and serves as a "precedent", we should see more successful coordination in phase 2 of INEXP than in phase 1 of EXP. However, that is not case.

Figure 2: Average group prices in collective action for inexperienced voters and in decentralized play (INEXP, vote in period 1)

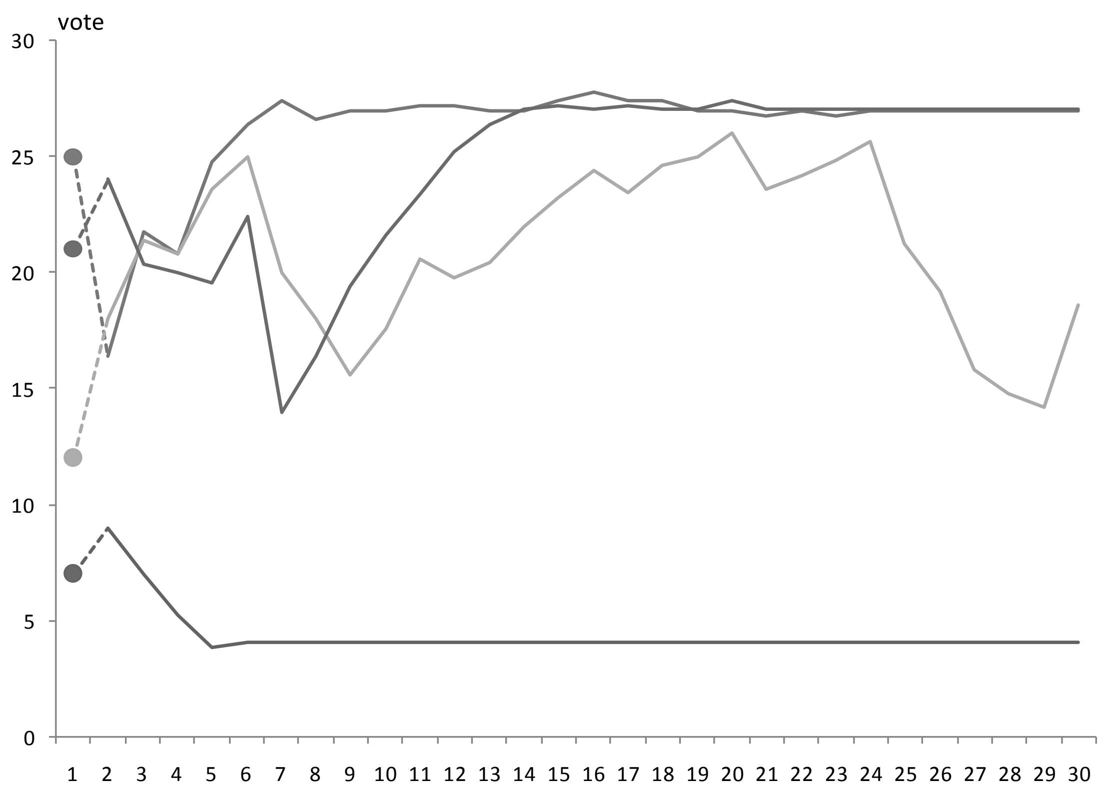


Figure 2 shows average group prices in INEXP over time. The main finding is that collective action by inexperienced voters fails to coordinate behavior on the superior equilibrium A. The dots in figure 2 show the median vote for each of the 4 groups, and we see that none of the groups is coordinated on $P_A = 4$. Only 25 percent of subjects vote $p_A = 4$, 5 percent for $p_B = 10$, and 0 percent vote $p_C = 27$. A full 70 percent of subjects vote for disequilibrium prices. Choices are not different in EXP and INEXP in period 1 (16.6 in both treatments, p = 0.83, MW). None of the groups coordinates on any equilibrium and while efficiency is higher in the presence than in the absence of collective action (8.9 in INEXP, 1.7 in period 1 of EXP, p =0.01) it remains low (31.8 percent of π_A). The massive coordination failure in period 1 of INEXP thus does not seem to be due to strategic uncertainty or failed equilibrium selection but rather to disequilibrium choices. We speculate that wide-spread disequilibrium play (in EXP) and voting (in INEXP) in period 1 is due to subjects' cognitive inability to identify any equilibrium. In line with FT, we speculate that this is due to the effect of the nominal representation (i.e. money illusion), and perhaps magnified by the large action space.

When comparing overall efficiency in the periods with decentralized play (periods 1-30 in EXP, 2-30 in INEXP), we find no difference (16.4 vs. 15.3, p = 0.36, MW). Thus, collective action in period 1 does not improve the efficiency of coordination in the consecutive decentralized game. Given that collective action failed to select any equilibrium, it is perhaps not surprising to note that collective action did not serve as a coordinating "precedent". When comparing voting of experienced (EXP in period 31) and inexperienced voters (INEXP in period 1), we note that votes in EXP are more often (71 percent) for equilibrium values than in INEXP (30 percent). The share of votes for A is similar in both treatments (25 vs. 37 percent), but the share of votes for C is higher in EXP (34 vs. 0 percent). Thus, extensive experience with playing the coordination game in a decentralized manner does not improve the coordinating effect of collective action.

A potential reason for the failure of majority voting to prevent (in INEXP) or break lock-in (in EXP) is that incentives for strategic voting are present for particular disequilibrium beliefs. To illustrate, assume that player i expects 2 of the other players to vote C and two to vote for a price between A and C. Given these beliefs, it is rational for player i to vote C to prevent the median to be between A and C implying a low disequilibrium payoff compared to π_C. As a consequence, majority voting may provide incentives not to vote for A if expectations are not coordinated. In other words, collective action in the guise of majority voting does not eliminate strategic uncertainty and this fact may partly explain it lack of success in improving coordination.

3. Does collective action coordinate expectations by "precedent"? -- Study 2

The purpose of study 2 is to test the coordinating effect of collective action by setting a "precedent". The idea is to see if successful equilibrium selection through collective action improves coordination in subsequent decentralized play by coordinating expectations. To test, we need a setup in which collective action is expected to be successful in selecting the efficient equilibrium in the first place. We make two modifications from study 1 to improve the effectiveness of collective action in equilibrium selection.

First, we reduce the action space to include only the 3 equilibrium actions. Thus, complexity of the payoff table is reduced which can be expected to reduces the potential for mis-coordination (see Ho and Weigelt 1996, Devetag and Ortmann 2007). Second, we introduce collective action in the guise of random dictator voting rather than majority voting as in study 1. Random dictator voting has the advantage of eliminating any strategic incentive not to vote A. Because of the reduced action space, all outcomes of collective action are equilibrium outcomes. The equilibrium selection problem is non-strategic since it is optimal to vote equilibrium A no matter what equilibrium other players vote. Except for these two changes, study 2 has the same parameters and procedures as study 1.

3.1. Design of study 2

Table 1 summarizes the design of study 2. Treatment NoVote has no collective action and the game is therefore played in a decentralized manner in all periods. That is, coordination can fail in NoVote because players pick different actions (are mis-coordinated) or because they coordinate on an inferior equilibrium. In decentralized play, we therefore expect coordination failure to loom large.

Treatment RepeatVote is the same as NoVote except that collective action in the guise of random dictator voting is introduced in period 1 when players are inexperienced, and in periods 11 and 21 when players are experienced. In these 3 periods, all players simultaneously vote for a price between 1 and 30 and the choice of one player is randomly determined to be implemented for the entire group.

Table 1: Experimental design of study 2

Treatment	**Description**	**Source of Coordination failure**
NoVote 49 players in 10 groups	**Decentralized action** in all 30 periods (with strategic uncertainty)	Players do not choose the same action (are mis-coordinated) Players coordinate on inferior equilibrium
RepeatVote 40 players in 8 groups	**Collective action** in periods 1, 11, 21 **Decentralized action** in periods 2-10, 12-20, 22-30	As in AllVote As in NoVote
AllVote 23 players in 5 groups	**Collective action** in all 30 periods (no strategic uncertainty)	Groups coordinate on inferior equilibrium (note: players are coordinated on some equilibrium by design)

We expect collective action to improve coordination in study 2 compared to study 1 for 2 reasons. First, because of the reduced action space (see table 2), voting in study 2 coordinates actions on the same *equilibrium* action, rather than just the same action as in study 1, by definition. The players' problem in study 2 is therefore reduced to equilibrium selection in periods with collective action. Second, voters are more likely to select equilibrium A because it is dominant for a rational and self-interested player *i* to vote for A, independent of *i*'s belief about voting by *-i*. Thus, in random dictator voting, unlike majority voting, any strategic incentive not to vote for A is absent in collective action. For the same reason we also expect collective action to improve efficiency compared to decentralized play in NoVote.

In RepeatVote, the game is played in a decentralized manner in periods 2-10, 12-20, and 22-30, exactly as in NoVote. In these periods, players face both aspects of the coordination problem, i.e. to pick the same and to pick the best equilibrium. Both of these problems involve strategic uncertainty in decentralized play. The purpose of collective action in RepeatVote is thus not only to test if collective action succeeds in selecting the efficient equilibrium A but also to test if it coordinates expectations in the decentralized game played after the vote by setting a "precedent".

Treatment AllVote is the same as RepeatVote except that voting takes place in all periods rather than only in 3 periods. In AllVote, outcomes are coordinated on one of the 3 equilibria by design and the coordination problem is reduced to a non-strategic equilibrium selection problem. Treatments AllVote and NoVote thus provide the benchmarks against which we evaluate the coordinating effect of collective action in RepeatVote.

Table 2: Payoff table for study 2 (note: table shows nominal payoffs)

	Average price of other firms		
Your price	**4**	**10**	**27**
4	112	11	28
10	7	50	29
27	4	10	567

Table 2 shows the payoff table used in all treatments and periods of study 2. As in study 1, the table shows nominal payoffs. While study 1 uses a payoff table with 30 actions (i.e. a 30 x 30 table, see Appendix C1), we reduce the action space to a 3 x 3 matrix in study 2 by eliminating all non-equilibrium actions. Absent collective action, this elimination facilitates coordination on the same action (since there are only 3 rather than 30 options), as well as coordination on an equilibrium (since all coordinated actions now result in equilibrium outcomes). The elimination of non-equilibrium actions has implications for the stability of equilibria in decentralized play. In study 2, equilibria A and C are unstable while equilibrium B is now stable in the sense that a deviation of one player from B does not provide incentives for others to deviate from equilibrium B.[10] While equilibrium B had the "watershed" (or separatrix) property in study 1, this property is in study 2 in some sense reversed to a strongly attractive equilibrium. Equilibrium B is quite robust since p_B is a best reply to about 80%[11] admissible beliefs, whereas the other two equilibria are fragile in the sense that unilateral deviations have large effects on the

[10] More specifically, if n-2 players choose action p_B and player i chooses p_A or p_C, player j has an incentive to choose p_B. As a consequence, each of the n-2 players' best replies is p_B.

[11] Table A1 in the appendix shows the best-replies given all possible beliefs in a group of n = 5 players. In this case, there are 81 admissible beliefs. In 67 of these cases, p_B = 10 is a best reply (including 6 cases with multiple best replies).

best-reply dynamics.[12] Since equilibrium B has a rather low real payoff of $\pi_B = 5$, we expect coordination in decentralized play to be rather inefficient which also means that collective action has a high potential to improve efficiency.

3.2. Results of study 2

In treatment NoVote coordination failure looms large in the early periods. While coordination improves with experience, it is far from perfect even after 30 periods. More specifically, efficiency, defined as the sum of payoffs relative to payoff in equilibrium A, is below 20 percent in the first 4 periods, increases to around 40 percent and remains at that level until period 20. In the last 10 periods, efficiency stabilizes at a level of about 60 percent (see figure 3). Average efficiency over all periods of NoVote is 43.7 percent.

We evaluate coordination failure in NoVote at the individual and at the group level. At the individual level, coordination failure prevails if (i) players do not receive any equilibrium payoff (they are uncoordinated) or (ii) if players do not receive the pareto-dominant equilibrium payoff π_A (they select an inferior equilibrium). At the group level, coordination failure prevails if (iii) a group (i.e. at least one of its members) is not coordinated on an equilibrium or (iv) if the group is not coordinated on the pareto-dominant equilibrium A. Note that coordination may fail in all four senses in decentralized play but can only occur in sense (iv) in all periods of AllVote and in RepeatVote in periods with collective action.[13]

According to these four measures, coordination in NoVote is far from perfect, to say the least. In period 1 of NoVote, 73 percent of players receive a non-equilibrium payoff and not a single player (0 percent) receives π_A (see first two columns of table 3). At the group level, coordination failure is total in period 1 of NoVote, in the sense that none of the 10 groups are coordinated on *any* equilibrium. Coordination improves over time in the sense that players coordinate on some equilibrium but coordination failure still looms large as they do not predominantly coordinate on the pareto-superior equilibrium A. For example, in period 11, 29 percent of players receive disequilibrium payoffs and only 35

---

[12] In addition to the 3 pure-strategy equilibria, there is a symmetric mixed-strategy equilibrium in which players choose prices p_A, p_B, and p_C with probabilities 0.11, 0.74, and 0.15, respectively. The expected payoff in the mixed strategy equilibrium is 9.9, resulting in an efficiency of 35.4 per cent (relative to equilibrium payoffs in A).

[13] Also note that the minimum payoffs are higher by design in periods with collective action (i.e. periods 1, 11, 21 of RepeatVote and in all periods of AllVote) because voting eliminates mis-coordination.

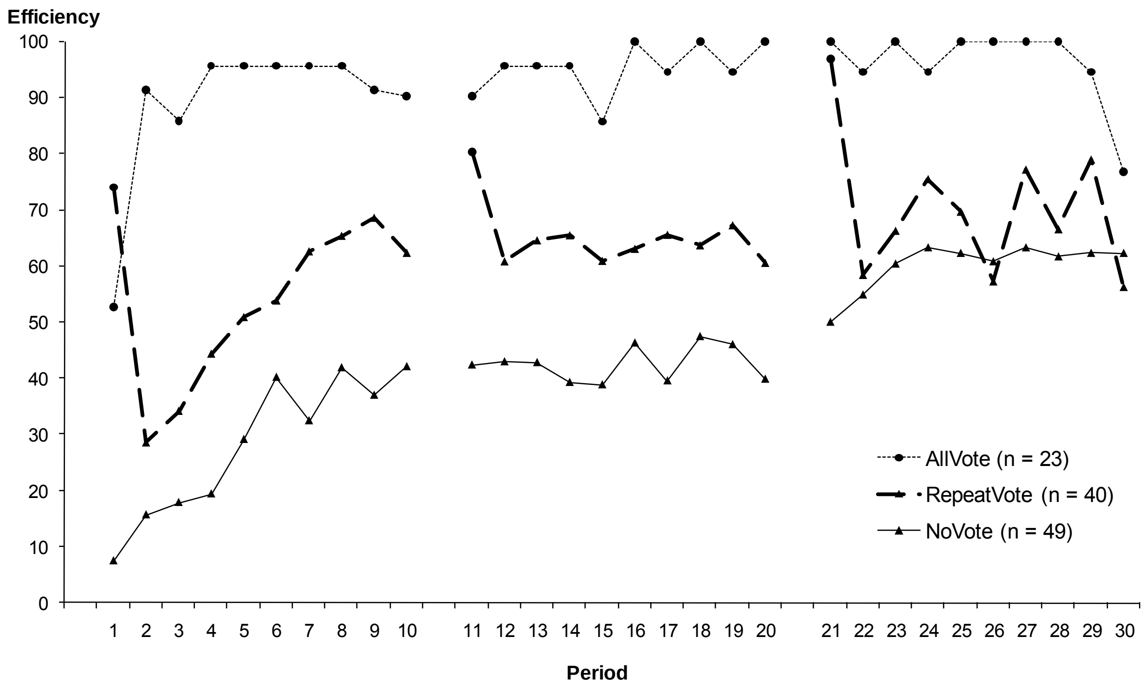
percent receive payoff π_A. Only 2 groups are coordinated on A while the other 8 groups are uncoordinated in period 11 of decentralized play. In period 21, still 22 percent of players receive disequilibrium payoffs and only 43 percent receive π_A. Even now only 3 groups are coordinated on A and 2 groups on B, while the other 5 groups are still uncoordinated. We conclude that coordination failure looms large in treatment NoVote and that our game therefore creates ideal preconditions for investigating if collective action improves coordination.

Table 3: Percentage of players receiving disequilibrium and equilibrium payoffs

	NoVote (n = 49)				**RepeatVote** (n = 40)				**AllVote** (n = 23)		
		Equilibrium				Equilibrium			Equilibrium		
Period	Diseq.	A	B	C	Diseq.	A	B	C	A	B	C
1	73	0	27	0	--	25	13	63	0	39	61
2-10	43	23	35	0	32	14	6	47	72	0	28
11	29	35	37	0	--	50	13	38	61	0	39
12-20	15	33	52	0	22	28	5	45	87	2	11
21	22	43	35	0	--	88	0	13	100	0	0
22-30	10	49	33	9	23	44	4	28	88	2	10

Treatment AllVote provides the benchmark to evaluate RepeatVote at the opposite end of the spectrum since choices are now made collectively in all periods. In AllVote, coordination failure is a priori less likely to occur because coordination is only a matter of non-strategic equilibrium selection. In fact, average efficiency over all periods in AllVote is about the double of NoVote (93.4 vs. 43.7 percent) and we observe that efficiency remains between 90 and 100 percent throughout the game with few exceptions (see fine dotted line in figure 3). Efficiency is higher in AllVote than in NoVote already in the first period (14.7 vs. 2.1, p = 0.000, MW), for all periods (p = 0.000, MW) and the last 10 periods jointly (p = 0.022, MW), but not in the very last period (p = 0.211, MW). Surprisingly, coordination is imperfect in period 1 of AllVote since 0 groups are coordinated in A, 2 in B and 3 in C. Thus, collective action does not prevent coordination failure even in the game with a much reduced scope for coordination failure. But coordination rapidly improves in AllVote. For example, In period 11, 3 groups are coordinated in A and 2 groups in C, and in period 21 coordination is perfect since all 5 groups are coordinated in A. Thus, comparing NoVote and AllVote suggests that this type of collective action is not successful in preventing, but quite successful in eliminating coordination failure after players have gained some experience.

Figure 3: Efficiency by treatment (sum of payoffs, as a percentage of maximum)



In RepeatVote, periods 1, 11 and 21 are with collective action (see dots in the bold-faced graph of figure 3) as in AllVote, while in all other periods, the game is played in a decentralized manner as in NoVote. The comparison of period 1 in RepeatVote and NoVote shows that collective action improved efficiency substantially for inexperienced voters. Figure 3 shows that efficiency in period 1 with collective action was 74 percent with collective action but only 8 percent without (p = 0.000, MW). While average efficiency was high in period 1 of RepeatVote, equilibrium selection is rather imperfect. For example, only 2 groups (with 25 percent of all players) coordinate on equilibrium A, while 1 and 5 groups cordinate on equilibrium B and C, respectively. Thus, we conclude, as we did in the discussion of AllVote, that collective action does not prevent coordination failure altogether in the sense that voters fail to select the superior equilibrium even in the reduced game.

Remarkably, collective action has only a modest effect on efficiency by setting a “precedent” in periods 2 to 9, i.e. when RepeatVote is played in the same way as NoVote. That is, the relatively successful coordination of actions through collective action in period 1 of RepeatVote was only partly successful in coordinating expectations in subsequent decentralized play of the game (see Hamman et al. 2007 for the limited effect of precedents through collective monetary rewards and punishments in minimum effort games). Efficiency falls from 74 percent in period 1 to 29 percent in period 2 of RepeatVote. Average efficiency in RepeatVote develops along a similar path in periods 2 to 9 as in NoVote but is at a higher level in RepeatVote (p = 0.000, MW). In period 2 of RepeatVote, only 2 groups that were coordinated in period 1 on C manage to remain coordinated, all other groups discoordinate. That is particularly surprising for the 2 groups that were coordinated on equilibrium A.

In the second vote of RepeatVote (see dot in period 11) efficiency jumps up (from 62 percent in period 10 to 80 percent in period 11; p = 0.006 Wilcoxon Sign Rank test, henceforth WSR). In period 11, 4 groups are coordinated on A, 1 on B, 3 on C. Average efficiency with collective action is again higher than without (compare RepeatVote and NoVote in period 11, p = 0.000 MW). However, the second instance of collective action had no lasting impact on decentralized play as the average efficiency in the 4 periods after the second vote is the same as the average in the 4 periods before (0.65 vs. 0.63, p = 0.781 WSR). In period 12, the same 2 groups that were coordinated in period 2 manage to coordinate on equilibrium C plus one group manages to remain coordinated on A. The other 3 groups that were coordinated by collective action in period 11 on A, 1 player deviates each. We conclude that the “precedent” effect of the second instance of collective action is rather weak.

The pattern repeats itself for the third vote. Again, voting in itself improved efficiency (97 percent in RepeatVote vs. 50 percent in NoVote of period 21, p = 0.000, MW), but average efficiency in the decentralized game is not different in period 22 to 30 across RepeatVote and NoVote (p = 0.993 MW). In period 21 we find that collective action coordinates 7 groups on the efficient equilibrium A and 1 on C. Yet, in period 22 with decentralized play, only 3 of the 7 groups that coordinated on A remain coordinated.

A comparison of RepeatVote and AllVote yields the following. In period 1, when participants are inexperienced, we find that average efficiency is different (p = 0.002, MW), but the distributions of payoffs are not different between RepeatVote and AllVote (p = 0.177 according to a two-sample Kolmogorov-Smirnov test which corrects for ties, henceforth KS). Averages and distributions are not different across AllVote and RepeatVote in period 11 (p = 0.243 MW, p = 0.958 KS) and period 21 (p = 0.080 MW, p = 0.958 KS) at the 5 percent level. However, AllVote has significantly higher efficiency when compared to the periods with decentralized decision making in RepeatVote in periods 2 to 10, 12 to 20 (p = 0.000 MW each), and significantly different distributions (p = 0.000 KS each for periods 2 to 10, 12 to 20, and 22 to 30).

We conclude that collective action is efficiency-increasing because it successfully coordinates actions but it does not perfectly solve the equilibrium selection problem, nor does it eliminate coordination failure in the decentralized game by setting an efficient precedent. We believe that the reason for this failure to persistently eliminate coordination failure is due to uncoordinated expectations.

Figure 4: Distribution of expectations on P_{-i} (average price of others)

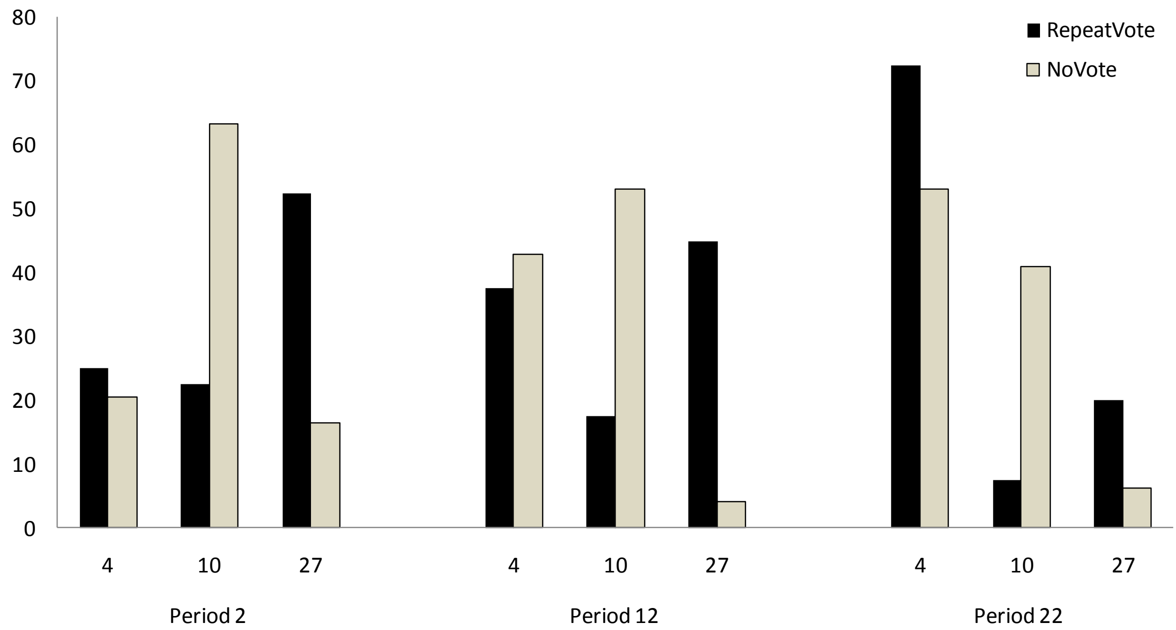


Figure 4 shows the distribution of expectations on P_{-i} , i.e. the average price chosen by of other firms, in the periods following the three votes in RepeatVote (see black bars). The grey bars show the distribution of expectations in NoVote for comparison. In period 2, the percentage of players expecting equilibrium A to prevail is very similar in the two treatments (25 vs. 20 percent). But expectations for equilibrium B (23 vs. 63 percent) and C (53 vs. 16 percent) are clearly different. Voting in period 1 of RepeatVote therefore seems to systematically have affected the level (p = 0.030, MW) and the distribution of expectations (p = 0.001, KS) but does no induce assurance that A will be played. A similar conclusion holds for the period after the second vote in RepeatVote (period 12). Participants do not expect equilibrium A to prevail at different rates (38 vs. 43 percent), but differences are pronounced for B and C. Again, the level (p = 0.001 MW) and the distribution (p = 0.014 KS) of expectations are different between NoVote and RepeatVote. The same pattern is also present after the third vote (period 22) but in this period, neither the overall level (p = 0.296 MW) nor the distribution of expectations (p = 0.272 KS) is different. Figure 4 also illustrates that collective action only very gradually coordinates expectations on the superior equilibrium A. The percentage of players expecting others on average to choose A increases from 25 percent in period 2 to 38 percent in period 12 to reach 73 percent in period 22.

In conclusion, it seems fair to say that collective action in one period helped reducing the prevalence of beliefs in the equilibrium B with particularly low efficiency (a belief which prevails if *i* expects the group to be coordinated on B or to be uncoordinated) in the subsequent period of decentralized play. However, voting did not increase beliefs in the efficient equilibrium A but, perhaps surprisingly, it did increase beliefs in equilibrium C. We therefore conclude that money illusion affected collective action to a remarkable extent. Thirteen percent of players vote in line with money illusion in period 21 of RepeatVote and about 20 percent expect others to pick prices in line with money illusion in period 22.[14]

[14] While the present paper is mainly concerned with investigating collective action as a cure for coordination failure rather than with what aspect of money illusion causes coordination failure, we still wanted to test if money illusion is also a cause in the simplified game of study 2 (we already know from Fehr and Tyran 2007 that it is a cause of coordination failure in study 1). To test, we ran a control treatment of NoVote (n = 21) in a *real* representation. We find that the effect of money illusion is strong, even in this simple game. Average payoffs over all periods in the real representation NoVote are roughly twice the payoffs in the nominal representation (24.2 vs. 12.2, p = 0.000, MW). The effect of money illusion is already pronounced in the first period, i.e. before players get feedback from other payers' choices (19.1 in real vs. 2.1 in nominal, p = 0.000, MW). The effect is persistent. It remains strong and significant up to the very last period of the game (28.0 in real vs. 17.5 nominal, p = 0.000, MW).

5. Concluding remarks

This paper has provided evidence that collective action may not be as successful in improving coordination as common intuition suggests. The evidence is clear and systematic and comes from games in which coordination failure looms large in decentralized play because at least some players are confused about the real payoffs and, therefore, about the pareto-ranking of equilibria. This confusion introduces additional strategic uncertainty into the game. One possible reason why collective action in the guise of majority voting in study 1 has only weak effects on coordination is that majority voting does not totally eliminate strategic uncertainty. In contrast, we have seen that collective action is rather effective in the guise of random dictator voting in which strategic uncertainty is totally eliminated. Thus, this difference between the two modes of collective action is a candidate explanation for the differential effectiveness of collective action in improving coordination. However, it is difficult to assess to what extent strategic uncertainty can be held responsible for these differences from comparing study 1 and 2 because these studies have not been designed for direct comparison. As a consequence, they differ by more than one aspect.

While we believe that our demonstration of the ineffectiveness of collective action to improve coordination is surprising and convincing, we would like to caution the reader from extrapolating our findings to other types of coordination games. One reason is that the present study is one of the first to study the effects on collective action in the literature, and little is therefore known about this potential institutional remedy against coordination failure in general. Another reason is that the game we have studied exhibits massive coordination failure in decentralized play. We believe that our choice of the game is appropriate from a methodological perspective because it provides a convenient workhorse and an almost ideal test-bed to study potential effects of collective action. However, we also know – from an extensive literature in experimental economics – that players' ability to coordinate can depend on a myriad of game-, presentation- and protocol-specific factors (see the contributions on coordination games in the special issue of *Experimental Economics* in 2007 or Devetag and Ortmann 2007 for a discussion). Accordingly, games with slightly different characteristics can be constructed in which coordination is much more efficient than in the games studied here (For example, coordination failure occurs at a much lower rate if the games used in studies 1 and 2 are played showing real, rather than nominal payoffs. More generally, subtle differences in payoffs can have a large effect on how the game is played, see Crawford, Gneezy, Rottenstreich 2008).

While alternative institutional remedies for coordination failure like financial incentives (e.g. Brands and Cooper 2006, Hamman, Rick and Weber 2007), and in particular pre-play communication have extensively been studied and shown to be effective in many cases (e.g. Charness 2000, Duffy and Feltovich 2002, Blume and Ortmann 2007), we believe that the role of collective action and coordination is under-researched. We therefore hope that our paper motivates further research on collective action and coordination.

References

Akerlof, G.A. and Shiller, R.J. (2009): *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton: Princeton University Press.

Blume, A. and Ortmann, A. (2007): The Effects of Costless Pre-Play Communication: Experimental Evidence from Games with Pareto-Ranked Equilibria. *Journal of Economic Theory* 132(1): 274-90.

Brandts, J. and Cooper, D.J. (2006): A Change Would Do You Good: An Experimental Study on How to Overcome Coordination Failure in Organizations. *American Economic Review* 96(3): 669-93.

Brunnermeier, M.K. and Juillard, C. (2008): Money Illusion and Housing Frenzies. *Review of Financial Studies* 21(1): 135-80.

Cabrales, A., Nagel, R. and Rodriguez Mora, J.V. (2006): It is Hobbes, not Rousseau: An Experiment on Social Insurance. Working paper 071808, University Carlos III, Madrid.

Camerer, C.F. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press, Ch. 7: Coordination: 336-407.

Capra, C.M., Tanaka, T., Camerer, C., Feiler, L., Sovero, V., and Noussair, C.N. (2009): The Impact of Simple Institutions in Experimental Economies with Poverty Traps. *Economic Journal*, forthcoming.

Charness, G. (2000): Self-Serving Cheap Talk: A Test of Aumann's Conjecture. *Games and Economic Behavior* 33(2): 177-194.

Cooper, R. (1999): *Coordination Games: Complementarities and Macroeconomics*. Cambridge, U.K. Cambridge University Press.

Crawford, V. (1997): Theory and Experiment in the Analysis of Strategic Interactions. In: D. Kreps and K. Wallis (eds.): *Advances in Economics and Econometrics: Theory and Applications*, 7th World Congress, Vol. 1, Cambridge: Cambridge University Press: 206-332.

Crawford, V., Gneezy, U., Rottenstreich, Y. (2008): The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failure. *American Economic Review* 98(4): 1443-58.

Dal Bo, P., Foster, A. and Putterman, L. (2008): Institutions and Behavior: Experimental Evidence on the Effects of Democracy. NBER working paper 13999,

Devetag, G. and Ortmann, A. (2007): When and Why? A Critical Survey of Coordination Failure in the Laboratory. *Experimental Economics* 10: 331-44.

Duffy, J. and Kim, M. (2005): Anarchy in the Laboratory (and the Role of the State). Journal of Economic Behavior and Organization 56: 297-329.

Duffy, J. and Feltovich, N. (2002): Do Actions Speak Louder than Words? Observation vs. Cheap Talk as Coordination Devices. *Games and Economic Behavior* 39(1), 1-27.

Fehr, E. and Tyran, J.-R. (2007): Money Illusion and Coordination Failure. *Games and Economic Behavior* 58(2): 246-68.

Feri, F., Irlenbusch, B. and Sutter, M. (2009): Efficiency Gains from Team-Based Coordination – Large-Scale Experimental Evidence. WP Max Planck Institute for research on Collective Goods, 2009/14.

Fischbacher, U.(2007): z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10(2): 171-78.

Greiner, B. (2004), An Online Recruitment System for Economic Experiments, in K. Kremer and V. Macho (eds.): Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung: 79–93.

Hamman, J., Rick, S. and Weber, R.A. (2007): Solving Coordination Failure with “All-or-None” Group-level Incentives. *Experimental Economics* 10(3): 285-303.

Heinemann, F., Nagel, R., and Ockenfels, P. (2009): Measuring Strategic Uncertainty in Coordination Games. *Review of Economic Studies* 76(1): 181-221.

Ho, T.-H. and Weigelt, K. (1996): Task Complexity, Equilibrium Selection, and Learning: An Experimental Study. *Management Science* 42(5): 659-79.

Noussair, C.N. and Tan, F. (2009): Voting on Punishment Systems Within a Heterogeneous Group. CentER Discussion Paper 2009-19.

Sefton, M. and Yava, A. (1996): Threat to Regulate and Coordination Failures: Experimental Evidence. *Journal of Real Estate Finance and Economics* 12(1): 97-115.

Shafir, E., Diamond, P., and Tversky, A. (1997): Money Illusion. *Quarterly Journal of Economics* 112(2): 341-74.

Tyran, J.-R. (2007a): Coordination Failure. *International Encyclopedia of the Social Sciences*, 2nd ed. McMillan, Vol. 2: 127-8.

Tyran, J.-R. (2007b): Money Illusion and the Market. *Science* 314:1042-43.

Tyran, J.-R. and Feld, L.P. (2005): Achieving Compliance when Legal Sanctions are Non-Deterrent. *Scandinavian Journal of Economics* 108(1): 135-56.

Van Huyck, J.B., Battalio, R.C. and Beil, R.O. (1991): Strategic Uncertainty, Equilibrium Selection and Coordination Failure in Average Opinion games. *Quarterly Journal of Economics* 106(3): 885-911.

Van Huyck, J.B., Battalio, R.C. and Cook, J. (1997): Adaptive Behavior and Coordination Failure. *Journal of Economic Behavior and Organization* 32: 483-503.

Weber, B., Rangel, A., Wibral, M. and Falk, A. (2009): The Medial Prefrontal Cortex Exhibits Money Illusion. *Proceedings of the National Academy of Science* 106(13): 5025-8.

Appendix A

Table A1: Best-replies for all possible expectations (illustration for groups of n = 5)

Expectations	Average	Best Reply
(4,4,4,4)	4	**4**
(4,4,4,10)	5.5	**4**
(4,4,10,10)	7	**10**
(4,10,10,10)	8.5	**10**
(4,4,4,27)	9.75	**10**
(10,10,10,10)	10	**10**
(4,4,10,27)	11.25	**10**
(4,10,10,27)	12.75	**10**
(10,10,10,27)	14.25	**10**
(27,27,4,4)	15.5	**10**
(4,10,27,27)	17	**10**
(10,10,27,27)	18.5	**10**
(4,27,27,27)	21.25	**27**
(10,27,27,27)	22.75	**27**
(27,27,27,27)	27	**27**

Appendix B: Instructions

[Common part for all treatments]

Welcome to the experiment. Please read these instructions carefully. When you read these instructions carefully, you can earn money. During the experiment, we calculate your income in points. All points you earn during the experiment will be converted into Danish Kroner according to the exchange rate:

10 points = 1.50 Kroner

Please **do not communicate** with other participants during the experiment. Please raise your hand if you have any questions.

This experiment has several phases. The first phase has 10 periods. We explain the details of phase 1 now, the details of the other phases are explained later.

All participants are randomly assigned to groups. The size of your group will be displayed on the screen. You do not know who is in your group but the group remains the same throughout the experiment. Only the decisions in your group are relevant for your earnings. Decisions by other groups are irrelevant to you.

[Treatment AllVote:]

All group members are in the role of firms. In each period, all firms must simultaneously choose a price of 4, 10 or 27. After all firms have made their choices, one firm from the group is chosen at random. The price chosen by this firm is implemented for all firms. Therefore, the price and the average price of others is the same for all firms. How much a firm earns depends on the price implemented and on the average price of the other firms in the group. Thus, the earnings are depends on the choice of one single firm.

[Treatment NoVote:]

All group members are in the role of firms. In each period, all firms must **simultaneously** choose a price of 4, 10 or 27. How much a firm earns depends on the price it chooses and on the average price of the **other** firms in the group. Note that the resulting average is rounded to the nearest of 4, 10 or 27. Therefore the average price relevant for determining your income can only be 4, 10 or 27.

[Treatment RepeatVote:]

All group members are in the role of firms. In each period, all firms must simultaneously choose a price of 4, 10 or 27. How much a firm earns depends on the price it chooses and on the average price of the other firms in the group. Note that the resulting average is rounded to the nearest of 4, 10 or 27. Therefore the average price relevant for your income calculation can only be 4, 10 or 27. These rules apply to periods 2 to 10.

[All Treatments:]
How your income is calculated

The **income table** (on the next page) shows your **nominal point income**. All firms have the same tables. To illustrate how you read the table, consider the following example.

Example: Suppose you choose a price of 10 and the other firms choose prices of 10 on average. In this case, your nominal point income is 50 points.

For the determination of your earnings, only the real point income is relevant. This holds for all firms. To calculate your real point income from your nominal point income, you have to divide the nominal point income by the average price of other firms. Therefore, the nominal and the real point income are related as follows:

$$\boxed{\text{Real point income} = \frac{\text{Nominal point income}}{\text{Average price of other firms}}}$$

[Treatment NoVote and RepeatVote]

If the resulting real point income is a non-integer, it is rounded to the closest whole number. In the example above, your nominal point income is 50 points, but your **real** point income is 5 points, since: 50 points / 10 = 5.

[Treatment AllVote]

To summarize: at the beginning of each period, you choose a price. Then, a random member of the group will be chosen, and his/her price is implemented for all. At the end of each period, you learn the average price of the others and your real income in this period.

[Treatment NoVote]

To summarize: at the beginning of each period, you choose a price. In addition, you indicate what average price of the other group members you expect. The choices by all firms determine the average price of other firms. The average price of other firms is rounded to the nearest of the three possible prices: 4, 10 and 27. At the end of each period, you learn the average price of the others and your real income in this period.

[Treatment RepeatVote]

To summarize: at the beginning of each period, you choose a price. In addition, you indicate what average price of the other group members you expect. In the first period, a random member of the group will be chosen, and his/her price is implemented for all. In the remaining 9 periods, the choices by all firms determine the average price of other firms. The average price of other firms is rounded to the nearest of the three possible prices: 4, 10 and 27. At the end of each period, you learn the average price of the others and your real income in this period.

[All treatments:]

You may write on these instructions and the table. Do you have any questions?

Income Table

	Average price of other firms		
Your price	**4**	**10**	**27**
4	112	11	28
10	7	50	29
27	4	10	567

Appendix C1: Payoff table for EXP and INEXP in study 1 (treatment NH with nominal representation in Fehr and Tyran 2007)

Selling price \ Average price of other firms	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	13	11	11	15	19	15	13	12	11	10	11	12	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
2	24	25	19	25	32	22	16	14	12	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
3	13	48	44	58	73	37	23	16	13	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
4	6	25	84	112	140	84	39	22	15	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
5	3	11	44	58	73	162	88	37	19	12	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
6	2	7	19	25	32	84	168	80	29	12	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
7	2	5	11	15	19	37	88	152	59	14	13	14	15	16	17	18	19	19	20	21	22	23	24	25	26	27	28	29	30	31
8	2	4	8	10	13	22	39	80	108	18	14	15	15	16	17	18	19	20	21	22	23	24	25	25	26	27	28	30	31	32
9	1	3	6	8	10	15	23	37	59	30	17	16	17	17	18	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
10	1	3	5	7	9	12	16	22	29	50	22	19	18	18	19	19	20	21	21	22	23	24	25	26	27	28	29	30	31	32
11	1	3	5	6	8	10	13	16	19	30	39	26	22	21	20	20	21	21	22	23	24	24	25	26	27	28	29	30	31	32
12	1	3	4	6	7	9	11	14	15	18	66	48	31	25	23	22	22	22	23	23	24	25	26	27	27	28	29	30	31	32
13	1	2	4	5	7	8	10	12	13	14	39	84	59	36	29	25	24	24	24	24	25	25	26	27	28	29	30	31	32	33
14	1	2	4	5	6	8	9	11	12	12	22	48	104	70	42	32	28	26	26	26	26	26	27	28	28	29	30	31	32	33
15	1	2	4	5	6	8	9	10	11	12	17	26	59	126	83	48	36	31	29	28	27	27	28	28	29	30	31	32	33	34
16	1	2	4	5	6	7	9	10	11	11	14	19	31	70	150	96	54	40	34	31	30	29	29	29	30	30	31	33	33	34
17	1	2	3	5	6	7	8	9	10	11	13	16	22	36	83	176	111	61	44	36	33	32	31	31	31	31	32	34	34	35
18	1	2	3	5	6	7	8	9	10	11	12	15	18	25	42	96	204	126	68	48	40	36	34	33	32	32	34	35	35	36
19	1	2	3	4	6	7	8	9	10	10	12	14	17	21	29	48	111	234	143	76	53	43	38	36	35	34	35	37	37	37
20	1	2	3	4	6	7	8	9	10	10	12	13	15	18	23	32	54	126	266	160	84	57	46	41	38	36	38	39	39	38
21	1	2	3	4	5	7	8	9	10	10	12	13	15	17	20	25	36	61	143	300	179	92	62	49	43	40	42	43	42	41
22	1	2	3	4	5	7	8	9	10	10	12	13	14	16	19	22	28	40	68	160	336	198	101	67	53	46	48	50	46	44
23	1	2	3	4	5	6	8	9	9	10	11	13	14	16	18	20	24	31	44	76	179	374	219	110	73	57	59	61	54	49
24	1	2	3	4	5	6	7	8	9	10	11	13	14	15	17	19	22	26	34	48	84	198	414	240	120	78	81	84	67	57
25	1	2	3	4	5	6	7	8	9	10	11	12	14	15	17	18	21	24	29	36	53	92	219	456	263	130	135	140	93	71
26	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	18	20	22	26	31	40	57	101	240	500	286	297	308	157	99
27	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	18	19	21	24	28	33	43	62	110	263	546	567	588	348	168
28	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	19	21	23	26	30	36	46	67	120	286	297	308	667	375
29	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	19	20	22	24	27	32	38	49	73	130	135	140	348	720
30	1	2	3	4	5	6	7	8	9	10	11	12	13	14	16	17	18	20	21	23	26	29	34	41	53	78	81	84	157	375

Appendix C2: Payoff table in Fehr and Tyran (2007) – real representation [*Note*: this table was not used in study 1]

Average price of other firms

Selling price	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	13	6	4	4	4	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	24	13	6	6	6	4	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	13	24	15	15	15	6	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	6	13	28	28	28	14	6	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	3	6	15	15	15	27	13	5	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	2	4	6	6	6	14	24	10	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	2	3	4	4	4	6	13	19	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	2	2	3	3	3	4	6	10	12	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	2	2	2	2	3	3	5	7	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	2	2	2	2	2	2	3	3	5	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	1	2	2	2	2	2	2	2	2	3	4	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	1	2	1	2	1	2	2	2	2	2	6	4	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	2	1	1	4	7	5	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	2	4	8	5	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	2	2	5	9	6	3	2	2	2	1	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1	1	2	2	5	10	6	3	2	2	2	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1	1	1	2	3	6	11	7	3	2	2	2	1	1	1	1	1	1	1	1	1
18	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	6	12	7	4	2	2	2	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	7	13	8	4	3	2	2	2	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	7	14	8	4	3	2	2	2	1	1	1	1	1
21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	8	15	9	4	3	2	2	2	2	2	1	1
22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	4	8	16	9	4	3	2	2	2	2	2	1
23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	4	9	17	10	5	3	2	2	2	2	2
24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	4	9	18	10	5	3	3	3	2	2
25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	4	10	19	11	5	5	5	3	2
26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	4	10	20	11	11	11	5	3
27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	5	11	21	21	21	12	6
28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	5	11	11	11	23	13
29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	5	5	5	12	24
30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	3	3	5	13

Coordination Failure Caused by Sunspots

Coordination Failure Caused by Sunspots

May 2009

We study whether a sunspot can lead to either coordination on an inferior equilibrium (mis-coordination) or to out-of equilibrium behavior (dis-coordination) in a coordination game with Pareto-ranked equilibria. While much of the literature searches for mechanisms to attain coordination on the efficient equilibrium, we consider sunspots as a potential reason for coordination failure. We conducted an experiment with a three player 2x2x2 game in which coordination on the efficient equilibrium is easy and should normally occur. In the control session, there was almost perfect coordination on the payoff-dominant equilibrium, but in the sunspot treatment, dis-coordination was frequent. Sunspots lead to significant inefficiency, and we conclude that sunspots indeed can cause coordination failure.

Julie Beugnot
Université Montpellier I
LAMETA

Zeynep Gürgüç
Universitat Pompeu
Fabra

Frederik R. Øvlisen
University of
Copenhagen

Michael W.M. Roos
Ruhr-Universität
Bochum & University of
East Anglia

Keywords: sunspots, coordination, multiple Pareto-ranked equilibria, experiment
JEL-codes: C92, C72, D81, E40 J52

1. INTRODUCTION

The purpose of this paper is to study experimentally whether a sunspot can lead to coordination failure, i.e. either coordination on an inferior equilibrium (mis-coordination) or to out-of equilibrium behavior (dis-coordination) in a coordination game with Pareto-ranked equilibria. Following the definition of Duffy and Fisher (2005) we think of a sunspot being an extrinsic random variable that does not directly affect economic fundamentals. While much of the relevant literature searches for mechanisms to attain coordination on the efficient equilibrium (e.g. Bornstein et al. 2002, Weber 2006, Brandts and Cooper 2006, Brandts et al. 2007), we focus on potential reasons for coordination failure. Is it possible that in a game in which coordination on the efficient equilibrium is easy and should normally occur, a sunspot can prevent subjects from coordinating on any equilibrium or even make them coordinate on an inferior equilibrium? If so, sunspots can explain coordination failures or collective choices of dominated equilibria observed in the real world such bank runs, stock market crashes or other financial turmoil (see Diamond and Dybvig 1983, Allen and Gale 2004, Harrison and Weder 2006). In addition to those macroeconomic coordination problems, coordination is important for large organizations, in which it is necessary to synchronize the efforts of individual workers in order to avoid production bottlenecks (see Van Huyck et al. 1990, Knez and Camerer 1994). In both cases, rumors or some external news could affect the behavior of agents and lead to coordination failure.

Previous work has shown that it is not easy to generate sunspots in the laboratory that affect subjects' behavior (Marimon et al. 1993). Duffy and Fisher (2005) were the first to experimentally establish that sunspots may influence economic choices. We modify their approach in two important ways. First, we consider Pareto-ranked equilibria whereas their model has equilibria which are not Pareto-ranked. Second, we simplify the game and its presentation. While Duffy and Fisher generated sunspots in a market setting, we use a simple three-player 2x2x2 game, in which coordination on the obvious Pareto-superior equilibrium is very easy. Similar to Duffy and Fisher (2005), our sunspot is an announcement determined by the roll of a dice. We find that sunspots influence choices and cause coordination failure even though the conditions of the experiment are such that we theoretically should not expect any effects of the sunspot. We thus show that sunspots can affect economic behavior, but also that they can do it in a very significant and welfare-decreasing way.

In the experiment, a sunspot consists of two announcements which correspond with the strategies available to subjects. Due to the very suggestive character of our sunspot, our work can be related to the literature on recommendation in games (e.g. Brandts and MacLeod 1995, Cason and Sharma 2007, Kuang et al. 2007) However, our sunspot differs from a recommendation because of its obvious random nature. The literature on recommendations shows that in some cases subjects follow a recommendation to play a strategy that leads to an inferior equilibrium. In other cases, recommendations create uncertainty about whether the other players will follow the recommendation which results in dis-coordination. Brandts and MacLeod (1995) show that if incentives are strong enough, subjects follow reasonable recommendations, but ignore unreasonable ones. In particular, recommendations to play an equilibrium that is not subgame-perfect are usually not followed. This is in contrast to our finding that subjects follow a sunspot, although it is not in their interest to do so.

Duffy and Fisher (2005) provide evidence that the semantics of the sunspot matters. Subjects are more inclined to coordinate on a sunspot, if the sunspot can be interpreted as being related to the economic problem. If the sunspot is abstract or difficult to relate to the economic problem, subjects seem to ignore it. For that reason we choose to frame our experiment as a real-work economic problem rather than using an abstract game. We consider workers' decisions to go on a strike. A strike can be seen as a coordination game, because a worker wants to join the others either striking or not striking. If a worker does not join his coworkers, he either foregoes benefits of a successful strike or he bears the costs of an unsuccessful one. According to Franzosi (1989), it is difficult to explain why strikes occur. Perfectly rational and informed workers and managers would negotiate and avoid strikes. Hence striking might be a dominated equilibrium in a coordination game and workers' decision to go on strike might be influenced by exogenous signals that either make coordination on not striking more difficult or even ease the coordination on striking. Kaufman (1982) presents empirical evidence that non-economic attitudinal or psychological factors such as the militancy of workers, the charisma of union leaders or public opinion towards organized labor have explanatory power for the annual number of strikes or the number of workers involved in strikes in the US. Therefore, we believe that such psychological motivations for strikes could be influenced by random events.

2. EXPERIMENTAL DESIGN AND PROCEDURE

We use a coordination game in which subjects in groups of 3 people are put in the role of workers, who choose between the two actions *work* (W) and *strike (S)*. The effect of sunspots is studied in a within-subject design, in which subjects play two different phases of 20 periods (a total of 40 periods) – one phase without sunspots and one with sunspots. The payment depends on performance of all the periods. The sunspot corresponds to an announcement which is made at the beginning of a period. There are two possible announcements – "work" and "strike". The realized announcement is determined randomly by the roll of a (6-sided) dice. The two possible announcements thus correspond to the action space of the workers. In this way it is very clear how the sunspot could be used as a coordination device. Given the results in Duffy and Fisher (2005) on the significance of the sunspots' semantics, we do not expect abstract announcements such as "green" and "red" to have an effect. The sunspot is determined by pure chance, and it is therefore not a recommendation, although it could appear so. Therefore, it should be obvious that choosing the action according to the announcement generally cannot be expected to lead to higher payoffs. In order to make the random determination of the announcement very salient to subjects, one of them rolls the dice herself and the experimenter makes the announcement about the number on the dice to the whole group.

The coordination game is very simple and the payoffs are shown in the following table:

Table 1: Payoff table

		Other Players' Decisions in Your Group		
		If **BOTH of the other** participants choose **WORK**	If **ONE of the other** participants chooses **WORK** and the other chooses **STRIKE**	If **BOTH of the other** participants choose **STRIKE**
Your Decision	**WORK**	**40**	**10**	**10**
	STRIKE	**0**	**20**	**20**

The game has the two pure strategy Nash equilibria (S,S,S) and (W,W,W) which are Pareto-ranked[1]. Notice that the payoff-dominant equilibrium (W,W,W) is also risk-dominant[2], which distinguishes our game from a typical stag-hunt game where the inefficient equilibrium is risk-dominant. This ensures that there is no conflict between risk- and payoff dominance. Moreover, expected payoff from playing W is higher than the expected payoff from playing S.

Under these circumstances, we expect subjects to coordinate easily on (W,W,W) in the absence of a contradictory signal as well as in the presence of one. Indeed, even with the randomly determined announcement "strike", conventional theory predicts that rational subjects (with common knowledge of rationality) should ignore the signal and coordinate on (W,W,W). However, "strike" announcement might create strategic uncertainty if some subjects believe that other subjects will follow the announcement. In that case, we will observe that at least some subjects choose S instead of W. Potentially, strategic uncertainty could be so strong that all subjects coordinate on (S,S,S), whenever they receive the strike announcement. This would be a sunspot equilibrium, in which every subject believes the other subjects to follow the announcement. By playing the best response to this belief, the belief would be self-confirming.

In order to avoid reputation effects, the game is played as a repeated one-shot game with random matching. Moreover, we vary the order of the sunspot phase to check for order effects. To do this, we run control/sunspot sessions (C/S session hereafter), where subjects start by playing a coordination game without announcements in the first 20 periods and play a coordination game with the sunspot in the last 20 periods, and sunspot/control sessions (S/C session hereafter), where, they start with the sunspot phase and play the pure coordination game without announcements after period 20.

---

[1] In addition, there is a symmetric mixed-strategy equilibrium in which players choose "work" and "strike" with probabilities $\sqrt{1/5}$ and $1-\sqrt{1/5}$, respectively. The expected payoff in the mixed strategy equilibrium is 16 resulting in an efficiency of 40 per cent (relative to equilibrium payoffs in (W,W,W)).

[2] As the product of the deviation losses for W,W,W is the highest (Harsanyi and Selten (1988)).

We use two devices in order to avoid that the sunspot destroys the expected coordination on the superior equilibrium. The first one consists of choosing a low probability of the "work" announcement. Actually, if the "strike" announcement occurs frequently, subjects have many opportunities to observe the behavior of the other players and to learn that coordination failure is costly. Hence, we announce "work" only if the dice shows 1 and "strike" otherwise. With the same rationale, the second device consists of providing the subjects with the complete history of the game. After each period subjects are informed about their own decision, the decisions of the other players in the group, the earned points and the announcement, if there was one. Displaying this information for all past periods should also facilitate learning and coordination on the superior equilibrium.

The sessions for this experiment were conducted in LEEX at Universitat Pompeu Fabra (UPF) and in LEE at University of Copenhagen (CPH), using z-Tree (Fischbacher 2007). 36 undergraduate students from various departments of UPF and 48 students from various departments of CPH participated in this experiment. After receiving the instructions[3], participants answered control questions to check their understanding of the game at hand at the beginning and filled in a questionnaire to collect their comments at the end of each session. As mentioned above, each session consisted of two different treatments, one with announcement (the sunspot treatment) and another without announcement (the control treatment). In the sunspot treatments, the throw of a dice was used in order to determine which announcement was realized. At the beginning of each period subjects were randomly matched with others and the dice was rolled by one of the participants. Next, the number on the dice was announced out loud. Finally, according to the number on the dice that was stated, subject received announcements of "work" or "strike" on their computer screen. All the subjects knew that they received the same announcement which was determined according to the roll of the dice. In order to get more independent observations from each treatment, we divided participants in each treatment into subgroups of 6 people, and the subjects from each subgroup were randomly matched with each other for the remaining periods. All participants were paid a show-up fee of 3 € (20 Danish kroner (DKK)) and moreover, received 1 € per 150 points (15 DKK per 150 points) at the end of the

[3] See appendix

experiment. All the subjects were paid in cash at the end of each session and the average earnings were €15 in Barcelona and €16 in Copenhagen.

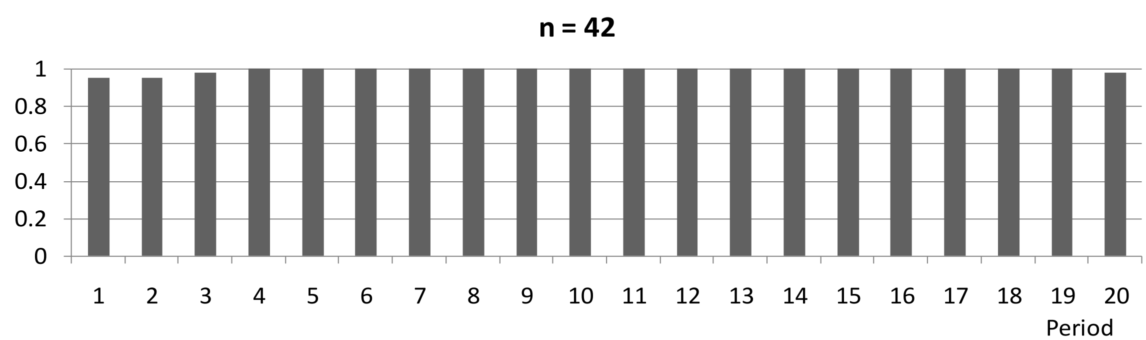
3. RESULTS

We begin this section with an analysis of the control treatments of both sessions. Firstly, we will see that without sunspots subjects coordinate on the Pareto efficient equilibrium. Afterwards, we return to the sunspot treatments, where we find that sunspots create coordination failure.

3.1 Control Treatment

Before we analyze the potential effects of the sunspot, we would like to know whether coordination took place in the absence of the sunspot. Figure 1 summarizes the share of W decisions in all treatments without announcements of all sessions (both C/S sessions and S/C sessions).

Figure 1: Share of "W" decisions in the control treatments



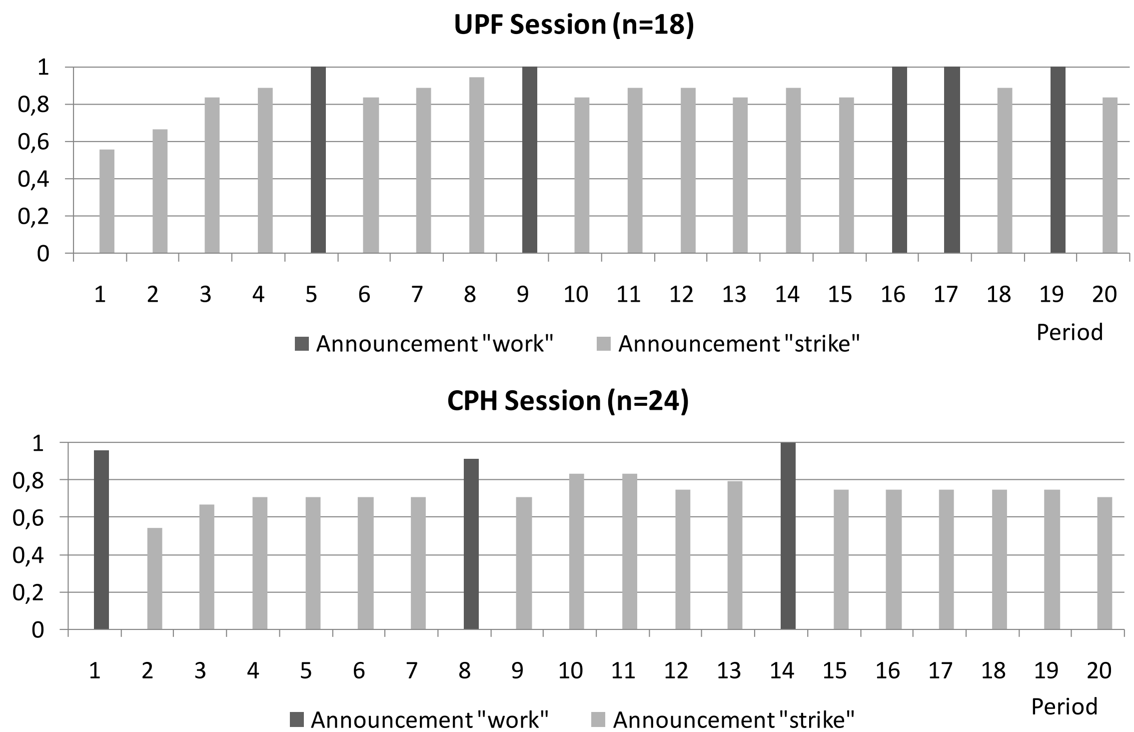
We clearly see that without sunspot announcements coordination was almost perfect. Indeed, 98.6% of all 1680 decisions were W and even in the first periods W was chosen in 95.2% of all cases. In brief, the subjects almost always managed to coordinate on the payoff-dominant equilibrium.

We take this finding as a confirmation that the subjects understood the game and that our game is easy enough so that coordination on the superior equilibrium (W,W,W) occurs right from the start if subjects are not influenced by sunspot announcements. In this paper we focus on sunspots as a potential explanation of coordination failure. Therefore, it is crucial, that the design is such that there is coordination in the absence of sunspots.

3.2 Sunspot Treatments

In this section, we analyze the potential effect of the sunspot on the coordination in our game. Figure 2 shows the share of W decisions in the treatments in which subjects received randomly determined announcements in the first 20 periods (S/C sessions).

Figure 2: Share of W decisions in the sunspot treatments of the S/C sessions



The difference to the control treatment is striking. In both the session at the University of Pompeu Fabra (UPF) and in that at the University of Copenhagen (CPH), a large percentage of subjects chose S in each period with a "strike" announcement. At the UPF session, full coordination on (W,W,W) was only achieved in the cases in which there was a "work" announcement and never with a "strike" announcement. At the CPH session, when there was a "work" announcement, all subjects chose W in only one period and failed to coordinate perfectly in the remaining periods.

Taking both sessions together, the total share of W decisions was 98.1% if the announcement was "work" and 77.1% if the announcement was "strike". As visible in Figure 2, the share of W choices was significantly lower in Copenhagen (73.0%) than in Barcelona (83.3%) if there was a "strike" announcement (t-test, $p<0.001$).

Averaging over periods and subjects within subgroups gives us five clearly independent observations in both the sunspot and the control treatment[4] for the S/C sessions. The Wilcoxon signed-rank test applied to these average shares of work decisions confirms that there is a significant difference between the two treatments ($z=-2.03$, $p=0.04$).

Another way to assess the importance of the sunspot is to estimate the following random effects panel probit model for each period t and each subject i (840 observations)

$$P(Y_{it}=1) = (\beta_0 + \beta_1 STRIKE_{it} + \beta_2 \, PERIOD_t + \boldsymbol{\gamma GD_i}),$$

where Y = 1 indicates the choice of W, STRIKE equals 1 if the announcement is "strike" and 0 otherwise, PERIOD is a time variable, and **GD** is a collection of dummy variables indicating the subgroups with one subgroup at UPF omitted. The time variable serves to check for learning effects and the subgroup dummies control for any group-specific effects.

We find[5] that the announcement of "strike" reduces the probability of W by 0.098 ($p=0.001$) and that as the experiment goes on, an additional period increases the probability of W by 0.004 ($p=0.008$). This means that the strike announcement has a significant effect at the individual level, but also that individuals learn to ignore it over time (even if only slowly).

By running different sessions with different order of treatments, we find a significant order effect. In the C/S session, in which subjects played the control treatment first and the sunspot one next, there is no difference between the treatments (Wilcoxon test, 7 observations, $z=0.78$, $p=0.43$). Thus, the order matters and subjects learn enough from the control treatment to avoid any inefficient decision when the sunspot is introduced.

[4] Unfortunately, there was an allocation problem of subjects to subgroups in the control group of the Copenhagen session, so that we do not want to use two subgroups and are left with 5 instead of 7 independent observations.

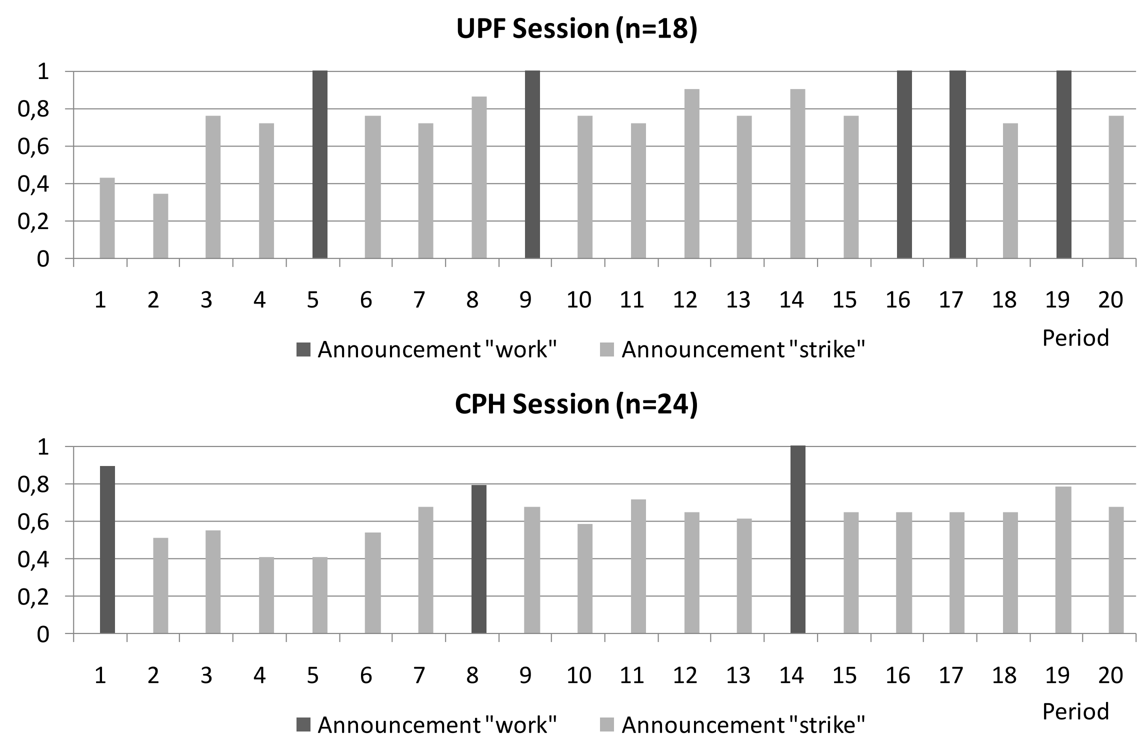
[5] A likelihood ratio test rejects the null hypothesis of no random effects at $p<0.001$. Of the six subgroup dummy variables, two for CPH and one for UPF are significantly negative.

While the sunspot appears to change the behavior of some subjects, it is not strong enough to generate a sunspot equilibrium. We never observe mis-coordination on the inferior (S,S,S) Nash equilibrium in any groups, but rather dis-coordination. Thus, in our experiment the sunspot does not appear as a coordination device but rather as a source of uncertainty. This is not surprising, because strategic uncertainty appears to be very low in the control treatment, in which subjects coordinate almost perfectly on the payoff-dominant equilibrium right from the start.

3.3 Efficiency

The dis-coordination caused by the sunspot is quite costly. We measure efficiency by the actual payoff per period divided by 40, which is the obtained payoff in the Pareto-dominant equilibrium (W,W,W). Figure 3 shows us the average levels of efficiency in the treatments in which the sunspot matters.

Figure 3: Average level of efficiency in the sunspot treatments of the S/C session



At the UPF session (upper panel), the average level of efficiency is always 1 if the announcement is "work", but only three times above 0.8 if the announcement is "strike". The mean of the level of efficiency in the UPF session with a "strike" announcement is equal to 0.728, which is significantly smaller than 1 (t-test, $p<0.001$).

At the Copenhagen session (lower panel), losses are even greater. Here the average level of efficiency is always below 0.8 if the announcement is "strike" and the mean of the average level of efficiency in the 17 periods with "strike" announcements is equal to 0.61, which is significantly smaller than the average efficiency of 0.896 in the three "work" periods (t-test, $p<0.001$).

Overall, efficiency is significantly reduced when the sunspot is introduced in the first periods compared to periods where it is absent (0.714 vs. 0.901, t-test, $p<0.001$). We conclude that the sunspot is relevant, as it can produce significant economic losses through dis-coordination when subjects have not learned enough about the game,.

4. Conclusion

We have demonstrated that a purely random signal unrelated to the fundamentals of the game – a sunspot – produces coordination failure (and efficiency losses) among individuals who almost perfectly coordinate themselves otherwise. This is especially remarkable because our coordination game is so simple. With only two Nash equilibria in pure strategies and with the payoff-dominant equilibrium also being risk-dominant, coordination on this superior equilibrium is made very easy, as confirmed by our control treatments. Yet the introduction of a – in terms of fundamentals – irrelevant signal has huge impact on behavior.

Our study is the second paper that shows the relevance of sunspots in a laboratory experiment. In contrast to the results in Duffy and Fisher (2005), the sunspots in our experiment do not facilitate coordination, but lead to dis-coordination. Furthermore, we introduce sunspots in a setting with Pareto ranked equilibria. In this study, rather than serving as a coordination device and reducing strategic uncertainty, the sunspot generates it and leads to coordination failure. This is a new potential impact of sunspots which has not been extensively discussed in the theoretical literature before. Thus, this study is consistent with the view that real life coordination failures can be caused by random, exogenous signals – sunspots – rather than fundamentals.

ACKNOWLEDGEMENTS

We like to thank organizers and participants of BLEES-M (Barcelona LEEX Experimental Summer School Macro 2007 and the IMEBE 2009 for helpful comments. Financial support from Rosemarie Nagel and Jean-Robert Tyran is gratefully acknowledged.

REFERENCES

Allen, F. and F. Gale (2004). Financial Fragility, Liquidity, and Asset Prices. Journal of the European Economic Association 2(6), 1015 – 1048.

Bornstein, G., U. Gneezy, and R. Nagel (2002). The effect of intergroup competition on group coordination: an experimental study. Games and Economic Behavior 41(1), 1-25.

Brandts, J. and W.B. MacLeod (1995). Equilibrium Selection in Experimental Games with Recommended Play. Games and Economic Behavior 11(1), 36 – 63.

Brandts, J. and D.J. Cooper (2006). A Change Would Do You Good... An Experimental Study on How to Overcome Coordination Failure in Organizations. American Economic Review 96(3), 669.693.

Brandts, J., D.J. Cooper, and E. Fatas (2007). Leadership and overcoming coordination failure with asymmetric costs. Experimental Economics 10(3), 269-284.

Cason, T.N. and T. Sharma (2007). Recommended Play and Correlated Equilibra: an Experimental Study. Economic Theory 33, 11-27.

Diamond, D. and P. Dybvig (1983). Bank Runs, Deposit Insurance, and Liquidity. Journal of Political Economy 91(3), 401 – 419.

Duffy, J. and E. O'N. Fisher (2005). Sunspots in the Laboratory. American Economic Review, 95(3), 510-529.

Fischbacher, U.(2007): z-Tree: Zurich Toolbox for Ready-made Economic Experiments. Experimental Economics 10(2): 171-78.

Franzosi, R. (1989). One Hundred Years of Strike Statistics: Methodological and Theoretical Issues in Quantitative Strike Research. Industrial and Labor Relations Review 42(3), 348 – 362.

Harrison, S.G. and M. Weder (2006). Did sunspot forces cause the Great Depression? Journal of Monetary Economics 53(7), 1327-1339.

Harsanyi, J.C. and Selten, R. (1988): A General Theory of Equilibrium Selection in Games, MIT Press, Cambridge.

Kaufman, B. (1982). The Determinants of Strikes in the United States, 1900 – 1977. Industrial and Labor Relations Review 35(4), 473 – 490.

Knez, M. and C. Camerer (1994). Creating Expectational Assets in the Laboratory: Coordination in 'Weakest- Link' Games. Strategic Management Journal 15, 101-119.

Kuang, J.X., R.A. Weber, and J. Dana (2007). How Effective is Advice from Interested Parties? An Experimental Test Using a Pure Coordination game. Journal of Economic Behavior and Organization 62(4), 591-604.

Marimon, R., S.E. Spear, and S. Sunder (1993). Expectationally Driven Market Volatility: An Experimental Study. Journal of Economic Theory 61(1), 74–103.

Van Huyck, J. B., R.C. Battalio, and R.O. Beil (1990). Tacit coordination games, strategic uncertainty, and coordination failure. American Economic Review 80(1), 234–248.

Weber, R. A. (2006), .Managing Growth to Achieve Efficient Coordination in Large Groups., American Economic Review 96(1), 114.126.

APPENDIX

S/C Session Instructions: (announcements in the first 20 periods)

Thank you for participating in this experiment. Your points and thus your payment at the end of this experiment depend on your decisions and the decisions of the other participants. Moreover, you will receive a show-up fee of 3 Euros. The amount you earn (that depends on the total points you receive) will be paid in cash immediately after the experiment is finished. It is very important that you read these instructions with care. From now on, you are not allowed to talk or communicate in any way with the other participants. If you have any questions, please raise your hand and one of the experimenters will answer them in private. Please do not ask your questions aloud.

This experiment consists of 40 rounds. The experiment will be divided into two sections of 20 rounds. When the first 20 rounds finish, you will receive the instructions for rounds 21 to 40. At the beginning of each round you will be randomly paired with two other participants to form a group of three. Hence, **the members of your group will be different** in each round. The members of a group are not necessarily sitting side by side. All groups in the experiment consist of three people.

Your Decision
In each round, you and the other members of your group will decide whether to **"WORK"** or **"STRIKE"**.

Announcement
Once you are assigned to a group for that round, **you and the other members of your group will receive an announcement at the beginning of each period.** The announcement will be either **"WORK"** or **"STRIKE"**, and it will be the same for all the participants in this experiment, hence for all the members of your group.

It is important that you understand that these announcements are random since they are determined by the throw of a dice. The experimenter will throw the dice in front of all the participants and will ask one of the participants to announce the number on the dice.

You will receive the announcement **"WORK"** if the dice is 1 and **"STRIKE"** otherwise.

Your Payoff
Your income will be determined by the points you earn according to the payoff below and it will based on **your decisions and the decisions of the other players in your group.** All groups in this experiment receive the same payoff table which is explained below.

Other Players' decisions in Your Group

		If **BOTH of the other** participants choose **WORK**	If **ONE of the other** participants chooses **WORK** and the other chooses **STRIKE**	If **BOTH of the other** participants choose **STRIKE**
Your Decision	**WORK**	**40**	**10**	**10**
	STRIKE	**0**	**20**	**20**

In this table, rows indicate your decision of **"WORK"** or **"STRIKE"**, and columns show the decisions of the other players. Each cell indicates the points you will receive depending on your decision and the decisions of the others in your group. For example, if you choose **"WORK";** and if both of the other members of your group choose **"STRIKE"** you receive **10**; whereas, if you choose **"WORK",** and if both of the other players choose **"WORK"**, you receive **40**. If you have any questions please raise your hand.

Information at Each Round
After you make your decision, you will be informed about your current decision, the current decisions of your group members and your points from that round. Moreover, in each round, you can see the information from previous periods (your decisions, the announcement, the decisions of the others and your points). **Please remember that you are randomly matched with different group members at the beginning of each period.**

Each row will give information about each round. The announcement given in that period will be in the second column. Your previous decisions will be in the third column and your previous points from previous rounds will be in the last column. Column 4 will display the previous decisions of the other members of your group. The screen you will see in each round will be similar to one of the following:

ROUND	Announcement	Your decision	Decisions of the other players in your group	Your points In Round
….	….	….	….	….

Payment
The total amount of points you collect after each round will be summed up to determine your total points at the end of the experiment. This final sum will be converted into Euros and will be paid out in cash immediately after the experiment is finished. For each 150 points you earn, you will receive 1 Euro. Moreover, as it is mentioned at the beginning of the experiment, you will receive a show-up fee of 3 Euros. The payment will be made individually and anonymous.

ROUNDS 21-40[6]

Your decisions and your payoffs are the same for these rounds. The points that you will receive are determined according to the same payoff table (p.1). However, once you are assigned to a group for that round, **you and the other members of your group will receive an announcement at the beginning of each period.** The announcement will be either **"WORK"** or **"STRIKE"**, and it will be the same for all the participants in this experiment, hence for all the members of your group.

[6] The instructions for this part were given to the subjects as a separate sheet in the second part of the experiment.

It is important that you understand that these announcements are random since they are determined by the throw of a dice. The experimenter will throw the dice in front of all the participants and will ask one of the participants to announce the number on the dice.

You will receive the announcement **"WORK"** if the dice is 1 and **"STRIKE"** otherwise.

Information at Each Round
As before, after you make your decision, you will be informed about your current decision, the current decisions of your group members and your points from that round. Moreover, in each round, you can see the information from previous periods (your decisions, the announcement, the decisions of the others and your points). **Please remember again that you are randomly matched with different group members at the beginning of each period.**

As before, each row will give information about each round. The announcement given in that period will be in the second column. Your previous decisions will be in the third column and your previous points from previous rounds will be in the last column. Column 4 will display the previous decisions of the other members of your group. The screen you will see in each round will be similar to one of the following:

ROUND	Announcement	Your decision	Decisions of the other players in your group	Your points In Round
….	….	….	….	….

Payment
The total amount of points you collect after each round will be summed up to determine your total points at the end of the experiment. This final sum will be converted into Euros and will be paid out in cash immediately after the experiment is finished. For each 150 points you earn, you will receive 1 Euro. Moreover, as it is mentioned at the beginning of the experiment, you will receive a show-up fee of 3 Euros. The payment will be made individually and anonymous.

C/S Session Instructions (announcements in the last 20 periods)

Thank you for participating in this experiment. Your points and thus your payment at the end of this experiment depend on your decisions and the decisions of the other participants. Moreover, you will receive a show-up fee of 3 Euros. The amount you earn (that depends on the total points you receive) will be paid in cash immediately after the experiment is finished. It is very important that you read these instructions with care. From now on, you are not allowed to talk or communicate in any way with the other participants. If you have any questions, please raise your hand and one of the experimenters will answer them in private. Please do not ask your questions aloud.

This experiment consists of 40 rounds. The experiment will be divided into two sections of 20 rounds. When the first 20 rounds finish, you will receive the instructions for rounds 21 to 40. At the beginning of each round you will be randomly paired with two other participants to form a group of three. Hence, **the members of your group will be different** in each round. The members of a group are not necessarily sitting side by side. All groups in the experiment consist of three people.

Your Decision
In each round, you and the other members of your group will decide whether to **"WORK"** or **"STRIKE"**.

Your Payoff
Your income will be determined by the points you earn according to the payoff below and it will based on **your decisions and the decisions of the other players in your group.** All groups in this experiment receive the same payoff table which is explained below.

Other Players' decisions in Your Group

		If **BOTH of the other** participants choose **WORK**	If **ONE of the other** participants chooses **WORK** and the other chooses **STRIKE**	If **BOTH of the other** participants choose **STRIKE**
Your Decision	**WORK**	**40**	**10**	**10**
	STRIKE	**0**	**20**	**20**

In this table, rows indicate your decision of **"WORK"** or **"STRIKE"**, and columns show the decisions of the other players. Each cell indicates the points you will receive depending on your decision and the decisions of the others in your group. For example, if you choose **"WORK";** and if both of the other members of your group choose **"STRIKE"** you receive **10**; whereas, if you choose **"WORK",** and if both of the other players choose **"WORK"**, you receive **40**. If you have any questions please raise your hand.

Information at Each Round
After you make your decision, you will be informed about your current decision, the current decisions of your group members and your points from that round. Moreover, in each round, you can see the information from previous periods (your decisions, the decisions of the others and your points). **Please remember that you are randomly matched with different group members at the beginning of each period.**

Each row will give information about each round. Your previous decisions will be in the second column and your previous points from previous rounds will be in the last column. Column 3 will display the previous decisions of the other members of your group. The screen you will see in each round will be similar to one of the following:

ROUND	Your decision	Decisions of the other players in your group	Your points In Round
….	….	….	….

Payment
The total amount of points you collect after each round will be summed up to determine your total points at the end of the experiment. This final sum will be converted into Euros and will be paid out in cash immediately after the experiment is finished. For each 150 points you earn, you will receive 1 Euro. Moreover, as it is mentioned at the beginning of the experiment, you will receive a show-up fee of 3 Euros. The payment will be made individually and anonymous.

ROUNDS 21-40[7]

Your decisions and your payoffs are the same for these rounds. The points that you will receive are determined according to the same payoff table (p.1). However, once you are assigned to a group for that round, **you and the other members of your group will receive an announcement at the beginning of each period.** The announcement will be either **"WORK"** or **"STRIKE"**, and it will be the same for all the participants in this experiment, hence for all the members of your group.

It is important that you understand that these announcements are random since they are determined by the throw of a dice. The experimenter will throw the dice in front of all the participants and will ask one of the participants to announce the number on the dice.

You will receive the announcement **"WORK"** if the dice is 1 and **"STRIKE"** otherwise.

Information at Each Round
As before, after you make your decision, you will be informed about your current decision, the current decisions of your group members and your points from that round. Moreover, in each round, you can see the information from previous periods (your decisions, the announcement, the decisions of the others and your points). **Please remember again that you are randomly matched with different group members at the beginning of each period.**

As before, each row will give information about each round. The announcement given in that period will be in the second column. Your previous decisions will be in the third column and your previous points from previous rounds will be in the last column. Column 4 will display the previous decisions of the other members of your group. The screen you will see in each round will be similar to one of the following:

ROUND	Announcement	Your decision	Decisions of the other players in your group	Your points In Round
….	….	….	….	….

[7] Ibid.

Control Questions for the C/S sessions:[8]

1) In all the rounds are you in a group with the same people?
2) What is your payoff, if you choose "work", and if one of the other members chooses "work" and the other chooses "strike"?
3) What is your payoff, if you choose "strike", and if one of the other members chooses "work" and the other chooses "strike"?
4) Will you always receive the same announcement?
5) Is the announcement you receive determined randomly?
6) In the rounds 21 to 40, did the payoff table that you are using change?

Questionnaire for the C/S sessions:[9]

1) First of all we would like to get some statistical information
    1. Studies
    2. Years of study
    3. Age
    4. Sex
2) Please answer the following questions:
    1. The experiment was boring/interesting: 1 2 3 4 5
    2. The experiment was too short/too long: 1 2 3 4 5
    3. Would you like to participate in more experiments?
    4. Did you take any class on game theory?
3) Please briefly explain your decisions of “work” or “strike” in the first part.
4) Please briefly explain your decisions of “work” or “strike” in the second part.
5) Did the announcements in the second part affect your decisions?
6) What was more important for your decisions?
    1. Gaining more points
    2. Maintaining a safe payment
    3. They were both equally important for me
7) Did the previous decisions of the other participants affect your decisions?
8) Do you have further comments?

The experiment is finished

Thank you participating

[8] The control questions for the S/C sessions are the same as the C/S sessions; however, the order of questions is different.

[9] The questionnaire for the S/C sessions are the same as the C/S sessions except for question 5. The question 5 for S/C sessions referred to the announcement in the “first” part.