Bounded Rationality and Endogenous Preferences

Robert Östling

STOCKHOLM SCHOOL OF ECONOMICS

EFI THE ECONOMIC RESEARCH INSTITUTE

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics

KEYWORDS: Bounded rationality, cognitive dissonance, cognitive hierarchy, endogenous preferences, ethnic diversity, income inequality, level-k, moral values, Poisson games, political polarization, pre-play communication, quantal response equilibrium, redistribution, self-serving bias, social identity, size of government.

ISBN 978-91-7258-757-1

PRINTED BY:
Elanders, Stockholm 2008

DISTRIBUTED BY:
EFI, The Economic Research Institute
Stockholm School of Economics
P O Box 6501, SE-113 83 Stockholm
www.hhs.se/efi

To all of you who let me play in my corner of the sandbox

Contents

Introduction 1

Papers 7

Paper 1. Communication and Coordination: The Case of Boundedly Rational Players 9

1. Introduction 9
2. Model 15
3. Extensions 24
4. Evidence 29
5. Concluding Remarks 30

Appendix 1: Characterization of Behavior 32
Appendix 2: Proofs 37
References 41

Paper 2. Strategic Thinking and Learning in the Field and Lab: Evidence from Poisson LUPI Lottery Games 45

1. Introduction 45
2. Theory 48
3. The Field LUPI Game 58
4. The Laboratory LUPI Game 66
5. Learning 73
6. Conclusion 77

Appendix A. The Symmetric Fixed-n Nash Equilibrium 80
Appendix B. Proof of Proposition 1 84
Appendix C. Computational and Estimation Issues 86
Appendix D. Additional Details About the Field LUPI Game 94
Appendix E. Additional Details About the Lab Experiment 99
References 106

Paper 3. Economic Influences on Moral Values 111

1. Introduction 111
2. Model 113
3. Empirical Analysis 118
4. Concluding Remarks 124

Appendix: Comparative Statics 125
Appendix: Description of Data 128
References 130

Paper 4. Identity and Redistribution 133
1. Introduction 133
2. General Model 137
3. Black and White Model 143
4. Ethnic Diversity 147
5. Income Inequality 151
6. Ethnic and Class Hostility 153
7. American Exceptionalism 155
8. Concluding Remarks 156
Appendix: Proofs 157
References 162

Paper 5. Political Polarization and the Size of Government 167
1. Introduction 167
2. Related Literature 169
3. Data and Measurement 172
4. Results 180
5. Other Questions 187
6. Conclusion 188
Appendix 189
References 194

Introduction

The standard approach in economics takes preferences as given and assumes that economic agents make rational decisions based on those preferences. In this thesis I relax both assumptions in a number of economic applications. These applications are somewhat diverse and belong to different fields of economics—game theory, microeconomics and political economics—but they can all be viewed against the backdrop of recent developments in behavioral economics.

The growth of behavioral economics took off in the late 1990s as a critique of neoclassical economics, which generally viewed man as self-interested and rational. Although few economists probably wholeheartedly ever believed that all of us are economic men, until recently, the core of economic theorizing relied mainly on that view of human behavior. This has changed and there are now plenty of economic models based on the alternative assumptions that individuals are imperfectly rational or care for others. The first generation of papers in behavioral economics focused on convincing other economists that self-interest and perfect rationality are not always the best descriptive assumptions. The second generation of papers in behavioral economics explores the implications of those alternative assumptions about human behavior in various economic settings. I hope that most of the papers in this thesis belong to the second generation and I think this is most clear in the first paper of this thesis.

In the first paper, which is written jointly with my much appreciated advisor Tore Ellingsen, we use a model of bounded rationality in order to better understand a classical question in game theory, namely how communication affects behavior in strategic interactions. Thomas Schelling discussed communication in games already in 1960 in his book *The Strategy of Conflict*. He described the cold war between the US and the Soviet Union as a simple two player game (see Chapter 9 in *The Strategy of Conflict*). Either side could choose to start a nuclear attack or not to attack. Both would be better off if no side attacked, but if one side attacks, the other side would like to do so as well. Game theorists recognize this game as a Stag Hunt—named after Jean-Jacques Rousseau's description of a hunt—a game that has two equilibria: one good, but risky equilibrium where both sides stay calm, and a bad equilibrium in which both sides attack. The most interesting—and challenging—question with this game is to understand which equilibrium that will prevail and whether communication between players can help them to obtain the good outcome. In retrospect, we know that there was no nuclear war, which may partly be due to the existence of the hot line between Moscow and Washington. In the last chapter of *Arms and Influence* from 1966, Thomas

Schelling discussed the hot line more in depth and concluded that communication increases the chances that a nuclear war can be avoided: “The hot line is not a great idea, only a good one. (p. 262)”

Imagine yourself in the shoes of John F. Kennedy and that you receive a call from Nikita Khrushchev who tells you that today he and his close friend Fidel Castro is *not* going to start a nuclear war against you. Should you believe him? Initially, the consensus among game theorists was that you should since the message is *self-committing*, that is, if Mr. Khrushchev thinks that you believe the message, it is in his best interest to do what he said and not attack. However, Robert Aumann pointed out early on that Mr. Khrushchev's message is not *self-signaling*. If he for some reason has decided to start a nuclear war, although that is a worse outcome than no war, then he would have told you that he wasn't going to attack (since a surprise attack is better than a full scale nuclear war right from the start). Although this line of reasoning seems theoretically sound, experiments have shown that communication often is successful in getting people to coordinate on the good outcome.

One of the findings in the first paper is that when players are boundedly rational it makes sense for President Kennedy to believe in Mr. Khrushchev. The notion of bounded rationality used in the first paper is based on the steps of reasoning that many people do when they think about how to act in strategic situations. The kind of thinking that we model goes roughly along the following lines:

> What if Khrushchev is completely irrational? Most likely, he's going to naïvely say what he is going to do, so I might as well believe him. But what if he is not completely irrational? Then he would figure out that I would believe his message, so he's going to be truthful and I better believe him.

This way of thinking about bounded rationality goes under the name of level-k reasoning and has been used to explain behavior in a wide range of experiments. The first paper applies that model to communication in games and argues that it provides a better account of how real human beings communicate than the perfectly rational model.

The second paper focuses on a different and much debated question in game theory, namely to what extent mixed equilibria are reasonable descriptions of behavior. In a mixed equilibrium, players attach probabilities to (some of) the strategies they have available and randomize based on those probabilities when they play. Typical games which have realistic mixed strategy equilibria are penalty kicks in soccer, Matching Pennies and Rock-Paper-Scissors. Although these are simple two player games, they share some features with the game studied in the second paper, which has thousands of players and strategies.

The second paper is written together with Colin Camerer, who generously hosted me during my stay at California Institute of Technology, as well as Joseph Tao-yi Wang and Eileen Chou. In the paper, we study the LUPI game that was introduced by the Swedish gambling monopoly in 2007. The rules of the game are simple: each person picks an integer between 1 and $99,999$ and the person that picked the lowest unique number, in other words, the lowest number that was only picked by one person, wins a fixed prize. The mixed equilibrium of this game is that each player plays 1 with highest probability and attaches a lower probability the higher the number is. The idea of playing lower numbers with higher probability is intuitive, but the exact magnitude of these probabilities is not (judge for yourself by looking at Figure 1 on page 52). Although the equilibrium is both difficult to compute and not particularly intuitive, players quickly learn to play close to the equilibrium prediction. To corroborate our findings, we also run classroom experiments in which students played the LUPI game with very similar results.

What more is there to say about the LUPI game if people play according to the equilibrium prediction *as if* they were perfectly rational? In my opinion, a theoretical model should not only be judged by its predictions, but also by the soundness of its assumptions. The assumptions that underlie the equilibrium prediction requires a great deal in terms of both rationality and computational power and we would therefore like to have a theory with more realistic assumptions that explain how people learn to play close to the equilibrium prediction. The answer we provide turns out to be remarkably simple: If people simply imitate numbers around previous winning numbers, they will soon learn to play something which is very similar to the equilibrium prediction. This learning dynamic requires almost no rationality of the players.

The final piece of the LUPI puzzle is to account for how people play the game the first time they play it, before they have had any opportunity to learn. Primarily to explain behavior in early rounds, we develop a model based on a similar notion of bounded rationality as in the first paper: the most naïve players pick numbers completely randomly, players that do one step of reasoning pick very low numbers and those that do two steps of reasoning therefore pick slightly higher numbers (continuing in a similar fashion for more steps of reasoning). This model combined with the learning model can account for how players play initially and then gradually learn to play close to the equilibrium prediction.

In the first two papers I try to develop more realistic descriptions of human behavior by relaxing the rationality assumption. In the third and fourth papers, I instead relax the assumption that people have stable and exogenous preferences. In some circumstances it is a valid simplification that preferences are exogenous, but in others it is not. Preferences do change and they sometimes do so in predictable ways, and that may have economic implications.

One area in which I believe preference changes to be of particular importance is with respect to moral preferences. The third paper focuses on moral preferences related to consumer goods. The paper builds on the psychological theory of cognitive dissonance which briefly stated says that whenever we experience contradicting cognitions, we experience a negative feeling that we are motivated to reduce. For example, if you are concerned about climate change, you might feel bad when you think about that you ought not to travel by plane at the same time as you buy a flight ticket. In order to reduce that feeling of dissonance you may rationalize the consumption decision, for example by convincing yourself that this particular trip is morally motivated.

In the paper, cognitive dissonance is combined with standard consumption theory. A consumer decides how to allocate his income between immoral and moral goods.

A lower price, or higher income, might be a temptation to buy more of the immoral good, which is assumed to be in conflict with the consumer's moral values. In order to reduce the resulting dissonance, the consumer rationalizes his consumption decision by changing moral values. For example, a decrease in the price of air travel may not only increase air travel, but may also lead to consumers adapting their moral values and becoming more tolerant towards going by air. The paper also contains an empirical analysis which shows that we are more tolerant toward goods and activities we tend to consume more. For example, rich people are more tolerant toward tax evasion than poor people, whereas poor people are more tolerant toward benefit fraud than rich people are.

The fourth paper in this thesis, written together with my colleague and friend Erik Lindqvist, concerns group identification. Like the third paper, it focuses on how people's motivation might change in response to changes in the environment they face. People tend to identify with groups, be it their ethnic group, gender, company or neighborhood, which in extreme cases can lead to violent conflicts. For economists, a natural way to understand group identification is that people join certain groups to form coalitions in order to extract more of some material resource. However, there are plenty of experiments, particularly by social psychologists, that suggest that the tendency to identify with groups is more fundamental and not always motivated by material incentives. The fourth paper incorporates some of the insights from social psychology into economic theory in order to better understand the determinants of the level of redistribution from rich to poor. We focus primarily on the interaction between ethnicity, social class and redistribution, which has interested social scientists throughout the 20th century. In some ways, we formalize ideas that go all the way back to Gunnar Myrdal's *An American Dilemma* and other scholars that have pointed out the black-white racial relationship as the reason for the difference between the US and Western Europe when it comes to redistribution.

In the simplest version of our theoretical model, individuals belong to one social class, rich or poor, as well as one ethnic group, black or white. Individuals choose whether to identify with their class or ethnic group, and that choice in turn determine their preferences and how they cast their vote over redistributive policies. For example, a poor white person in the model chooses between identifying with the white or with the poor. If he identifies himself as white, he becomes altruistic toward the white group which contains both rich and poor whites. If he instead identifies himself as poor, he becomes altruistic towards the poor. This means that he supports lower levels of redistribution from rich to poor if he identifies with the white group than if he identifies with the poor. There are two kinds of equilibria in the model. In the "European equilibrium", the poor identify as poor and favor high taxes and the level of redistribution is high. In the low-tax "US equilibrium", the poor whites identify with the white and redistribution is low.

An implication of the model is that an increase in the size of an ethnic minority, for example as a result of immigration, might lead to the ethnic majority switching to identifying with their ethnic group, which reduces the level of redistribution. This is in line with several empirical studies that have found that more ethnically diverse societies have lower levels of redistribution.

The third and fourth papers both focus on how the social and economic environment affects people's preferences. Both papers imply that preferences are likely to be heterogeneous across individuals. For example, the rich are more tolerant toward tax evasion than the poor and someone who belongs to a poor ethnic group is likely to prefer more redistribution than an equally poor person that belongs to a rich ethnic group. In the fifth and final paper, which is also written jointly with Erik Lindqvist, we take preference heterogeneity as given and study its economic implications. More specifically, we empirically study the relationship between polarization of citizens' preferences and the size of government.

Why should we expect a relationship between polarization and government size? Suppose that you live in a heterogeneous society in which people have widely different ideas about what the most appropriate policies are. In such a society, it is quite likely that the policies the government implements will differ from your preferred ones. Irrespective of your own ideological position, you are therefore likely to prefer a smaller government the more polarized the society is.

To test these ideas, we derive a measure of the level of polarization in a country based on responses to survey questions about economic policy. We show that there is a strong negative relationship between political polarization and the size of government. The more polarized a country is, the smaller is the government. The effect is only present in the most democratic countries and the results are therefore consistent with a political mechanism like the one just described.

The remainder of this thesis consists of the five papers introduced above. The papers are self-contained and written with the purpose of eventually being published as separate articles in scientific journals. Although the topics covered are disparate I hope that this introduction has inspired you to continue reading the parts that interest you the most.

There is a long list of people without whom this thesis would have been in much worse shape. My co-author and advisor, Tore Ellingsen, has played a particularly important role. My other co-authors have also been crucial: Erik Lindqvist, Colin Camerer, Joseph Tao-yi Wang and Eileen Chou. Although Magnus Johannesson is not (yet) a co-author, he has provided much support throughout my graduate studies. There are many other colleagues, friends and family members that have played different and important roles for me and this thesis, but they are too many to be mentioned here and they deserve more attention than can be given on a few lines in this thesis. Stockholm School of Economics and the Jan Wallander and Tom Hedelius Foundation paid for the five inspiring, enjoyable and productive years.

Papers

PAPER 1

Communication and Coordination: The Case of Boundedly Rational Players

with Tore Ellingsen

Abstract. Communication about intentions facilitates coordination. It has been suggested that the analysis of costless pre-play communication makes more sense if players are boundedly rational than if they are perfectly rational. Using the level-k model of boundedly rational interaction, we fully characterize the effects of pre-play communication in symmetric 2×2 games. One-way communication weakly increases coordination on Nash equilibrium outcomes, although average payoffs sometimes decrease. Two-way communication further improves payoffs in some games, but is detrimental in others. More generally, communication facilitates coordination in all two-player common interest games, but there are games in which any type of communication hampers coordination.

1. Introduction

According to biologist Martin Nowak (2006, Chapter 13), language is the most interesting innovation of the last 600 million years. Sociobiologists hypothesize that human language first evolved as a response to an environment that rewarded cooperation, notably among hunters of large animals (see, e.g., Pinker and Bloom 1990, Section 5.3). This explanation emphasizes the value of language in coordinating behavior among individuals with common interests. However, communication also plays a crucial role in preventing conflict between individuals that have partially conflicting interests. Consider for example the analysis of communication between military leaders in Schelling (1966, pages 260–264), which ends as follows:

> The most important measures of arms control are undoubtedly those that limit, contain, and terminate military engagements. Limiting war

This paper is an extensive revision of Ellingsen and Östling (2006). We have benefited greatly from the detailed comments by Vincent Crawford and three anonymous referees. We are also grateful for helpful discussions with Colin Camerer, Drew Fudenberg, Joseph Tao-yi Wang and seminar participants at the Arne Ryde Symposium 2007, California Institute of Technology, Lund University and Stockholm School of Economics. Financial support from the Torsten and Ragnar Söderberg Foundation and the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

> is at least as important as restraining the arms race, and limiting or terminating a major war is probably more important in determining the extent of destruction than limiting the weapon inventories with which it is waged. There is probably no single measure more critical to the process of arms control than ensuring that if war should break out the adversaries are not precluded from communication with each other.

While there is still considerable uncertainty about how language evolved, it is evident that language can sometimes be used truthfully in order to attain coordination on jointly desirable outcomes, or at least prevent the costliest coordination failures. Language can also be used deceptively in order to favor one party at the expense of others.[1] But under exactly what circumstances is communication (privately or socially) valuable? When does deception occur? When does communication ensure coordination on jointly desirable outcomes? Answers to these basic questions are of interest to all social sciences, yet they have proven to be surprisingly elusive, both theoretically and empirically. In this paper, we argue that answers are easier to come by once we realize that humans have not evolved to perfection, and that communication is only boundedly rational.

The idea that bounded rationality is the key to understand deception may seem trivial. Fooling a fool is easier than fooling a genius. However, simple ideas can be surprisingly complicated to articulate, and the first satisfactory game theoretic analysis of deception is due to Crawford (2003). Our main contribution is to demonstrate that bounded rationality can also explain why communication improves coordination. In a nutshell, we argue that bounded rationality furnishes players with a major reason to listen as well as to speak. Listening becomes interesting because players believe that they can infer a boundedly rational opponent's intentions, and speaking becomes interesting because they believe that they can affect a boundedly rational opponent's behavior.

Before providing additional details, it is useful to consider where the literature stands. Game theoretic analysis of costless communication (cheap talk) started late. The seminal works, Crawford and Sobel (1982) and Farrell (1987, 1988) are decades younger than many other core applications of game theory. However, the relative recency hardly explains the limited success of cheap talk models. Compared to other deep ideas in game theory, including the closely related idea of costly communication (signaling), the cheap talk literature has had a modest influence on the disciplines of economics and political science.

[1] Indeed, Pinker and Bloom (1990) suggest that, once language was established, language and intelligence co-evolved in a cognitive arms race between cheaters and cheating detectors.

The fundamental problem associated with the analysis of costless communication in games was identified already by Farrell (1988, page 209): "What solution concept do we use for the extended game of communication followed by play?" Using Nash equilibrium will not eliminate any of the Nash equilibria in the underlying game, so if we assume that players are rational, language must have some property that helps to refine the set of Nash equilibria. Introducing a notion of "sensible" messages, Farrell (1987, 1988), and later Rabin (1990, 1994), argue that cheap talk can communicate intentions and thereby entails two primary benefits.[2] The first function is to *improve determinacy.* When the game has many efficient pure strategy equilibria, communication helps players coordinate partly (under multilateral communication) or fully (under unilateral communication) on some profile of efficient equilibrium actions. The second function is to *provide reassurance.* When an efficient equilibrium entails greater strategic risk than some inefficient outcome, communication helps to assure listeners about the speaker's intention to behave in accordance with the efficient equilibrium.[3] Again, expected payoffs rise relative to the outcome without communication.

While theorists broadly agree that cheap talk can improve determinacy in mixed motive games like the Battle of the Sexes, they disagree about the extent to which cheap talk provides reassurance in coordination games like Stag Hunt, the prototype representation of the hunting games from which language may plausibly have emerged. Notably, Aumann (1990) argues that cheap talk among rational players should not suffice to provide reassurance in the Stag Hunt game depicted in Figure 1.

FIGURE 1. Stag Hunt

	H(igh)	L(ow)
H(igh)	$9, 9$	$0, 8$
L(ow)	$8, 0$	$7, 7$

In this game, the two players both prefer the (H, H) equilibrium to the (L, L) equilibrium. Yet, without communication many theories predict the (L, L) equilibrium, since L is considerably less risky than H in case the player is uncertain about what the opponent will do. Farrell (1988) suggests that one-way communication suffices to solve the problem, because the message is *self-committing.* If sending the message "H" convinces the receiver that the sender intends to play H, the best response is for

[2] As emphasized by Myerson (1989) cheap talk can communicate both own intended actions ("promises") and desires about others' actions ("requests"). Like most of the literature, we focus on the former.

[3] For a non-technical introduction to the literature on cheap talk about intentions, see Farrell and Rabin (1996), especially pages 110–116.

the receiver to play H, and thus the sender has an incentive to play according to the own message. Aumann (1990) objects that even a sender who has decided to play L has an incentive to induce the opponent to play H. That is, the message "H" is not *self-signaling*.

Farrell and Rabin (1996, page 114) acknowledge that communication in Stag Hunt may not work perfectly in theory, but they suggest that it will—at least to some extent—work in practice: "[A]lthough we see the force of Aumann's argument, we suspect that cheap talk will do a good deal to bring Artemis and Calliope to the stag hunt." Are Farrell and Rabin right? The experimental evidence on behavior in Stag Hunt games is somewhat conflicting, but it strongly suggests that communication matters. For example, in an experiment by Charness (2000) one-way communication induces substantial coordination on the efficient equilibrium. In the prior experiment by Cooper et al. (1992) one-way communication is rather ineffective, whereas two-way communication often suffices to create efficient coordination. If the fully rational model cannot account for these patterns—or for the intuitions of Farrell and Rabin—is there any other model that can do so?

Crawford (2003) argues that communication frequently makes more sense if people are boundedly rational than if they are fully rational. His analysis considers a special class of zero-sum games, namely Hide and Seek games, with one-way communication. Our work adapts Crawford's approach in order to study a different (and larger) class of games, while also considering a larger set of communication protocols.[4] More precisely, we follow in Crawford's footsteps by using the level-k model of bounded rationality, but we study both one-way and two-way pre-play communication. We consider the class of all symmetric and generic 2×2 games and also study several games outside of this large class. The model's predictions are broadly consistent with the available evidence and suggest several new avenues for empirical work.

The level-k model is a structural non-equilibrium model of initial responses that was introduced by Stahl and Wilson (1994, 1995) and Nagel (1995) and that has been shown to outperform equilibrium models in a range of one-shot games.[5] The level-k model has the feature that players differ in the sophistication that they ascribe to their opponent. The most primitive player type that is assigned a positive probability in our model, the level-1 player, assumes that the opponent plays a random action (is "level-0") and best responds given this belief. A level-2 player assumes that the

[4] Previously Cai and Wang (2006) have adapted Crawford's model to study one-sided cheap talk in sender-receiver games. See also the ongoing work by Crawford (2007) and Wengström (2007).

[5] See for example Stahl and Wilson (1994, 1995), Nagel (1995), Costa-Gomes et al. (2001), Camerer et al. (2004), Costa-Gomes and Crawford (2006) and Crawford and Iriberri (2007) for various normal form game applications.

opponent is a level-1 player, and so on.[6] Like all level-k models, our model of pre-play communication is primarily a model of how people play a game the first time they play it. If a game is played repeatedly, players are likely to learn from previous rounds which raises a number of issues that are not captured in our model.

For parameter choices that are typical in the level-k literature, our main results are the following: (i) One-way communication improves average payoffs in Stag Hunt games with a conflict between efficiency and strategic risk, such as that in Figure 1, and in some but not all mixed motive (Chicken) games. (ii) Two-way communication may yield higher average payoffs than one-way communication, but only in Stag Hunt games with a conflict between efficiency and strategic risk and in mixed motive games with high miscoordination payoffs. (iii) In mixed motive games with high miscoordination payoffs, average payoffs can be lower with communication than without. An additional remarkable finding is that if players are sufficiently sophisticated, both one-way and two-way communication suffices to attain the efficient outcome in Stag Hunt. This conclusion holds not only in the limit as sophistication goes two infinity; it suffices that both players perform at least two thinking steps.

A key to our results is that players are assumed to communicate their true intentions whenever they are indifferent between messages. In other words, they have a lexicographic preference for honesty, as in Demichelis and Weibull (2008).[7] This tie-breaking assumption is innocuous enough from a psychological point of view, but has a powerful effect in our model. It directly implies that level-0 players are telling the truth (or more precisely, that level-1 players believe that their opponent will be honest).[8] Even though we focus our analysis on the case in which there is actually never any level-0 player, the indirect impact on more advanced player types is significant: Level-0 behavior constitutes the level-1 player's model of the opponent. A level-1 receiver will thus play a best response to the received message. Since level-1 behavior constitutes the level-2 player's model of the world, a level-2 sender will therefore send a message that corresponds to the sender's favorite Nash equilibrium. Indeed, it is straightforward to check that all player types will communicate their intentions honestly under one-way communication. (However, this does not imply that one-way communication suffices

[6] A natural extension of the level-k model is to assume that a level-k player believes that the opponent is drawn from a distribution of more primitive player types; see Camerer et al. (2004) for an analysis of the ensuing cognitive hierarchy model. In an earlier version of our paper, we also considered the cognitive hierarchy model. Since the main insights are robust to the choice of model, we only develop the simple level-k model here.

[7] There is considerable experimental evidence that many people assign strictly positive utility to behaving honestly (e.g., Ellingsen and Johannesson 2004*b* and the references therein), but the analysis becomes simpler if the preference is lexicographically small.

[8] As we shall see, in some games a lexicographic preference for truthfulness also has a direct effect on the behavior of more sophisticated players.

to induce an efficient outcome. For example, in the Stag Hunt game above, level-1 players would send and play L.)

Extending our analysis to larger games and/or relaxing the symmetry assumption, we find that both one-way and two-way communication facilitates coordination in all two-player common interest games: When both players make at least two thinking steps, there is always coordination on (the best) Nash equilibrium in these games. This result is simple to prove, but nonetheless remarkable in view of the fact that coordination may require unrealistically many thinking steps when players cannot communicate.

On the other hand, we also identify games in which communication erodes coordination. The reason is that players have an incentive to deceive the opponent by misrepresenting their intentions. Even if the game has a unique pure strategy equilibrium, players can obtain large non-equilibrium payoffs if they successfully fool their opponent. When players are similar and not too sophisticated, they end up playing non-equilibrium strategies that may be either more or less profitable than the equilibrium.

Observe that we take for granted that players have access to a common language. That is, we take an eductive approach to communication. A substantial fraction of the literature on cheap talk starts from the presumption that messages are not inherently meaningful; instead, messages may or may not acquire meaning in equilibrium—where equilibrium is typically depicted, implicitly or explicitly, as a steady state of an evolutionary process of random matches between pre-programmed players; see, for example, Matsui (1991), Wärneryd (1991), Kim and Sobel (1995), Anderlini (1999) and Banerjee and Weibull (2000). The eductive and evolutionary approaches are complementary, and our assumption of bounded rationality closes part of the gap between them. However, while the evolutionary approach can explain how language emerges in "old" games, the eductive approach asks how an existing language will be used in "new" games.

Within the evolutionary cheap talk literature, we are only aware of one contribution that emphasizes the distinction between one-way and two-way communication. In a paper quite closely related to ours, Blume (1998) proves that two-way communication can be superior to one-way communication in games with strategic risk, such as Stag Hunt. Interestingly, Blume's result requires that messages have some small a priori information content. For example, players may have a slight preference for playing (H, H) if both players sent the message "H" and the expected payoffs to playing H and L are otherwise equal. As Blume notes, his assumption amounts to assuming some small amount of gullibility on the part of receivers. In our eductive model, honesty of level-0 senders is instead what drives the superiority of two-way communication in the Stag Hunt game. More recently, in a paper that is contemporaneous with ours, Demichelis and Weibull (2008) find that both one-way and two-way communication induces efficient equilibria in an evolutionary model when players have lexicographic preferences for truthfulness.

2. Model

Let G denote some symmetric and generic 2×2 game.[9] The two strategies are labeled H and L. We refer to G as an *action game* and H and L as *actions*. In the action game G preceded by one-way communication, $\Gamma_I(G)$, one of the players is allowed to send one of two messages, h and l, before the action game G is played. These messages are assumed to articulate a statement about the sender's intention (rather than for example a statement about which action the sender desires from the receiver). Nature decides with equal probability which of the players that is allowed to send a message. We assume that players share a common language and that h corresponds to the intention to take action H and l to action L. Since the message is observed before the action game is played, the actions chosen by the receivers can be made conditional on the received message. A strategy s_i for a player i of the full game $\Gamma_I(G)$ prescribes what message m_i to send and action a_i to take in the sender role, and a mapping $f_i : \{h, l\} \to \{H, L\}$ from received messages to actions in the receiver role. We write a pure strategy of player i (given the received message m_j) as

$$s_i = \langle m_i, a_i, f_i(m_j = h), f_i(m_j = l) \rangle .$$

For example, $s_1 = \langle h, H, L, L \rangle$ means that player 1 sends the message h and takes the action H if he is the sender, while playing L whenever acting as receiver.

In the game with two-way communication, $\Gamma_{II}(G)$, both players simultaneously send a message $m_i \in \{h, l\}$ before the action game G is played.[10] Since messages are observed before the action game G is played, the actions chosen can be made conditional on messages sent. A strategy s_i for player i of the full game is therefore given by a message m_i and a mapping $f_i : \{h, l\} \to \{H, L\}$ from the opponent's message to actions. A pure strategy of player i (given the message m_j sent by player j) can

[9] There is a tension between genericity and symmetry, but none of our results are knife-edge with respect to symmetry. For the purpose of this paper, we consider a game to be generic if no player obtains exactly the same payoff for two different pure strategy profiles. We restrict attention to symmetric and generic games merely in order to keep down the number of cases under consideration. In section 3.2, however, we discuss an asymmetric 2×2 game.

[10] Simultaneous messages may appear to be an artificial assumption. However, besides preserving symmetry, the case of simultaneous messages may capture the notion from models with sequential communication that the first and the last speaker may both have an impact. At any rate, as Rabin (1994, page 390) has argued, the simultaneous communication assumption appears to put a useful lower bound on the amount of coordination that is attainable through cheap talk.

thus be written

$$s_i = \langle m_i, f_i(m_j = h), f_i(m_j = l) \rangle .$$

For example, $s_1 = \langle h, H, L \rangle$ means that player 1 sends the message h, but plays according to the received message (i.e., plays H if player 2 sends message h and L if player 2 plays message l).

Observe that we neglect unused strategy components by restricting attention to the reduced normal form. In other words, we do not specify what action a player would take in the counterfactual case when he sends another message than the message specified by his strategy. The reason is that interesting counterfactuals cannot arise in our model, as will soon become clear.

Players' behavior depends on their degree of sophistication. A player of type 0 (or level-0), henceforth called a T_0 player, is assumed to understand only the set of strategies, and not how these strategies map into payoffs. Thus, T_0 makes a uniformly random action plan, sticking to this plan independently of any message from the opponent. (Hearing the opponent's intended action is of little help to a player who does not understand which game is being played.) Importantly, since T_0 players do not understand how their own or their opponent's actions map into payoffs, or how their messages may affect their opponent's action, they are indifferent concerning their own messages.

For positive integers k, a T_k player chooses a best response to (the behavior that the T_k player expects from) a T_{k-1} opponent. In particular, T_1 plays a best response to T_0. When $k \geq 2$, T_k players will sometimes observe unexpected messages. In this case T_k assumes that the message comes from a T_{k-l} player, where $l \leq k$ is the smallest integer that makes T_k's inference consistent. (As we shall see, T_0 sends all messages with positive probability, so $l \in \{1, .., k\}$ always exists.) Let p_k denote the proportion of type k in the player population. As we shall see, players who perform more than one thinking step often, but not always, behave alike. Therefore, it is convenient to let T_{k+} denote player types that perform at least k thinking steps.

When a player is indifferent about actions in G, we assume that the player randomizes uniformly. However, when the player is indifferent about what pre-play message to send, we assume that there is randomization only in case the player is unable to predict the own action—which can only happen under two-way communication. Otherwise, indifferent players send truthful messages (or more precisely, a message that conveys the action that the player expects to be playing). The assumption reflects the notion that people are somewhat averse to lying, but it does so without incurring the notational burden of introducing explicit lying costs into the model. (Our results are preserved under small positive costs of lying.) While such lexicographic preference for truthfulness is an apparently weak assumption, one of its immediate implications is that the message by T_0 reveals the intended action. Or to put it even more starkly, T_1 believes in received messages. (In Section 2.3 we explore alternative assumptions regarding how T_0 treat messages.)

In Appendix 1 we explicitly characterize the strategies of all player types. However, it is common to argue that T_0 does not accurately describe the behavior of any significant portion of real adult people and that actual players are best described by a distribution with support only on T_1, T_2 and T_3 (e.g. Costa-Gomes et al. 2001 and Costa-Gomes and Crawford 2006). For some of our results we thus refer to type distributions consisting exclusively of players of these three types. To have shorthand definition, we say that $p = (p_0, p_1, ...)$ is a *standard type distribution* if $p_k > 0$ for all $k \in \{1, 2, 3\}$ and $p_k = 0$ for all $k \notin \{1, 2, 3\}$.

2.1. Examples. Consider the Stag Hunt game in Figure 1. Absent communication, T_1 best responds to the uniformly randomizing T_0 by playing the risk dominant action L. Understanding this, the best response of T_2 is to play L as well. Indeed, by induction any player T_{1+} plays L. For any type distributions with $p_0 = 0$, the unique outcome is the risk dominant equilibrium (L, L).[11] The level-k model hence provides a rationale for why players play the risk dominant equilibrium in coordination games without communication.

If players can communicate, one-way communication suffices to induce play of H by all types T_{2+}. The analysis starts by considering the behavior of T_0 (as imagined by T_1). By assumption, a T_0 sender randomizes uniformly over L and H, while sending the corresponding truthful message. A T_0 receiver randomizes uniformly over L and H. As a sender, T_1 best responds by playing the risk dominant action L, and due to the lexicographic preference for truthfulness sends the honest message l. As a receiver, T_1 believes that messages are honest and thus plays L following the message l and H following the message h. Consider now T_2. A T_2 sender believes to be facing a T_1 receiver who best responds to the message, so T_2 sends h and plays H. A T_2 receiver, expects to receive an l message and therefore play L. If receiving a counterfactual h message, T_2 thinks it is sent by a truthful T_0 sender and therefore plays H. It is easily checked that all T_{2+} behave like T_2, implying that there will be coordination on the payoff dominant equilibrium whenever two T_{2+} players meet and communicate. In other words, the level-k model not only shows that it is feasible for advanced players to coordinate on the payoff dominant equilibrium, but that the *unique* outcome is that

[11] Note that this is not about equilibrium selection in the ordinary sense. Players do not select among the set of equilibria, but best-respond to the behavior of lower-step thinkers. Their behavior ultimately results from the uniform randomization of T_0, which explains the parallel to risk dominance.

they will do so. Note in particular how reassurance plays a crucial role in the example. When a receiver gets a message h, the receiver is reassured that the sender will play H, and is therefore also willing to play H. Even if the message h is actually only self-signaling for (the non-existing) level-0 senders, it is self-committing for all other types, and this suffices to attain efficient coordination as long as both parties perform at least two thinking steps.

In Stag Hunt, the reassurance role of communication is strengthened even more when both players send messages. Under such two-way communication, T_1 trusts the received message and responds optimally to it. Expecting to play either action with equal probability, T_1 sends both messages with equal probability. T_2 believes that the opponent listens to messages, and therefore sends h and plays H irrespective of the received message. T_{3+} players believe that the opponent will play H and they therefore play H and send an h message. If they receive an unexpected l message, they believe it comes from T_1 and therefore play H anyway (as T_1 will respond to the received h message by playing H). Note that under two-way communication, T_{2+} players are so certain that the opponent will play H that they play H irrespective of the received message.

Table 1 summarizes the action profiles that will result in the Stag Hunt under one-way and two-way communication. The notation $1S$ indicates a player of type 1 in the role of sender, and so on. "Uniform" indicates that all four outcomes are equally likely.

Table 1. Action profiles played in Stag Hunt with communication

$\Gamma_I\,(G)$ (*one-way communication*)				$\Gamma_{II}(G)$ (*two-way communication*)			
	$0R$	$1R$	$\geq 2R$		0	1	≥ 2
$0S$	Uniform	$\frac{1}{2}HH, \frac{1}{2}LL$	$\frac{1}{2}HH, \frac{1}{2}LL$	0	Uniform	$\frac{1}{2}HH, \frac{1}{2}LL$	$\frac{1}{2}HH, \frac{1}{2}LH$
$1S$	$\frac{1}{2}LL, \frac{1}{2}LH$	LL	LL	1	$\frac{1}{2}HH, \frac{1}{2}LL$	Uniform	HH
$\geq 2S$	$\frac{1}{2}HH, \frac{1}{2}HL$	HH	HH	≥ 2	$\frac{1}{2}HH, \frac{1}{2}HL$	HH	HH

Communication entails perfect coordination on the payoff dominant equilibrium whenever T_{2+} players meet. However, one-way and two-way communication differ in two respects whenever T_1 players are involved. With one-way communication, T_1 senders play L and the risk dominant equilibrium therefore results whenever T_1 senders play (since T_0 does not exist). Under two-way communication, however, there is mis-coordination in half of the cases when two T_1 players meet. Thus, there is a trade-off when choosing the optimal communication structure between coordination on either equilibria and achieving the payoff dominant equilibrium more often. For standard type distributions, two-way communication entails higher expected payoffs than one-way communication as long as $p_1 \in (0, 2/3)$.

In the Stag Hunt, communication increases players payoff because it brings sufficiently much reassurance for players to coordinate on the risky but payoff dominant equilibrium. In mixed motive games such as Battle of the Sexes and Chicken, communication instead serves the role of symmetry-breaking. To see this, consider the mixed motive game depicted in Figure 2, where $a < 3$ and $a \neq 2$. If $a = 0$, then this is a Battle of the Sexes, whereas it is a Chicken game if $a > 0$. The outcome for this game depends on whether L or H is the risk dominant action, i.e., whether $a \gtrless 2$. For simplicity, we disregard the possibility that $a = 2$, but allow the "Battle of the Sexes" possibility that $a = 0$ (although this makes the game non-generic).

FIGURE 2. Mixed motive game

	H	L
H	$0, 0$	$3, 1$
L	$1, 3$	a, a

First consider the case of no communication. T_1 then plays the risk dominant action, i.e., L if $a > 2$ and H if $a < 2$. T_2 responds optimally by playing H if $a > 2$ and L if $a < 2$. The behavior of more advanced players continues to alternate, odd types playing L if $a > 2$ and H otherwise, whereas even types play H if $a > 2$ and L otherwise. The outcome therefore depends on the type distribution, but there will generally be many instances of miscoordination.[12]

One-way communication powerfully breaks the symmetry inherent in such games with two pure asymmetric equilibria. If H is the risk dominant action, then T_{1+} senders send h and play H, whereas T_{1+} receivers optimally respond to messages. If instead L is risk dominant, a T_1 sender sends l and plays L, whereas T_{2+} senders continue to send h and play H. One-way communication therefore implies that T_{1+} players always coordinate on an equilibrium. Except in the case when L is risk dominant and the sender is of type T_1, coordination is on the sender's preferred equilibrium.

It is unsurprising that one-way communication can break the symmetry and achieve coordination in games with two asymmetric equilibria. However, our analysis also reveals the novel possibility that in some versions of Chicken some players propose and play their least favorite equilibrium. T_1 senders play their risk-dominant action which may not correspond to their preferred equilibrium, whereas T_2 senders are confident in reaching their preferred equilibrium. Table 2 shows the outcomes that will result

[12] The outcome without communication does generally not resemble the symmetric mixed strategy equilibrium, but may happen to do so for certain combinations of payoff configurations and type distributions.

without communication and with one-way communication, demonstrating the improved coordination on equilibrium outcomes.

Table 2. Action profiles played in mixed motive games (a > 2)

G (*no communication*)				$\Gamma_I(G)$ (*one-way communication*)			
	0	Odd	Even		$0R$	$1R$	$\geq 2R$
0	Uniform	$\frac{1}{2}HL, \frac{1}{2}LL$	$\frac{1}{2}LH, \frac{1}{2}HH$	$0S$	Uniform	$\frac{1}{2}HL, \frac{1}{2}LH$	$\frac{1}{2}HL, \frac{1}{2}LH$
Odd	$\frac{1}{2}LH, \frac{1}{2}LL$	LL	LH	$1S$	$\frac{1}{2}LH, \frac{1}{2}LL$	LH	LH
Even	$\frac{1}{2}HL, \frac{1}{2}HH$	HL	HH	$\geq 2S$	$\frac{1}{2}HL, \frac{1}{2}HH$	HL	HL

Although one-way communication entails more equilibrium coordination than no communication, more coordination need not raise players' average payoffs. If $a > 2$, then players prefer the (L, L) outcome to ending up in either equilibrium with equal probability. If the type distribution is such that the (L, L) outcome results sufficiently often without communication, average payoffs are thus higher without communication. For example, when $a = 5/2$ and there is a standard type distribution with $p_2 < 1/3$, then average payoffs are lower under one-way communication than under no communication.

Suppose players could choose whether to engage in communication or not, and that the allocation of roles is random. Each player type k would then consider the own expected payoff in each regime conditional on meeting a player of type $k - 1$. To illustrate that players may prefer not to communicate, we consider the case when $a = 0$, i.e., the Battle of the Sexes. Absent communication, T_3 believes that the opponent will play L and thus obtains the preferred equilibrium payoff. With one-way communication and a random allocation of roles, however, T_3 expects to end up in either equilibrium with equal probability. That is, T_3 expects to be better off if communication is impossible.

2.2. Results. In this section we generalize the findings from the previous section to all symmetric and generic 2×2 games, disregarding (the measure zero class of) games in which neither action is risk dominant. There are three broad classes of such games. The first class of games are the dominance solvable ones, like Prisoners' Dilemma. We use the convention of labelling the dominant action of these games H(igh). The second class are coordination games, where we follow the example above and label the actions corresponding to the payoff dominant equilibrium H(igh). The third class of games are mixed motive games like the one in Figure 2. For this class of games, we label the action corresponding to a player's preferred equilibrium H(igh). In Appendix 1, we completely characterize behavior of all player types $k \in \mathbb{N}$ for these three classes of games. These characterizations provide the foundation for the results in this section, where we focus on average outcomes under standard type distributions.

Our first result states the conditions under which one-way communication serves to increase players' average payoffs relative to no communication.

PROPOSITION 1. *Given a standard type distribution, the average payoff associated with $\Gamma_I(G)$ exceeds the average payoff associated with G if and only if (i) G is a coordination game with a conflict between risk and payoff dominance, or (ii) G is a mixed motive game that satisfies either*

a. L is risk dominant and

$$\left(\frac{1}{2} - p_2\left(1 - p_2\right)\right)\left(u_{HL} + u_{LH}\right) > p_2^2 u_{HH} + \left(1 - p_2\right)^2 u_{LL},$$

or

b. H is risk dominant and

$$\left(\frac{1}{2} - p_2\left(1 - p_2\right)\right)\left(u_{HL} + u_{LH}\right) > \left(1 - p_2\right)^2 u_{HH} + p_2^2 u_{LL}.$$

PROOF. In Appendix 2. □

If we replace p_2 by p_E, the probability that players think an even number of steps, Proposition 1 generalizes straightforwardly to all type distributions in which $p_0 = 0$. In our examples, we have already explained why one-way communication improves average payoffs in Stag Hunt, and indicated why it sometimes fails to improve payoffs in mixed motive games. A straightforward implication of Proposition 1 is that one-way communication raises the average payoff in the Battle of the Sexes.[13] (To see this, recall that in Battle of the Sexes $0 = u_{HH} = u_{LL} < u_{LH} < u_{HL}$, which implies that H is risk dominant and that condition (b) in Proposition 1 is satisfied.) Proposition 1 also implies that communication does not improve average payoffs in dominance solvable games. For Chicken, the impact of communication hinges more delicately on parameters, and communication may even serve to reduce payoffs.

COROLLARY 1. *Given a standard type distribution, the average payoff associated with $\Gamma_I(G)$ is smaller than the average payoff of G if and only if G is a game of Chicken that satisfies either*

[13] Note that this does not contradict the statement at the end of Section 2.1 that T_3 prefers not to communicate in the Battle of the Sexes. Proposition 1 refers to payoffs averaged across player types, while the earlier remark referred only to T_3's payoff given that he is certain that he faces a T_2 opponent.

a. L is risk dominant and

$$\left(\frac{1}{2}-p_2\left(1-p_2\right)\right)\left(u_{HL}+u_{LH}\right)<p_2^2u_{HH}+\left(1-p_2\right)^2u_{LL},$$

or

b. H is risk dominant and

$$\left(\frac{1}{2}-p_2\left(1-p_2\right)\right)\left(u_{HL}+u_{LH}\right)<\left(1-p_2\right)^2u_{HH}+p_2^2u_{LL}.$$

Proof. In Appendix 2. □

Since H is risk dominant in Battle of the Sexes, one-way communication suffices to attain perfect coordination on the speaker's preferred equilibrium outcome. Thus, we here have a case in which the prediction from the level-k model coincides with the prediction from fully rational models. Likewise, the ineffectiveness of cheap talk in dominance solvable games is the same as in the fully rational model. At a deeper level, the two approaches also share the property that communication, if anything, pulls players towards Nash equilibria.

Proposition 2. *For any distribution of types, the frequency of coordination on pure strategy Nash equilibrium profiles is weakly greater in $\Gamma_I(G)$ than in G.*

Proof. In Appendix 2. □

The pull towards Nash equilibria is so strong that one-way communication results in equilibrium play whenever T_{1+} meet. Moreover, T_{2+} always play the action corresponding to the sender's preferred equilibrium.

Corollary 2. *For type distributions with $p_0=0$, players in $\Gamma_I(G)$ always coordinate on pure strategy Nash equilibrium profiles. If in addition $p_1=0$, players in $\Gamma_I(G)$ always coordinate on the sender's preferred equilibrium.*

Proof. Follows directly from Tables A1 to A4 in the proof of Proposition 2. □

In contrast to one-way communication, two-way communication may destroy not only average payoffs but also coordination on equilibrium outcomes. For example, suppose there are only T_1 players and let G be a coordination game in which payoff and risk dominance coincide. Then $\Gamma_{II}(G)$ entails miscoordination in half of the cases, because T_1 sends random messages while listening to received messages. By contrast, in G and in $\Gamma_I(G)$ two T_1 players always play the (payoff and risk) dominant equilibrium.

Our model therefore captures the intuition that two-way communication can bring noise in the form contradictory messages.

Nevertheless, there are important classes of games in which two-way communication outperforms one-way communication.

PROPOSITION 3. *Given a standard type distribution, the average payoff associated with $\Gamma_{II}(G)$ exceeds the average payoff associated with $\Gamma_I(G)$ if and only if (i) G is a coordination game in which L is the risk dominant action and $(4-3p_1)\,u_{HH} + p_1\,(u_{LH}+u_{HL}) > (4-p_1)\,u_{LL}$, or (ii) G is a mixed motive game with a type distribution satisfying the following condition:*

$$1+\frac{2\,(p_1-1)\,(p_1-1+2p_3)}{p_1^2+4p_3^2}<\frac{u_{LL}-u_{HH}}{u_{LH}+u_{HL}-2u_{HH}}.$$

PROOF. In Appendix 2. □

The Stag Hunt game in Figure 1 belongs to the first class of games identified by Proposition 3. For that particular game, two-way communication yields higher expected payoff than one-way communication whenever $p_1 \in (0, 2/3)$. The second class of games identified in Proposition 3 is harder to specify because of the cycling patterns of behavior under two-way communication in mixed motive games. However, for two-way communication to be beneficial, the payoff when both players play L must be sufficiently high (at least $(u_{HL}+u_{LH})/2$) and in addition the type distribution has to be such that the miscoordination outcome (L, L) happens sufficiently often with two-way communication. For example, with only type T_3 players, the outcome is (L, L) under two-way communication, whereas such players coordinate on an asymmetric equilibrium with one-way communication.

2.3. Robustness. How robust are our results to the assumptions that we have made about players' behavior?

The largest difference in comparison with other level-k applications is that we assume that players have a weak preference for truthfulness. If players have no preference for truthfulness, communication ceases to have any effect whatsoever in our model: behavior is the same in $\Gamma_I(G)$, $\Gamma_{II}(G)$ and G. This specification is strongly at odds with the evidence that communication matters in many game experiments.

Another alternative hypothesis is that all players prefer to be truthful, but that the most primitive types also respond systematically to received messages. The idea is that (if the actions of both players have the same label), the receiver could imitate or differentiate based on the sender's message. The most natural way to account for such imitation is to allow heterogeneous T_1 players, some believing that receivers randomize, others believing that receivers imitate.[14] With one-way communication, this implies that some T_1 players believe T_0 receivers randomize, whereas others believe that they imitate. With two-way communication, some T_1 players believe that opponents are truthful, whereas other believe they imitate. Let us now consider the consequences of this specification.

First consider the Stag Hunt in Figure 1. Under one-way communication, T_1 senders who believe that receivers imitate send the message h and play H. This in turn implies that T_2 receivers respond to messages as if they were truthful irrespective of which kind of T_1 sender they think they face. Under one-way communication, the only difference compared to our original assumption is that there will be somewhat more coordination on (H, H) since some T_1 senders now play H. Under two-way communication, T_1 players who believe that opponents imitate send h and play H instead of responding to received messages. T_2 players therefore optimally send h and play H irrespective of which type of T_1 player they meet. Since miscoordination only occurs whenever two T_1 players that send random messages meet, there will now be more equilibrium coordination compared to the standard case.

Second, consider one-way communication in the Battle of the Sexes. While T_1 receivers, and hence T_2 senders, behave as before, T_1 senders that believe they face imitators now send l and play H. In the previous footnote, we have already argued that this behavior is implausible and that the fraction of such T_1 players must therefore be small. However, irrespective of how small a proportion they constitute, T_2 receivers now play L irrespective of what message they receive. This implies that T_3 senders send h and play H. Under a standard type distribution, the outcome in terms of observed action profiles is thus the same as before.

Although some details of the analysis change with the introduction of heterogeneous T_1 players, we conclude that the main mechanisms are robust to this modification.

3. Extensions

So far, we have confined attention to 2×2 games. In principle it is straightforward to extend the analysis to games with more players and strategies.[15] In this section, we

[14] An alternative is to let T_1 assume that some fraction of T_0 imitates rather than randomizes. In this case, T_1 is sophisticated enough to consider heterogeneity among T_0. We do not think this is plausible, and the consequences are counterfactual too: Consider one-way communication in the Battle of the Sexes. If there is heterogeneity among T_0, T_1 will send l and play H—believing that some opponents ignore their message, whereas others imitate their message and play L. Since p_1 is typically estimated to be quite high, the implication is that sending l and playing H would be a relatively common practice. Cooper et al. (1989) studies one-way communication in Battle of the Sexes. They find that only 2 percent of all senders even sent an l message.

[15] We restrict attention to two-player games here, but an earlier version of this paper find similar results for some n-player games.

show that the reassurance property of communication extends to two-player games in which players' interests are sufficiently well aligned. When attractive non-equilibrium outcomes are present, however, senders might try obtain these by deceiving the opponent. The possibility of deception implies that one-way communication may hamper coordination on Nash equilibria.

3.1. Common interest games. The Stag Hunt example illustrates that pre-play communication facilitates the play of a risky payoff dominant equilibrium. Since our model does not assume equilibrium play, it is also applicable to situations in which players realistically fail to play a unique and efficient Nash equilibrium—such as the High Risk game, devised by Gilbert (1990) and reproduced in Figure 3 (in which best replies are marked with asterisks).[16] Absent communication, the level-k model predicts that two T_{5+} players coordinate on the unique pure strategy equilibrium (U, X) ,whereas all less sophisticated players fail to do so.[17] In contrast, one-way and two-way communication implies that T_{2+} coordinate on equilibrium. That is, much less sophistication is required to reach equilibrium with communication than without.[18]

FIGURE 3. High Risk game

	X	Y	Z
U	$5^*, 5^*$	$-50, -50$	$2, 4$
V	$-50, -50$	$2, 4^*$	$4^*, 3$
W	$4, 4^*$	$3^*, 3$	$3, 3$

The positive effect of communication in the High Risk game extends to all finite and normal form two-player games which has a payoff dominant equilibrium that gives strictly higher payoffs to both players than all other outcomes of a game, i.e., to all *common interest games.* For this class of games it is straightforward to show that T_{2+} coordinate on the payoff dominant equilibrium. The underlying mechanism is that

[16] Experimental results of Burton and Sefton (2004) confirm the prevalence of coordination failure in one-shot play of the High Risk game, but demonstrate that players learn to play the equilibrium after having played a number of practice rounds with the same opponent.

[17] To see this, note that T_1 plays W and Z since these are the risk dominant actions. Using the best responses indicated in Figure 3 it follows that T_2 plays V and X, T_3 plays U and Y, T_4 plays W and X, and finally that T_{5+} plays U and X.

[18] To see this, first consider one-way communication. A T_1 row sender sends w and plays W, while a column sender sends z and plays Z. A T_1 receiver best responds to messages. A T_{2+} row sender therefore sends u and plays U, while a column sender sends x and plays X, while a T_{2+} receiver best responds to messages. Now consider two-way communication. T_1 believes the opponent is truthful and therefore best responds to messages and randomize what message to send. A T_{2+} row player therefore sends u and plays U while a column player sends x and plays X.

since T_1 listens and best responds to messages, T_2 can achieve the best possible outcome by sending and playing the payoff dominant equilibrium.

Proposition 4. *Let G be a two-player common interest game. For type distributions with $p_0 = p_1 = 0$, players in $\Gamma_I(G)$ and $\Gamma_{II}(G)$ always coordinate on the payoff dominant Nash equilibrium.*

Proof. First consider $\Gamma_I(G)$. A T_1 sender sends and plays the action that is optimal given that the opponent randomizes uniformly over actions. If there are several such actions, T_1 plays each of them with equal probability and sends a truthful message. As a receiver, T_1 best responds to messages. Since the payoff dominant equilibrium gives the highest possible payoff, T_2 sends and plays the corresponding action as sender, while best responding to messages as receiver. It follows that T_{3+} behaves as T_2. Now consider $\Gamma_{II}(G)$. T_1 believes the opponent is truthful and therefore best responds to messages, but sends a random message. T_{2+} believes the opponent best responds and therefore sends and plays the payoff dominant equilibrium irrespective of the received message. □

3.2. Other games. In common interest games and in symmetric 2×2 games with one-way communication, players always represent their intentions truthfully. In other classes of games, however, this is not necessarily the case. Crawford (2003) already shows how deception arises naturally in a level-k model of communication in Hide-and-Seek games. We observe that deception can also arise in an asymmetric dominance solvable 2×2 game with a unique pure strategy equilibrium. Consider the game in Figure 4.

Figure 4. Asymmetric 2×2 game

	Y	Z
W	$3^*, 2^*$	$4^*, 0$
X	$0, 0$	$0, 1^*$

The game's unique pure strategy equilibrium is (W, Y). Since W and Y are the risk dominant actions, T_{1+} players coordinate on the (W, Y) equilibrium if they are not allowed to communicate. Now consider one-way communication. Suppose that the row player acts as sender and the column player acts as receiver. The T_1 sender sends w and plays W, while a T_1 receiver best responds to received messages. A T_2 sender therefore sends x, but plays W, while a T_2 receiver best responds to messages. T_3 sends x but plays W, while a T_3 receiver ignores messages and always plays Y. Whenever T_{3+}

players meet, the resulting outcome is the sender's preferred equilibrium, but not when less sophisticated players play. In contrast to Proposition 2, one-way communication leads to less equilibrium coordination than no communication unless all players carry out three or more thinking steps.

Proposition 2 does not generalize to symmetric two-player games with more than two actions either. To see this, consider the game in Figure 5.[19] This symmetric 3 × 3 game has a unique pure strategy equilibrium, (H, H), for all $n > 1$, but the game also has the asymmetric outcomes (H, L) and (L, H) that are attractive either to the row or column player. Since there is a third strategy, D, which has L as its best response, some senders will try to use this strategy to deceive the other player into playing L.

FIGURE 5. Symmetric 3 × 3 game

	H	L	D
H	$4/n^*, 4/n^*$	$(4+1/n)^*, 0$	$0, 0$
L	$0, (4+1/n)^*$	$0, 0$	$1^*, 1$
D	$0, 0$	$1, 1^*$	$0, 0$

Specifically, consider the case when $n = 1$ and pre-play communication is not possible. In that case T_1 would play H since it is the best action to take if the opponent randomizes uniformly, and T_{2+} would best respond by playing H. One-way communication, however, makes it more difficult to reach equilibrium. A T_1 sender sends h and plays H, while a T_1 receiver best responds (as indicated by the asterisks in Figure 5) to the received message. A T_2 sender sends d, but plays H, while a T_2 receiver best responds to received messages. A T_3 sender sends d and plays H, while a T_3 receiver plays H irrespective of the received message. A T_{4+} sender is indifferent about what message to send and is thus truthful, sending h and playing H; a T_{4+} receiver ignores messages and plays H. We conclude that T_{3+} coordinate on (H, H) and that one-way communication consequently lowers equilibrium coordination unless all players make three or more thinking steps.

A modification of the game illustrates how the number of thinking-steps required to reach equilibrium may increase linearly with the size of the game. Consider the $3N \times 3N$ game shown in Figure 6. It has the game in Figure 5 on the main diagonal and zero payoffs elsewhere.

Let messages be denoted m_n, with $m \in \{h, l, d\}$ and $n \in \{1, 2, ..., N\}$. Without communication, T_{1+} plays H_1 as in the 3 × 3 game. However, when one-way communication is allowed, all players must make at least $2N + 2$ thinking steps in order to

[19] This game is non-generic, but the analysis is analogous in the generic case.

coordinate on the unique equilibrium (H_1, H_1). To see why, note first that T_1 through T_3 will behave as in the 3×3 game, but that receivers will best-respond to all messages m_n with $n \in \{2, 3, ..., N\}$, believing those messages to come from T_0. A T_4 sender therefore sends d_2 and plays H_2 in order to get the outcome (H_2, D_2) which is preferred over (H_1, H_1). T_5 receivers do not believe in d_2 messages and therefore play H_2 if either h_2, l_2 or d_2 is played. In turn, T_6 senders send d_3 and play H_3 in order to induce the (H_3, L_3) outcome. The inductive argument continues like this up until T_{2N+1} sends d_N and plays H_N. A T_{2N+2} sender cannot hope to get anything better than (H_1, H_1) and therefore sends h_1 and plays H_1, whereas a T_{2N+2} receiver plays H_n whenever h_n, d_n or l_n is played (for all n).

FIGURE 6. Symmetric $3N \times 3N$ game

	H_1	L_1	D_1	H_2	L_2	D_2	$\cdots$	H_N	L_N	D_N
H_1	$4,4$	$5,0$	$0,0$	$0,0$	$0,0$	$0,0$	$\cdots$	$0,0$	$0,0$	$0,0$
L_1	$0,5$	$0,0$	$1,1$	$0,0$	$0,0$	$0,0$	$\cdots$	$0,0$	$0,0$	$0,0$
D_1	$0,0$	$1,1$	$0,0$	$0,0$	$0,0$	$0,0$	$\cdots$	$0,0$	$0,0$	$0,0$
H_2	$0,0$	$0,0$	$0,0$	$2,2$	$4.5,0$	$0,0$	$\cdots$	$0,0$	$0,0$	$0,0$
L_2	$0,0$	$0,0$	$0,0$	$0,4.5$	$0,0$	$1,1$	$\cdots$	$0,0$	$0,0$	$0,0$
D_2	$0,0$	$0,0$	$0,0$	$0,0$	$1,1$	$0,0$	$\cdots$	$0,0$	$0,0$	$0,0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$0,0$	$0,0$	$0,0$
H_N	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$\frac{4}{N},\frac{4}{N}$	$4+\frac{1}{N},0$	$0,0$
L_N	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$0,4+\frac{1}{N}$	$0,0$	$1,1$
D_N	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$0,0$	$1,1$	$0,0$

This example illustrates that the degree of sophistication required to play equilibrium increases with the size of the game. Since the degree of sophistication required is unrealistically high, in these games players coordinate better if they are unable to communicate.

3.3. Other communication protocols. Like much of the cheap talk literature, we have here considered communication of intentions. Messages are of the form "I plan to play...". What would happen if players communicated requests instead, that is if messages were of the form "I want you to play..."? While the model still admits a notion of truthfulness, the analysis would be quite different. For example, it is no longer clear that T_1 players should care about the messages that they receive, since T_0 players' requests may reveal nothing about their intentions. We thus expect that credulity will play a more important role than truthfulness in this case. Specifically, communication might now affect behavior if T_1 senders believe that receivers are credulous in the sense that they fulfill requests. Preliminary investigations suggest that the ensuing analysis offers a perspective on how cheap talk may be used to understand cheating in games, but we leave a fuller analysis for a separate paper.

4. Evidence

The level-k model of pre-play communication is primarily a model to explain initial responses, i.e., the behavior of players that play a game for the first time. If players gain experience of the game and the population of players, they are likely to change their model of opponents' behavior or perhaps think further and proceed to higher levels of reasoning. In experimental work on pre-play communication, players typically play the same game in several rounds. Strictly speaking, most of the available evidence is thus inadequate for our purposes. With this caveat in mind, let us briefly discuss some of the most relevant communication experiments.

Two papers contrast one-way and two-way communication in Stag Hunt games. Cooper et al. (1992) report that average coordination on the payoff dominant equilibrium is 0 percent without communication, 53 percent with one-way communication and 91 percent with two-way communication. This study therefore strongly suggests that communication plays a reassurance role.[20] Burton et al. (2005) on the other hand find that one-way communication results in 52 percent coordination on the payoff dominant equilibrium, whereas two-way communication led to average coordination on the payoff dominant equilibrium of only 34 percent. Both papers find that behavior varies substantially across sessions, indicating that heterogeneity in early rounds of the game affect players choices in later rounds.

In addition to these two studies, there is also a few studies of the Stag Hunt game that investigate either one-way or two-way communication. Duffy and Feltovich (2002) finds that one-way communication entails coordination on the payoff dominant equilibrium in 84 percent of the cases with one-way communication and in 61 percent of the cases without communication. Charness (2000) studies the effect of one-way communication in three versions of the Stag Hunt and finds 86 percent coordination on the payoff dominant equilibrium with one-way communication. Clark et al. (2001) study two-way communication in two different Stag Hunt games. In the first game, playing L yields the same payoff irrespective of the opponents behavior. In this game, coordination on the payoff dominant equilibrium is 2 percent without communication and 70 with two-way communication. In a standard Stag Hunt game, they find that

[20] Relatedly, Ellingsen and Johannesson (2004a) identifies a reassurance role of communication in hold-up games with multiple equilibria.

coordination on the payoff dominant equilibrium occurs in only 19 percent of the cases with two-way communication.

It is difficult to draw clear conclusions regarding pre-play communication in the Stag Hunt based on these studies. The degree of coordination on the payoff dominant equilibrium varies greatly and does not seem to systematically depend on the communication technology. Our analysis suggests that the precise interpretation of messages in terms of intentions or requests as well as the composition of the player population might cause some of the differences, but the only reliable way to find out is to conduct new experiments that systematically manipulate the communication design and subject pool.

For mixed motive games the picture seems clearer, although this may be due to fewer studies. Cooper et al. (1989) find that one-way communication results in a high degree of coordination in Battle of the Sexes. Averaged over several rounds of play, Cooper et al. (1989) report that one-way communication increases coordination from 48 percent without communication to 95 percent with one-way communication. With one round of two-way communication, coordination is 55 percent.[21] For a comparison of this evidence with the prediction of the rational cheap talk model, see Costa-Gomes (2002).

To summarize, we believe that more experimental work is needed in order to test the theory laid out in this paper. Such a test should focus on players' initial responses to several different games, which would allow a clearer separation of types. Costa-Gomes and Crawford (2006) illustrates how this can be done. It would also be useful to directly test the assumption about T_0 players. Since T_0 players mainly exist in the minds of other players, we need data on players' beliefs. Such data can be generated not only through belief elicitation (e.g., Costa-Gomes and Weizsäcker 2007), but also by response time measurement (e.g., Camerer et al. 1993 and Rubinstein 2007), information search (e.g., Camerer et al. 1993, Costa-Gomes et al. 2001 and Costa-Gomes and Crawford 2006) and through neuroimaging (e.g., Bhatt and Camerer 2005).

5. Concluding Remarks

The level-k model of bounded rationality captures many long-held intuitions both about the plausibility of Nash equilibrium play and about equilibrium selection. If

[21] It should be noted, however, that Cooper et al. (1989) allow the players to be silent and that 27 percent of the players in the two-way treatment, and 5 percent in the one-way treatment, choose to do so. We have not allowed silence in our analysis. It is of course possible to extend the message space to allow for silence, but we have chosen not to do so. Since players are assumed to have a slight preference for truthfulness, they might want to be silent when they don't know what action they are going to take in the action game (as T_1 under two-way communication in coordination games).

players cannot communicate, the model provides a precise sense in which equilibrium is unlikely in the High Risk game, and it also correctly predicts play of the risk dominant equilibrium in Stag Hunt. Our analysis demonstrates that the level-k model also allows a number of sharp and non-trivial predictions concerning the role of communication in non-zero-sum games.

For pre-play communication in the class of symmetric 2×2 games, we are able to characterize precisely the outcomes in all games and for all type distributions. Arguably, our most remarkable result is the proof that communication can create reassurance in coordination games even if messages are highly unlikely to be self-signaling. When players are sufficiently sophisticated, the mere belief that some player type thinks that some player type thinks...etc...that a message is self-signaling suffices to uniquely select the efficient outcome in Stag Hunt with communication. When there are relatively unsophisticated (level-1) players in the population, we moreover find that two-way communication may yield higher expected payoff in Stag Hunt than does one-way communication. The latter result is typically reversed in mixed motive games: when players rank equilibria differently, average payoffs are usually higher under one-way communication.

While we show that communication also has beneficial effects in general two-player common interest games, not all results from our analysis of symmetric 2×2 games extend readily to other classes of games. In particular, we demonstrate by example that one-way communication sometimes hampers coordination, unless players think implausibly many steps.

Appendix 1: Characterization of Behavior

We here characterize behavior in all symmetric and generic 2 × 2 games using the level-k model. Consider the symmetric 2 × 2 game in Figure A1.

Figure A1. Symmetric 2 × 2 game

	H	L
H	u_{HH}, u_{HH}	u_{HL}, u_{LH}
L	u_{LH}, u_{HL}	u_{LL}, u_{LL}

We assume that this game is generic in the sense that none of the four different payoffs (u_{HH}, u_{HL}, u_{LH} and u_{LL}) are identical. Depending on the relations $u_{HH} \lessgtr u_{LH}$ and $u_{LL} \lessgtr u_{HL}$, we can divide the class of generic 2×2 games into three familiar types of games as shown in Figure A2.[22]

If we were only interested in Nash equilibria, there would be only one prediction for each of these games. For the level-k model, however, these games will be divided into subclasses with different predictions. The most important distinction is indicated by the dashed line in the figure. This condition corresponds to whether $u_{LL} - u_{HL} \lessgtr u_{HH} - u_{LH}$, i.e., whether $u_{LH} + u_{LL} \lessgtr u_{HH} + u_{HL}$. This means that action H is risk dominant above the dashed line in Figure A2, whereas action L is risk dominant below it. For tractability, we disregard the cases when neither action is risk dominant throughout the paper.

Dominance solvable games. Dominance solvable games are easiest to analyze, but also least interesting. In a dominance solvable game, players always have an incentive to play the dominant action, and neither one-way or two-way communication affect the actions players take.

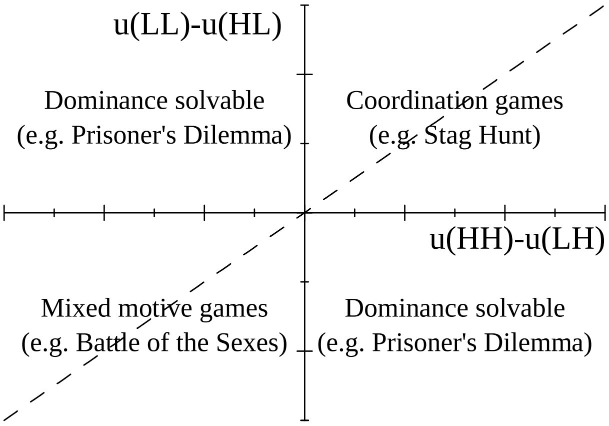
We assume $u_{HL} > u_{LL}$ and $u_{HH} > u_{LH}$ so that H(igh) is the dominant action. The case when L is the dominant action is symmetric.

[22] The classification of symmetric games follows Weibull (1995) closely. To understand how this classification arises, note that if we were only interested in Nash equilibria of 2×2 games, we could have substracted u_{LH} from both action H and L when the other player plays H and u_{HL} from both actions when the other player plays L. This would leave the equilibria of the game unchanged, whereas it affects the prediction for level-k models. The main reason is that in a level-k model, strategic uncertainty plays a role due to the randomization of level-0 players and we can therefore not use the sure-thing principle to transform the game. After the transformation, the game is the following.

	H	L
H	$u_{HH} - u_{LH}, u_{HH} - u_{LH}$	$0, 0$
L	$0, 0$	$u_{LL} - u_{HL}, u_{LL} - u_{HL}$

From this game it is clear why the class of symmetric games can be classified by two real numbers, $u_{HH} - u_{LH}$ and $u_{LL} - u_{HL}$.

FIGURE A2. The four types of generic and symmetric 2×2 games



OBSERVATION 1. *If players cannot communicate, T_{1+} plays the dominant action H. If players can communicate, then both one-way and two-way communication implies that T_{1+} sends h and plays H irrespective of any received messages.*

PROOF. Since H is a dominant action, T_{1+} players play H irrespective of the believed behavior of the opponent. With the possibility to communicate, this also implies that there are no players that respond to messages, and T_{1+} players are therefore indifferent about sending h or l. (Sending l would have been beneficial if some players responded to messages and $u_{HL} > u_{HH}$ as in the Prisoner's Dilemma.) However, since players have a lexicographic preference for truthfulness, they send h. □

For dominance solvable 2×2 games, communication plays no role. Except for some miscoordination due to T_0 playing the dominated action, all players play the dominant action. Since the proof only relies on the fact that each player has a strictly dominant strategy, the result extends to all normal form two-player games in which both players have a strictly dominant action.

Coordination games. Behavior in coordination games depends crucially on payoff and risk dominance. Since we restrict attention to generic games, one of the equilibria has to be payoff dominant. Let us without loss of generality assume that H(igh) is the payoff dominant equilibrium, i.e., $u_{HH} > u_{LL}$.

OBSERVATION 2. *(No communication) T_{1+} plays the risk dominant action.*

PROOF. T_1 players believe that the opponent randomizes uniformly and therefore plays the risk dominant action. T_2 players best respond and play the same risk dominant action, and so on. □

Absent communication, T_1 plays the best response to a uniformly randomizing T_0 opponent, which is the risk dominant action. Since this is a coordination game, more advanced players best respond by playing the same action.

OBSERVATION 3. *(One-way communication) If H is the risk dominant action, T_{1+} sends h and plays H as sender and responds to messages as receiver. If L is the risk dominant action, T_1 sends l and plays L as sender and responds to messages as receiver. T_{2+} sends h and plays H as sender and responds to messages as receiver.*

PROOF. First consider the case when H is risk dominant. T_1 plays $\langle h, H, H, L\rangle$ (facing randomizing T_0 receivers and truthful T_0 senders). A T_2 sender believes that the receiver best-responds to the sent message and therefore sends h and plays H. A T_2 receiver believes that the sender will send h and play H, but if T_2 receives message l, he believes it comes from a truthful T_0 sender. T_{2+} therefore plays $\langle h, H, H, L\rangle$.

Now consider the case when L is risk dominant. Then, T_1 plays $\langle l, L, H, L\rangle$. T_{2+} believes that the opponent responds to messages and that all messages are truthful and therefore play $\langle h, H, H, L\rangle$. □

When risk and payoff-dominance coincide, one-way communication is sufficient to achieve coordination among T_{1+} players. When there is a conflict between risk and payoff dominance, there is still perfect coordination among T_{1+} players, but there is more play of the risk dominant equilibrium (since a T_1 sender plays the action corresponding to that equilibrium).

OBSERVATION 4. *(Two-way communication) T_1 randomizes messages and responds to received messages, whereas T_{2+} sends h and plays H.*

PROOF. T_1 believes that the opponent is truthful and therefore best responds to the received message, while sending random messages (not knowing what action will be taken). T_2 believes that the opponent responds to messages and therefore sends and plays H irrespective of the message received (since T_1 sends a random message). T_3 therefore sends h and plays H. Receiving an unexpected L message, T_3 also plays H, believing the opponent to be T_1. More advanced players reason in the same way and thus also play $\langle h, H, H\rangle$. □

Mixed motive games. Two common examples of 2×2 mixed motive games are Chicken or Hawk-Dove and Battle of the Sexes. In order for the game to have mixed motive, we assume $u_{HL} > u_{LL}$ and $u_{LH} > u_{HH}$. Without loss of generality, we further assume that $u_{HL} > u_{LH}$ so that each player prefer the equilibrium where he is the one to play H(igh). If $u_{LL} = u_{HH} = 0$, then this game is the Battle of the Sexes, whereas it is a Chicken game if $u_{LL} > u_{HH}$. Battle of the Sexes is a non-generic game, but the results in this section hold also for the Battle of the Sexes.

OBSERVATION 5. *(No communication) If H is the risk dominant action, then T_k plays H if k is odd and L if k is even. If L is the risk dominant action, then T_k plays L if k is odd and H if k is even.*

PROOF. T_1 plays the risk dominant action and T_k best-responds to the behavior of T_{k-1}, which generates the alternating behavior. □

With no possibility to communicate, there is little players can do to coordinate on either of the asymmetric equilibria and behavior therefore alternates over thinking steps. One-way communication, on the other hand, provides a way to break the symmetry inherent in the game.

OBSERVATION 6. *(One-way communication) If H is the risk dominant action, then T_{1+} sends h and plays H as sender and responds to messages as receiver. If L is the risk dominant action, then T_1 sends l and plays L as sender and responds to messages as receiver. T_{2+} sends h and plays H as sender and responds to messages as receiver.*

PROOF. First let H be the risk dominant action. A T_1 sender faces a randomizing receiver and therefore plays H and sends h. A T_1 receiver, on the other hand, responds to the sent message, believing it comes from a truthful T_0 opponent. T_{2+} can get the preferred equilibrium as sender and therefore sends h and plays H, while responding to messages as receiver. If instead L is the risk dominant action, a T_1 sender instead sends and plays L, but otherwise behavior is unchanged. □

In general, senders play their preferred equilibrium and receivers yield and play their least preferred equilibrium. However, if the preferred equilibrium does not coincide with the risk dominant action, T_1 senders send and play their least preferred equilibrium.[23]

OBSERVATION 7. *(Two-way communication) T_1 sends h and l with equal probabilities and responds to messages. The behavior of T_{2+} players cycles in thinking steps of six as follows: $\langle h, H, H\rangle$,$\langle l, L, L\rangle$,$\langle h, L, H\rangle$,$\langle h, H, H\rangle$,$\langle l, L, H\rangle$,$\langle h, L, H\rangle$.*

[23] The result when L is risk dominant is sensitive to the assumption that T_{1+} players have lexicographic preferences for truthfulness. Without that preference, level-1 senders would send random messages. Then, the behavior of more advanced players would alternate and entail many instances of miscoordination.

PROOF. T_1 believes that the opponent is truthful and therefore sends random messages, but responds to the message sent. T_2 believes that the opponent responds to messages and therefore plays $\langle h, H, H\rangle$. T_3 expects to receive a truthful h message, and thus sends l and plays L. If receiving an l message, T_3 believes it comes from a T_1 opponent and therefore plays L (believing the opponent will play H). T_4 expects to play H and therefore sends h. If receiving the message h, T_4 believes it comes from a T_2 opponent and therefore responds by playing L. T_5 thinks the opponent responds to messages and therefore plays H and sends h. Believe an l message comes from a T_2 opponent, T_5 subsequently plays H. T_6 expects to play L and therefore sends l, but plays H upon receiving an l message (believing it comes from a T_2 opponent). T_7 expects to play H and sends an h message, playing L if receiving an h message. T_8 sends h and plays H, playing H if he receives an l message, just like T_2. T_9 plays $\langle l, L, L\rangle$ just like T_3. Since the behavior of eight and nine-level players is just like two- and three-level players, and the rationale for T_{4+} did not depend on the behavior of T_0 or T_1, behavior continues to cycle like this. □

Note that the behavior of T_0, T_1, T_2, and T_3 is identical to Crawford (2007). However, T_4 responds to received messages in our model, but always plays H in Crawford (2007). The difference stems from the fact that we assume that whenever T_4 receives the message h, the inference is that it comes from a T_2 player that will actually play H, whereas Crawford (2007) assumes that T_4 believes an h message is a mistake by a T_3 opponent who will play H anyway.[24]

Comparing one-way and two-way communication, it is clear that two-way communication will lead to several instances of miscoordination. However, as pointed out by Crawford (2007), the degree of coordination may still be higher than predicted by Farrell (1987) and Rabin (1994).

Finally, note the parallel to coordination games that risk-dominance only plays a role with one-way communication. The underlying reason is the strategic uncertainty resulting from randomizing T_0 receivers.

[24] Also note that although our T_3 behaves as in Crawford (2007), the rationale for their behavior is slightly different. T_3 in our framework believes an l message comes from a T_1 opponent that sends random messages. Since T_3 sent the message l, the player believes that the opponent will play H and they therefore play L. In Crawford (2007), a T_3 player that receives the counterfactual message l believes that it was a mistake by the T_2 opponent and therefore plays L anyway.

Appendix 2: Proofs

Proof of Proposition 1. From Observation 1 we know that communication has no effect in dominance solvable games. Similarly, for coordination games when H is risk dominant, Observation 2 and 3 show that communication has no effect. In coordination games when L is risk dominant, however, Observation 2 and 3 show that one-way communication results in either (L, L) or (H, H), whereas no communication results in (L, L). As long as there is a positive fraction of T_{2+} players, one-way communication therefore results in higher expected payoffs.

For mixed motive games, first suppose L is risk dominant. From Observation 6 we know that one-way communication always induces coordination when T_{1+} play, so the expected payoff for a player playing the game is $(u_{HL} + u_{LH})/2$. However, as noted in Observation 5, no communication results in miscoordination when two odd-level players meet as well as when two even-level players meet. Under the standard type distribution, a player's average payoff is

$$p_2^2 u_{HH} + p_2 (1 - p_2) u_{HL} + (1 - p_2) p_2 u_{LH} + (1 - p_2)^2 u_{LL}.$$

One-way communication results in higher expected payoff whenever

$$\left(\frac{1}{2} - p_2 (1 - p_2)\right) (u_{HL} + u_{LH}) > p_2^2 u_{HH} + (1 - p_2)^2 u_{LL}.$$

A sufficient condition is that $u_{LL} < u_{HL}$ (we already know that $u_{HH} < u_{LH}$), but the necessary condition depends on p_2. Now let H be the risk dominant outcome. The expected payoff for communicating players is unchanged, whereas the condition for one-way communication to result in higher expected payoff is

$$\left(\frac{1}{2} - p_2 (1 - p_2)\right) (u_{HL} + u_{LH}) > (1 - p_2)^2 u_{HH} + p_2^2 u_{LL}.$$

Proof of Corollary 1. From the proof of Proposition 1 it follows directly that one-way communication only decreases average payoffs if one of the conditions hold with opposite inequality. To see why the corresponding game is a Chicken, suppose first that L is risk dominant. The first condition in Proposition 1 for one-sided communication to decrease expected payoffs is

$$\text{(A1)} \qquad \left(\frac{1}{2} - p_2 (1 - p_2)\right) (u_{HL} + u_{LH}) < p_2^2 u_{HH} + (1 - p_2)^2 u_{LL}.$$

We know that $u_{HL} > u_{LL}$, $u_{LH} > u_{HH}$ and $u_{HL} > u_{LH}$. This implies that $u_{HH} < (u_{LH} + u_{HL})/2$. Suppose that $u_{LL} \leq (u_{LH} + u_{HL})/2$. Then the right hand side of

(A1) satisfies

$$\begin{aligned} p_2^2 u_{HH} + (1-p_2)^2 u_{LL} &< p_2^2 \frac{1}{2}(u_{LH}+u_{HL}) + \frac{1}{2}(1-p_2)^2 (u_{LH}+u_{HL}) \\ &= \left(\frac{1}{2} - p_2(1-p_2)\right)(u_{LH}+u_{HL}). \end{aligned}$$

This implies that (A1) cannot hold, and therefore the condition must fail unless $u_{LL} > \frac{1}{2}(u_{LH}+u_{HL})$. This implies that $u_{LL} > u_{HH}$, which implies that it is a Chicken. An analogous argument can be made when H is risk dominant.

Proof of Proposition 2. From Observation 1 we know that communication has no effect in dominance solvable games.

From Observation 2 and 3, we know that the outcomes of coordination games in which L is the risk dominant action. These are given in Table A1. Pairwise comparison of the cells in Table A1 reveals that one-way communication entails weakly more coordination.

Table A1. Action profiles played in coordination games (L risk dominant)

G (no communication)			$\Gamma_I(G)$ *(one-way communication)*			
	0	≥ 1		$0R$	$1R$	$\geq 2R$
0	Uniform	$\frac{1}{2}LL, \frac{1}{2}LH$	$0S$	Uniform	$\frac{1}{2}HH, \frac{1}{2}LL$	$\frac{1}{2}HH, \frac{1}{2}LL$
≥ 1	$\frac{1}{2}LL, \frac{1}{2}LH$	LL	$1S$	$\frac{1}{2}LL, \frac{1}{2}LH$	LL	LL
			$\geq 2S$	$\frac{1}{2}HH, \frac{1}{2}HL$	HH	HH

If instead H is risk dominant, the outcomes are given in Table A2. The degree of coordination is again the same or higher with one-way communication than without communication.

Table A2. Action profiles played in coordination games (H risk dominant)

G (no communication)			$\Gamma_I(G)$ *(one-way communication)*		
	0	≥ 1		$0R$	$\geq 1R$
0	Uniform	$\frac{1}{2}HH, \frac{1}{2}HL$	$0S$	Uniform	$\frac{1}{2}HH, \frac{1}{2}LL$
≥ 1	$\frac{1}{2}HH, \frac{1}{2}HL$	HH	$\geq 1S$	$\frac{1}{2}HH, \frac{1}{2}HL$	HH

Now consider mixed motive games. Observations 5 and 6 yield the outcomes reported in Table A3 when L is risk dominant. Pairwise comparisons of cells reveal that the degree of coordination is higher with one-way communication.

Table A3. Action profiles played in mixed motive games (L risk dominant)

G (no communication)				$\Gamma_I(G)$ *(one-way communication)*			
	0	Odd	Even		$0R$	$1R$	$\geq 2R$
0	Uniform	$\frac{1}{2}HL, \frac{1}{2}LL$	$\frac{1}{2}LH, \frac{1}{2}HH$	$0S$	Uniform	$\frac{1}{2}HL, \frac{1}{2}LH$	$\frac{1}{2}HL, \frac{1}{2}LH$
Odd	$\frac{1}{2}LH, \frac{1}{2}LL$	LL	LH	$1S$	$\frac{1}{2}LH, \frac{1}{2}LL$	LH	LH
Even	$\frac{1}{2}HL, \frac{1}{2}HH$	HL	HH	$\geq 2S$	$\frac{1}{2}HL, \frac{1}{2}HH$	HL	HL

Finally, when H is risk dominant, the outcomes are given in Table A4. Again the degree of coordination is the same or higher for one-way communication for all combinations of types.

Table A4. Action profiles played in mixed motive games (H risk dominant)

G (no communication)				$\Gamma_I(G)$ *(one-way communication)*		
	0	Odd	Even		$0R$	$\geq 1R$
0	Uniform	$\frac{1}{2}LH, \frac{1}{2}HH$	$\frac{1}{2}HL, \frac{1}{2}LL$	$0S$	Uniform	$\frac{1}{2}HL, \frac{1}{2}LH$
Odd	$\frac{1}{2}HL, \frac{1}{2}HH$	HH	HL	$\geq 1S$	$\frac{1}{2}HL, \frac{1}{2}HH$	HL
Even	$\frac{1}{2}LH, \frac{1}{2}LL$	LH	LL			

Proof of Proposition 3. As Observation 1 shows, communication plays no role in dominance solvable games, so two-way communication cannot increase expected payoffs. In coordination games in which H is risk dominant, Observation 3 and 4 imply that $\Gamma_I(G)$ and $\Gamma_{II}(G)$ yield identical outcomes unless two T_1 players meet. In $\Gamma_I(G)$, players then coordinate on (H, H), whereas there is miscoordination in $\Gamma_{II}(G)$. Thus $\Gamma_I(G)$ is weakly better than $\Gamma_{II}(G)$ in this case. When instead L is the risk dominant action, T_1 senders always play L. The average payoff associated with $\Gamma_I(G)$ is thus

$$p_1(1-p_1)u_{LL} + p_1(1-p_1)u_{HH} + (1-p_1)(1-p_1)u_{HH} + p_1^2 u_{LL}.$$

The average payoff associated with $\Gamma_{II}(G)$ is

$$2p_1(1-p_1)u_{HH} + (1-p_1)(1-p_1)u_{HH} + \frac{1}{4}p_1^2(u_{LL} + u_{HH} + u_{LH} + u_{HL}).$$

Two-way communication thus yields higher payoff whenever

$$(4-3p_1)u_{HH} + p_1(u_{LH} + u_{HL}) > (4-p_1)u_{LL}.$$

Now consider mixed motive games. Observation 6 shows that for T_{1+} players, $\Gamma_I(G)$ entails perfect coordination, implying an average payoff of $(u_{LH} + u_{HL})/2$. As shown in Observation 7, matters are generally more complicated for $\Gamma_{II}(G)$ since behavior cycles over six thinking steps. Table A5 provides the resulting outcomes when confining attention to standard type distributions.

Table A5. Action profiles played in mixed motive games

$\Gamma_{II}(G)$ *(two-way communication)*			
	1	2	3
1	Uniform	LH	HL
2	HL	HH	HL
3	LH	LH	LL

We know that $(u_{LH}+u_{HL})/2 > u_{HH}$. However, if $u_{LL} > (u_{LH}+u_{HL})/2$ then two-way communication might be preferable. Two-way communication is preferable to one-way communication whenever

$$\left(p_2p_1+p_1p_3+p_2p_3+\frac{1}{4}p_1^2\right)(u_{HL}+u_{LH})+\left(p_3^2+\frac{1}{4}p_1^2\right)u_{LL}+\left(p_2^2+\frac{1}{4}p_1^2\right)u_{HH}$$
$$>\frac{1}{2}(u_{LH}+u_{HL}).$$

Letting $p_2=(1-p_1-p_3)$ we can rewrite this as

$$\frac{u_{LL}-u_{HH}}{u_{LH}+u_{HL}-2u_{HH}} > 1+\frac{2(p_1-1)(p_1-1+2p_3)}{p_1^2+4p_3^2}.$$

A necessary condition for this inequality to hold is that $u_{LL} > (u_{LH}+u_{HL})/2$. This follows from the fact that the minimum of the right hand side is $1/2$, whereas the left hand side can only be larger than $1/2$ if $u_{LL} > (u_{LH}+u_{HL})/2$.

References

Anderlini, L. (1999), "Communication, Computability, and Common Interest Games", *Games and Economic Behavior* 27, 1–37.

Aumann, R. J. (1990), "Nash Equilibria are not Self-Enforcing", in J. Gabszewicz, J. F. Richard and L.Wolsey, eds, *Economic Decision-Making: Games, Econometrics and Optimization*, Elsevier, Amsterdam, chapter 10, pp. 201–206.

Banerjee, A. and Weibull, J. W. (2000), "Neutrally Stable Outcomes in Cheap-Talk Coordination Games", *Games and Economic Behavior* 32(1), 1–24.

Bhatt, M. and Camerer, C. F. (2005), "Self-Referential Thinking and Equilibrium as States of Mind in Games: fMRI Evidence", *Games and Economic Behavior* 52(2), 424–459.

Blume, A. (1998), "Communication, Risk, and Efficiency in Games", *Games and Economic Behavior* 22(2), 171–202.

Burton, A., Loomes, G. and Sefton, M. (2005), "Communication and Efficiency in Coordination Game Experiments", in J. Morgan, ed., *Experimental and Behavioral Economics*, Vol. 13 of Advances in Applied Microeconomics, JAI Press, pp. 63–85.

Burton, A. and Sefton, M. (2004), "Risk, Pre-play Communication and Equilibrium", *Games and Economic Behavior* 46(1), 23–40.

Cai, H. and Wang, J. T.-Y. (2006), "Overcommunication in Strategic Information Transmission Games", *Games and Economic Behavior* 56(1), 7–36.

Camerer, C. F., Ho, T.-H. and Chong, J.-K. (2004), "A Cognitive Hierarchy Model of Games", *Quarterly Journal of Economics* 119(3), 861–898.

Camerer, C., Johnson, E., Rymon, T. and Sen, S. (1993), "Cognition and Framing in Sequential Bargaining for Gains and Losses", in K. Binmore, A. Kirman and P. Tani, eds, *Frontiers of Game Theory*, MIT Press, Boston, pp. 27–48.

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

Clark, K., Kay, S. and Sefton, M. (2001), "When are Nash Equilibria Self-Enforcing? An Experimental Analysis", *International Journal of Game Theory* 29(4), 495–515.

Cooper, R., DeJong, D. V., Forsythe, R. and Ross, T. W. (1989), "Communication in the Battle of the Sexes Game: Some Experimental Results", *RAND Journal of Economics* 20(4), 568–587.

Cooper, R., DeJong, D. V., Forsythe, R. and Ross, T. W. (1992), "Communication in Coordination Games", *Quarterly Journal of Economics* 107(2), 739–771.

Costa-Gomes, M. (2002), "A Suggested Interpretation of Some Experimental Results on Pre-play Communication", *Journal of Economic Theory* 104, 104–136.

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.

Costa-Gomes, M. A. and Weizsäcker, G. (2007), "Stated Beliefs and Play in Normal-form Games", *Review of Economic Studies* (forthcoming).

Costa-Gomes, M., Crawford, V. and Broseta, B. (2001), "Cognition and Behavior in Normal-Form Games: An Experimental Study", *Econometrica* 69(5), 1193–1235.

Crawford, V. (2003), "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions", *American Economic Review* 93(1), 133–149.

Crawford, V. (2007), "Let's talk it over: Coordination via preplay communication with level-k thinking", mimeo.

Crawford, V. P. and Iriberri, N. (2007), "Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games", *American Economic Review* 97(5), 1731–1750.

Crawford, V. and Sobel, J. (1982), "Strategic Information Transmission", *Econometrica* 50, 1141–1152.

Demichelis, S. and Weibull, J. W. (2008), "Language, meaning and games – A model of communication, coordination and evolution", *American Economic Review*, forthcoming.

Duffy, J. and Feltovich, N. (2002), "Do Actions Speak Louder Than Words? An Experimental Comparison of Observation and Cheap Talk", *Games and Economic Behavior* 39(1), 1–27.

Ellingsen, T. and Johannesson, M. (2004*a*), "Is There a Hold-Up Problem?", *Scandinavian Journal of Economics* 106, 475–494.

Ellingsen, T. and Johannesson, M. (2004*b*), "Promises, Threats, and Fairness", *Economic Journal* 114, 397–420.

Ellingsen, T. and Östling, R. (2006), "Organizational Structure as the Channeling of Boundedly Rational Pre-play Communication", *SSE/EFI Working Paper Series in Economics and Finance* No 634, Stockholm School of Economics.

Farrell, J. (1987), "Cheap Talk, Coordination, and Entry", *RAND Journal of Economics* 18(1), 34–39.

Farrell, J. (1988), "Communication, Coordination and Nash Equilibrium", *Economic Letters* 27(3), 209–214.

Farrell, J. and Rabin, M. (1996), "Cheap Talk", *Journal of Economic Perspectives* 10(3), 103–118.

Gilbert, M. (1990), "Rationality, Coordination, and Convention", *Synthese* 84(1), 1–21.

Kim, Y.-G. and Sobel, J. (1995), "An Evolutionary Approach to Pre-play Communication", *Econometrica* 63(5), 1181–1193.

Matsui, A. (1991), "Cheap Talk and Cooperation in Society", *Journal of Economic Theory* 54(2), 245–258.

Myerson, R. (1989), "Credible Negotiation Statements and Coherent Plans", *Journal of Economic Theory* 48(1), 264–303.

Nagel, R. (1995), "Unraveling in Guessing Games: An Experimental Study", *American Economic Review* 85(5), 1313–1326.

Nowak, M. (2006), *Evolutionary Dynamics*, Belknap Press of Harvard University Press, Cambridge.

Pinker, S. and Bloom, P. (1990), "Natural Language and Natural Selection", *Behavioral and Brain Sciences* 13(4), 707–784.

Rabin, M. (1990), "Communication Between Rational Agents", *Journal of Economic Theory* 51(1), 144–170.

Rabin, M. (1994), "A Model of Pre-Game Communication", *Journal of Economic Theory* 63(2), 370–391.

Rubinstein, A. (2007), "Instinctive and Cognitive Reasoning: A Study of Response Times", *Economic Journal* 117, 1243–1259.

Schelling, T. C. (1966), *Arms and Influence*, Yale University Press, New Haven.

Stahl, D. O. and Wilson, P. W. (1994), "Experimental evidence on players' models of other players", *Journal of Economic Behavior and Organization* 25(3), 309–327.

Stahl, D. O. and Wilson, P. W. (1995), "On Players' Models of Other Players: Theory and Experimental Evidence", *Games and Economic Behavior* 10(1), 33–51.

Wärneryd, K. (1991), "Evolutionary Stability in Unanimity Games with Cheap Talk", *Economic Letters* 36(4), 375–378.

Weibull, J. W. (1995), *Evolutionary Game Theory*, MIT Press.

Wengström, E. (2007), "Setting the Anchor: Price Competition, Level-n Theory and Communication", *Department of Economics Working Paper Series* No 2007:6, Lund University.

PAPER 2

Strategic Thinking and Learning in the Field and Lab: Evidence from Poisson LUPI Lottery Games

with Joseph Tao-yi Wang, Eileen Chou and Colin F. Camerer

Abstract. Game theory is usually difficult to test precisely in the field because predictions typically depend sensitively on features that are not controlled or observed. We conduct a rare such test using field data from the lowest unique positive integer (LUPI) game. In the LUPI game, players pick positive integers and the player who chose the lowest unique number (not chosen by anyone else) wins a fixed prize. We derive theoretical equilibrium predictions, assuming fully rational players with Poisson-distributed uncertainty about the number of players. We also derive predictions for boundedly rational players using quantal response equilibrium, a cognitive hierarchy of rationality steps with quantal responses, as well as a simple learning model based on imitation. The theoretical predictions are tested using both field data from a Swedish gambling company, and laboratory data from a scaled-down version of the field game. The field and lab data show similar patterns: players choose very low and very high numbers too often, and avoid focal ("round") numbers. However, there is learning and a surprising degree of convergence toward equilibrium. The cognitive hierarchy model with quantal responses can account for some of the basic discrepancies between the equilibrium prediction and the data, and the learning model can account for the adaptation towards equilibrium.

1. Introduction

Game theory seeks to explain decision-making in interactive situations. However, clear tests of game theoretical predictions using field data are rare because predictions are

The first two authors, Joseph Tao-yi Wang and Robert Östling, contributed equally to this paper. We are grateful for helpful comments from Tore Ellingsen, Magnus Johannesson, Botond Köszegi, David Laibson, Erik Lindqvist, Stefan Molin, Noah Myung, Rosemarie Nagel, Charles Noussair, Carsten Schmidt, Dmitri Vinogradov, Mark Voorneveld, Jörgen Weibull, seminar participants at California Institute of Technology, Stockholm School of Economics, Mannheim Empirical Research Summer School 2007, UC Santa Barbara Cognitive Neuroscience Summer School 2007, Research Institute of Industrial Economics (Stockholm), London School of Economics, Institute of International Economic Studies (Stockholm), Carnegie Mellon University, Kellogg School of Management, Copenhagen University and Oslo University. Robert Östling acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation. Colin Camerer acknowledges support from the NSF HSD program, HSFP, and the Betty and Gordon Moore Foundation.

often sensitive to details about strategies, information and payoffs that are difficult to observe in the field. As Robert Aumann pointed out: “In applications, when you want to do something on the strategic level, you must have very precise rules; [...] An auction is a beautiful example of this, but it is very special. It rarely happens that you have rules like that (cited in van Damme 1998, p. 196).”

In this paper we exploit such a “rare happening”: field data from a Swedish lottery game created in 2007. In the lottery, players simultaneously choose positive integers from 1 to K. The winner is the player who chooses the lowest number that nobody else picked. We call this the LUPI game, because the *l*owest *u*nique *p*ositive *i*nteger wins.[1] This paper analyzes LUPI theoretically and reports data from the Swedish field experience and from parallel lab experiments.

In addition to testing equilibrium theory using field data in an unusually straightforward way, special properties of the LUPI data enable us to make four other contributions:

Applying Poisson game theory: The number of players is not fixed. Normally, finding equilibria with many strategies and an unknown number of players is extremely difficult computationally. However, we apply the theory of Poisson games which assumes that the number of players is Poisson-distributed (Myerson 1998).[2] Remarkably, assuming a variable number of players rather than a fixed number makes computation of equilibrium *simpler* (provided the number of players is Poisson-distributed). The LUPI data provide the first empirical test of Poisson-Nash equilibrium.

Measuring learning: Every day 53,783 people played (on average) and the lottery was played each day for 49 consecutive days. The large number of players gives enough statistical power to study the rate of learning across the time series, which most other field studies of can not.[3]

[1] The Swedish company called the game Limbo, but we think LUPI is more mnemonic, and more apt because in the typical game of limbo, two players who tie in how low they can slide underneath a bar do not lose.

[2] This also distinguishes our paper from the ongoing research on unique bid auctions by Eichberger and Vinogradov (2007), Raviv and Virag (2007) and Rapoport et al. (2007) which all assume that the number of players is fixed and commonly known.

[3] A few studies have tested mixed-strategy equilibrium using field data from sports where mixing is expected to occur, like tennis and soccer (Walker and Wooders 2001, Chiappori et al. 2002, Palacios-Huerta 2003 and Hsu et al. 2007). These studies use highly experienced players and the studies on soccer pool data across substantial spans of time to be able to test the mixed equilibrium prediction powerfully. They do not study how players learn to play equilibrium. However, Chiappori et al. (2002) provide some suggestive evidence by noting that among the kickers with the most experience in their sample (those with eight or more kicks) only one of nine fails a randomness test at the 10% level. However, this is a very crude test for learning effects compared to our data which compare a much larger sample of choices over a longer span with day-by-day comparisons. There are also some studies of randomization in naturally-occurring risky choices (e.g., Sundali and Croson 2006) which are not strategic.

Comparing models of bounded rationality: The simple LUPI structure allows us to compare Poisson-Nash equilibrium predictions with predictions of two parametric models of boundedly rational play—quantal response equilibrium, and a level-k or cognitive hierarchy approach. These theories have been developed and refined using experimental data. The LUPI data allows us to study these models using both field and laboratory data.

Lab-field parallelism: It is easy to run a lab experiment that matches the key features of the game played in the field. This close match adds to a small amount of evidence of how well experimental lab data can generalize to a particular field setting when the experiment was specifically intended to do so.

While LUPI is not an exact model of anything that social scientists usually care about, it combines strategic features of interesting naturally-occurring games. For example, in games with congestion, a player's payoffs are lower if others choose the same strategy. Examples include choices of traffic routes and research topics, or buyers and sellers choosing among multiple markets. LUPI has the property of an extreme congestion game, in which having even one other player choose the same number reduces one's payoff to zero.[4] Indeed, LUPI is similar to a game in which being first matters (e.g., in a patent race), but if players are tied for first they do not win. One close market analogue to LUPI is the lowest unique bid auction (see the ongoing research by Eichberger and Vinogradov 2007, Houba et al. 2008, Raviv and Virag 2007 and Rapoport et al. 2007). In these increasingly popular auctions, an object is sold to the lowest bidder whose bid is unique (or in some versions, to the highest unique bidder). LUPI is simpler because winners don't have to pay the amount they bid and there are no private valuations and beliefs about valuations of others, but contains the same essential strategic conflict: players want to choose low numbers, in order to be the lowest, but also want to avoid numbers others will choose, in order to be unique.

In sum, our contribution is that the LUPI game permits an unusually sharp test of game theory and of the speed and nature of learning in the field, provides an initial field test of Poisson-Nash equilibrium, can be used to compare models of bounded rationality, and can be recreated closely in a lab experiment.

The next section provides a theoretical analysis of a simple form of the LUPI game, including the (symmetric) Poisson-Nash equilibrium, quantal response equilibrium and cognitive hierarchy behavioral models. Section 3 and 4 analyze the field and lab data, respectively. Section 5 discusses learning. Section 6 concludes the paper.

[4] Note, however, that LUPI is not a congestion game as defined by Rosenthal (1973) since the payoff from choosing a particular number does not only depend on how many other players that picked that number, but also on how many that picked lower numbers.

2. Theory

In the simplest form of LUPI, the number of players, N, has a known distribution, the players choose integers from 1 to K simultaneously, and the lowest unique number wins. The winner earns a payoff of 1, while all others earn 0.[5]

We first analyze the game when players are assumed to be fully rational, best-responding, and have equilibrium beliefs. We focus on symmetric equilibria since players are generally anonymous to each other. We also assume that the number of players is a random variable that has a Poisson distribution, which is much easier to work with and is a plausible approximation in the field (and can be exactly implemented in the lab).[6] Appendix A discusses the fixed-n equilibrium and why it is so much more difficult to compute than the Poisson-Nash equilibrium. We then discuss the quantal response equilibrium (QRE), and predictions from a cognitive hierarchy model with quantal response.

2.1. Properties of Poisson Games.

In this section, we briefly summarize the theory of Poisson games developed by Myerson (Myerson 1998, 2000), which is then used in the next section to characterize the Poisson-Nash equilibrium in the LUPI game.

Games with population uncertainty relax the assumption that the exact number of players is common knowledge. In particular, in a *Poisson game* the number of players N is a random variable that follows a Poisson distribution with mean n. We have

$$N \sim \text{Poisson}(n): \quad N = k \text{ with probability } \frac{e^{-n}n^k}{k!}$$

and, in the case of a Bayesian game, players' types are independently determined according to the probability distribution $r = (r(t))_{t \in T}$ on some type space T.[7] Let a type profile be a vector of non-negative integers listing the number of players of each type t in T, and let $Z(T)$ be the set of all such type profiles in the game. Combining

[5] In this stylized case, we assume that if there is no lowest unique number there is no winner. This simplifies the analysis because it means that only the probability of being unique must be computed. In the Swedish game, if there is no unique number then the players who picked the smallest and least-frequently-chosen number share the top prize. This is just one of many small differences between the simplified game analyzed in this section and the game as played in the field, which are discussed further below.

[6] Players did not know the number of total bets in both the field and lab versions of the LUPI game. Although players in the field could get information about the current number of bets that had been made so far during the day, players had to place their bets before the game closed for the day and therefore could not know with certainty the total number of players that would participate in that day.

[7] The LUPI game itself is not a Bayesian game. However, in the cognitive hierarchy model (developed in Section 2.4), there are players with different degree of strategic sophistication and we therefore include types in our presentation of Poisson games in this section.

N and r can describe the *population uncertainty* with the distribution $y \sim Q(y)$ where $y \in Z(T)$ and $y(t)$ is the number of players of type $t \in T$.

Players have a common finite action space C with at least two alternatives, which generates an *action profile* $Z(C)$ containing the number of players that choose each action. Utility is a bounded function $U : Z(C) \times C \times T \to \mathbb{R}$, where $U(x, b, t)$ is the payoff of a player with type t, choosing action b, and facing an opponent action profile of x. Let $x(c)$ denote the number of other players playing action $c \in C$.

Myerson (1998) shows that the Poisson distribution has two important properties that are relevant for Poisson games and simplify computations dramatically. The first is the *decomposition property*, which in the case of Poisson games imply that the distribution of type profiles for any $y \in Z(T)$ is given by

$$Q(y) = \prod_{t \in T} \frac{e^{-nr(t)}(nr(t))^{y(t)}}{y(t)!}.$$

Hence, $\tilde{Y}(t)$, the random number of players of type $t \in T$, is Poisson with mean $nr(t)$, and is independent of $\tilde{Y}(t')$ for any other $t' \in T$. Moreover, suppose each player independently plays the mixed strategy σ, choosing action $c \in C$ with probability $\sigma(c|t)$ given his type t. Then, by the decomposition property, the number of players of type t that chooses action c, $\overline{Y}(c, t)$, is Poisson with mean $nr(t)\sigma(c|t)$ and is independent of $\overline{Y}(c', t')$ for any other c', t'.

The second property of Poisson distributions is the *aggregation property* which states that any sum of independent Poisson random variables is Poisson distributed. This property implies that the number of players (across all types) who choose action c, $\tilde{X}(c)$, is Poisson with mean $\sum_{t \in T} nr(t)\sigma(c|t)$, independent of $\tilde{X}(c')$ for any other $c' \in C$. We refer to this property of Poisson games as the *independent actions* (IA) property.

Myerson (1998) also shows that the Poisson game has another useful property: *environmental equivalence* (EE). Environmental equivalence means that conditional on being in the game, a type t player would perceive the population uncertainty as an outsider would, i.e., $Q(y)$.[8] If the strategy and type spaces are finite, Poisson games are the only games with population uncertainty that satisfy both IA and EE (Myerson 1998). EE is a surprising property. Take a Poisson LUPI game with 27 players on average. In our lab implementation, a large number of players are recruited and are told that the number of players who will be active in each period varies. Consider a player who is told she is active. On the one hand, she might then act as if she is playing against the number of opponent players is Poisson-distributed with a mean of

[8] In particular, for a Poisson game, the number of opponents he faces is also a random variable of Poisson(n).

26 (since her active status has lowered the mean of the number of remaining players). On the other hand, the fact that she is active is a clue that the number of players in that period is large, not small. If N is Poisson-distributed the two effects exactly cancel out so all active players in all periods act as if they face a Poisson-distributed number of opponents. EE, combined with IA, makes the analysis rather simple.

A (symmetric) *equilibrium* for the Poisson game is defined as a strategy function σ such that every type assigns positive probability only to actions that maximize the expected utility for players of this type; that is, for every action $c \in C$ and every type $t \in T$,

$$\text{if } \sigma(c|t) > 0 \text{ then } \overline{U}(c|t,\sigma) = \max_{b \in C} \overline{U}(b|t,\sigma)$$

for the expected utility

$$\overline{U}(b|s,\sigma) = \sum_{x \in Z(C)} \prod_{c \in C} \left(\frac{e^{-n\tau(c)}(n\tau(c))^{x(c)}}{x(c)!} \right) U(x,b,s)$$

where

$$\tau(c) = \sum_{t \in T} r(t)\sigma(c|t)$$

is the marginal probability that a random sampled player will choose action c under σ.

Myerson (1998) proves existence of equilibrium under all games of population uncertainty with finite action and type spaces, which includes Poisson games.[9] Note that the equilibria in games with population uncertainty must be symmetric in the sense that each type plays the same strategy. This existence result provides the basis for the following characterization of the Poisson-Nash equilibrium and the cognitive hierarchy model with quantal responses.

2.2. Poisson Equilibrium for the LUPI Game. In the symmetric Poisson equilibrium, all players employ the same mixed strategy $\mathbf{p} = (p_1, p_2, \cdots, p_K)$ where $\sum_{i=1}^{K} p_i = 1$. Let the random variable $X(k)$ be the number of players who pick k in equilibrium. Then, $Pr(X(k) = i)$ is the probability that the number of players who pick k in equilibrium is exactly i. By environmental equivalence (EE), $Pr(X(k) = i)$ is also the probability that i opponents pick k. Hence, the expected payoffs for choosing

[9] For infinite types, Myerson (2000) proves existence of equilibrium for Poisson games alone.

different numbers are:[10]

$$\begin{aligned}
\pi(1) &= Pr(X(1)=0) = e^{-np_1} \\
\pi(2) &= Pr(X(1)\neq 1)\cdot Pr(X(2)=0) \\
\pi(3) &= Pr(X(1)\neq 1)\cdot Pr(X(2)\neq 1)\cdot Pr(X(3)=0) \\
&\ \vdots \\
\pi(k) &= \left(\prod_{i=1}^{k-1} Pr(X(i)\neq 1)\right)\cdot Pr(X(k)=0) \\
&= \left(\prod_{i=1}^{k-1}\left[1-np_i e^{-np_i}\right]\right)\cdot e^{-np_k}
\end{aligned}$$

for all $k > 1$. If both k and $k+1$ are chosen with positive probability in equilibrium, then $\pi(k) = \pi(k+1)$. Rearranging this equilibrium condition implies

$$e^{np_{k+1}} = e^{np_k} - np_k. \tag{2.1}$$

In addition to this condition, the probabilities must sum up to one and the expected payoff from playing numbers not in the support of the equilibrium strategy cannot be higher than the numbers played with positive probability.

The three equilibrium conditions allows us to characterize the equilibrium and show that it is unique.

Proposition 1. *There is a unique equilibrium* $\mathbf{p} = (p_1, p_2, \cdots, p_K)$ *of the Poisson LUPI game that satisfies the following properties:*

1. *Full support:* $p_k > 0$ *for all* k.
2. *Decreasing probabilities:* $p_{k+1} < p_k$ *for all* k.
3. *Convexity/concavity:* $(p_k - p_{k+1})$ *is increasing in* k *for* $p_k < 1/n$ *and decreasing in* k *for* $p_k > 1/n$.
4. *Convergence to uniform play with many players: for any fixed* K, $n \to \infty$ *implies* $p_{k+1} \to p_k$. [11]

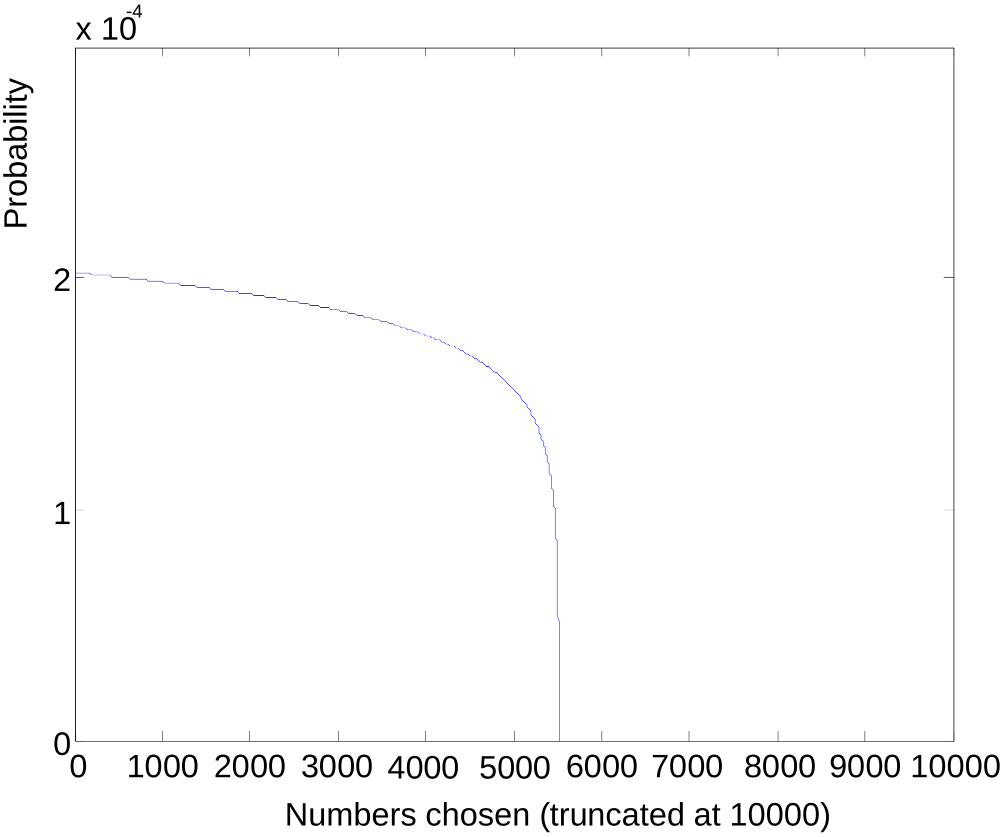
Proof. See Appendix B. □

In the Swedish game the average number of players was $N = 53,783$ and number choices were positive integers up to $K = 99,999$. As Figure 1 shows, the equilibrium

[10] Recall that winner's payoff is normalized to 1, and others are 0.

[11] To illustrate the convergence to uniform distribution as $n \to \infty$ numerically, when $K = 100$ and $N = 500$ the mixture probabilities start at $p_1 = 0.0124$ and end with $p_{97} = 0.0043, p_{98} = 0.0038, p_{99} = 0.0031, p_{100} = 0.0023$; so the ratio of highest to lowest probabilities is about six-to-one. When $K = 100$ and $N = 5,000$, all mixture probabilities for numbers 1 to 100 are 0.01 (up to two-decimal precision).

FIGURE 1. Poisson-Nash equilibrium for the LUPI game ($n = 53783$, $K = 99999$).



involves mixing with substantial probability between 1 and 5500, starting from $p_1 = 0.0002025$. The predicted probabilities drop off very sharply at around 5500. Figure 1 shows only the predicted probabilities for 1 to 10,000, since probabilities for numbers above 10,000 are positive but minuscule.

The central empirical question that will be answered later is how well actual behavior in the field matches the equilibrium prediction in Figure 1. Keep in mind that the simplified game analyzed in this section differs in some potentially important ways from the actual Swedish game. Computing the equilibrium is complicated and its properties are not particularly intuitive. It would therefore be surprising if the actual data matched the equilibrium closely.

2.3. Logit QRE. As described in McKelvey and Palfrey (1995) and Chen et al. (1997), the quantal response equilibrium (QRE) replaces best responses by quantal responses, allowing for either error in actions or uncertainty about payoffs. QRE has been applied to hundreds of experimental data sets and can often account for both behavior close to equilibrium and behavior that deviates from equilibrium (e.g. Goeree and Holt 2001, Goeree et al. 2002, Levine and Palfrey 2007, and Goeree and Holt 2005).

As in stochastic consumer choice models, QRE can fit any pattern of data if the error structure is general enough (Haile et al. 2008). Therefore, as is always done in empirical work we use a particular restriction, logit QRE. In the logit QRE response form, a vector $\mathbf{p} = (p_1, p_2, \cdots, p_K)$ is a symmetric equilibrium if all probabilities satisfy

$$p_k = \frac{\exp(\lambda\pi(k))}{\sum_{j=1}^{K}\exp(\lambda\pi(j))},$$

where payoffs are expected payoffs given the equilibrium probabilities.

If we assume that the number of players are Poisson distributed, we can use the expression for the payoff from playing the k^{th} number from the previous section. This gives the following symmetric QRE probabilities of the game:

$$p_k = \frac{\exp\left(\lambda\prod_{i=1}^{k-1}\left[1 - np_ie^{-np_i}\right]e^{-np_k}\right)}{\sum_{j=1}^{K}\exp\left(\lambda\prod_{i=1}^{j-1}\left[1 - np_ie^{-np_i}\right]e^{-np_j}\right)}.$$

Note that in a logit QRE, as in the Poisson equilibrium, all numbers are played with positive probability and larger numbers are chosen less often ($p_{k+1} \leq p_k$, for $\lambda > 0$).[12]

Some intuition about how QRE behaves[13] can be obtained from the case implemented in the lab experiments, which has an average of $N = 26.9$ players and number choices from 1 to $K = 99$. Figure 2 shows a 3-dimensional plot of the QRE probability distributions for many values of λ, along with the Poisson-Nash equilibrium. When λ is low, the distribution is approximately uniform. As λ increases more probability is placed on lower numbers 1-12. When λ is high enough the QRE closely approximates the Poisson-Nash equilibrium, which puts roughly linear declining weight on numbers 1

[12] To see why this is the case, suppose by contradiction that $p_{k+1} > p_k$, i.e., $p_{k+1}/p_k > 1$. From the expression for the ratio p_{k+1}/p_k we know that this implies that

$$\left(\lambda\prod_{i=1}^{k-1}\left[1 - np_ie^{-np_i}\right]\left[\left(1 - np_ke^{-np_k}\right)e^{-np_{k+1}} - e^{-np_k}\right]\right) > 0.$$

Dividing by λ (assuming that $\lambda > 0$) and the multiplicative operator and rearranging we get

$$\left(1 - np_ke^{-np_k}\right)e^{np_k} > e^{np_{k+1}}.$$

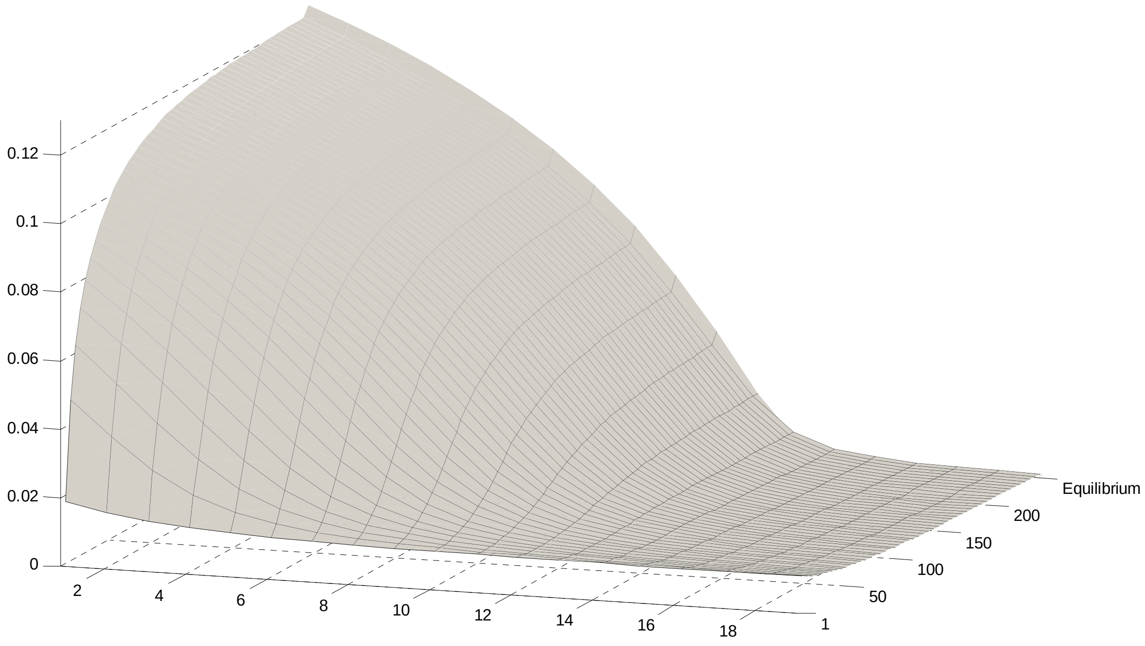
Taking logarithms

$$\frac{1}{n}\ln\left(1 - np_ke^{-np_k}\right) > p_{k+1} - p_k.$$

Since $p_{k+1} > p_k$, the right hand side is positive. The left hand side, however, is always negative since $1 - np_ke^{-np_k} = P(X(k) \neq 1)$ (which is a probability between zero and one). This is a contradiction, and we can therefore conclude that $p_k > p_{k+1}$ whenever $\lambda > 0$.

[13] We have not shown that the symmetric logit QRE is unique, but no other symmetric equilibria have emerged during numerical calculations.

FIGURE 2. Probability of choosing numbers 1 to 20 in symmetric logit QRE ($n = 26.9$, $K = 99$, $\lambda = 1, ..., 250$) and in the Poisson-Nash equilibrium ($n = 26.9$, $K = 99$).



to 15 and infinitesimal weight on higher numbers. (There is a discrete jump up from the highest λ value used and the Poisson-Nash equilibrium distribution.) We conjecture that logit QRE always approaches the Poisson-Nash equilibrium in this way, shifting weight from higher numbers to lower numbers in the transition from random ($\lambda = 0$) to Poisson-Nash ($\lambda \to \infty$) behavior but have not been able to prove the conjecture.

2.4. Cognitive Hierarchy with Quantal Response. A natural way to model limits on strategic thinking is by assuming that different players carry out different numbers of steps of iterated strategic thinking in a cognitive hierarchy (CH). This idea has been developed in behavioral game theory by several authors (e.g., Nagel 1995, Stahl and Wilson 1995, Costa-Gomes et al. 2001, Camerer et al. 2004 and Costa-Gomes and Crawford 2006) and applied to many games of different structures (e.g., Crawford 2003, Camerer et al. 2004 and Crawford and Iriberri 2007*b*). A precursor to these models was the insight, developed much earlier in the 1980's by researchers studying negotiation, that people often 'ignore the cognitions of others' in asymmetric-information bidding and negotiation games (Bazerman et al. 2000).

These models require a specification of how k-step players behave and the proportions of players for various k. We follow Camerer et al. (2004) and assume that the proportion of players that do k thinking steps is Poisson distributed with mean τ, i.e., the proportion of players that think in k steps is given by

$$f(k) = e^{-\tau}\tau^k/k!.$$

We assume that k-step thinkers correctly guess the proportions of players doing 0 to $k-1$ steps. Then the conditional density function for the belief of a k-step thinker about the proportion of $l < k$ step thinkers is

$$g_k(l) = \frac{f(l)}{\sum_{h=0}^{k-1} f(h)}.$$

The IA and EE properties of Poisson games (together with the general type specification described earlier) imply that the number of players that a k-step thinker believes will play strategy i is Poisson distributed with mean

$$nq_i^k = n\sum_{j=0}^{k-1} g_k(j)\, p_i^j.$$

Hence, the expected payoff for a k-step thinker of choosing number i is

$$\pi^k(i) = \prod_{j=1}^{i-1}\left[1 - nq_j^k e^{-nq_j^k}\right]\cdot e^{-nq_i^k}.$$

To fit the data well, it is necessary to assume that players respond stochastically (as in QRE) rather than always choose best responses (see also Camerer et al. 2007).[14] We assume that level 0 players randomize uniformly across all numbers 1 to K, and higher-step players best respond with probabilities determined by a power function.[15] The probability that a k step player plays number i is given by

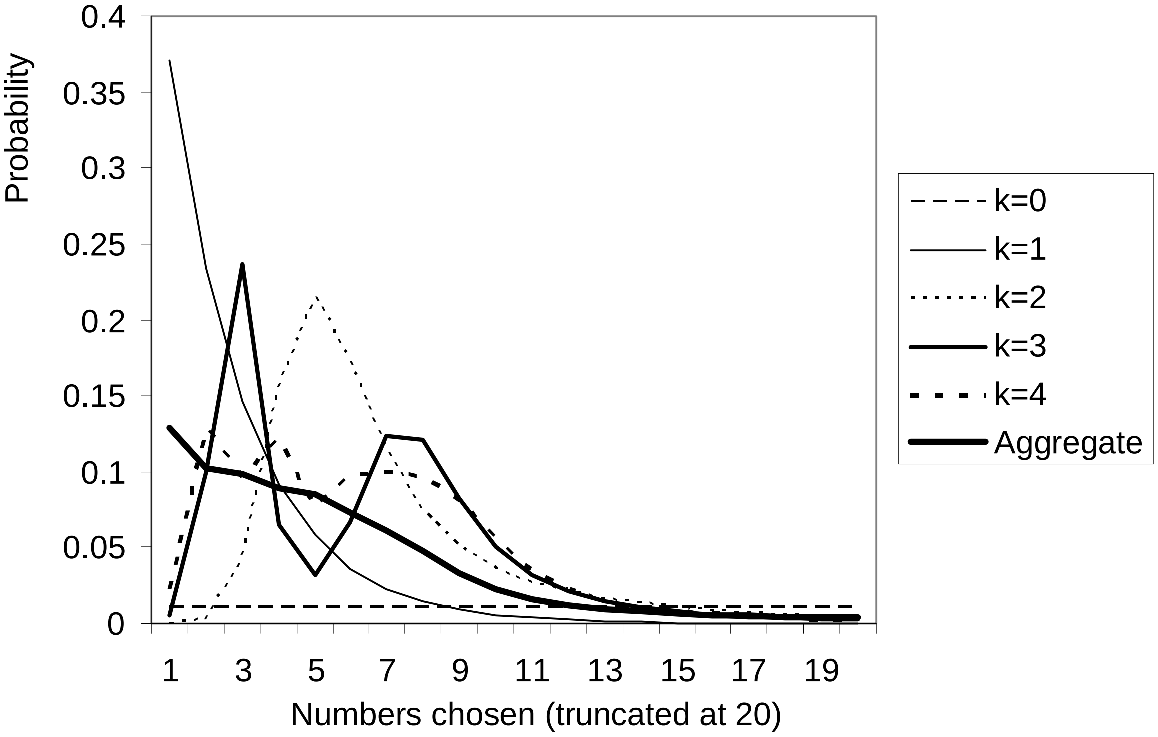
$$p_i^k = \frac{\left(\prod_{j=1}^{i-1}\left[1 - nq_j^k e^{-nq_j^k}\right] e^{-nq_i^k}\right)^\lambda}{\sum_{l=1}^{K}\left(\prod_{j=1}^{i-1}\left[1 - nq_j^k e^{-nq_j^k}\right] e^{-nq_l}\right)^\lambda},$$

for $\lambda > 0$. Since q_j^k is defined recursively—it only depends of what lower step thinkers do—it is straightforward to compute the predicted choice probabilities numerically for each type of k-step thinker (for given values of τ and λ) using a loop, then aggregating

[14] The CH model with best-response piles up most predicted responses at a very small range of the lowest integers (1-step thinkers choose 1, 2-step thinkers choose 2, and k-step thinkers will never pick a number higher than k). Assuming quantal response smoothes out the predicted choices over a wider number range.

[15] A logit choice function fits substantially worse in this case.

FIGURE 3. Probability of choosing numbers 1 to 20 in cognitive hierarchy model ($n = 26.9$, $K = 99$, $\tau = 1.5$, $\lambda = 2$).



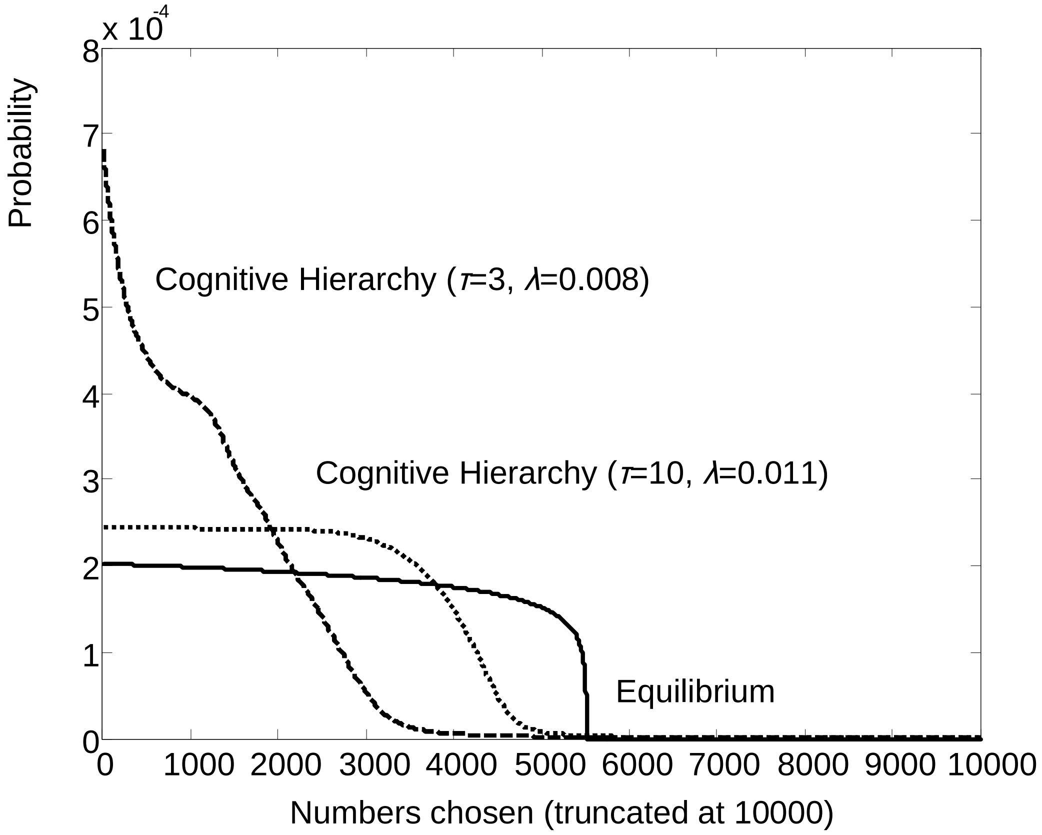
the estimated p_k^k across steps k. Apart from the number of players and the numbers of strategies, there are two parameters: the average number of thinking steps, τ, and the precision parameter, λ.

To illustrate how the CH model behaves, consider the parameters of our lab experiments, in which $N = 26.9$ and $K = 99$, with $\tau = 1.5$ and $\lambda = 2$. Figure 3 shows how 0 to 5 step thinkers play LUPI and the predicted aggregate frequency, summing across all thinking steps. In this example, 1-step thinkers put most probability on number 1, 2-step thinkers put most probability on number 5, and 3–step thinkers put most probability on numbers 3 and 7.[16]

Figure 4 shows the prediction of the cognitive hierarchy model for the parameters of the field LUPI game, i.e., when $N = 53,783$ and $K = 99,999$. The dashed line corresponds to the case when players do relatively few steps of reasoning and their

[16] Remarkably, these predictions put more overall weight on odd numbers, which is evident in the field data too, but that is likely to be a numerical coincidence rather than a basic property of the game.

FIGURE 4. Probability of choosing numbers 1 to 10000 in the Poisson-Nash equilibrium and the cognitive hierarchy model ($n = 53783$, $K = 99999$).



responses are very noisy ($\tau = 3$ and $\lambda = 0.008$). The dotted line corresponds to the case when players do more steps of reasoning and respond more precisely ($\tau = 10$ and $\lambda = 0.011$). Increasing τ and λ creates a closer approximation to the Poisson-Nash equilibrium, although even with a high τ there are too many choices of low numbers.

There is a clear contrast between the ways in which QRE and CH models deviate from equilibrium. QRE predicts number choices will be more evenly spread across the entire range than what equilibrium predicts, so it predicts too *few* low numbers compared to equilibrium. CH predicts there will be too *many* low numbers (see Figure 4). This distinction in how the two theories deviate from equilibrium is useful for comparing them because the deviations they predict from equilibrium often coincide (see Camerer et al. 2007).

3. The Field LUPI Game

The field version of LUPI, called Limbo, was introduced by the government-owned Swedish gambling monopoly Svenska Spel on the 29th of January 2007.[17] This section describes its essential elements; additional description is in Appendix D.

In Limbo, players chose up to six integers between 1 and 99,999, and each number bet costs 10 SEK (approximately 1 EURO). The game was played daily and the winning number was presented on TV in the evening and on the Internet. The winner received 18 percent of the total sum of bets, with the prize guaranteed to be at least 100,000 SEK (approximately 10,000 EURO). If no number was unique the prize was shared evenly among those who chose the smallest and least-frequently chosen number. There were also smaller second and third prizes (1000 SEK and 20 SEK) for being close to the winning number.

During the first three to four weeks, it was only possible to play the game at physical branches of Svenska Spel by filling out a form (Figure A9). The form allowed players to bet on up to six numbers[18], to play the same numbers for up to 7 days in a row, and to let a computer choose random numbers for them (a “HuxFlux” option).

Daily data were downloaded for the first seven weeks, ending on the 18th of March 2007. The game was stopped on March 24th, one day after a newspaper article claimed that some players had colluded in the game, but it is unclear whether collusion actually occurred or how it could be detected.

Unfortunately, we have only gained access to aggregate daily frequencies, not to individual-level data. We also do not know how many players used the randomization HuxFlux option. However, because the operators told us how HuxFlux worked, we can estimate that approximately 19 percent of players were randomizing in the first week.[19]

Note that the theoretical analysis of the LUPI game in the previous section differs from the field LUPI game in three ways. First, the theory used a tie-breaking rule in which nobody wins if there is no uniquely chosen number, while in the field game players who choose the smallest and least-frequently chosen number share the prize. This is a minor difference because the probability that there is no unique number is very small and it never happened during the 49 days we have data for. A second, more

[17] Stefan Molin at Svenska Spel told us that he invented the game in 2001 after taking a game theory course from the Swedish theorist and experimenter Martin Dufwenberg.

[18] The rule that players could only pick up to six numbers a day was enforced by the requirement that players had to use a “gambler’s card” linked to their personal identification number when they played. Colluding in LUPI can conceivably increase the probability of winning but would require a remarkable degree of coordination across a large syndicate, and is also risky if others might be colluding in a similar way.

[19] In the first week, the randomizer chose numbers from 1 to 15,000 with equal probability. The drop in numbers just below and above 15,000 implies the 19 percent figure.

important, difference is that we assume that each player can only pick one number. In the field game, players are allowed to bet on up to six numbers. This does play a role for the theoretical predictions, since it allows players to "knock out" a low-number winner by choosing the same number as the winner and then bet on a higher number hoping that number will win. Finally, we do not take the second and third prizes present in the field version into account, but this is unlikely to make a big difference for the strategic nature of the game.

Nevertheless, these three differences between the game analyzed theoretically and the field game as played is an important motivation for running laboratory experiments with single bets, no opportunity for direct collusion, and only a first prize, which match the game analyzed theoretically more closely.

3.1. Descriptive Statistics. Table 1 reports summary statistics for the first 49 days of the game. To get some notion of possible learning over time (discussed further below), two additional columns display the corresponding daily averages for the first and last weeks. The last column displays the corresponding statistics that would result from play according to Poisson equilibrium.

Overall, the average number of bets was 53,783, but there was considerable variation over time. There is no apparent time trend in the number of participating players, but there is less participation on Sundays and Mondays (see Figure A11). The variation of the number of bets across days is therefore much higher than what the Poisson distribution predicts (its standard deviation is 232), which is one more reason to expect the equilibrium prediction to not fit very well.

Despite the many differences between the simplified theory and the way the field lottery game was implemented, the average number chosen overall was 2835, which is close to the equilibrium prediction of 2595. Winning numbers, and the lowest numbers not chosen by anyone, also varied a lot over time. *All* the aggregate statistics converge reasonably closely to equilibrium from the first week to the last week. For example, in equilibrium essentially nobody should choose a number above 10,000. In the first week 12 percent chose these high numbers, but in the last week that fraction is only 1 percent.

An interesting feature of the data is a tendency to avoid round or focal numbers and choose quirky numbers that are perceived as "anti-focal" (as in hide-and-seek games, see Crawford and Iriberri 2007*a*). Even numbers were chosen less often than odd ones (46.75% vs. 53.25%). Numbers divisible by 10 are chosen a little less often than predicted. Strings of repeating digits (e.g., 1111) are chosen too often.[20] Players also

[20] Similar behavior can be found in the federal tax evasion case of Joe Francis, the founder of "Girls Gone Wild." Mr. Francis was indicted on April 11, 2007 for claiming false business expenses

Table 1. Descriptive statistics and Poisson-Nash equilibrium predictions for field LUPI game data

	All days				1^{st} week	7^{th} week	Eq.
	Avg.	Std.	Min	Max	Avg.	Avg.	Avg.
# Bets	53783	7782	38933	69830	57017	47907	53783
Average number played	2835	813	2168	6752	4512	2484	2595
Median number played	1674	348	435	2272	1203	1935	2541
Winning number	2095	1266	162	4465	1159	1982	2585
Lowest number not played	3103	929	480	4597	1745	3462	4077
Below 100 (%)	6.08	4.84	2.72	2.97	15.16	3.19	2.02
Below 1000 (%)	32.31	8.14	21.68	63.32	44.91	27.52	20.05
Below 5000 (%)	92.52	6.44	68.26	97.74	78.75	96.44	93.34
Below 10000 (%)	96.63	3.80	80.70	98.94	88.07	98.81	100.00
Even numbers (%)	46.75	0.58	45.05	48.06	45.91	47.45	49.99
Divisible by 10 (%)	8.54	0.466	7.61	9.81	8.43	9.01	9.99
Proportion 1900–2010 (%)	71.61	4.28	67.33	87.01	79.39	68.79	49.78
11, 22,...,99 (%)	12.2	0.82	10.8	14.4	12.4	11.4	9.00
111, 222,...,999 (%)	3.48	0.65	2.48	4.70	4.27	2.78	0.90
1111, 2222,...,9999 (1/1000)	4.52	0.73	2.81	5.80	4.74	3.95	0.74
11111, 22222,... (1/1000)	0.76	0.84	0.15	5.45	2.26	0.21	0

Proportion of numbers between 1900 and 2010 refers to the proportion relative to numbers between 1844 and 2066. "11, 22,...,99" refers to the proportion relative to numbers below 100, "111,222,...,999" relative to numbers below 1000, and so on. The "eq. avg" predictions refers to the prediction of the Poisson-Nash equilibrium with $n = 53,783$ and $K = 99,999$.

overchoose numbers that represent years in modern time (perhaps their birth years). If players had played according to equilibrium, the fraction of numbers between 1900 and 2010 divided by all numbers between 1844 and 2066 should be 49.78 percent, but the actual fraction was 70 percent.[21]

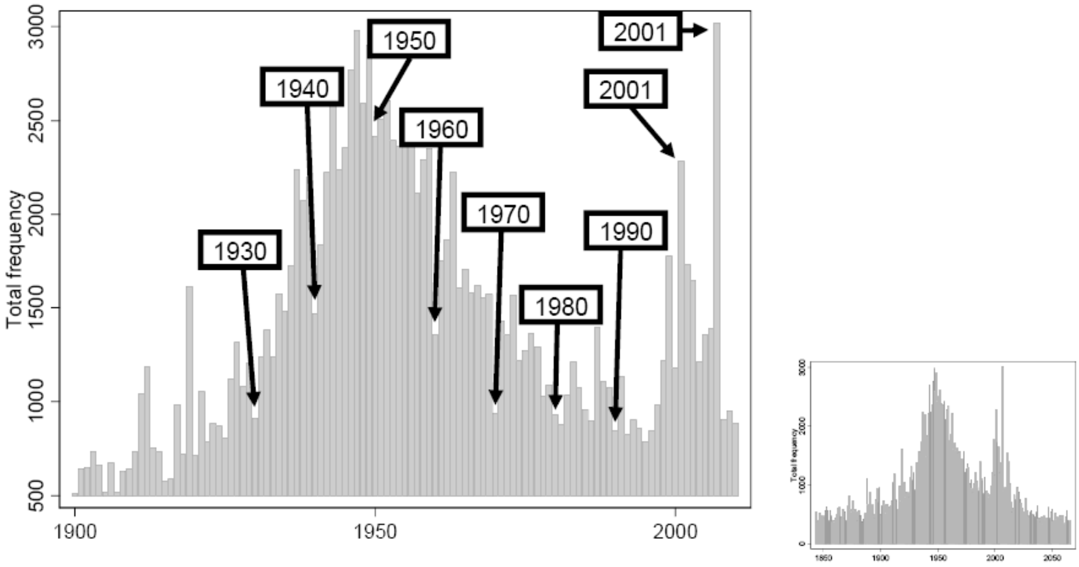
Figure 5 shows a histogram of numbers between 1900 and 2010 (aggregating all 49 days). Note that although the numbers around 1950 are most popular, there are noticeable dips at focal years that are divisible by ten.[22] Figure 5 also shows the aggregate distribution of numbers between 1844 and 2066, which clearly shows the

such as $333,333.33 and $1,666,666.67 in insurance, which were too suspicious *not* to attract attention. See http://consumerist.com/consumer/taxes/girls-gone-wild-tax-indictment-teaches-us-not-to-deduct-funny+looking-numbers-252097.php for the proposed tax lesson.

[21] We compare the number of choices between 1900 and 2010 to the number of choices between 1844 and 2066 since there are twice as many strategies to choose from in the latter range compared to the first. If all players randomized uniformly, the proportion of numbers between 1900 and 2010 would be 50 percent.

[22] Note that it would be unlikely to observe these dips reliably with typical experimental sample sizes. It is only with the large amount of data available from the field, 2.5 million observations, that these dips are visually obvious and different in frequency than neighboring unround numbers.

FIGURE 5. Numbers chosen between 1900 and 2010, and between 1844 and 2066, during all days in the field.



popularity of numbers around 1950 and 2007. There are also spikes in the data for special numbers like 2121, 2222 and 2345. Explaining these "focal" numbers with CH and QRE models is not easy (unless the 0-step player distribution is defined to include focality) so we will not comment on them further (though see Crawford and Iriberri 2007*a* for a successful application in simpler hide-and-seek games).

3.2. Results. Do subjects in the field LUPI game play according to the equilibrium prediction? In order to investigate this, we assume that the number of players is Poisson distributed with mean equal to the empirical daily average number of numbers chosen $(53,783)$. As noted, this assumption is wrong because the variation in number of bets across days is much higher than what the Poisson distribution predicts.

Figure 6 shows the average daily frequencies from the first week together with the equilibrium prediction (the dashed line), for all numbers up to 99,999 and for the restricted interval up to 10,000. Recall that in the Poisson-Nash equilibrium, probabilities of choosing higher numbers first decrease slowly, drop quite sharply at around 5500, and asymptotes to zero after $p_{5513} \approx 1/n$ (recall Proposition 1 and Figure 1). Compared to equilibrium, there is overshooting at numbers below 1000 and undershooting at numbers between between 2000 and 5500. It is also noteworthy how spiky the data is compared to the equilibrium prediction, which is a reflection of clustering on special numbers, as described above. Nonetheless, the ability of the very complicated Poisson-Nash equilibrium to capture the basic features of the data is surprisingly good.

Figure 6. Average daily frequencies and Poisson-Nash equilibrium prediction for the first week in the field ($n = 53783$, $K = 99999$).

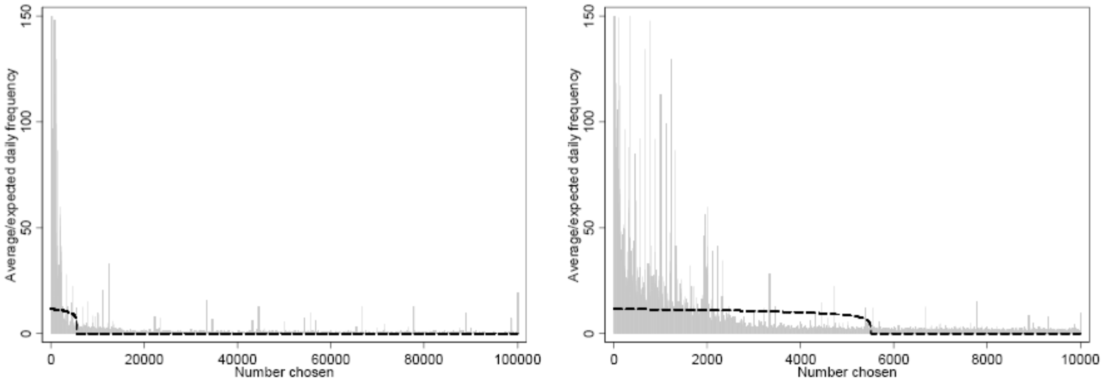


Figure 7. Average daily frequencies and Poisson-Nash equilibrium prediction for week 2-7 in the field ($n = 53783$, $K = 99999$).

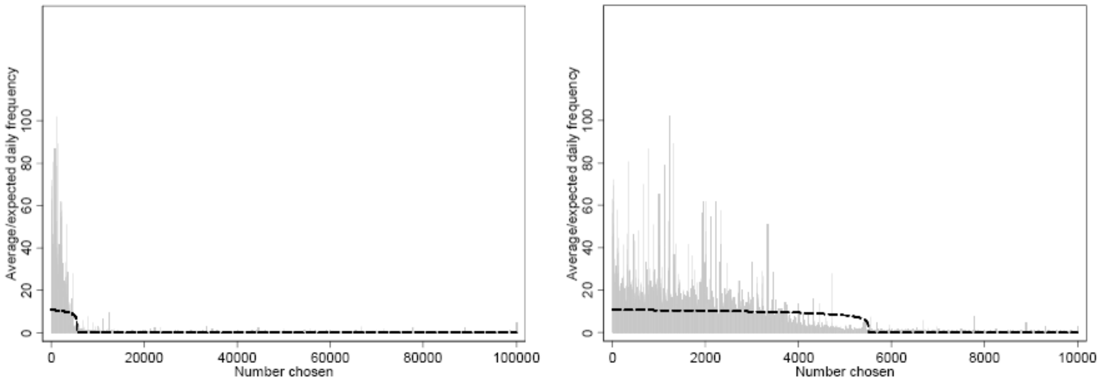


Figure 7 shows average daily frequencies of choices from the second through the last (7th) week. Behavior in this period is even closer to equilibrium than in the first week. However, when only numbers below 10,000 are plotted, the overshooting of low numbers and undershooting of intermediate numbers is still clear (although the undershooting region shrinks to numbers between 4000 and 5500) and there are still many spikes of clustered choices.

The next question is whether alternative theories can explain both the degree to which the equilibrium prediction is surprisingly accurate *and* the degree to which there is systematic deviation.

3.3. Rationalizing Non-Equilibrium Play. In this section, we investigate if the cognitive hierarchy model can account for the main deviations from equilibrium just described in the previous section. The QRE model is not estimated for two reasons: First, it is very computationally challenging to estimate for the large-scale field data.[23] Second, if we are correct that the QRE approaches the Poisson-Nash equilibrium smoothly from random to Poisson-Nash, then it cannot account for overshooting of low numbers. Indeed, it is conceivable that the best-fitting QRE function is very close to Poisson-Nash, since most of the choices are below 5000 and there is substantial overshooting in that region which QRE can only fit by approximating Poisson-Nash.

Table 2 reports the results from the maximum likelihood estimation of the data using the cognitive hierarchy model.[24] The best-fitting estimates week-by-week, shown in Table 2, suggest that both parameters increase over time. The average number of thinking steps that people carry out, τ, increases from about 3 in the first week—an estimate reasonably close to estimates from 1.5 to 2.5 that typical fit experimental data sets well (Camerer et al. 2004)—to 10 in the last week.

Table 2. Maximum likelihood estimation of the cognitive hierarchy model for field data

Week	1	2	3	4	5	6	7
τ	2.98	5.83	7.32	7.2	7.82	10.27	10.27
λ	0.0080	0.0094	0.0103	0.0108	0.0110	0.0108	0.0107

Figure 8 shows the average daily frequencies from the first week together with the cognitive hierarchy estimation and equilibrium prediction. The cognitive hierarchy model does a reasonable job of accounting for the over- and undershooting tendencies at low and intermediate numbers (with the estimated $\hat{\tau} = 2.98$). Furthermore, while the CH model does have two degrees of freedom which the Poisson equilibrium prediction does not, there is so much data that the good explanation of the deviations is not due to overfitting.

In later weeks, the week-by-week estimates of τ seem to drift upward (and λ increases slightly), which is a reduced-form model of learning as an increase of thinking steps (see more details below). In the last week the cognitive hierarchy prediction is much closer to equilibrium (because τ is around 10) but is still consistent with the smaller amounts of over- and undershooting (see Figure 9).

[23] Keep in mind that the CH model includes a quantal response component as well. However, because the CH model is recursive (level-k behavior is determined by lower-level behavior and λ) it is much easier to estimate.

[24] It is difficult to guarantee that these estimates are global maxima since the likelihood function is not smooth and concave. We also used a relatively coarse grid search, so there may be other parameter values that yield slightly higher likelihoods and different parameter values.

FIGURE 8. Average daily frequencies, cognitive hierarchy (solid line) and Poisson-Nash equilibrium prediction (dashed line) for the first week in the field ($n = 53783$, $K = 99999$, $\tau = 2.98$, $\lambda = 0.008$).

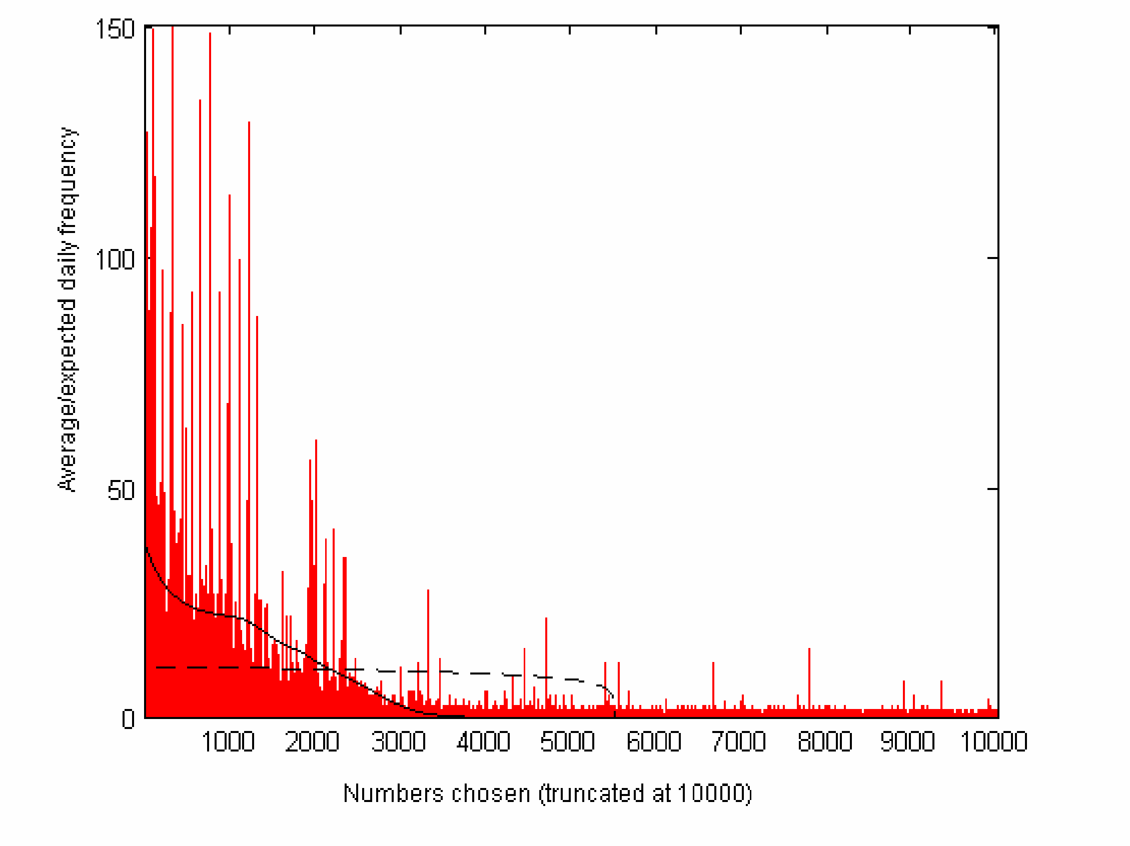


TABLE 3. Goodness-of-fit for cognitive hierarchy and equilibrium for field data

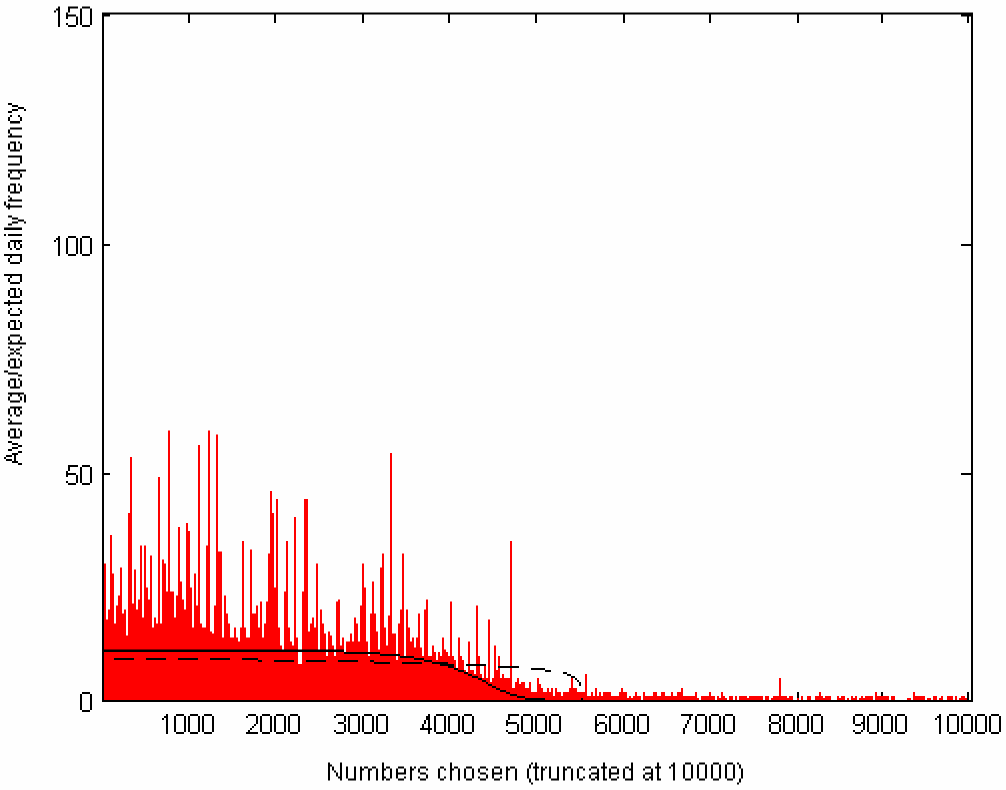
Week	1	2	3	4	5	6	7
Log-likelihood CH	-63956	-36390	-23716	-20546	-20255	-19748	-18083
Proportion below CH (%)	61.08	72.50	77.69	79.87	81.86	82.63	81.94
Proportion below equil. (%)	49.56	61.82	67.66	67.70	70.23	76.79	76.61

The proportion below the theoretical prediction refers to the fraction of the empirical density that lies below the theoretical prediction, or one minus the fraction of overshooting.

To get some notion of how close to the data the fitted cognitive hierarchy model is, Table 3 displays two goodness-of-fit statistics. First, the log-likelihoods reveal that the cognitive hierarchy model does better in explaining the data toward the last week and is always much better than Poisson-Nash.[25] Second, in order to compare the CH model

[25] Since the computed Poisson-Nash equilibrium probabilities are zero for $k > 5518$, the likelihood is always zero for the equilibrium prediction. In Appendix C, however, we compute the log-likelihood

Figure 9. Average daily frequencies, cognitive hierarchy (solid line) and Poisson-Nash equilibrium prediction (dashed line) for the last week in the field ($n = 53783$, $K = 99999$, $\tau = 10.27$, $\lambda = 0.0107$).



with the equilibrium prediction, we calculate the proportion of the empirical density that lies below the predicted density. This measure is one minus the summed "miss rates", the differences between actual and predicted frequencies, for numbers which are chosen more often than predicted. If there is a lot of overshooting this statistic is low and if there is very little overshooting this statistic is close to 1. The cognitive hierarchy model does better than the equilibrium prediction in all seven weeks based on this statistic. For example, in the first week, 61 percent of players' choices were consistent with the cognitive hierarchy model, whereas only 50 percent were consistent with equilibrium. However, both models improve substantially across the weeks.

4. The Laboratory LUPI Game

We conducted a parallel lab experiment for two reasons.

for low numbers. Based on Schwarz (1978) information criterion, the cognitive hierarchy model still performs better in all weeks.

First, the rules of the field LUPI game do not exactly match the theoretical assumptions used to generate the Poisson-Nash equilibrium prediction. (The field data included some choices made by a random number generator, some players might have chosen multiple numbers or colluded, and there were multiple prizes.) In the lab, we can more closely implement the assumptions of the theory. If the theory fits poorly in the field and closely in the lab, then that suggests the theory is on the right track when its underlying assumptions are most carefully controlled. If the theory fits closely in both cases, that suggests that the additional factors in the field that are excluded from the theory do not matter.

Second, because the field game is rather simple, it is possible to design a lab experiment which closely matches the field in its key features. How closely the lab and field data match provides some evidence in ongoing debate about how well lab results generalize to comparable field settings (e.g., Levitt and List 2007 and Camerer 2008).

In designing the laboratory game, we compromise between two goals: to create a simple environment in which theory should apply (theoretical validity), and to recreate the features of the field LUPI game in the lab. Because we use this opportunity to create an experimental protocol that is closely matched to a particular field setting, we often sacrificed theoretical validity for field replication.

The first choice is the scale of the game: The number of players (N), possible number choices (K), and stakes. We choose to scale down the number of players and the largest payoff by a factor of 2000. This implies that there were on average 26.9 players and the prize to the winner in each round was \$7. We chose $K = 99$ since the shape of the equilibrium distribution with that value has some of the basic features of the field data distribution. Since the field data span 49 days, the experiment has 49 rounds in each session. (We typically refer to experimental rounds as "days" and seven-"day" intervals as "weeks" for semantic comparability between the lab and field descriptions.)

Because the number of players is endogenous in the field, in the lab experiment the number of players in each round was also determined randomly so that the average number of subjects participating in a round was 26.9.[26] In contrast to the field game, each player was allowed to choose only one number and there was only one prize per round, in the amount of \$7. There was no option to use a random number generator and in the case there was no number that only one player picked, nobody won in that

[26] Unfortunately, the number of participants in the laboratory experiments were not Poisson distributed due to a technical mistake in the lab implementation. The variance was 8.2, compared to 26.9 in a Poisson distribution. However, behaviorally we believe this plays a minor role since 1) we only told subjects about the average number of players and 2) subjects were not told how many players that were selected to play in each round.

round. These rules implement theoretical assumptions but depart from the rules in the field game.

Two design choices deliberately limited the information subjects had in order to maintain parallelism with the field. While the winning number was announced in each field-game day, we do not know how much Swedish players learned about the full number distribution (which was only available online and partially reported on a TV show). Therefore, we chose to announce only the winning number in the lab. And because players in the field did not necessarily know the number of players each day, we did not tell the lab subjects the process by which the number of players in each round was determined or the number of subjects who played in each specific round, although they knew that on average 26.9 subjects played.

Three laboratory sessions were conducted at the California Social Science Experimental Laboratory (CASSEL) at University of California Los Angeles on the 22nd and 25th of March 2007. The experiments were conducted using the Zürich Toolbox for Ready-made Economic Experiments (zTree) developed by Urs Fischbacher, as described in Fischbacher (2007). Within each session, 38 graduate and undergraduate students were recruited, through CASSEL's web-based recruiting system. All subjects knew that their payoff will be determined by their performance. We made no attempt to replicate the demographics of the field data, which we unfortunately know very little about. However, the players in the laboratory are likely to differ in terms of gender, age and ethnicity compared to the Swedish players. In all three sessions, we had more female than male subjects, with all of them clustered in the age bracket of 18 to 22, and the majority spoke a second language. The majority of the subjects had never participated in any form of lottery before. Subjects had various levels of exposure to game theory, but very few had seen or heard of a similar game prior to this experiment.

4.1. Experimental Procedure. At the beginning of each session, the experimenter first explained the rules of the LUPI game. The instructions were based on a version of the lottery ticket for the field game translated from Swedish to English (see Appendix E). Subjects were then given the option of leaving the experiment, in order to see how much self-selection influences experimental generalizability. None of the recruited subjects chose to leave, which indicates a limited role for self-selection (after recruitment and instruction).

To avoid an end-game effect, subjects were told that the experiment would end at a predetermined, but non-disclosed time (also matching the field setting, which ended abruptly and unexpectedly). Subjects were told that participation was randomly determined at the beginning of each round, with 26.9 subjects participating on average.

In the beginning of each round, subjects were informed whether they would actively participate in the current round (i.e., if they had a chance to win). They were required to submit a number in each round, even if they were not selected to participate. The difference between behavior of selected and non-selected players gives us some information about the effect of marginal incentives.

When all subjects had submitted their chosen numbers, the lowest unique positive integer was determined. If there was a lowest unique positive integer, the winner earned \$7; if no number was unique, no subject won. Each subject was privately informed, immediately after each round, what the winning number was, whether they had won that particular round, and their payoff so far during the experiment. This procedure was repeated 49 times, with no practice rounds (as is the case of the field). After the last round, subjects were asked to complete a short questionnaire which allowed us to build the demographics of our subjects and a classification of strategies used. In one of the sessions, we included the cognitive reflection test as a way to measure cognitive ability (to be described below). All sessions lasted for less than an hour, and subjects received a show-up fee of \$8 or \$13 in addition to prizes from the experiment (which averaged \$8.6).

Screenshots from the experiment are shown in Appendix E.

4.2. Lab Descriptive Statistics. Behavior in the laboratory differs slightly among the three sessions when all subjects' choices are included, but do not significantly differ when using the choices of subjects selected to actively participate, so from now on we use only the active participants' data. (See Appendix E for details.)

Figure 10 shows the data for the choices of participating players (together with the Poisson-Nash equilibrium prediction). There are very few numbers above 20 so the numbers 1 to 20 are the focus in subsequent graphs. In line with the field data, players have a predilection for certain numbers, while others are avoided. Judging from Figure 10, subjects avoid some even numbers, especially 2 and 10, while they endorse the odd (and prime) numbers 3, 11, 13 and 17. Interestingly, no subject played 20, while 19 was played five times and 21 was played six times.

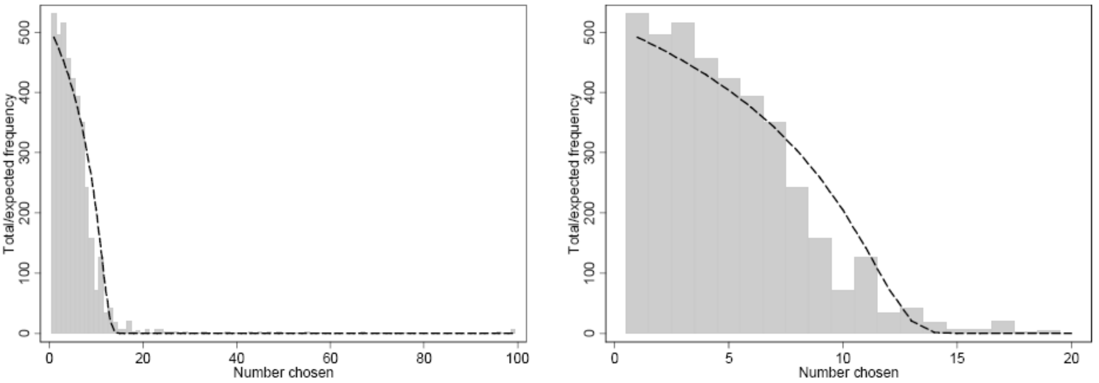
Table 4 shows some descriptive statistics for the participating subjects in the lab experiment. As in the field, some players in the first week tend to pick very high numbers (above 20) but the percentage shrinks by the seventh week. The average number chosen in the last week corresponds closely to the equilibrium prediction (5.3 vs. 5.2) and the medians are identical (5.0). The average winning numbers are too high compared to equilibrium play, which is consistent with the observation that players pick very low numbers too much, creating non-uniqueness among those numbers so that unique numbers are unusually high. The tendency to pick odd numbers decreases over

Table 4. Descriptive statistics for laboratory data

	All rounds				R 1-7	R. 43–49	Eq.
	Avg.	Std.dev.	Min	Max	Avg.	Avg.	Avg.
Average number played	5.7	1.6	4.2	13.1	9.0	5.3	5.2
Median number played	4.8	1.0	3.5	8.0	6.0	5.0	5
Below 20 (%)	98.13	3.43	78.05	100.00	92.81	98.83	100.00
Even numbers (%)	44.07	5.84	29.47	56.94	39.79	47.16	46.86
Session 1							
Winning number	6.0	9.3	1	67	13.0	2.5	2.9
Lowest number not played	8.1	2.5	1	12	4.9	8.1	3.3
Session 2							
Winning number	5.1	2.6	1	10	5.8	5.1	2.9
Lowest number not played	7.5	2.9	1	12	6.3	8.4	3.3
Session 3							
Winning number	5.6	3.2	1	14	6.1	5.7	2.9
Lowest number not played	7.5	2.7	2	13	7.4	10.0	3.3

Summary statistics are based only on choices of subjects who are selected to participate. The equilibrium column refers to what would result if all players played according to equilibrium ($n = 26.9$ and $K = 99$)

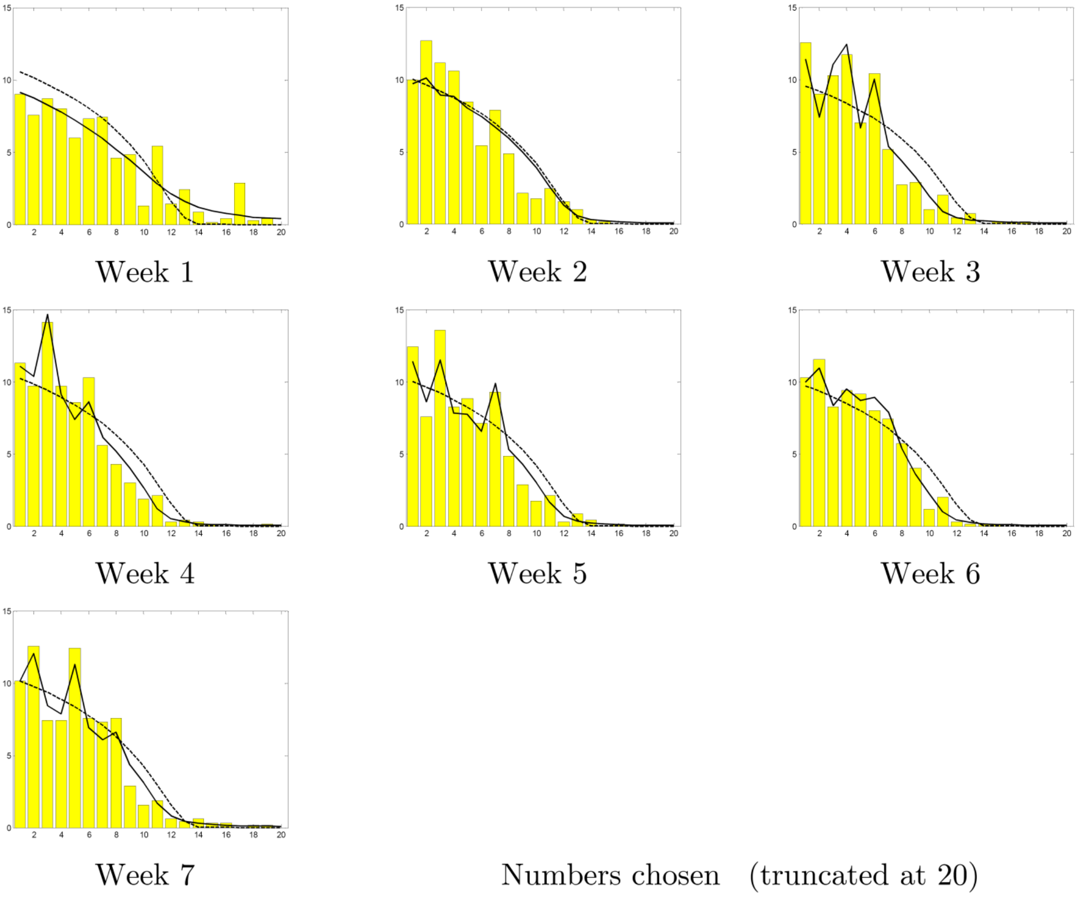
Figure 10. Laboratory total frequencies and Poisson-Nash equilibrium prediction (all sessions, participating players only, $n = 26.9$, $K = 99$).



time—40 percent of all numbers are even in the first week, whereas 47 percent are even in the last week (which coincides with the equilibrium proportion of even numbers). As in the field data, the overwhelming impression from Table 4 is that convergence to equilibrium is quite rapid over the 49 periods (despite receiving feedback only about the winning number).

4.3. Aggregate Results. In the Poisson equilibrium with 26.9 average number of players, strictly positive probability is put on numbers 1 to 16, while other numbers

FIGURE 11. Average daily frequencies in the laboratory, Poisson-Nash equilibrium prediction (dashed lines) and estimated cognitive hierarchy (solid lines), week 1 to 7 ($n = 26.9$, $K = 99$).



Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

have probabilities numerically indistinguishable from zero. Figure 11 shows the average frequencies played in week 1 to 7 together with the equilibrium prediction (dashed line) and the estimated week-by-week results using the cognitive hierarchy model (solid line). These graphs clearly indicates that learning is quicker in the laboratory than in the field. Despite that the only feedback given to players in each round is the winning number, behavior is remarkably close to equilibrium already in the second week. However, we can also observe the same discrepancies between the equilibrium prediction and observed behavior as in the field. The distribution of numbers is too spiky and there is overshooting of low numbers and undershooting at numbers just below the equilibrium cutoff (at number 16).

Figure 11 also displays the estimates from a maximum likelihood estimation of the cognitive hierarchy model presented in the theoretical section (solid line). The cognitive hierarchy model can account both for the spikes and the over- and undershooting. Table 5 shows the estimated parameters.[27] There is no clear time trend in the two parameters, and in some rounds the average number of thinking steps is unreasonably large compared to other experiments showing τ around 1.5. Since there are two free parameters with relatively few choice probabilities to estimate, we might be over-fitting by allowing two free parameters. We therefore estimate the precision parameter λ while keeping the average number of thinking steps fixed. We set the average number of thinking steps to 1.5, which has been shown to be a value of τ that predicts experimental data well in a large number of games (Camerer et al. 2004). The estimated precision parameter is considerably lower in the first week, but is then relatively constant.[28]

Table 5 also displays the maximum likelihood estimate of λ for the logit QRE. The precision parameter is relatively high in all weeks, but particularly from the second week and onwards. Recall from Figure 2 that the QRE prediction for such high λ is very close to the Poisson-Nash equilibrium.

TABLE 5. Maximum likelihood estimation of the cognitive hierarchy model and QRE for laboratory data

Week	1	2	3	4	5	6	7
τ	8.15	13.14	6.48	5.31	11.52	5.05	9.00
λ	1.19	11.27	10.85	14.92	13.53	14.67	8.73
λ ($\tau = 1.5$)	1.08	2.37	2.85	2.82	2.76	2.34	2.16
λ_{QRE}	123.40	526.83	396.24	430.83	523.30	517.25	309.89

Table 6 provides some goodness-of-fit statistics for the cognitive hierarchy model, QRE and the equilibrium prediction. Based the proportion of the empirical density that lies below the predicted density, the equilibrium prediction does remarkably well. The equilibrium prediction does better than the cognitive hierarchy model with $\tau = 1.5$ in all weeks, but the cognitive hierarchy model (with two free parameters) does better than the equilibrium prediction in all but the second week. The logit QRE performs better than equilibrium in the first week, but is practically indistinguishable from equilibrium after the first week (due to high λ). The log-likelihood of the cognitive hierarchy model (with two parameters) is higher than the QRE during all weeks, but

[27] The log-likelihood function is neither smooth nor concave, so the estimated parameters may not reflect a global maximum of the likelihood.

[28] Figure A4 shows the fitted cognitive hierarchy model when τ is restricted to 1.5. It is clear that the model with $\tau = 1.5$ can account for the undershooting also when the number of thinking steps is fixed, but it has difficulties in explaining the overshooting of low numbers. The main problem is that with $\tau = 1.5$, there are too many zero-step thinkers that play all numbers between 1 and 99 with uniform probability.

the QRE performs better than the cognitive hierarchy model with $\tau = 1.5$ based on the log-likelihood values.[29]

Table 6. Goodness-of-fit for cognitive hierarchy, QRE and equilibrium for laboratory data

Week	1	2	3	4	5	6	7
Log-likelihood CH	-150.9	-75.9	-67.5	-65.0	-64.4	-60.7	-68.5
Log-likelihood CH $\tau = 1.5$	-204.1	-180.8	-171.6	-179.8	-177.8	-178.4	-185.8
Log-likelihood logit QRE	-172.2	-76.8	-94.8	-88.5	-82.0	-76.9	-88.9
Proportion below CH (%)	86.06	88.02	92.26	93.13	91.41	94.99	92.60
Proportion below CH $\tau = 1.5$ (%)	81.11	76.53	79.00	76.79	78.23	76.22	77.18
Proportion below logit QRE (%)	84.95	87.94	83.64	86.88	86.13	90.21	86.61
Proportion below eq. (%)	81.71	88.16	83.60	87.19	86.13	90.79	86.88

The proportion below the theoretical prediction refers to the fraction of the empirical density that lies below the theoretical prediction.

On the aggregate level, behavior in the lab is remarkably close to equilibrium from the second to the last week. The cognitive hierarchy model can rationalize the tendencies that some numbers are played more, as well as the undershooting below the equilibrium cutoff. The value-added of the cognitive hierarchy model is not primarily that it gives a slightly better fit, but that it provides a plausible story for how players manage to play so close to equilibrium. Most likely, few players would be capable of calculating the equilibrium during the course of the experiment, whereas many of them should be able to carry out a few steps of reasoning along the lines of the cognitive hierarchy model.

4.4. Individual Results. An advantage of the lab over the field, in this case, is that the behavior of individual subjects can be tracked over time and we can gather more information about them to link to choices. Appendix E discusses some details of these analyses but we summarize them here only briefly.

In a post-experimental questionnaire, we asked people to state why they played as they did. We coded their responses into four categories (sometimes with multiple categories): "Random", "stick" (with one number), "lucky", and "strategic" (explicitly mentioning response to strategies of others). The four categories were coded 50%, 40%, 15% and 70% of the time. These categories had some relation to actual choices because "stick" players chose fewer distinct numbers and "lucky" players had number

[29] In Appendix C we calculate the log-likelihoods using data from numbers 1 to 16, which allows us to compare the equilibrium prediction with the other models. Based on Schwarz (1978) information criterion, both QRE and cognitive hierarchy (with two parameters) outperforms equilibrium.

choices with a higher mean and higher variance. The only demographic variable with a significant effect on choices and payoffs was "exposure to game theory"; those subjects chose a significantly lower average number with less variation across rounds. A measure of "cognitive reflection" (Frederick 2005), a short-form IQ test, did not correlate with choice measures or with payoffs.

As is often seen in games with mixed equilibria, there is some evidence of "purification" since subjects chose only 9.46 different numbers on average (see Appendix E), compared to 10.9 expected in Poisson-Nash equilibrium.

TABLE 7. Panel data regressions explaining individual play in the laboratory

	All periods	Week 1	Week 2	Week 3-7
Round (1-49)	-0.011	-0.529	-0.102	0.0144
	(-1.09)	(-0.58)	(-0.47)	(1.10)
$t-1$ winner	0.188**	0.154**	0.376*	0.089*
	(10.55)	(3.55)	(2.20)	(1.98)
$t-2$ winner	0.140**	0.111*	0.323	0.056
	(7.43)	(1.99)	(1.28)	(1.26)
$t-3$ winner	0.082**	0.078	-0.057	0.036
	(4.10)	(1.13)	(-0.26)	(0.83)
Fixed effects	Yes	Yes	Yes	Yes
Observations	3156	319	483	2354
R^2	0.05	0.12	0.01	0.00

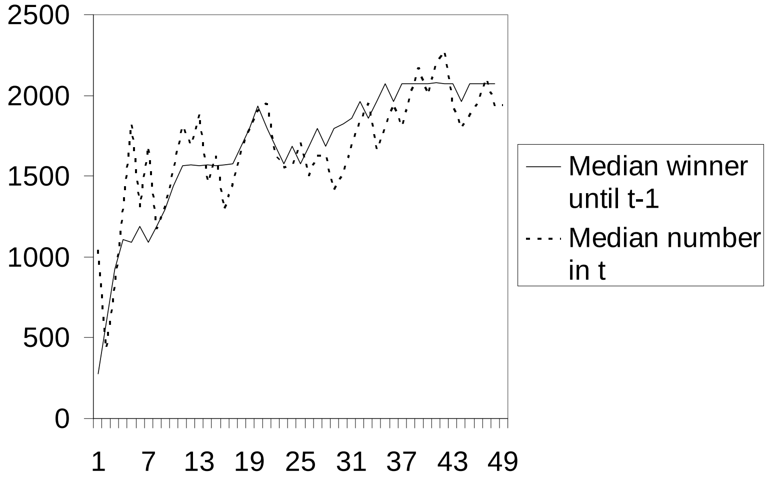
*=5 percent and **=1 percent significance level. The table report results from a linear fixed effects panel regression. Only actively participating subjects are included. $t-$statistics within parentheses.

In the post-experimental questionnaire, several subjects said that they responded to previous winning numbers. To measure the strength of this learning effect we regressed players' choices on the winning number in the three previous periods. Table 7 shows that the winning numbers in previous rounds do affect players' choices early on, but this tendency to respond to previous winning numbers is considerably weaker in later weeks (3 to 7). The small round-specific coefficients in Table 7 also show that there does not appear to be any general trend in players' choices over the 49 rounds.

5. Learning

The LUPI game is challenging for traditional models of learning. Although a wide range of learning dynamics are likely to converge to equilibrium in the limit, it is more difficult to explain how players can learn to play close to equilibrium in only 49 rounds. For example, reinforcement learning is unlikely to match the speed at which people

Figure 12. Median winner and median choices in the field



learn since players win rarely (and hence, their strategies are rarely reinforced). Belief-based models like fictitious play, on the other hand, are also likely to have a hard time in explaining the speed of learning. In the field, there is typically no number below the winning number that wasn't chosen by anyone (that happened only in 6 out of the 49 days), so all numbers are most often best-responses to the empirical distribution. In the lab, on the other hand, it happens more often that there are unpicked numbers below the winnning number, but there is no way for players to figure out what numbers these are and use that information to update beliefs as hypothesized by fictitious play. Hybrid models like EWA (Camerer and Ho 1999, Ho et al. 2007) require the same information as fictitious play and therefore do not apply well to this information environment.

To explain the learning pattern in both the field and lab we therefore need a model that 1) does not rely on best responses to the full empirical distribution, that 2) does not only consider a player's own payoff and 3) is not based on any other information than the structure of the game, a player's own experience and winning numbers. We therefore propose a simple learning model in which all players imitate numbers around previous winning numbers.[30] Such a model is empirically motivated by the fact that players seem to change strategies in the direction of previous winners. Figure 12 shows how closely the median number chosen in period t in the field is related to the median

[30] We conjecture that imitation is a theoretically sound model of learning in the LUPI game in the sense that a learning model that only reinforces previous winning numbers converges to the equilibrium with fixed number of players if the speed of learning is sufficiently low. Note that in explaining learning in weak-link games (Roth 1995) and proposer competition ultimatum games ("market games", Roth and Erev 1995), Roth and Erev change from reinforcement according to chosen strategies to a model based on imitating the most successful players. Our model continues in this tradition.

winning numbers from period 1 until $t-1$. Similarly, the regression analysis reported in Table 7 shows that players' choices in the lab depend on previous winners (at least in early rounds).

Let $A_k(t)$ denote the attraction of strategy k in period t. Based on these attractions, players probabilistically pick numbers in the next period using a power function so that the probability of picking number k in the next period is

$$p_k(t+1) = \frac{A_k(t)^{\lambda}}{\sum_{j=1}^{K} A_j(t)^{\lambda}}. \tag{5.1}$$

Note that $\lambda = 0$ means uniform randomization and $\lambda \to \infty$ means playing only the strategy with the highest attraction.

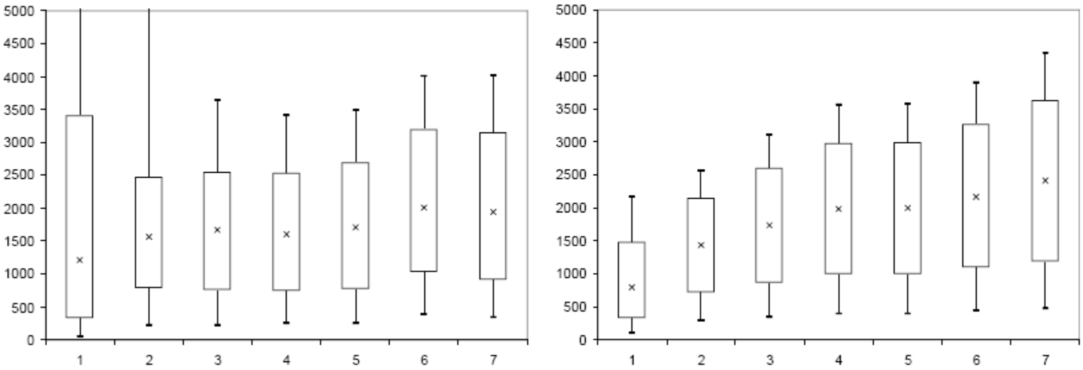
Any learning model requires an assumption about the choice probabilities in the first period, $p_k(1)$. We use the empirical frequencies to create choice probabilities in the first period. Given these probabilties and λ, we determine $A(1)$ so that equation (5.1) gives the assumed choice probabilities $p_k(1)$. Since the power choice function is invariant to scaling, we determine the attractions in the first period so that they sum to one, i.e., $\sum_{k=1}^{K} A_k(1) = 1$. From the second period onwards, strategies are reinforced by a factor $r_k(t)$, which depends on the winning number in period $t-1$. For the empirical estimation of the learning model we use the actual winning numbers from the field and lab. Attractions in period $t > 1$ are given by

$$A_k(t) = \frac{A_k(t-1) + r_k(t)}{1 + \sum_{j=1}^{K} r_j(t)}.$$

The reinforcement factors are determined by the winning number in the previous period (if there is no winning number, the same attractions carry over to the next period). However, since the strategy sets are so large, only reinforcing the previous winning number would predict learning that is too slow and too tightly clustered on previous winners. We therefore follow Sarin and Vahid (2004) by assuming that numbers that are "similar" to the winning number are also reinforced. We use the triangular Bartlett similarity function proposed by Sarin and Vahid (2004), which puts reinforcement on strategies near the previous winner that declines linearly with distance from that winning number. Let W denote the size of the "similarity window" and $k^*(t-1)$ the winning number in the previuos round. Then the reinforcement factors in period t are given by

$$r_k(t) = \frac{\max\{0, (1 - |k - k^*(t-1)|/W\}}{\sum_{j=1}^{K} r_j(t)}.$$

FIGURE 13. Weekly box plots of data (left) and estimated learning model (right) (10-25-50-75-90 percentile box plots, $W = 344$, $\lambda = 0.0085$).



Note that we scale the reinforcement factors so that they sum to one, just as the first period attractions were scaled to sum to one.[31]

The learning model has two parameters: the size of the similarity window, W, and the precision of the choice function, λ. We estimate the best-fitting values by minimizing the squared deviation between predicted choice densities and empirical densities summed over all numbers, rounds and sessions (in the laboratory). The estimated values for the field data are $W = 344$ and $\lambda = 0.0085$. For the laboratory data, we divide the estimated window size from the field by 100 and fix $W = 3$. The estimated λ for the laboratory data is 0.31.[32]

To see how the learning model fits the data, Figure 13 shows box plots of the field data and the prediction of the learning model averaged over weeks. The learning model captures the shift toward higher numbers in later weeks, but it does not explain the extent of very high numbers in the first week. That the model captures the upward shift of the empirical distribution quite well is also shown in Figure 14, which displays the average weekly predicted densities of the learning model for numbers up to 6000.[33]

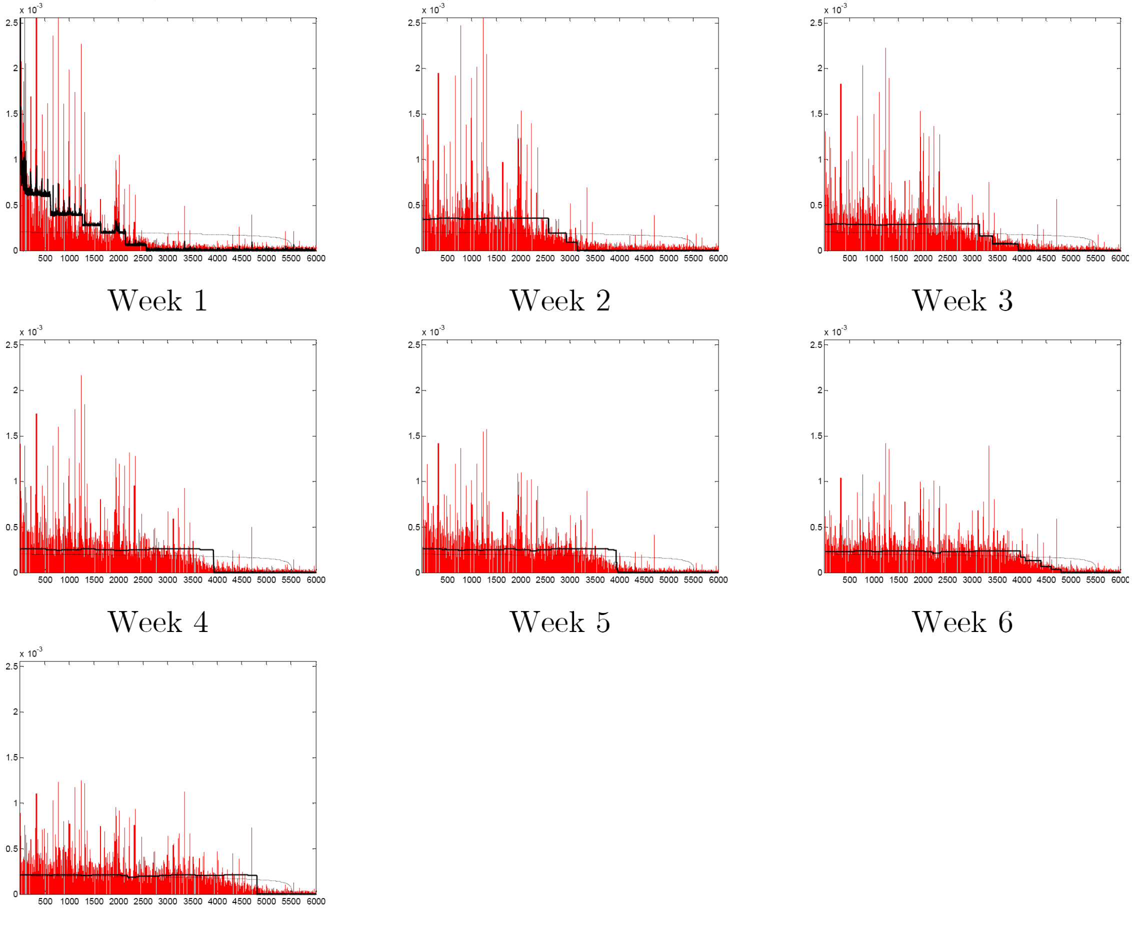
As was discussed in the previous section, players in the laboratory seem to learn to play the game much quicker and there is not so much learning to be explained by the learning model. The learning model can explain some of the ups and downs during the

[31] Figure A7 shows an example of the reinforcement factors when $k^* (t - 1) = 10$ and $W = 3$.

[32] Estimating both W and λ for the laboratory data gives $W = 10$ and $\lambda = 1.62$. However, the fit is nearly identical with the smaller window size. For the lab data, W and λ largely play inverse roles. Higher window sizes W combined with higher response sensitivities λ often generate very close squared deviations (since higher W is generating a wider spread of responses and higher λ is tightening the response). The higher W is, the higher is λ, but the overall fit is nearly unchanged as W varies between 3 and 12. See Appendix C for details.

[33] Note that the learning model fits extremely well in week 1 by construction because it was initialized using actual data from week 1.

Figure 14. Average weekly empirical densities (bars), estimated learning model (lines) and Poisson-Nash equilibrium (dotted lines) for the field ($W = 344$, $\lambda = 0.0085$).



Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

first 14 rounds in the laboratory, as well as the shrinking dispersion of numbers over time, but there is no trend toward higher numbers as seen in the field data.[34]

6. Conclusion

It is difficult to test game theory using field data because equilibrium predictions depend so sensitively on strategies, information and payoffs, which are usually not observable in the field. This paper exploits an empirical opportunity to test game theory

[34] Figure A8 displays box plots for the 14 first rounds in the three sessions. Note that the learning model predicts much more dispersion of numbers in the early rounds in the first session. This is explained by the fact that players played very high numbers in the first round in that session and that a very high number, 67, won in the fourth period. The imitation-based model is substantially affected by that outlying win.

in a field setting which is simple enough that clear predictions apply (with some approximations). The game is a LUPI lottery, in which the lowest unique positive integer wins a fixed prize. LUPI is a close relative of auctions in which the lowest unique bid wins.

One contribution of our paper is to characterize the Poisson-Nash equilibrium and analyze people's behavior in this game using both a field data set, including more than two million choices, and parallel laboratory experiments which closely match the field setting. In both the field and lab, players quickly learn to play close to equilibrium, but there are some diagnostic discrepancies between players' behavior and equilibrium predictions.

Another contribution is measuring learning week-by-week. Most field studies that compare mixed-strategy equilibrium predictions with field data combine data across a long time span in order to test the theory with statistical power. The large amount of data we have enable us to study behavior week-by-week, which permits the study of how learning works. Since the subjects have only the winning number to learn from, fictitious play and hybrid EWA models do not apply well, and simpler reinforcement models (in which players learn only from reinforcement of successful strategies) predict essentially no learning. Therefore, we apply a model in which players imitate successful strategies by shifting reinforcement (and hence, choice probability) to strategies in a window around the previous winning number. This model does a reasonable job of explaining the time path of change in the field data. It does a less impressive job in the lab data, largely because choices are so close to the equilibrium in early periods that there is little to learn.

Because the game is simple, it is also possible to see whether models of bounded rationality—cognitive hierarchy and quantal response equilibrium (QRE)—can explain short-run deviations from the Poisson-Nash equilibrium. The cognitive hierarchy approach can explain overshooting of low numbers (when coupled with quantal response); in the first week of field data, the best-fitting value of τ, the number of average thinking steps, is 2.98, close to estimates derived from experimental data. Numerical computations (reported in Figure 2) indicate that QRE converges from random choices to Poisson-Nash, so it cannot explain why there are too many low numbers chosen in the field data (compared to equilibrium).

Finally, because the LUPI field game is simple, it is possible to do a lab experiment that closely replicates the essential features of the field setting (which most experiments are not designed to do). This close lab-field parallelism in design adds evidence to the ongoing debate about when lab findings generalize to parallel field settings (e.g., Levitt and List 2007). The lab game was described very much like the Swedish lottery (controlling context), experimental subjects were allowed to select out of the experiment after it was described (allowing self-selection), and lab stakes were made equal to the field stakes. Basic lab and field findings are quite close: In both settings, choices are close to equilibrium, but there are too many large numbers and too few agents choose numbers at the high end of the equilibrium range. We interpret this as a good example of close lab-field generalization, when the lab environment is designed to be close to a particular field environment.

Appendix A. The Symmetric Fixed-n Nash Equilibrium

Let there be a finite number of n players that each pick an integer between 1 and K. If there are numbers that are only chosen by one player, then the player that picks the lowest such number wins a prize, which we normalize to 1, and all other players get zero. If there is no number that only one player chooses, everybody gets zero.

To get some intuition for the equilibrium in the game with many players, we first consider the cases with two and three players. If there are only two players and two numbers to choose from, the game reduces to the following bimatrix game.

	1	2
1	$0,0$	$1,0$
2	$0,1$	$0,0$

This game has three equilibria. There are two asymmetric equilibria in which one player picks 1 and the other player picks 2, and one symmetric equilibrium in which both players pick 1.

Now suppose that there are three players and three numbers to choose from (i.e., $n = K = 3$). In any pure strategy equilibrium it must be the case that at least one player plays the number 1, but not more than two players play the number 1 (if all three play 1, it is optimal to deviate for one player and pick 2). In pure strategy equilibria where only one player plays 1, the other players can play in any combination of the other two numbers. In pure strategy equilibria where two players play 1, the third player plays 2. In total there are 18 pure strategy equilibria. To find the symmetric mixed strategy equilibrium, let p_1 denote the probability with which 1 is played and p_2 the probability with which 2 is played. The expected payoff from playing the pure strategies if the other two players randomize is given by

$$\begin{aligned}
\pi(1) &= (1-p_1)^2\,, \\
\pi(2) &= \left[(1-p_1-p_2)^2 + p_1^2\right], \\
\pi(3) &= \left[p_1^2 + p_2^2\right].
\end{aligned}$$

Setting the payoff from the three pure strategies yields $p_1 = 2\sqrt{3} - 3 = 0.464$ and $p_2 = p_3 = 2 - \sqrt{3} = 0.268$.

In the game with n players, there are numerous asymmetric pure strategy equilibria as in the three-player case. For example, in one type of equilibrium exactly one player picks 1 and the other players pick the other numbers in arbitrary ways. In order to find symmetric mixed strategy equilibria, let p_k denote the probability put on number k.[35] In a symmetric mixed strategy equilibrium, the distribution of guesses will follow the multinomial distribution. The probability of x_1 players guessing 1, x_2 players guessing 2 and so on is given by

$$f\left(x_1, ..., x_K; n\right) = \begin{cases} \frac{n!}{x_1! \cdots x_K!} p_1^{x_1} \cdots p_K^{x_K} & \text{if } \sum_{i=1}^{K} x_i = n, \\ 0 & \text{otherwise,} \end{cases}$$

where we use the convention that $0^0 = 1$ in case any of the numbers is picked with zero probability. The marginal density function for the k^{th} number is the binomial distribution

$$f_k\left(x_k; n\right) = \frac{n!}{x_k! \left(n - x_k\right)!} p_k^{x_k} \left(1 - p_k\right)^{n - x_k}.$$

Let $g_k\left(x_1, x_2, ..., x_k; n\right)$ denote the marginal distribution for the first k numbers. In other words, we define g_k for $k < K$ as

$$g_k\left(x_1, x_2, ..., x_k; n\right) = \sum_{x_{k+1} + x_{k+2} + \cdots + x_K = n - \left(x_1 + x_2 + \cdots + x_k\right)} \frac{n!}{x_1! x_2! \cdots x_K!} p_1^{x_1} p_2^{x_2} \cdots p_K^{x_K}.$$

Using the multinomial theorem we can simplify this to[36]

$$g_k\left(x_1, x_2, ..., x_k; n\right) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \frac{\left(p_{k+1} + p_{k+2} + \cdots + p_K\right)^{n - \left(x_1 + x_2 + \cdots + x_k\right)}}{\left(n - \left(x_1 + x_2 + \cdots + x_k\right)\right)!}.$$

If $k = K$, then $g_k\left(x_1, x_2, ..., x_k; n\right) = f\left(x_1, x_2, ..., x_k; n\right)$. Finally, let $h_k\left(n\right)$ denote the probability that nobody guessed k and there is at least one number between 1 to $k - 1$ that only one player guessed. This probability is given by (again if $k < K$)

$$h_k\left(n\right) = \sum_{\substack{\left(x_1, ..., x_{k-1}\right):\ \text{some } x_i = 1 \\ \&\ x_1 + \cdots + x_{k-1} \leq n}} g_k\left(x_1, x_2, ..., x_{k-1}, 0; n\right).$$

If $k = K$, then this probability is given by

$$h_K\left(n\right) = \sum_{\substack{\left(x_1, ..., x_{k-1}\right):\ \text{some } x_i = 1 \\ \&\ x_1 + \cdots + x_{k-1} = n}} f\left(x_1, x_2, ..., x_{K-1}, 0; n\right).$$

[35] We have not been able to show that there is a unique symmetric equilibrium, but when numerically solving for a symmetric equilibrium we have not found any other equilibria than the ones reported below. Existence of a symmetric equilibrium is guaranteed since players have finite strategy sets. (A straightforward extension of Proposition 1.5 in Weibull 1995 shows that all symmetric normal form games with finite number of strategies and players have a symmetric equilibrium.)

[36] The multinomial theorem states that the following holds

$$\left(p_1 + p_2 + \cdots + p_K\right)^n = \sum_{x_1 + x_2 + \cdots + x_K = n} \frac{n!}{x_1! x_2! \cdots x_K!} p_1^{x_1} p_2^{x_2} \cdots p_K^{x_K},$$

given that all $x_i \geq 0$.

The probability of winning when guessing 1 and all other players follow the symmetric mixed strategy is given by

$$\pi\left(1\right)=f_{1}\left(0;n-1\right)=\left(1-p_{1}\right)^{n-1}.$$

The probability of winning when playing $1<k<K$ is given by[37]

$$\begin{aligned}\pi\left(k\right)&=f_{k}\left(0;n-1\right)-h_{k}\left(n-1\right),\\&=\left(1-p_{k}\right)^{n-1}-h_{k}\left(n-1\right).\end{aligned}$$

Similarly, the probability of winning when playing $k=K$ is given by

$$\pi\left(K\right)=f_{K}\left(0;n-1\right)-h_{K}\left(n-1\right).$$

In a symmetric mixed strategy equilibrium, the probability of winning from all pure strategies in the support of the equilibrium must be the same. In the special case when $n=K$ and all numbers are played with positive probability, we can simply solve the system of $K-2$ equations where each equation is

$$\left(1-p_{k}\right)^{n-1}-h_{k}\left(n-1\right)=\left(1-p_{1}\right)^{n-1},$$

for all $2<k<K$ and the Kth equation

$$\left(1-p_{K}\right)^{n-1}-h_{K}\left(n-1\right)=\left(1-p_{1}\right)^{n-1}.$$

In principle, it is straightforward to solve this system numerically. However, computing the h_k function is computationally explosive because it requires the summation over a large set of vectors of length $k-1$. The number of combinations explodes as n and K gets large and it is non-trivial to solve for equilibrium for more than 8 players. As an illustration, when $n=K=7$, $h_7(6)$ involves the summation over 391 vectors, and when $n=K=8$ computing $h_8\left(7\right)$ involves 1520 vectors. To understand the magnitude of the complexity, suppose we want to compute $h_K\left(n-1\right)$.

[37] The easiest way to see this is to draw a Venn diagram. More formally, let $A=\{\text{No other player picks } k\}$ and let $B=\{\text{No number below } k \text{ is unique}\}$, so that $P(A)=f_k\left(0;n-1\right)$ and $P\left(B\right)=h_k\left(n-1\right)$. We want to determine $P(A\cap B)$, which is equal to

$$P(A\cap B)=P(A)+P(B)-P(A\cup B).$$

To determine $P\left(A\cup B\right)$, note that it can be written as the union between two independent events

$$P\left(A\cup B\right)=P\left(B\cup\left(B'\cap A\right)\right).$$

Since B and $B'\cap A$ are independent,

$$P\left(A\cup B\right)=P(B)+P(B'\cap A).$$

Combining this with the expression for $P\left(A\cap B\right)$ we get

$$P(A\cap B)=P(A)-P(A\cap B').$$

This involves the summation over all vectors $(x_1, ..., x_{K-1})$ such that some $x_i = 1$ and $x_1 + \cdots + x_{K-1} = n - 1$. Only a small subset of all these vectors are the ones where $x_1 = 1$. How many such vectors are there? For those vectors there must be $n - 2$ players that play numbers $x_2, ..., x_{K-1}$, i.e., potentially $K - 2$ different strategies. The total number of such vectors are

$$\frac{(K + n - 5)!}{(n - 2)!(K - 3)!},$$

where we have used the fact that the number of sequences of n natural numbers that sum to k is $(n + k - 1)!/(k!(n - 1)!)$. For example, when $n = 27$ and $K = 99$, the number of vectors in which $x_1 = 1$ is larger than 10^{25}. Note that this number is much lower than the actual total number of vectors since we have only counted vectors such that $x_1 = 1$.

Assuming $n = K$, the table below show the equilibrium for up to eight players.[38]

	3x3	**4x4**	**5x5**	**6x6**	**7x7**	**8x8**
1	0.4641	0.4477	0.3582	0.3266	0.2946	0.2710
2	0.2679	0.4249	0.3156	0.2975	0.2705	0.2512
3	0.2679	0.1257	0.1918	0.2314	0.2248	0.2176
4		0.0017	0.0968	0.1225	0.1407	0.1571
5			0.0376	0.0216	0.0581	0.0822
6				0.0005	0.0110	0.0199
7					0.0004	0.0010
8						0.0000

These probabilities are close to the Poisson-Nash equilibrium probabilities. To see this, the table below shows the Poisson-Nash equilibrium probabilities when n is equal to K for 3 to 8 players. Note that all the fixed-n and Poisson-Nash probabilities for all strategies in the 5x5 game and larger are within 0.02.

	3x3	**4x4**	**5x5**	**6x6**	**7x7**	**8x8**
1	0.4773	0.4057	0.3589	0.3244	0.2971	0.2747
2	0.3378	0.3092	0.2881	0.2701	0.2541	0.2397
3	0.1849	0.1980	0.2046	0.2057	0.2030	0.1983
4		0.0870	0.1129	0.1315	0.1430	0.1492
5			0.0355	0.0575	0.0775	0.0931
6				0.0108	0.0234	0.0385
7					0.0020	0.0064
8						0.0002

[38] See Appendix C for details about how these probabilites were computed.

Appendix B. Proof of Proposition 1

We first prove the four properties and then prove that the equilibrium is unique.

(1) We prove this property by induction. For $k = 1$, we must have $p_1 > 0$. Otherwise, deviating from the proposed equilibrium by choosing 1 would guarantee winning for sure. Now suppose that there is some number $k + 1$ that is not played in equilibrium, but that k is played with positive probability. We show that $\pi(k+1) > \pi(k)$, implying that this cannot be an equilibrium. To see this, note that the expressions for the expected payoffs allows us to write the ratio $\pi(k+1)/\pi(k)$ as

$$\frac{\pi(k+1)}{\pi(k)} = \frac{\prod_{i=1}^{k} Pr(X(i) \neq 1) \cdot Pr(X(k+1) = 0)}{\prod_{i=1}^{k-1} Pr(X(i) \neq 1) \cdot Pr(X(k) = 0)}$$
$$= \frac{Pr(X(k) \neq 1) \cdot Pr(X(k+1) = 0)}{Pr(X(k) = 0)}.$$

If $k + 1$ is not used in equilibrium, $Pr(X(k+1) = 0) = 1$, implying that the ratio is above one. This shows that all integers between 1 and K are played with positive probability in equilibrium.

(2) Rewrite equation (2.1) as

$$e^{np_{k+1}} - e^{np_k} = -np_k.$$

By the first property, both p_k and p_{k+1} are positive, so that the right hand side is negative. Since the exponential is an increasing function, we conclude that $p_k > p_{k+1}$.

(3) First rearrange equation (2.1) as

$$p_{k+1} = p_k + \frac{1}{n}\ln\left(1 - np_k e^{-np_k}\right). \tag{A1}$$

We want to determine $(p_k - p_{k+1})/(p_{k+1} - p_{k+2})$. Using (A1) we can write this ratio as

$$\frac{p_k - p_{k+1}}{p_{k+1} - p_{k+2}} = \frac{\ln(1 - np_k e^{-np_k})}{\ln(1 - np_{k+1}e^{-np_{k+1}})} = \frac{\ln(Pr(X(k) \neq 1))}{\ln(Pr(X(k+1) \neq 1))}.$$

The derivative of $Pr(X(k) \neq 1)$ with respect to p_k is positive if $p_k > 1/n$ and negative if $p_k < 1/n$. We therefore have shown that $(p_k - p_{k+1})$ is increasing in k when $p_k > 1/n$, whereas the difference is decreasing for $p_k > 1/n$.

(4) Taking the limit of (A1) as $n \to \infty$ implies that $p_{k+1} = p_k$.

In order to show that the equilibrium $\mathbf{p} = (p_1, p_2, \cdots, p_K)$ is unique, suppose by contradiction that there is another equilibrium $\mathbf{p}' = (p'_1, p'_2, \cdots, p'_K)$. By the equilibrium condition (2.1), p_1 uniquely determines all probabilities $p_2, ..., p_K$, while p'_1

uniquely determines $p'_2, ..., p'_K$. Without loss of generality, we assume $p'_1 > p_1$. Since in any equilibrium, p_{k+1} is strictly increasing in p_k by condition (2.1), it must be the case that all positive probabilities in $\mathbf{p}'$ are higher than in $\mathbf{p}$. However, since $\mathbf{p}$ is an equilibrium, $\sum_{k=1}^{K} p_k = 1$. This means that $\sum_{k=1}^{K} p'_k > 1$, contradicting the assumption that $\mathbf{p}'$ is an equilibrium.

Appendix C. Computational and Estimation Issues

This appendix provides details about the numerical computations and estimations that are reported in the paper. We have used MATLAB 7.4.0 for all computations and estimations. Both the data and all MATLAB programs that have been used for the paper can be obtained from the authors upon request.

Poisson-Nash Equilibrium. The Poisson-Nash equilibrium was computed in MATLAB through iteration of the equilibrium condition (2.1). Unfortunately, MATLAB cannot handle the extremely small probabilities that are attached to high numbers in equilibrium, so the estimated probablities are zero for high numbers (17 and above for the laboratory and 5519 and above for the field).

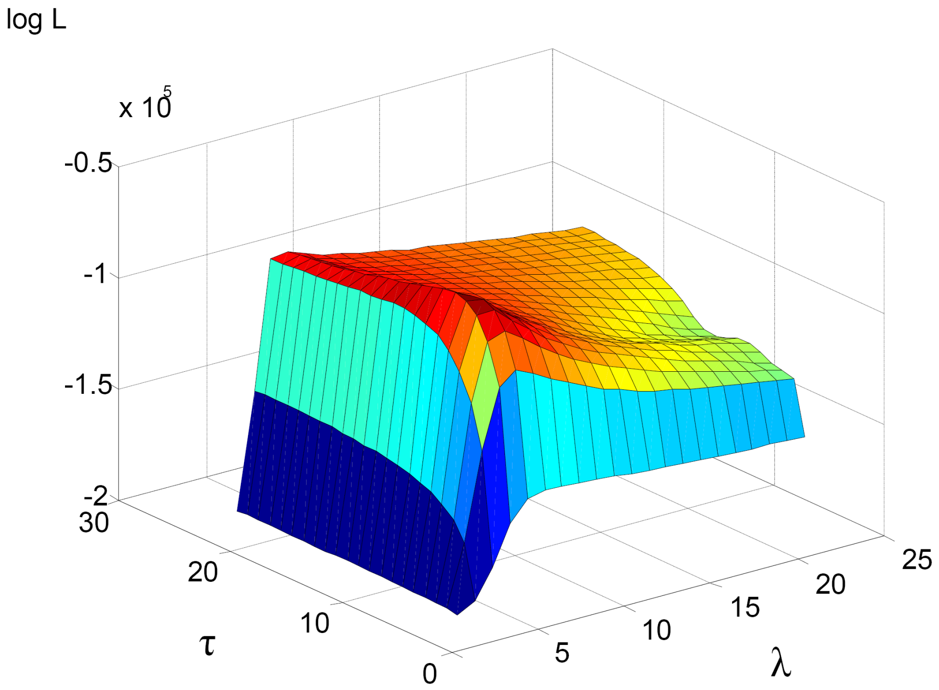
Fixed-n Equilibrium. To compute the equilibrium when the number of players is fixed and commonly known, we programmed the functions f_k, f_K, h_k and h_K in MATLAB and then solved the system of equations characterizing equilibrium using MATLAB's solver fsolve. However, the h_k function includes the summation of a large number of vectors. For high k and n the number of different vectors involved in the summation grows explosively and we only managed solve for equilibrium for up to 8 players.

Cognitive Hierarchy with Quantal Response. Calculating the cognitive hierarchy prediction for a given τ and λ is straightforward. However, the cognitive hierarchy prediction is non-monotonic in τ and λ, implying that the log-likelihood function isn't generally smooth.

In order to calculate the log-likelihood, we assume that all players play according to the same aggregate cognitive hierarchy prediction, i.e., the log-likelihood function is calculated using the multinomial distribution as if all players played the same strategy. For the field data, we calculated the log-likelihood for the daily average frequency for each week, but the frequency was rounded to integers in order to be able to calculate the log-likelihood. For the lab data, we instead calculated the log-likelihood by summing the frequencies for each week since we didn't want unnecessary estimation errors due to rounding off to integers.

Maximum likelihood estimation for the field data is computationally demanding so we used a relatively coarse two-dimensional grid search. We used a 20x20 grid and restricted τ to be between 0.05 and 12, and restricted λ to be between 0.0001 and 0.05. We tried wider bounds on the parameters as well, but that didn't change the results. The log-likelihood function is shown in Figure A1. The log-likelihood appears

FIGURE A1. Log-likelihood for cognitive hierarchy in the field (first week)

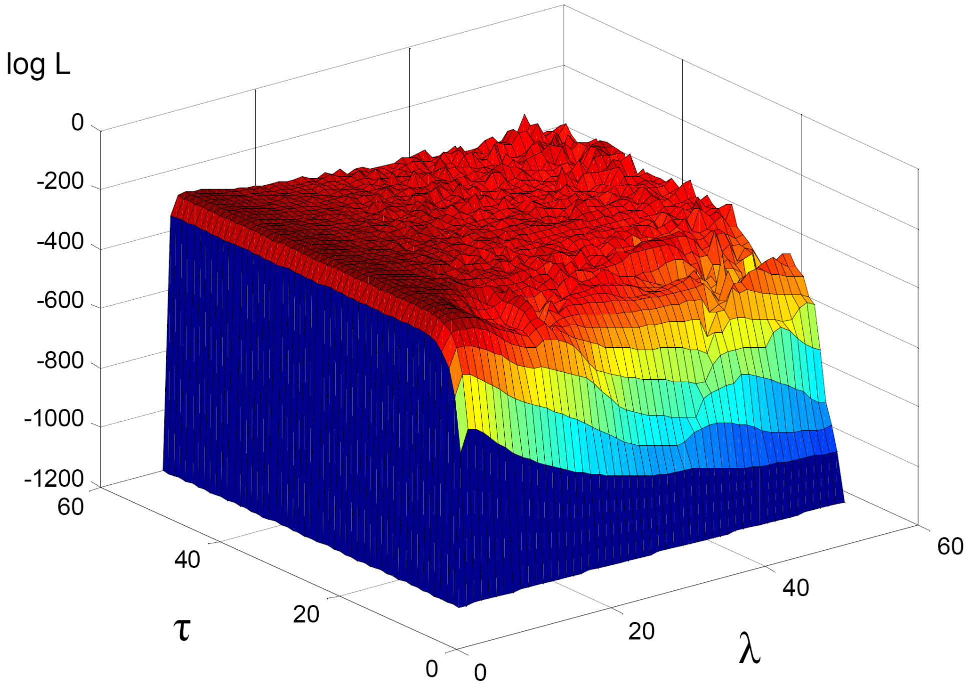


relatively smooth, but since we have been forced to use a very coarse grid we might not have found the global maximum.

For the maximum likelihood estimation of the lab data, we used a two-dimensional 300x300 grid search. We tried different bounds on τ and λ, then let both parameters vary between 0.001 and 20. The three-dimensional log-likelihood function is shown in Figure A2. It is clear that the log-likelihood function isn't smooth and that it is very flat with respect to λ when λ is low. There is therefore no guarantee that we have found a global maximum, but we have tried different grid sizes and bounds on the parameters which resulted in the same estimates.

When τ is fixed at 1.5, the maximum likelihood estimation is simpler. We used a grid size of 300 and tried different bounds for λ with unchanged results. The log-likelihood function for $\lambda = 0.001$ to $\lambda = 100$ from the first week is shown in Figure A3. The log-likelihood function is not globally concave, but seems to be concave around the global maximum, so it is likely that we have found a global maximum. Figure A4 shows the cognitive hierarchy prediction week-by-week for the laboratory data when τ is 1.5.

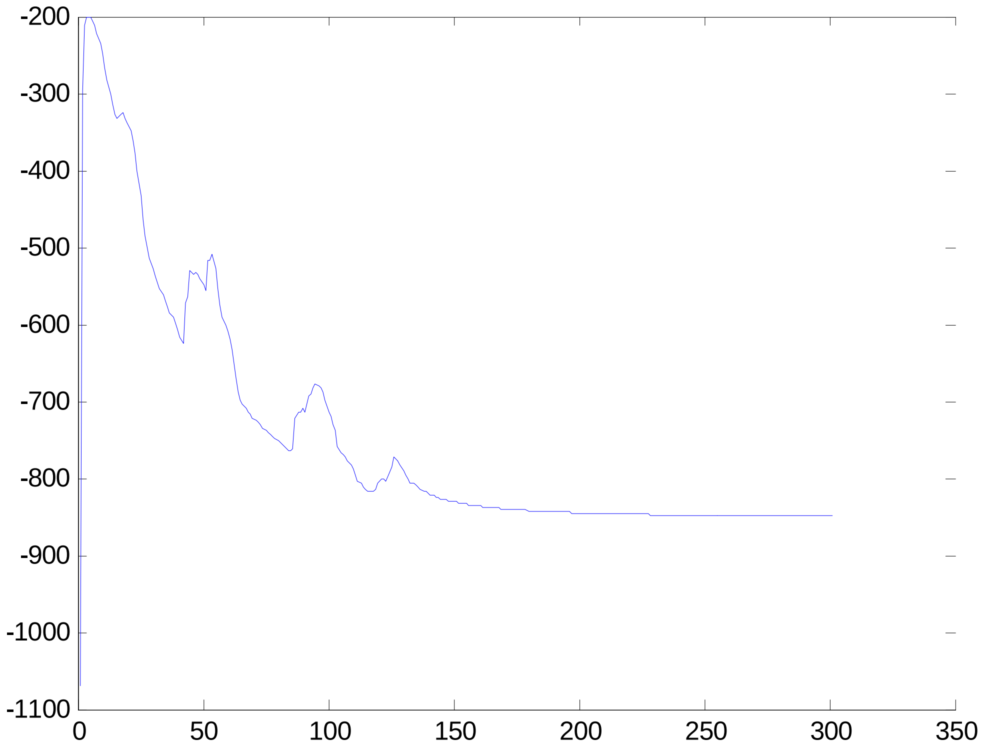
FIGURE A2. Log-likelihood for cognitive hierarchy in the laboratory (first week)



QRE. In order to calculate the QRE for a given level of λ, we used MATLAB's solver fsolve to solve the fixed-point equation that characterizes the QRE. In the ML estimation for the laboratory data we allowed λ between 0.001 and 700. To find the optimal value we used a grid search with a grid size of 50. The log-likelihood function for the first week is shown in Figure A5. The log-likelihood function is smooth and concave, indicating that we have are likely to have found a global maximum. In some of the cases the estimated λ is very high, in which case there might be a computational problem when calculating the QRE. However, for such high λ, the QRE is practically indistinguishable from the Poisson equilibrium anyway (as shown in Figure 2).

Learning. To estimate the learning model, we use the actual winning numbers in the field and in each laboratory session. The predicted choice probabilities are evaluated based on the sum of squared distances from the empirical densities, summed over numbers, days and sessions (in the laboratory). For the field data, we estimated λ through a grid search (with a grid size of 15) for window sizes between 100 and 400 and λ between 0.005 and 0.5. The sum of squared deviations with respect to both W and λ appears to be relatively smooth and convex, so it is likely that we have find

FIGURE A3. Log-likelihood function for cognitive hierarchy in the laboratory (first week, $\tau = 1.5$)



the best-fitting values. For the laboratory data, we estimated λ through grid search (with a grid size of 1000) for window sizes between 1 and 13 and λ between 0.01 and 2. Figure A6 shows the sum of the squared deviations for the laboratory data. As can be seen from the graph, the fit is relatively flat with respect to both W and λ when both parameters are increased proportionally. We have tried different bounds on the parameters and grid sizes and the estimated parameters appears robust. Figure A7 shows an example of a Bartlett similarity window and Figure A8 shows box plots with the data and learning model for the first 14 rounds in the laboratory.

Model Selection. Since the Poisson-Nash equilibrium probabilities are zero for high numbers, the likelihood of the equilibrium prediction is always zero. However, to be able to compare the equilibrium prediction with the cognitive hierarchy model and QRE, we calculate the log-likelihoods using only data on numbers up to 5518 (field) and 16 (laboratory). These log-likelihoods cannot be directly compared with the log-likelihoods in Table 3 and 6, however, since those are calculated using data

Figure A4. Average daily frequencies in the laboratory, Poisson-Nash equilibrium prediction (dashed lines) and estimated cognitive hierarchy (solid lines) when $\tau = 1.5$ (line), week 1 to 7

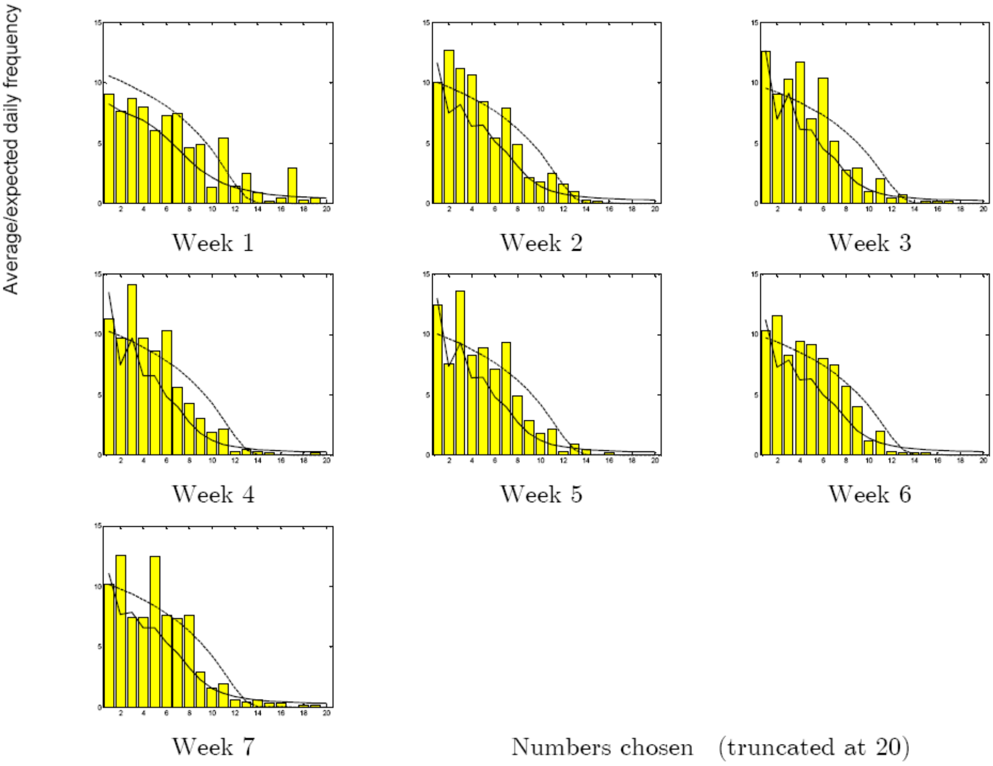


Figure A5. Log-likelihood function for QRE in the laboratory (first week)

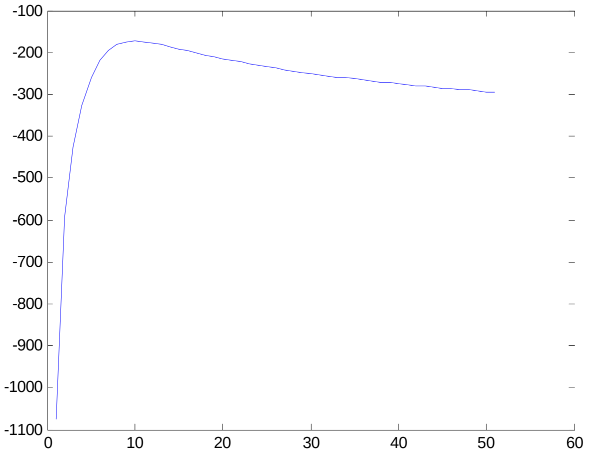


FIGURE A6. Sum of squared deviation for learning model in the laboratory ($W = 1, ..., 13$, $\lambda = 0.01, ..., 2$)

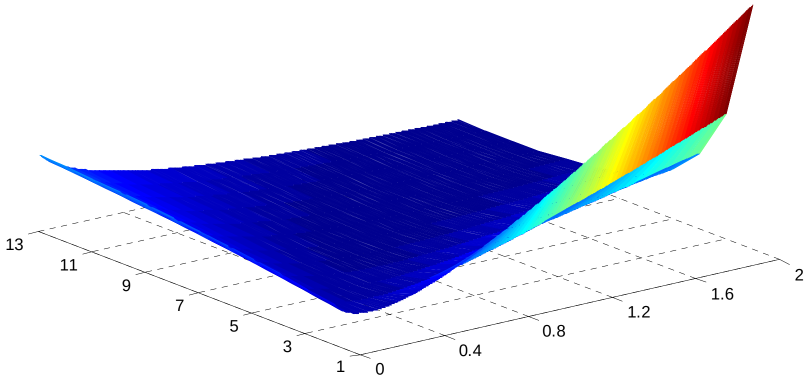
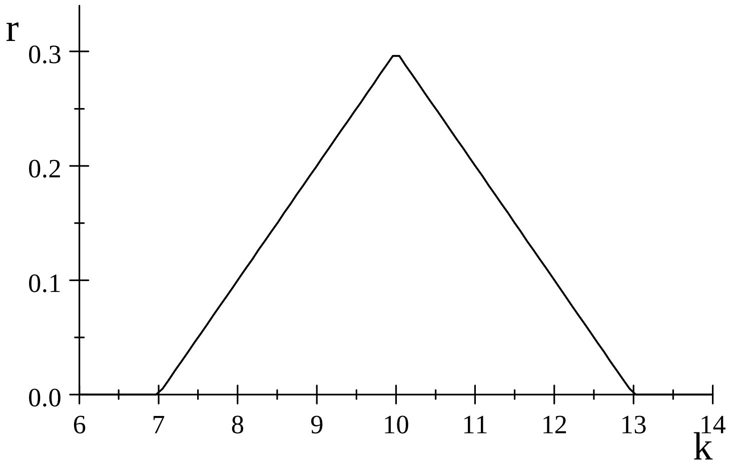
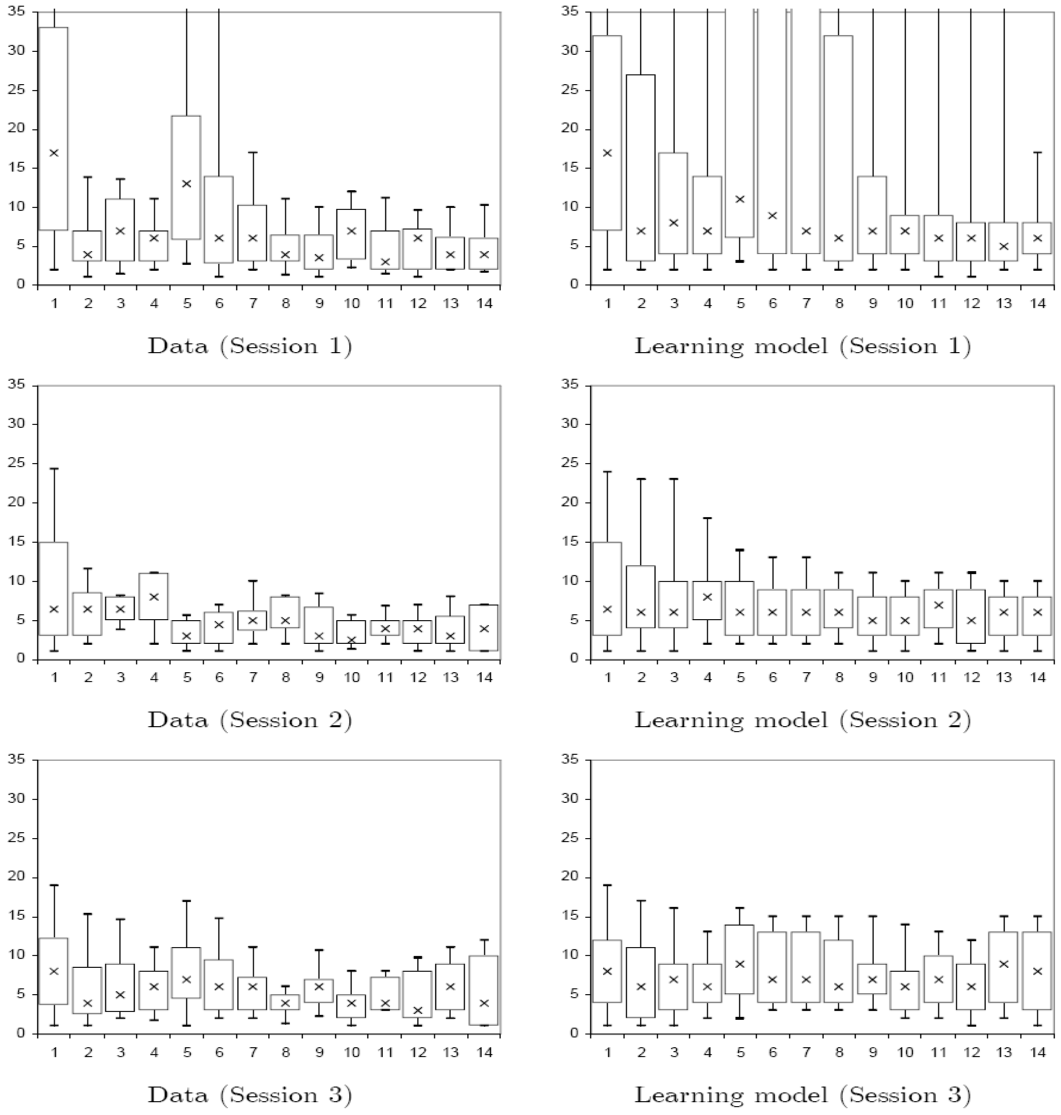


FIGURE A7. Bartlett similarity window ($k^* = 10, W = 3$)



on all numbers. For comparison, we therefore compute the log-likelihoods for the cognitive hierarchy model (as well as QRE for the laboratory) in the same way as for the equilibrium prediction. In order for these probabilites to sum up to one, we divide the probabilities by the total probability attach to numbers up to the threshold (5518 or 16). Using the estimated parameters reported in Table 2, Table A1 shows the log-likehoods only based on numbers up to 5518.

Figure A8. Box plots of data (left) and estimated learning model (right) for round 1-14 in the three laboratory sessions (10-25-50-75-90 percentile box plots, $W = 3, \lambda = 0.31$).



Data (Session 1)

Learning model (Session 1)

Data (Session 2)

Learning model (Session 2)

Data (Session 3)

Learning model (Session 3)

Table A1. Log-likelihoods for cognitive hierarchy and equilibrium for field data (up to 5518)

Week	1	2	3	4	5	6	7
Log-likelihood eq. (<5519)	-43365	-32073	-28453	-27759	-28087	-21452	-19719
Log-likelihood CH (< 5519)	-25307	-21606	-18630	-16253	-16123	-15829	-15010

The log-likelihoods are higher for the cognitive hierarchy model in all weeks. The cognitive hierarchy model is estimated with two parameters, while the equilibrium prediction has no free parameters. One way to compare the models is to use Schwarz (1978) information criterion which penalizes a model depending on the number of estimated parameters by substracting a factor $\log(n) \times m/2$ from the log-likelhood value, where n is the number of observations and m the number of estimated parameters. The log-likelihoods in Table A1 are calculated based on daily averages, so the penalty for the cognitive hierarchy model is approximately $\log(53783) = 10.9$, indicating that the cognitive hierarchy model is the better model in all weeks. Schwarz information criterion penalizes the number of estimated parameters more harshly than for example Aikake's information criterion. However, it should be kept in mind that the two parameters in cognitive hierarchy model are estimated using the data, whereas the equilibrium prediction is not estimated at all, so any comparison based on information criteria is likely to be unfair.

Table A2. Log-likelihood and Schwarz information criterion (BIC) for the cognitive hierarchy, QRE and equilibrium models in the laboratory (up to 16)

Week	1	2	3	4	5	6	7
Log-likelihood eq. (<17)	-143.9	-68.0	-95.0	-71.0	-85.7	-58.6	-117.4
Log-likelihood CH (<17)	-59.0	-52.1	-47.7	-40.7	-57.1	-37.9	-43.6
Log-likelihood CH $\tau = 1.5$ (<17)	-65.3	-55.2	-56.3	-54.2	-54.5	-56.8	-59.2
Log-likelihood QRE (<17)	-65.4	-58.1	-72.7	-62.8	-64.4	-53.2	-63.1
BIC eq. (<17)	-143.9	-68.0	-95.0	-71.0	-85.7	-58.6	-117.4
BIC CH (<17)	-65.2	-58.4	-54.0	-47.0	-63.5	-44.2	-49.9
BIC CH $\tau = 1.5$ (<17)	-68.4	-58.4	-59.5	-57.4	-57.6	-59.9	-62.4
BIC QRE (<17)	-68.6	-61.3	-75.9	-66.0	-67.5	-56.3	-66.3

Table A2 reports the restricted log-likelihoods and the corresponding values of the Schwarz information criterion for the laboratory data. Based on Schwarz information criterion, both the cognitive hierarchy model and QRE outperforms equilibrium in all weeks, but the equilibrium prediction does better than the cognitive hierachy model with $\tau = 1.5$ in the sixth week.

Appendix D. Additional Details About the Field LUPI Game

This part of the Appendix provides some additional details about the field game that was not discussed in the main text.

The prize guarantee for the winner of 100,000 SEK was first extended until the 11th of March and then to the 18th of March, so the prize guarantee covered all days for which we have data. The thresholds for the second and third prizes were determined so that the second prizes constituted 11 percent of all bets and the third prizes 17.5 percent. The winner of the first prize also won the possibility to participate in a "final game".[39] The final game ran weekly and had four to seven participants. The "final game" consisted of three rounds where the participants chose two numbers in each round. The rules of this game were very similar to the original game, but what happened in this game did not depend on what number you chose in the main game, so we leave out the details about this game.

The Hux Flux randomization option involved a uniform distribution where the support of the distribution was determined by the play during the 7 previous days.[40] It became possible to play the game on the Internet sometime between the 21st and 26th of February 2007. The web interface for online play is shown in Figure A10. This interface also included the option HuxFlux, but in this case players could see the number that was generated by the computer before deciding whether to place the bet.

We use daily data from the first seven weeks. The reason is that the game was withdrawn from the market on the 24th of March 2007 and we were only able to access data up to the 18th of March 2007.

Figure A11 shows histograms for the total number of daily bets separately for all days and for Sundays and Mondays. Figure A12 shows empirical frequencies together with the Poisson-Nash equilibrium for the last week in the field.

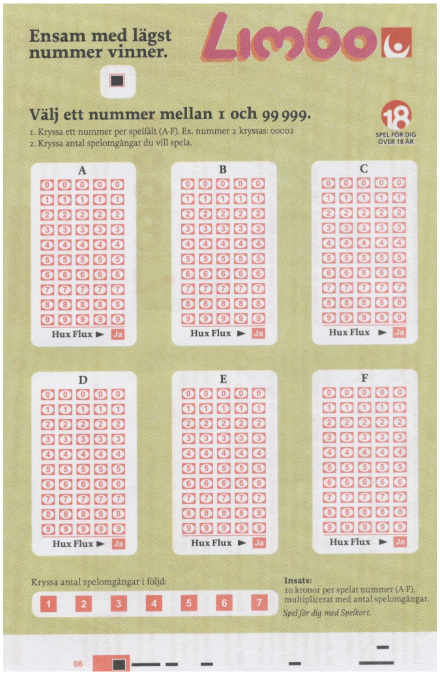
The game was heavily advertised around the days when it was launched and the main message was that this was a new game where you should be alone with the lowest number. The winning numbers (for the first, second, and third prizes) were reported on TV, text-TV and the Internet every day. In the TV programs they reported not only the winning numbers, but also commented briefly about how people had played previously.

The richest information about the history of play was given on the home page of Svenska Spel. People could display and download the frequencies of all numbers played

[39] 3.5 percent of all daily bets were reserved for this "final game".

[40] In the first week HuxFlux randomized numbers uniformly between 1 and 15000. After seven days of play, the computer randomized uniformly between 1 and the average 90th percentile from the previous seven days. However, the only information given to players about HuxFlux was that a computer would choose a number for them.

FIGURE A9. The paper entry form for the Swedish LUPI (Limbo) game



Ensam med lägst nummer vinner.

Välj ett nummer mellan 1 och 99 999.

1. Kryssa ett nummer per spelfält (A-F). Ex. nummer 2 kryssas: 00002
2. Kryssa antal spelomgångar du vill spela.

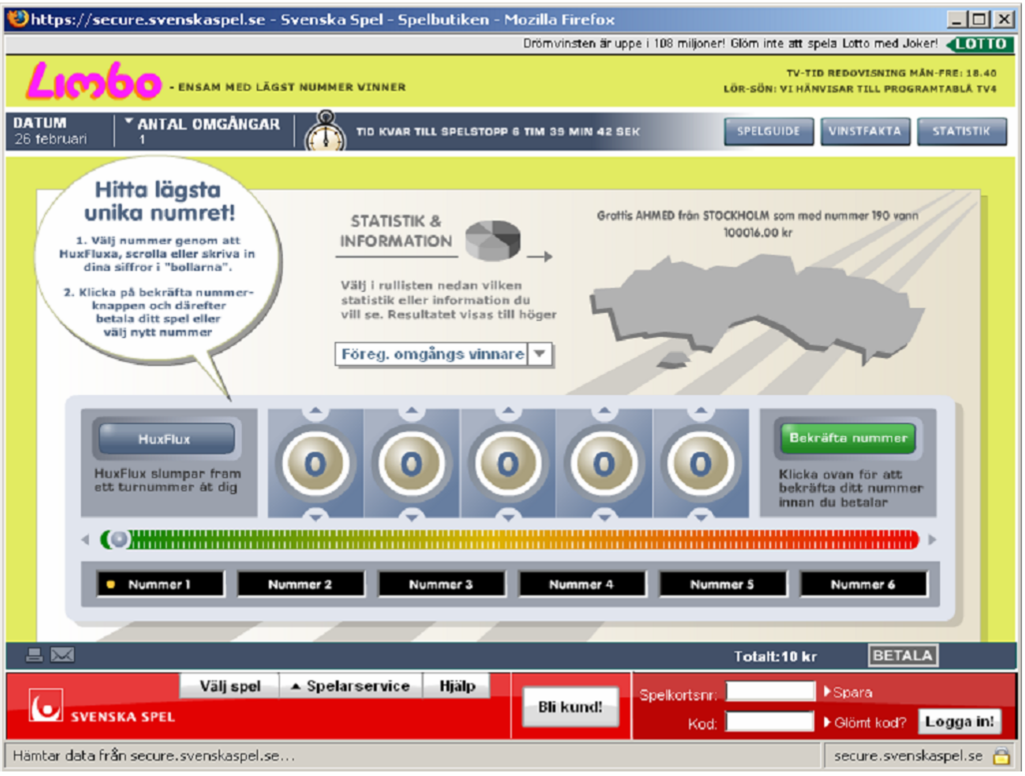
Kryssa antal spelomgångar i följd:

Insats:
10 kronor per spelat nummer (A-F), multiplicerat med antal spelomgångar.

Spel för dig med Spelkort.

Limbo
18
SPEL FÖR DIG ÖVER 18 ÅR
A
B
C
D
E
F
0 1 2 3 4 5 6 7 8 9
Hux Flux ► Ja
1 2 3 4 5 6 7
06

FIGURE A10. Online entry interface for the Swedish LUPI (Limbo) game



https://secure.svenskaspel.se - Svenska Spel - Spelbutiken - Mozilla Firefox

Drömvinsten är uppe i 108 miljoner! Glöm inte att spela Lotto med Joker!

Limbo - ENSAM MED LÄGST NUMMER VINNER

TV-TID REDOVISNING MÅN-FRE: 18.40
LÖR-SÖN: VI HÄNVISAR TILL PROGRAMTABLÅ TV4

DATUM: 26 februari

ANTAL OMGÅNGAR: 1

TID KVAR TILL SPELSTOPP 6 TIM 39 MIN 42 SEK

Hitta lägsta unika numret!

1. Välj nummer genom att HuxFluxa, scrolla eller skriva in dina siffror i "bollarna".

2. Klicka på bekräfta nummer-knappen och därefter betala ditt spel eller välj nytt nummer

STATISTIK & INFORMATION

Välj i rullisten nedan vilken statistik eller information du vill se. Resultatet visas till höger

Föreg. omgångs vinnare

Grattis AHMED från STOCKHOLM som med nummer 190 vann 100016.00 kr

HuxFlux slumpar fram ett turnummer åt dig

0 0 0 0 0

Klicka ovan för att bekräfta ditt nummer innan du betalar

Totalt: 10 kr

Spelkortsnr:

Kod:

Hämtar data från secure.svenskaspel.se...

secure.svenskaspel.se

Figure A11. Total number of daily bets on all days (left) and Sundays and Mondays (right)

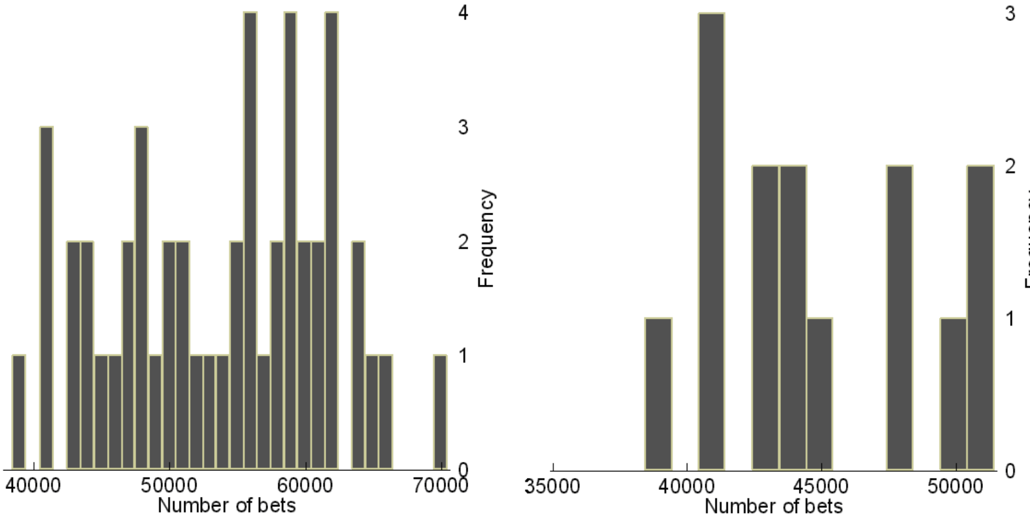
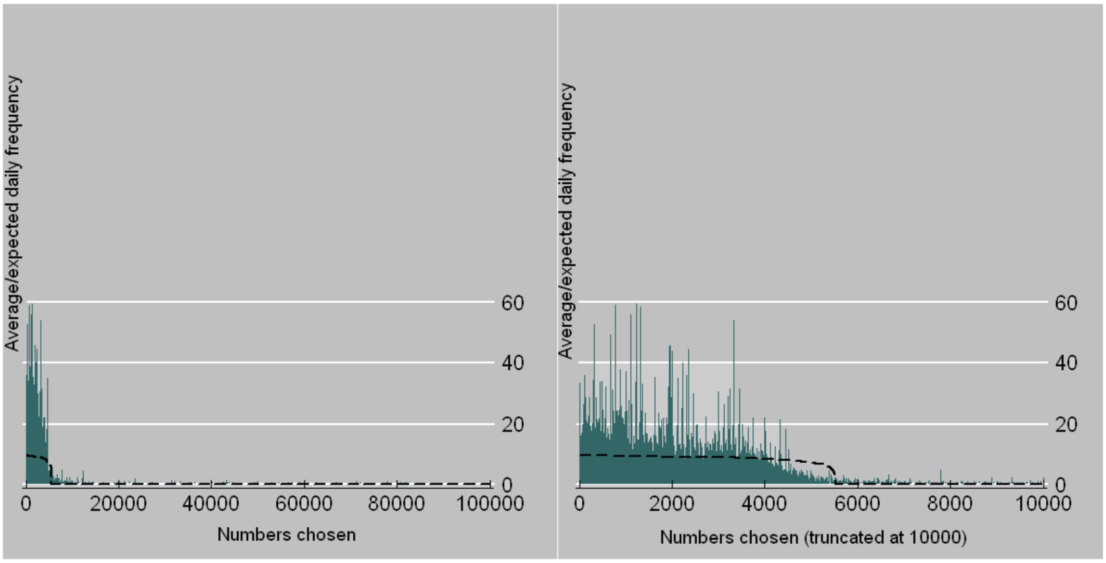
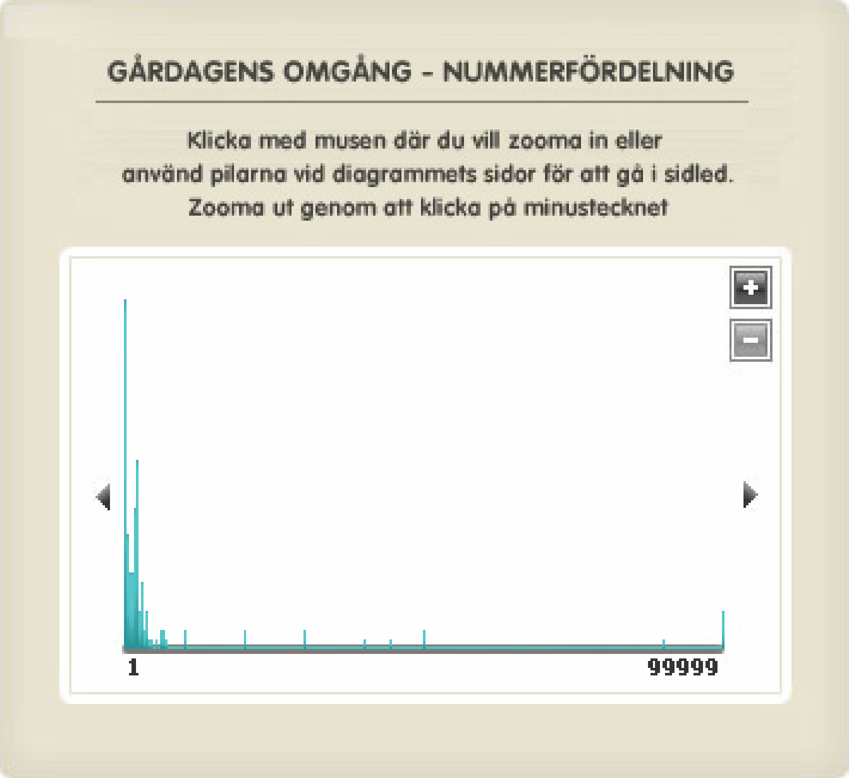


Figure A12. Average daily frequencies and equilibrium prediction for the last week in the field



for all previous days. However, this data was presented in a raw format and therefore not very accessible. The homepage also displayed a histogram of yesterday's guesses which made the data easier to digest. An example of how this histogram looked is shown in Figure A13. The homepage also showed the total number of bets that had been made so far during the day.

Figure A13. Histogram of yesterday's bets as shown online



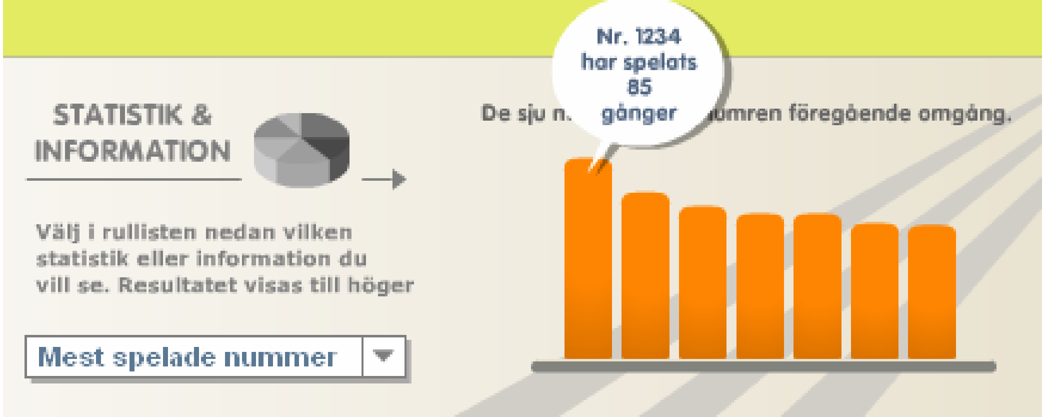
GÅRDAGENS OMGÅNG - NUMMERFÖRDELNING

Klicka med musen där du vill zooma in eller
använd pilarna vid diagrammets sidor för att gå i sidled.
Zooma ut genom att klicka på minustecknet

The web interface for online play also contained some easily accessible information. Besides links to the data discussed above as well as information about the rules of the game, there were some pieces of statistics that could easily be displayed from the main screen. The default information shown was the first name and home town of yesterday's first prize winner and the number that that person guessed. By clicking on the pull-down menu in the middle, you could also see the seven most popular guesses from yesterday. This information was shown in the way shown in Figure A14. By moving the mouse over the bars you can see how many people guessed that number. In this example, the most popular number was 1234 with 85 guesses! Note that this information was not easily available before online play was possible. From the same pull-down menu, you could also see the total number of distinct numbers people guessed on during the last seven days. Finally, you could display the numbers of the second- and third prize winners of yesterday.

In addition to this information, Svenska Spel also published posters with summary statistics for previous rounds of the game (see Figure A15). The information given on these posters varied slightly, but the one in Figure A15 shows the winning numbers, the number of bets, the size of the first prize and if there was any numbers below

FIGURE A14. Most popular numbers yesterday as shown online



STATISTIK & INFORMATION

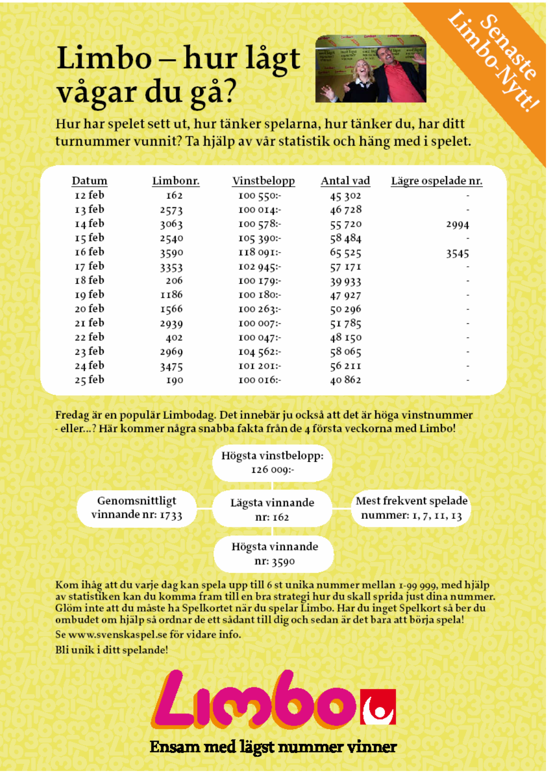
Välj i rullisten nedan vilken statistik eller information du vill se. Resultatet visas till höger

Mest spelade nummer

De sju n... ...ümren föregående omgång.

Nr. 1234 har spelats 85 gånger

FIGURE A15. Example of Limbo poster



Senaste Limbo-Nytt!

Limbo – hur lågt vågar du gå?

Hur har spelet sett ut, hur tänker spelarna, hur tänker du, har ditt turnummer vunnit? Ta hjälp av vår statistik och häng med i spelet.

Datum	Limbonr.	Vinstbelopp	Antal vad	Lägre ospelade nr.
12 feb	162	100 550:-	45 302	-
13 feb	2573	100 014:-	46 728	-
14 feb	3063	100 578:-	55 720	2994
15 feb	2540	105 390:-	58 484	-
16 feb	3590	118 091:-	65 525	3545
17 feb	3353	102 945:-	57 171	-
18 feb	206	100 179:-	39 933	-
19 feb	1186	100 180:-	47 927	-
20 feb	1566	100 263:-	50 296	-
21 feb	2939	100 007:-	51 785	-
22 feb	402	100 047:-	48 150	-
23 feb	2969	104 562:-	58 065	-
24 feb	3475	101 201:-	56 211	-
25 feb	190	100 016:-	40 862	-

Fredag är en populär Limbodag. Det innebär ju också att det är höga vinstnummer - eller...? Här kommer några snabba fakta från de 4 första veckorna med Limbo!

Högsta vinstbelopp: 126 009:-

Genomsnittligt vinnande nr: 1733

Lägsta vinnande nr: 162

Mest frekvent spelade nummer: 1, 7, 11, 13

Högsta vinnande nr: 3590

Kom ihåg att du varje dag kan spela upp till 6 st unika nummer mellan 1-99 999, med hjälp av statistiken kan du komma fram till en bra strategi hur du skall sprida just dina nummer. Glöm inte att du måste ha Spelkortet när du spelar Limbo. Har du inget Spelkort så ber du ombudet om hjälp så ordnar de ett sådant till dig och sedan är det bara att börja spela!

Se www.svenskaspel.se för vidare info.

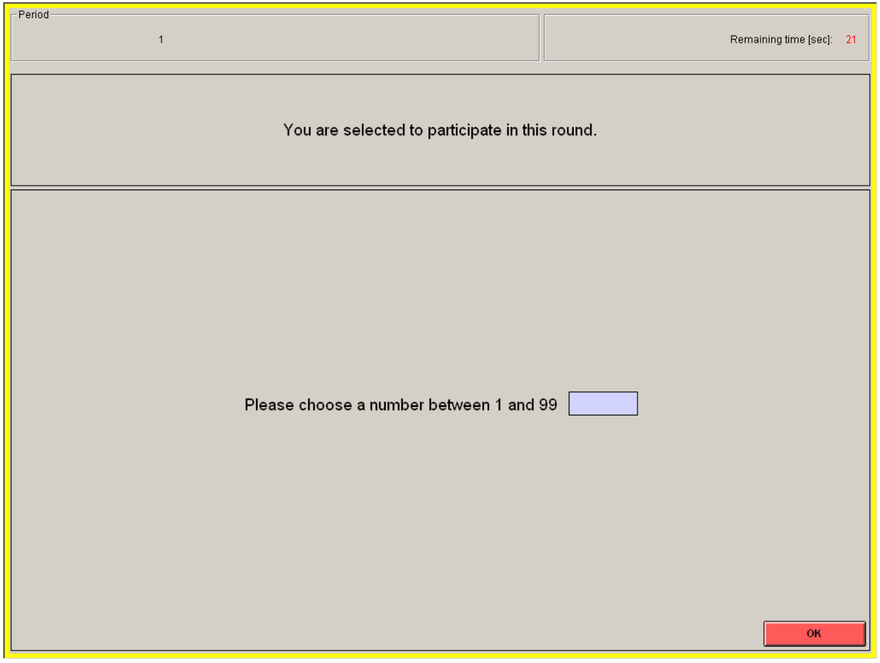
Bli unik i ditt spelande!

Limbo

Ensam med lägst nummer vinner

the winning number that no other player chose. It also shows the average, lowest and highest winning number, as well as the most frequently played numbers.

Figure A16. Screenshot of input screen in the laboratory experiment



You are selected to participate in this round.

Please choose a number between 1 and 99

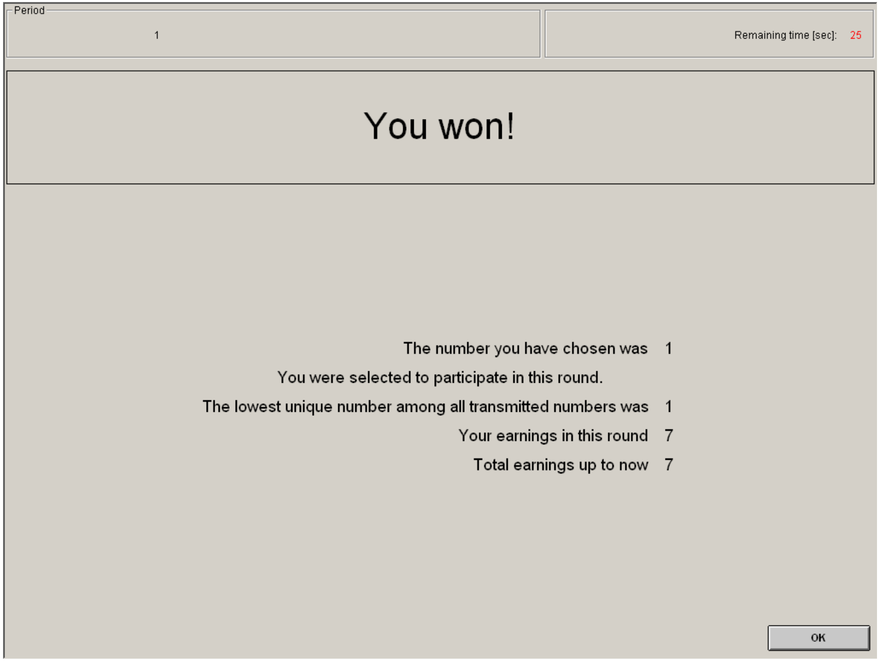
Appendix E. Additional Details About the Lab Experiment

Screenshots from the input and results screens of the laboratory experiment are shown in Figure A16 and A17. Figure A18 shows screenshots from the post-experimental questionnaire and Figure A19 a screenshot from the CRT.

Behavior in the laboratory differs slightly among the three sessions. We cannot reject that the first two sessions are different (the p-value using a Mann-Whitney test is 0.44), but the third session is statistically different from the pooled data from the other two sessions (Mann-Whitney p-value 0.009). However, if we only use the choices of players who were selected to participate in each round, we cannot reject that the distribution of the data is the same in all sessions at $p < 0.05$.[41] It should be noted, that we cannot reject that participating and non-participating players' behavior differ when pooling data from all sessions (Mann-Whitney p-value 0.16). Figure A20 displays the aggregate data from non-selected and selected subjects' choices. Subjects are slightly more likely to play high numbers above 20 when they are not selected to participate, but overall the pattern looks very similar. This implies that subjects'

[41] Using only selected players' choices, a Mann-Whitney test of the null hypothesis that the first two sessions are the same results in a p-value of 0.22. Separately comparing the third session with the first two sessions with the field distribution of players result in p-values of 0.06 and 0.46. Comparing the third session with the pooled data from the first two sessions results in a p-value of 0.13.

Figure A17. Screenshot of result screen in the laboratory experiment



1

Remaining time [sec]: 25

You won!

The number you have chosen was 1

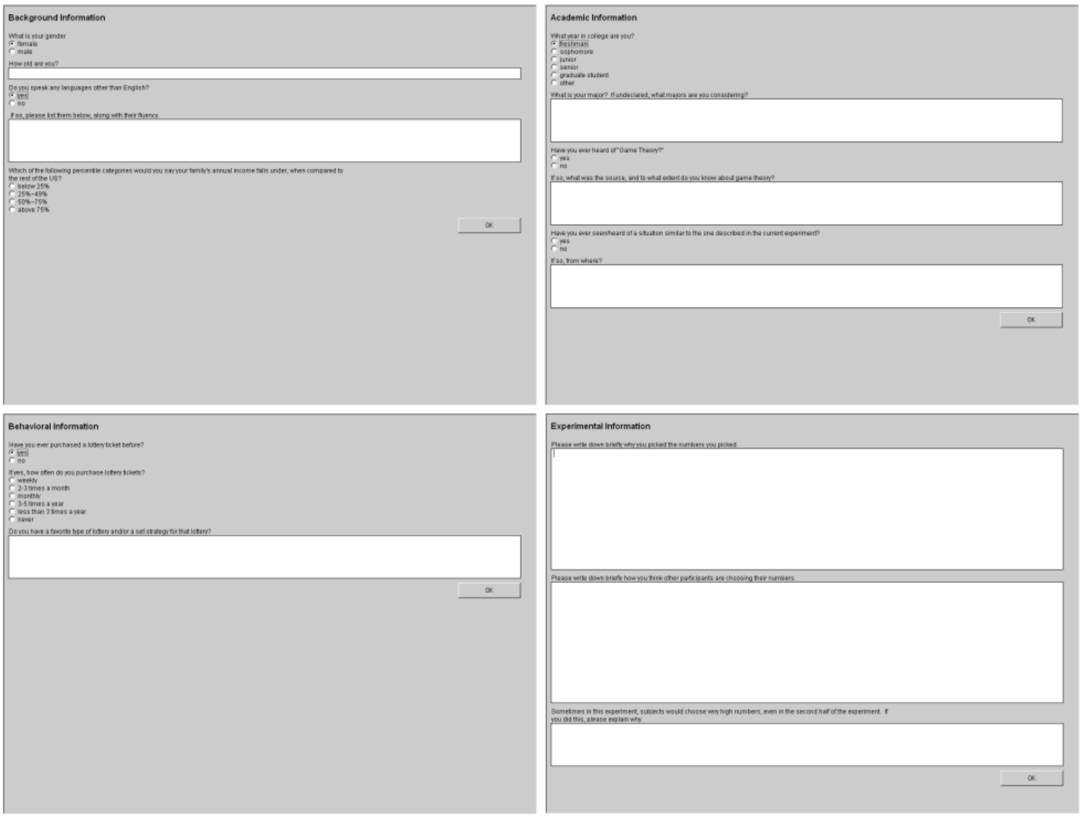
You were selected to participate in this round.

The lowest unique number among all transmitted numbers was 1

Your earnings in this round 7

Total earnings up to now 7

Figure A18. Screenshots of questionnaire in the laboratory experiment



Background Information

What is your gender
● female
○ male

How old are you?

Do you speak any languages other than English?
● yes
○ no

If so, please list them below, along with their fluency.

Which of the following percentile categories would you say your family's annual income falls under, when compared to the rest of the US?
○ below 25%
○ 25%~49%
○ 50%~75%
○ above 75%

Academic Information

What year in college are you?
● freshman
○ sophomore
○ junior
○ senior
○ graduate student
○ other

What is your major? If undeclared, what majors are you considering?

Have you ever heard of "Game Theory?"
○ yes
○ no

If so, what was the source, and to what extent do you know about game theory?

Have you ever seen/heard of a situation similar to the one described in the current experiment?
○ yes
○ no

If so, from where?

Behavioral Information

Have you ever purchased a lottery ticket before?
● yes
○ no

If yes, how often do you purchase lottery tickets?
○ weekly
○ 2-3 times a month
○ monthly
○ 3-5 times a year
○ less than 3 times a year
○ never

Do you have a favorite type of lottery and/or a set strategy for that lottery?

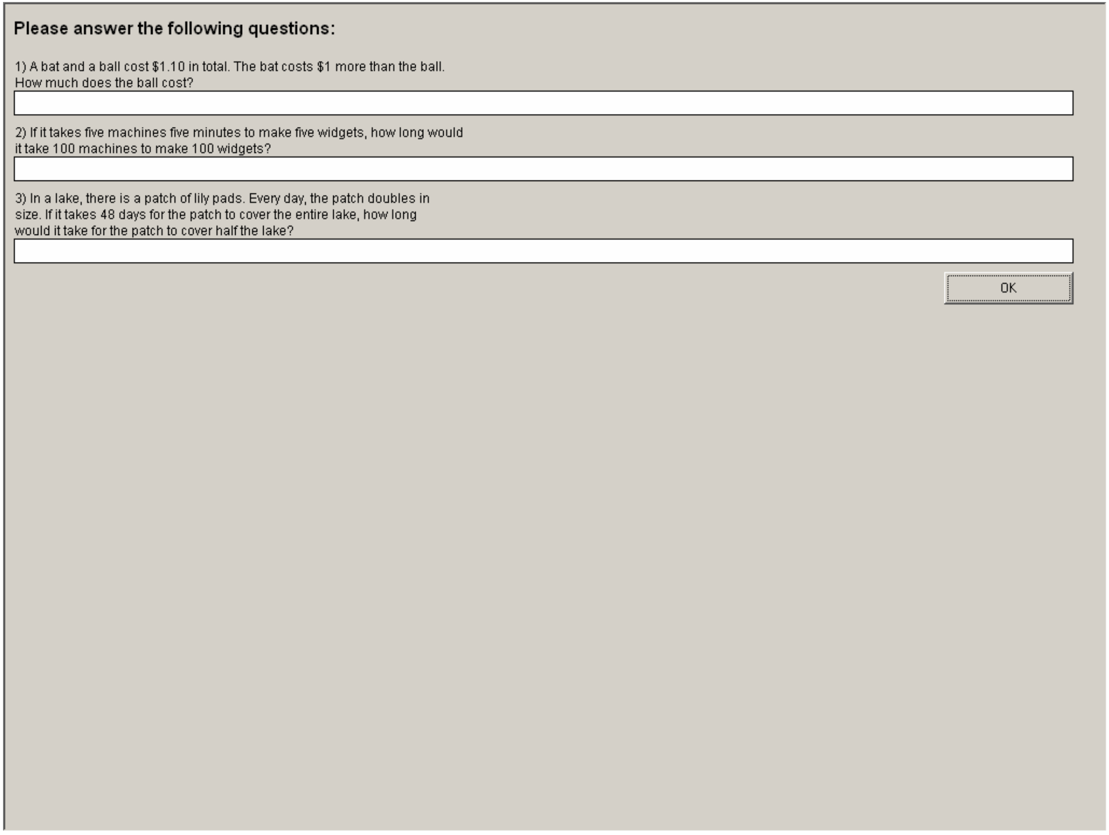
Experimental Information

Please write down briefly why you picked the numbers you picked.

Please write down briefly how you think other participants are choosing their numbers.

Sometimes in this experiment, subjects would choose very high numbers, even in the second half of the experiment. If you did this, please explain why.

Figure A19. Screenshot of CRT in the laboratory experiment



Please answer the following questions:

1) A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?

2) If it takes five machines five minutes to make five widgets, how long would it take 100 machines to make 100 widgets?

3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half the lake?

behavior in a particular round is almost unaffected depending on whether they had marginal monetary incentives or not.

Experimental Instructions. Instructions for the laboratory experiment are as follows (translated directly by author Robert Östling from the Swedish field instructions, but modified in order to fit the laboratory game):

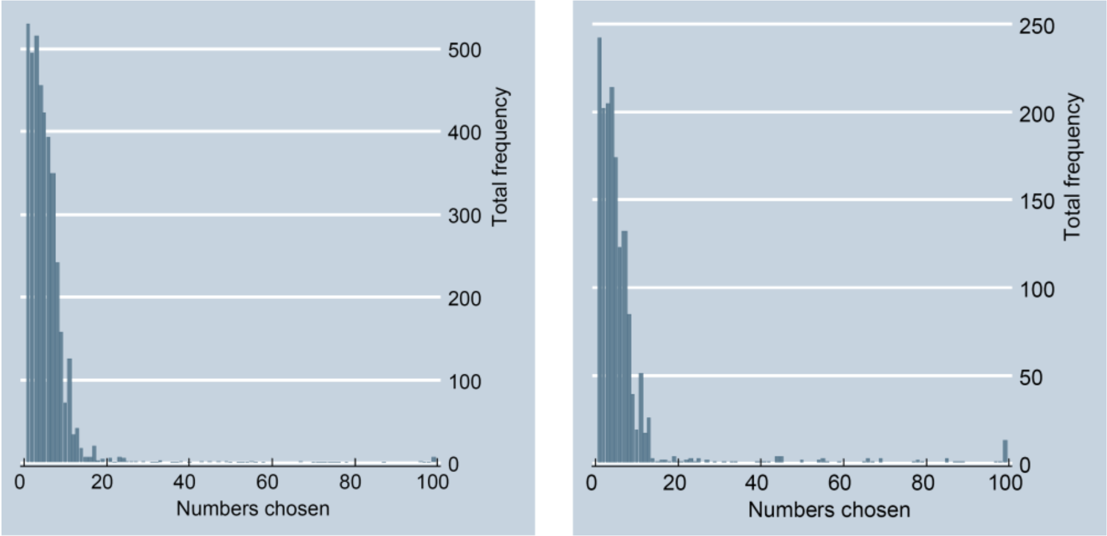
Instruction for Limbo[42]

Limbo is a game in which you choose to play a number, between 1 and 99, that you think nobody else will play in that round. The lowest number that has been played only once wins.

The total number of rounds will not be announced. At the beginning of each round, the computer will indicate whether you have been selected to participate in that round. The computer selects participating players randomly so that the average number of participating players in each round is 26.9. Please choose a number even if you are not selected to participate in that round.

[42] In order to mirror the field game as closely as possible, we referred to the LUPI game as "Limbo" in the lab.

FIGURE A20. Laboratory total frequencies, selected (left) vs non-selected (right) subjects



After all participating players have selected a number, the round is closed and all bets are checked. The lowest unique number that has been received is identified and the person that picked that number is awarded a prize of 7$.

The winning number is reported on the screen and shown to everybody after each round.

Prizes are paid out to you at the end of the experiment.

If you have any questions, raise your hand to get the experimenter's attention. Please be quiet during the experiment and do not talk to anybody except the experimenter.

Individual Lab Results. The regression results in Table 7 mask a considerably degree of heterogeneity between individual subjects. Based on the responses in the post-experimental questionnaire, we coded four variables depending on whether they mentioned each aspect as a motivation for their strategy.

Random: All subjects who claimed that they played numbers randomly were coded in this category.[43]

[43] For example, one subject motivated this strategy choice in a particular sophisticated way: "First I tried logic, one number up or down, how likely was it that someone else would pick that, etc. That wasn't doing any good, as someone else was probably doing the exact same thing. So I started mentally singing scales, and whatever number I was on in my head I typed in. This made it rather random. A couple of times I just threw curveballs from nowhere for the hell of it. I didn't pay any attention to whether or not I was selected to play that round after the first 3 or so."

Stick: All subjects who stated that they stuck to one number throughout parts of the experiment were included in this category. Many of these subjects explained their choices by arguing that if they stuck with the same number, they would increase the probability of winning.

Lucky: This category includes all subjects who claimed that they played a favorite or lucky number.

Strategic: This category includes all players who explicitly motivated their strategy by referring to what the other players would do.[44]

Several subjects were coded into more than one category.[45] The fraction of subjects within each combination of categories are reported in Table A3.

Table A3. Classification of self-reported strategies

(%)	Random	Stick	Lucky	Strategic
Random	35.1	7.0	1.8	7.0
Stick		34.2	3.5	15.8
Lucky			10.5	4.4
Strategic				41.2

How well does the classification based on the self-reported strategies explain behavior? Table A4 reports regressions where the dependent variables are four summary statistics of subjects' behavior—the number of distinct choices, the mean number, the standard deviation of number, and the total payoff. In the first column for each measure of individual play only the four categories above are included as dummy variables. There are few statistically significant relationships. Subjects coded into the "Stick" category did tend to choose fewer numbers, and subjects coded as "Lucky" tend to pick higher and more highly varied numbers (high standard deviation). Table A4 also report regressions for the same dependent variables and some demographic variables.[46] The only statistically significant relationship is that subjects familiar with game theory tend to pick lower and less dispersed numbers (though their payoffs are not higher). Note that the explanatory power is very low and that there are no significant coefficients in the regressions on the total payoff from the experiment. This suggests that it

[44] For example, one subject stated the following: "I tried to pick numbers that I thought other people wouldn't think of—whatever my first intuition was, I went against. Then I went against my second intuition, then picked my number. After awhile, I just used the same # for the entire thing."

[45] For example, the following subject was classified into all but the "Lucky" category: "At first I picked 4 for almost all rounds (stick) because it isn't considered to be a popular number like 3 and 5 (strategic). Afterwards, I realized that it wasn't helping so I picked random numbers (random)."

[46] Including demographic variables and the four categories in the same regressions does not affect any of the results reported here.

is hard to affect the payoff by using a particular strategy, which is consistent with the fully mixed equilibrium (where payoffs are the same for all strategies).

Table A4. Linear regressions explaining individual behavior

	# Distinct		Mean		Std. dev.		Payoff	
Random	0.529		-0.12		-0.93		-1.97	
	(0.97)		(-0.23)		(-0.85)		(-1.37)	
Stick	-1.14^*		-0.43		-1.62		-0.65	
	(-2.19)		(-0.86)		(-1.55)		(-0.48)	
Lucky	0.79		2.00**		3.22*		0.39	
	(1.01)		(2.64)		(2.04)		(0.19)	
Strategic	0.33		-0.40		-1.04		0.24	
	(0.64)		(-0.81)		(-1.00)		(0.18)	
Age		-0.19		-0.05		-0.03		0.34
		(-0.23)		(-0.59)		(-0.20)		(1.60)
Female		-0.09		-0.37		-1.17		-0.39
		(-0.19)		(-0.79)		(-1.19)		(-0.31)
Income (1-4)		-0.33		-0.06		-0.37		0.53
		(-1.30)		(-0.25)		(-0.72)		(0.81)
Lottery player		0.05		-0.24		-0.00		-0.17
		(0.10)		(-0.50)		(-0.00)		(-0.13)
Game theory		-0.04		-1.17^*		-2.09^*		-0.89
		(-0.08)		(-2.43)		(-2.08)		(-0.68)
R^2	0.07	0.02	0.08	0.07	0.07	0.06	0.02	0.03
Obs.	113	113	113	113	113	113	113	113

Only selected choices are included in the calculation of the dependent variables. $t-$statistics within parentheses. Constant included in all regressions. *=5 percent and **=1 percent significance level.

The questionnaire in one of the sessions also contained the three-question Cognitive Reflection Test (CRT) developed by Frederick (2005).[47] The purpose with collecting subjects' responses to the CRT is to get some measure of cognitive ability. In line with the results reported in Frederick (2005), a majority of the UCLA subjects answered only zero or one questions correctly. Interestingly, there does not appear to any relation between player's behavior or payoff in the LUPI game and the number of correctly answered questions, but the sample size is small ($n = 38$). The number of correctly answered CRT questions is not significant when the four measures in Table A4 are regressed on the CRT score.

[47] The CRT consists of three questions, all of which would have an instinctive answer, and a counterintuitive, but correct, answer. See Frederick (2005) or the screenshot in Figure A19 for the questions that we used.

FIGURE A21. Histogram of the number of distinct numbers chosen by subjects (selected subjects' choices from all sessions) and the corresponding simulated number of distinct numbers if subjects were playing the Poisson-Nash equilibrium

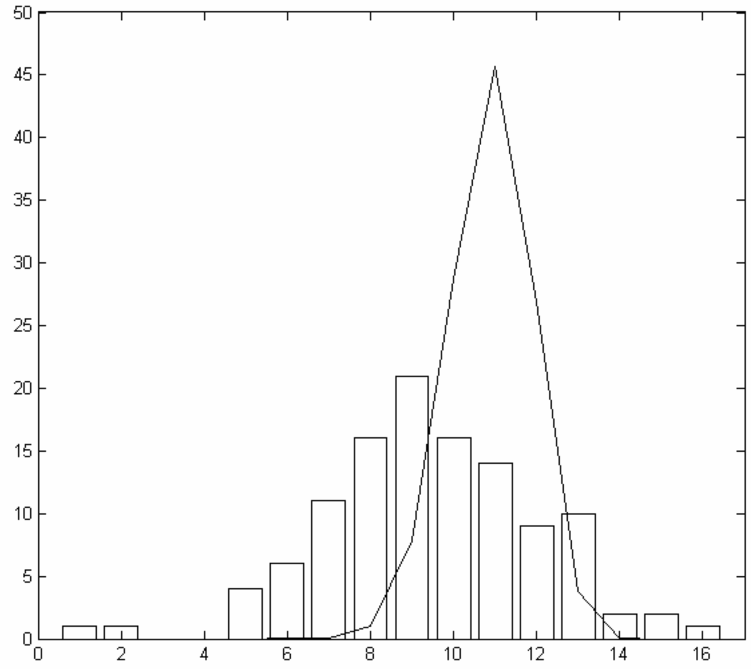


Figure A21 shows a histogram of the number of distinct numbers that subjects played during the experiments. Based only on choices when players were selected to participate, subjects played on average 9.46 different numbers, compared to 10.9 expected in Poisson-Nash equilibrium. Figure A21 also shows a simulated distribution of how many distinct numbers players would pick if they played acoording to the equilibrium distribution.

References

Bazerman, M. H., Curhan, J. R., Moore, D. A. and Valley, K. L. (2000), “Negotiation”, *Annual Review of Psychology* 51, 279–314.

Camerer, C. (2008), “The irrelevance and success of lab-fiel generalizability in economics experiments: A reply to Levitt and List”, mimeo, California Institute of Technology.

Camerer, C. F. and Ho, T. H. (1999), “Experience-weighted Attraction Learning in Normal Form Games”, *Econometrica* 67(4), 827–874.

Camerer, C. F., Ho, T.-H. and Chong, J.-K. (2004), “A Cognitive Hierarchy Model of Games”, *Quarterly Journal of Economics* 119(3), 861–898.

Camerer, C. F., Palfrey, T. R. and Rogers, B.W. (2007), “Heterogeneous Quantal Response Equilibrium and Cognitive Hierarchies”, mimeo, California Institute of Technology.

Chen, H.-C., Friedman, J.W. and Thisse, J.-F. (1997), “Boundedly Rational Nash Equilibrium: A Probabilistic Choice Approach”, *Games and Economic Behavior* 18(1), 32–54.

Chiappori, P. A., Levitt, S. D. and Groseclose, T. (2002), “Testing Mixed Strategy Equilibrium When Players are Heterogeneous: The Case of Penalty Kicks”, *American Economic Review* 92(4), 1138–1151.

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.

Costa-Gomes, M., Crawford, V. and Broseta, B. (2001), “Cognition and Behavior in Normal-Form Games: An Experimental Study”, *Econometrica* 69(5), 1193–1235.

Crawford, V. (2003), “Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions”, *American Economic Review* 93(1), 133–149.

Crawford, V. P. and Iriberri, N. (2007*a*), “Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games”, *American Economic Review* 97(5), 1731–1750.

Crawford, V. P. and Iriberri, N. (2007*b*), "Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?", *Econometrica* 75(6), 1721–1770.

Eichberger, J. and Vinogradov, D. (2007), "Least Unmatched Price Auctions", mimeo.

Fischbacher, U. (2007), "z-Tree: Zürich Toolbox for Readymade Economic Experiments", *Experimental Economics* 10(2), 171–178.

Frederick, S. (2005), "Cognitive Reflection and Decision Making", *Journal of Economic Perspectives* 19(4), 25–42.

Goeree, J. and Holt, C. (2005), "An Explanation of Anomalous Behavior in Models of Political Participation", *American Political Science Review* 99(2), 201–213.

Goeree, J., Holt, C. and Palfrey, T. (2002), "Quantal Response Equilibrium and Overbidding in Private-Value Auctions", *Journal of Economic Theory* 104(1), 247–272.

Goeree, J. K. and Holt, C. A. (2001), "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions", *American Economic Review* 91(5), 1402–1422.

Haile, P. A., Hortaçsu, A. and Kosenok, G. (2008), "On the Empirical Content of Quantal Response Equilibrium", *American Economic Review* 98(1), 180–200.

Ho, T. H., Camerer, C. F. and Chong, J.-K. (2007), "Self-tuning experience weighted attraction learning in games", *Journal of Economic Theory* 133(1), 177–198.

Houba, H., van der Laan, D. and Veldhuizen, D. (2008), "The Unique-lowest Sealed-bid Auction", mimeo.

Hsu, S.-H., Huang, C.-Y. and Tang, C.-T. (2007), "Minimax Play at Wimbledon: Comment", *American Economic Review* 97(1), 517–523.

Levine, D. and Palfrey, T. R. (2007), "The Paradox of Voter Participation? A Laboratory Study", *American Political Science Review* 101(1), 143–158.

Levitt, S. D. and List, J. A. (2007), "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World", *Journal of Economic Perspectives* 21(2), 153–174.

McKelvey, R. D. and Palfrey, T. R. (1995), "Quantal Response Equilibria for Normal Form Games", *Games and Economic Behavior* 10, 6–38.

Myerson, R. B. (1998), "Population Uncertainty and Poisson Games", *International Journal of Game Theory* 27, 375–392.

Myerson, R. B. (2000), "Large Poisson Games", *Journal of Economic Theory* 94, 7–45.

Nagel, R. (1995), "Unraveling in Guessing Games: An Experimental Study", *American Economic Review* 85(5), 1313–1326.

Palacios-Huerta, I. (2003), "Professionals Play Minimax", *Review of Economic Studies* 70(2), 395–415.

Rapoport, A., Otsubo, H., Kim, B. and Stein, W. E. (2007), "Unique Bid Auctions: Equilibrium Solutions and Experimental Evidence", mimeo.

Raviv, Y. and Virag, G. (2007), "Gambling by Auctions", mimeo.

Rosenthal, R. W. (1973), "A Class of Games Possessing Pure-strategy Nash Equilibria", *International Journal of Game Theory* 2(1), 65–67.

Roth, A. E. (1995), "Introduction to Experimental Economics", in A. E. Roth and J. Kagel, eds, *Handbook of Experimental Economics*, Princeton University Press, Princeton, chapter 1, pp. 3–109.

Roth, A. and Erev, I. (1995), "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term", *Games and Economic Behavior* 8(1), 164–212.

Sarin, R. and Vahid, F. (2004), "Strategy Similarity and Coordination", *Economic Journal* 114, 506–527.

Schwarz, G. (1978), "Estimating the Dimension of a Model", *Annals of Statistics* 6(2), 461–464.

Stahl, D. O. and Wilson, P. (1995), "On Players' Models of Other Players: Theory and Experimental Evidence", *Games and Economic Behavior* 10(1), 33–51.

Sundali, J. and Croson, R. (2006), "Biases in casino betting: The hot hand and the gambler's fallacy", *Judgement and Decision Making* 1(1), 1–12.

van Damme, E. (1998), “On the State of the Art in Game Theory: An Interview with Robert Aumann”, *Games and Economic Behavior* 24, 181–210.

Walker, M. and Wooders, J. (2001), “Minimax Play at Wimbledon”, *American Economic Review* 91(5), 1521–1538.

Weibull, J. W. (1995), *Evolutionary Game Theory*, MIT Press.

PAPER 3

Economic Influences on Moral Values

Abstract. This paper extends standard consumer theory to account for endogenous moral motivation. Building on cognitive dissonance theory, I show how moral values are affected by changes in prices and income. The key insight is that changes in prices and income that lead to higher consumption of an immoral good also affect the moral values held by the consumer so that the good is considered less immoral. An empirical analysis based on the World Values Survey confirms the model's predictions with respect to income.

1. Introduction

> It is easier for a camel to go through the eye of a needle than for a rich man to enter the kingdom of God. (Matthew 19:24)

People hold moral values that influence their behavior. Economists have long recognized this and the observation has been used to explain a wide range of economic phenomena. Many have also realized that the reverse is true – economic factors influence moral values (e.g., Lindbeck 1995, 1997, Bowles 1998 and the papers cited below). In contrast to most previous economic research, this paper focus on the individual determinants of internalized moral values. I show how standard consumer theory can be extended with a simple and yet plausible psychological mechanism to study changes in moral motivation. More specifically, the model shows how consumers that view a certain good or activity as immoral may self-servingly change their moral values as income and prices change.

The idea that moral motivation is affected by changes in prices and income has several important implications. The prevalence of moral motivation in consumer markets is demonstrated by the demand for environmental friendly products, ethical investments, organic foods and fair trade labelled goods. Policy makers might be interested

I would like to thank Colin Camerer, Tore Ellingsen, Ernst Fehr, Erik Lindqvist, Erik Mohlin, Robertas Zubrickas and seminar participants at the Stockholm School of Economics, SUDSWEc 2007, First Nordic Workshop in Behavioral Economics 2006, CEU Behavioral Economics Workshop 2004 and NNE Summer School in Behavioral Economics 2004 for valuable comments. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

in increasing demand for such products and therefore need to be aware of the impact of economic policies on the moral motivation of consumers. For example, the expansion of low-cost airlines might increase consumers' moral tolerance of carbon emissions, which may counteract measures taken to combat climate change. Relatedly, Shleifer (2004) and others have suggested that higher income implies higher willingness to pay for ethical behavior. This does not necessarily hold if moral values are endogenous. Higher incomes may also increase consumption of "immoral goods", for example air travel, which is likely to affect moral attitudes. The framework for endogenous moral values laid out here can help to explain moral attitudes regarding "immoral" consumer goods, but it can also explain attitudes toward tax evasion and benefit fraud.

In the model, prices and income affect the incentives for immoral behavior, entailing a conflict between narrow self-interest and moral values. This conflict gives rise to cognitive dissonance, which the consumer can reduce by exerting effort in order to modify her moral values. The main prediction of the model is that higher consumption of a good implies that the consumer will view that good more favorably from a moral point of view. For normal goods, higher incomes therefore lead to higher moral acceptability, whereas the opposite is true for inferior goods. An empirical analysis using data from the World Values Survey supports the prediction regarding the effects of income on moral motivation.

The model aims to capture an idea that Elliot Aronson, a leading social psychologist, presents as hypothetical advice from a modern Machiavelli: "If you want people to soften their moral attitudes toward some misdeed, tempt them so that they perform that deed" (Aronson 2003, p. 162). This type of self-justification of moral attitudes is discussed by Aronson (2003, chapter 5), but the underlying psychological idea is based on the cognitive dissonance theory developed by Festinger (1957).

Cognitive dissonance was first introduced in economics by Akerlof and Dickens (1982) and has since been used in several economic applications.[1] This paper's combination of cognitive dissonance theory and standard consumer choice theory is most closely related to Rabin (1994) and Konow (2000). My model extends their framework in two respects. First, I allow the consumer to choose between two, rather than one, consumption good. Second, the consumer faces a budget constraint and there are prices attached to both goods. This links their approach to standard consumer theory and allows comparative statics in terms of easily observable variables such as prices

[1] For example, economics of crime (Dickens 1986), moral behavior and social change (Rabin 1994), biased fairness norms in the dictator game (Konow 2000), mobility-reducing norms of low-productivity farmers (Haagsma and Koning 2002), formation of underclass attitudes (Oxoby 2003, 2004), redistributive politics (Bénabou and Tirole 2006*a*) and changes in political attitudes after elections (Beasley and Joslyn 2001 and Mullainathan and Washington 2006).

and income. In particular, having two consumption goods in the model is required to distinguish between normal and inferior goods, which is critical for the empirical identification of the model's main predictions.

Apart from Rabin (1994) and Konow (2000), there have also been some other attempts to model moral motivation endogenously. For example, Brekke et al. (2003) study moral motivation in a public goods provision model where a commonly shared moral norm can be affected by policy. Specifically, they assume that players are utilitarian and apply Kantian reasoning, i.e., the moral norm is determined by the action that maximizes the sum of all players' utility given that everybody takes the same action. This paper instead takes moral values as given and focuses on a psychological mechanism that may change moral motivation irrespective of what moral philosophical principles that underlie consumers' moral values.[2] Van de Ven (2003) uses cognitive dissonance theory to study how preferences for environmental-friendly goods are affected by subsidies, but his approach is only vaguely related to this paper.[3] Frey (1997) makes a distinction between intrinsic and extrinsic motivation and uses it to explain the crowding out of blood supply suggested by Titmuss (1970). Frey (1997) discusses how extrinsic motivation may affect intrinsic motivation, but he provides no formal model of this interaction.[4] In my model, there is a tension between extrinsic (prices) and intrinsic (moral) motivation, but the effects of extrinsic motivation cannot be reversed through crowding out of moral motivation.

2. Model

Consider the familiar utility maximization problem of a consumer with a fixed endowment w that she spends on two consumption goods. In addition, the consumer decides how much energy to spend on self-deception. We call the first of the consumption goods the *moral good* and denote the quantity consumed by x_M, while the second good is

[2] In other words, the approach taken in this paper is completely agnostic as to what moral values consumers hold. The paper studies the positive question how *given* moral values change, which distinguishes it from the normative question what these moral values *should* be (which has been studied thouroughly by moral philosophers, see Rachels 1999 for an excellent introduction).

[3] Van de Ven (2003) focuses on a discrete choice between two alternatives and interprets cognitive dissonance as indifference aversion, i.e., dissonance is assumed to be lower the more the two alternatives differ. This dissonance in turn motivates the consumer to rationalize her decision so that the alternative chosen appears better than she initially considered it to be.

[4] By now there are some economic models that potentially can explain this and other related phenomena. Bénabou and Tirole (2003) show how motivation might change if a person has imperfect self-knowledge and therefore may be sensitive to signals from a more informed player. Bénabou and Tirole (2006*b*) and Ellingsen and Johannesson (2008) instead assume that players are concerned about signalling their character traits. This paper differs from these theoretical models since I don't assume any form of interaction with other players – the focus is on a psychological mechanism on the individual level.

termed the *immoral good* and the consumed quantity x_I. The two consumption goods give the consumer material utility $u(x_M, x_I)$, which is a standard utility function that is twice continuously differentiable, strictly concave and increasing in both goods.

The consumer not only cares about material utility, but also has an original moral value $m_0 > 0$ that measures how immoral the consumer considers the immoral good to be. The original moral value is exogenous, but the consumer can choose to change moral value when making her consumption choice. The chosen moral value is denoted m.[5]

Consuming the immoral good creates cognitive dissonance, which is measured by a non-negative function $d(m, x_I)$. The dissonance function is twice continuously differentiable, strictly convex and increasing in m and x_I. Furthermore, there is no dissonance if the immoral good isn't consumed or if the consumer considers both goods to be equally moral, i.e., we assume $d(0, x_I) = 0$ and $d(m, 0) = 0$. Deviating from the original moral value, however, comes at a utility cost. This cost of self-deception is increasing and strictly convex in $|m_0 - m|$. For simplicity, we assume that this cost is $\delta(m_0 - m)^2$ with $\delta > 0$. Overall utility is additively separable in its three components.

The consumer maximizes utility by simultaneously choosing consumption and the moral value.[6] The utility maximization problem is therefore

$$\max_{x_M, x_I, m} \left[u(x_M, x_I) - d(m, x_I) - \delta(m_0 - m)^2 \right],$$

subject to

$$p_M x_M + p_I x_I \leq w,$$
$$x_M, x_I, m \in \mathbb{R}_+,$$

where p_M and p_I are the prices of moral and immoral goods, respectively, and w denotes consumer income.[7] For the sake of simplicity, we assume throughout that the solution

[5] Throughout the paper I discuss a decrease in m in terms of the consumer becoming more immoral. From a normative point of view it is of course impossible to judge whether this is "good" or "bad" without knowing anything about the original moral values of the consumer. In the extreme case, the consumer's moral good may be our immoral good, so that we are indeed very happy about the consumer becoming more immoral.

[6] It may appear more intuitive that consumers first choose consumption quantities and then rationalize their consumption decision by adapting moral values. However, Lieberman et al. (2001) provides suggestive evidence that this intution is likely to be false and that attitude change is a highly automated process that is hard to temporally separate from the behavioral decision.

[7] The maximization problem differs from from Rabin (1994) mainly in two respects. First, there is not only an immoral good, but also another consumption good. Second, the consumer has limited resources and there are prices attached to each of the goods, so prices and income enter the decision problem in a natural way. In addition, the dissonance function is slightly more general whereas the self-deception function has an explicit functional form, but both these differences are unimportant for the results.

to the utility maximization problem is interior, i.e., $x_M, x_I, m > 0$.[8] Furthermore, the marginal dissonance with respect to consumption of the immoral good is increasing in the moral value.[9] Note that if $m_0 = 0$ both goods are equally moral and the consumer's problem is simply to maximize material utility.[10]

Under these assumptions, the solution to the consumer's problem can be characterized by the first-order conditions of the Lagrangian. To guarantee that these conditions give a unique and optimal solution, we also require that a sufficient condition for a unique optimum is satisfied.[11] The solution to the utility maxmization problem is denoted x_M^*, x_I^* and m^*. The propositions below are proven in the Appendix using standard comparative static analysis of the first-order conditions of the utility maximization problem.

Proposition 1 establishes the effects of income changes on moral values.

PROPOSITION 1. *The chosen moral value, m^*, is decreasing in income if the immoral good is normal and increasing in income if the immoral good is inferior.*

The result in Proposition 1 can be derived directly from the first-order conditions of the problem. Rearranging the third first-order condition (in the Appendix) and differentiating with respect to income gives

$$\frac{dm^*}{dw} = -\left[\frac{\partial^2 d(m, x_I)/\partial m \partial x_I}{2\delta + \partial^2 d(m, x_I)/\partial m^2}\right]\frac{dx_I^*}{dw}.$$

Since we have assumed that all terms within brackets are positive, this implies that if the immoral good is a normal good, then the moral value is decreasing in income, while if the immoral good is an inferior good, then m^* increases with income. The intuition for this result is straightforward. When the immoral good is normal, higher income leads to higher consumption. Higher consumption of the immoral good creates cognitive dissonance, which can be reduced (at the margin) by changing moral values so that the immoral good is believed to be less immoral than before. In other words, if we consume more of goods that we believe it is immoral to consume, then we adjust

[8] Alternatively, we can assume that $\lim_{x_M \to 0} u(x_M, \cdot) = -\infty$ and $\lim_{x_I \to 0} u(\cdot, x_I) = -\infty$ so that the consumer always consumes positive quantities of both goods. In order for m to be guaranteed to be positive, we could, for example, assume that m_0 or δ is sufficiently high (the exact condition is given by the expression for the optimal moral value m at the end of this section).

[9] Formally this means that the cross-derivatives of the dissonance function are positive, $\partial^2 d(m, x_I)/\partial m \partial x_I = \partial^2 d(m, x_I)/\partial x_I \partial m > 0$.

[10] To see this, note that if $m_0 = 0$, then it is optimal to choose $m^* = 0$, which in turn implies that utility becomes $U = u(x_M, x_I)$.

[11] Technically this means that the naturally ordered principal minors of the bordered Hessian alternate in sign, which is a standard sufficient condition for a solution to the utility maximization problem in consumer theory (e.g. Varian 1992).

our values in order to reduce the dissonance that the increase in consumption gives rise to.

There are of course other ways in which higher incomes affect moral values. For example, as argued by Shleifer (2004), higher income also provides greater opportunity to behave morally when it is costly to do so. Moreover, when a higher income is observed by others, there could be a change in social pressures to behave morally. Nevertheless, the model points at another rather general mechanism. Higher income leads to higher consumption, which has consequences for our moral attitudes. Most people can't stand considering themselves to be immoral persons, and so they need to adjust their moral values to be compatible with their consumption pattern. A historical example of this effect is when the Catholic Church lifted their ban on eating meat on Fridays in the mid-1960s, supposedly because incomes had grown and meat had become relatively cheaper.[12]

The second proposition states how moral values are affected by changes in prices.

Proposition 2. *(i) If the immoral good is normal, then the chosen moral value, m^*, is increasing in the price of the immoral good, p_I. (ii) If the immoral good is inferior, then the chosen moral value, m^*, is decreasing in the price of the moral good, p_M.*

The intuition for this result is straightforward. If the immoral good is normal, then the income effect is negative. An increase in the price of the immoral good therefore leads to lower consumption and an upward adjustment of the moral value, i.e., the immoral good is considered more immoral. This result suggests that the failure of policy-makers to correct for externalities might be associated with an additional "moral cost". For example, due to international agreements there is currently no tax on air fuel, which means that the negative environmental externality of air travel is not reflected in the market price. Because air travel is cheaper than without the tax, more people are travelling and their attitudes toward pollution is less negative than they would have been if the tax was in place.

To clarify the effects of prices changes on moral values further, it is useful to rephrase the results in terms of gross substitutes and complements.

Proposition 3. *(i) If the two goods are gross substitutes, i.e., $dx_I^*/dp_M > 0$, then the chosen moral value, m^*, is decreasing in the price of the moral good, p_M. (ii) If*

[12] A related example is that older generations typically complain more about throwing away food than current generations. This is in part an effect of economic growth which has led us to afford more food (and other goods) than previous generations and hence to value food less. But most probably there is also the effect indicated by this paper. Grandmothers and grandfathers not only complained because food have a higher economic value to them, but often also argued that it is morally wrong to throw away food.

the two goods are gross complements, i.e. $dx_I^*/dp_M < 0$*, then the chosen moral value,* m^**, is increasing in the price of the moral good,* p_M*.*

If a change in the price of the moral good leads to higher consumption of the immoral good, the immoral good is considered less immoral. Hence, if the two goods are gross substitutes we have a crowding-in effect of a price change of the moral good, while there is a crowding-out effect if the two goods are gross complements.

As an example of the crowding-out effect, suppose that the two goods are gross complements. A price decrease of the moral good therefore leads both to higher consumption and higher moral acceptance of the immoral good. Hence, extrinsic motivation—a lower price on the moral good—can crowd out intrinsic moral motivation. Although a lower price on the moral good might lower moral motivation, i.e., lead to a lower m^*, it cannot imply lower demand of the moral good (unless it is a Giffen good). This is a kind of motivational crowding-out effect, although the effect goes via a change in consumption of the immoral good, and not directly from extrinsic motivation (lower price) to intrinsic motivation (moral values). The crowding-out effect illustrated by the model is relevant for policy. Consider the case of subsidies to consumption of organic food. For many people the choice of organic food is to some extent motivated by moral concerns, and policy makers might be interested in stimulating such moral values (for example if they believe it spills over to other areas, e.g., attitudes toward recycling and littering). The model illustrates that whether such subsidies will stimulate the preferred moral values depends on how demand for other goods is affected by the lowering of prices on organic food. It could be argued that lower prices on organic food will result in higher consumption of other more environmentally harmful goods, which affects moral attitudes in favor of these other goods.

The above discussion hides one complication. The terms normal and inferior goods are used as if consumption of the two goods are chosen independently of the moral value. However, since the moral value and consumption are determined simultaneously, and cognitive dissonance is not only affected by the moral value, but also by the consumption of the immoral good, the effect of changes in income and prices on consumption is slightly more complicated than in the standard utility maximization problem with two goods. Therefore modified Slutsky equations are derived in the Appendix to show that the standard interpretation of income and substitution effects carries over to this setting.

The model presented above is static and silent about what happens to the consumer's moral value after the consumption decision. The model predicts that the chosen moral value, m^*, will always be different from the original value, m_0. This can easily be seen by rearranging the third first-order condition (found in the Appendix):

$$m^* = m_0 - \frac{\partial d(m, x_I)}{\partial m} \frac{1}{2\delta}.$$

Since the dissonance function is increasing in m and $\delta > 0$, this implies that $m^* < m_0$. The fact that the consumer always deceives herself in the consumption decision is an example of self-servingly biased moral values.[13] In order to develop a dynamic version of the model there are two issues that need to be resolved. First, does the self-deception investment in reduction of moral values have a transitory or permanent effect? On the one extreme, m_0 may be constant over time and consumers merely deceive themselves at the time of consumption. On the other hand, self-deception could be a one-time investment – the chosen moral value in one period is m_0 in the next consumption decision. In the latter extreme case, moral values would tend to erode over time unless there are other factors that affect m_0.[14] Second, the predictions of a dynamic model would also depend on to what extent people are forward-looking and manage to predict their own changing preferences. There is little empirical evidence to guide these modelling assumptions and I have therefore abstained from developing a dynamic version of the model. The empirical analysis in the next section is based on a cross-section of individuals and does therefore not depend on the details of the dynamics.

3. Empirical Analysis

In order to test the predictions of Proposition 1, I use data from the latest wave (1999-2004) of the World Values Survey (WVS).[15] The 1999-2004 wave of the WVS contains responses to survey questions from 101,000 individuals in 70 countries.[16] Respondents are, among other things, asked about their moral attitudes toward certain behaviors.

[13] See Babcock and Loewenstein (1997) for an introduction to self-serving biases in economic problems.

[14] Note that this does not imply that we should expect all consumers not hold any moral values. Even in a dynamic model that has the implication that moral values are zero in the long run, real consumers do not live forever and certain consumption decision may be taken irregularly (implying that consumers do not live long enough to experience sufficiently many consumption decisions to erode moral values completely).

[15] The data has been obtained from www.worldvaluessurvey.org and the latest wave of the survey has been extracted from the following integrated data file: European Values Study Group and World Values Survey Association. EUROPEAN AND WORLD VALUES SURVEYS FOUR-WAVE INTEGRATED DATA FILE, 1981-2004, v. 20060423, 2006. Aggregate File Producers: Análisis Sociológicos Económicos y Políticos (ASEP) and JD Systems (JDS), Madrid, Spain; Tilburg University, Tilburg, The Netherlands. Data Files Suppliers: Analisis Sociologicos Economicos y Politicos (ASEP) and JD Systems (JDS), Madrid, Spain; Tillburg University, Tillburg, The Netherlands; Zentralarchiv fur Empirische Sozialforschung (ZA), Cologne, Germany. Aggregate File Distributors: Análisis Sociológicos Económicos y Políticos (ASEP) and JD Systems (JDS), Madrid, Spain; Tillburg University, Tilburg, The Netherlands; Zentralarchiv fur Empirische Sozialforschung (ZA) Cologne, Germany.

[16] Northern Ireland has been treated as a separate country in the survey which I consequently do also in my analysis. A more detailed description of the data can be found in the Appendix.

These questions are phrased as follows: "Please tell me for each of the following statements whether you think it can always be justified, never be justified, or something in between" and the respondents were asked to answer on a scale from 1 to 10 where 1 means "always justifiable" and 10 means "never justifiable" for different types of activities.[17]

The effect of income on moral values can be identified by estimating the following regression:

$$m_i = \alpha + \beta y_i + \boldsymbol{X}_i \boldsymbol{\gamma} + \varepsilon_i,$$

where m_i is the stated moral value, y_i the income of the respondent and $\boldsymbol{X}_i$ a vector with country dummies and individual characteristics. The individual characteristics are sex and age in the "short" specification, whereas the "long" specification in addition controls for educational level, employment status, profession, marital status, number of children and size of home town. All these characteristics are included as dummies using the response alternatives available in the WVS (see Appendix for details). The income data in the WVS refers to household income and is measured in ten country-specific income brackets based on self-reports. Income is consequently measured with error, but there is little reason to expect that the measurement error is correlated with true income. The estimated income coefficients are therefore likely to be biased toward zero. Since several of the control variables are strongly correlated with income, the inclusion of these variables most likely exacerbates the attenuation bias. We should therefore expect smaller income coefficients in the long than in the short specification.

Proposition 1 predicts that the income coefficient β is negative for normal goods and positive for inferior goods. But moral values may be correlated with other individual characteristics that are related to income. Although many such characteristics are included as controls in the long regressions, the estimated income coefficients might be biased due to omitted variables. However, these omitted variables are likely to be correlated with income and moral values in the same way for both normal and inferior goods. This implies that the estimated income coefficients not necessarily are expected to have opposite signs, but that they should be higher for inferior than for normal goods.

Some questions in the WVS refer to goods or activities that are difficult to relate to income and have therefore been left out.[18] The remaining goods and activities the

[17] Readers that are familiar with WVS may notice that I have reversed the scale in order to make the responses consistent with the interpretation of m.

[18] These questions are about homosexuality, abortion, divorce, casual sex, euthanasia, suicide, lying, adultery, sex under legal age of consent, littering, political assassination, experiments on human embryos and genetic manipulation of food. Furthermore, two other questions have also been left out although they might be related to income: accepting bribes and buying stolen goods. Although poorer people have stronger economic incentives to engage in these activities (given diminishing marginal

questions refer to are classified into inferior and normal goods based on a priori concerns and available empirical evidence.[19]

One potential problem with the empirical analysis is that income might depend on moral values for two of the questions used. People that are more tolerant toward benefit fraud and tax evasion will probably cheat more and might therefore report a higher income. Although this can rationalize a negative relationship between income and tax morale, it cannot explain a positive relationship between benefit fraud and income.[20]

Table 1 reports the income coefficients from the two different specifications with moral values as the dependent variables. The top section of Table 1 refers to activities that are likely to be inferior goods, and the bottom section refers to normal goods. A subset of the WVS is the European Values Survey (EVS) which contains some extra moral values questions for 32 European countries. Table 1 reports income coefficients estimated using the whole sample and the EVS countries separately.

Activities like benefit fraud, stealing cars and avoiding public transport fares are likely to be inferior since the incentive to engage in these activities is higher the lower is the income (given diminishing marginal utility of money). In line with the prediction of Proposition 1, all significant income coefficients for these questions are positive. Smoking is also an inferior good, at least in industrialized countries (Chaloupka and Warner 2000), and we might therefore expect richer people to be less tolerant of smoking in public buildings. As can be seen from Table 1, this is not supported by the data, but the negative coefficients are not statistically significant.

The bottom six questions in Table 1 refer to goods and activities that are likely to be positively related to income. Although I have found no data on sex buyers, it seems most plausible that it is a normal good.[21] Alcohol and marijuana consumption

utility of money), they also have less money to spend on stolen goods and they might be less likely to be offered bribes. Finally, a question regarding alcohol consumption that was only asked in Muslim countries has been left out.

[19] Although some of these goods and activities are not typical consumption goods, the theoretical model can be seen as a reduced form of richer models that model each situation in more detail. For example, for several of the activities there is a risk of legal sanctions, but the price p_I can be interpreted as a reduced-form form representation of the expected material cost of punishment.

[20] In addition, the income from benefit fraud and tax evasion is likely to consitute a neglible fraction of reported incomes for most respondents.

[21] In the case of prostitution, it is typically poor women that work as prostitutes. Some people working as prostitutes are likely to be included in the sample, and the results might therefore by affected if prostitutes are more tolerant toward prostitution. However, buyers outnumber sellers by far, so this is likely to be a limited problem. Moreover, excluding women from the sample leads to somewhat stronger effects in three regressions and a somewhat weaker effect in one regression, but the coefficients remain negative and strongly significant in all four cases.

Table 1. Moral values regressions

	European countries		All countries	
	Short	Long	Short	Long
Benefit fraud	**4.65*****	**1.78*****	**4.25*****	**1.52*****
	10.34	3.04	12.59	3.46
	32/33161	32/24334	66/77673	56/54102
Avoiding fare	0.13	−0.62	**2.33*****	**1.27*****
	0.20	0.72	6.22	2.56
	17/19802	17/14970	53/66522	42/45388
Joyriding	**0.98*****	0.35		
	3.65	1.00		
	32/33729	32/24764		
Smoking in public	−0.55	−0.41		
	0.93	0.53		
	32/33201	32/24365		
Prostitution	−**7.26*****	−**3.62*****	−**6.05*****	−**3.57*****
	10.25	3.90	16.96	7.40
	19/19789	19/14420	53/65096	41/42675
Taking soft drugs	−**1.05*****	−0.23		
	2.71	0.48		
	32/33522	32/24601		
Drink and drive	−**0.86*****	−**1.06*****		
	3.11	2.92		
	32/33759	32/24783		
Overspeeding	−**5.06*****	−**3.23*****		
	12.03	5.80		
	32/33580	32/24655		
Pay cash for services	−**3.86*****	−**3.24*****		
	6.70	4.28		
	32/32357	32/23691		
Cheating on taxes	−0.52	−0.69	−0.04	−**1.03****
	1.00	1.20	0.14	2.33
	32/33263	32/24393	66/78561	55/53446

The table reports income coefficients multiplied by 100, absolute t values and the number of countries/observations used in estimating the coefficient. Controls in the short specification are country dummies, age and sex. The long specification in addition controls for marital status, educational level, employment status, occupation, size of home town and number of children. *=10%, **=5% and ***=1% significance level.

is positively related to income, at least in the US (Saffer and Chaloupka 1999).[22] Since

[22] As with prostitution it is possible to make a supply-side argument. If drug dealers are on average poorer we would expect these to be more morally tolerant of soft drugs. But this effect is probably marginal since buyers are likely to outnumber sellers.

driving a car is also a normal good, the propensity to drive under the influence of alcohol is probably increasing in income. Similarly, Shinar et al. (2001) show that there is a positive relationship between overspeeding and income in the US. The evidence on tax evasion is somewhat mixed, but several studies point at a positive relationship (see Andreoni et al. 1998 for a discussion). As can be seen from Table 1, all income coefficients for the normal goods have the predicted negative sign, but they are not statistically significant for all questions in all specifications.

It is clear that the income coefficients reported in Table 1 generally have the predicted signs and are statistically significant in most of the regressions. For example, based on the short specification for all countries, an increase in income from the lowest to the highest income category implies that individuals on average believe that prostitution is 0.6 more morally justifiable on a 1 to 10 scale. This corresponds to an increase of the moral value of one fourth of a standard deviation. Although this is a relatively small effect, there are several reasons why we should expect it to be small. First, income is poorly measured which implies that the coefficients are likely to be biased toward zero. The coefficients are generally smaller in the long specification, suggesting that the inclusion of the extra controls exacerbates the attenuation bias. Second, consumption of most of the goods listed in Table 1 are not particularly strongly related to income. Moreover, many people are likely to never have consumed some of the goods listed there, implying that only a subset of the population is used to identify the effect. Since the relation between income and consumption is weak, the relationship between income and moral values should also be weak.

Due to omitted variables the income coefficients should not necessarily have opposite signs. To test whether coefficients are significantly different between normal and inferior goods, I run the same specification as in Table 1 but use the moral value regarding an inferior good minus the moral value with respect to a normal good as the dependent variable. This is also a way to control for the differences in the sample used in estimating the coefficient for different goods. For simplicity, Table 2 only reports the results from two pairwise such regressions, which are chosen because they are available for the largest number of countries.

Table 2 confirms our earlier conclusion. The richer people are, the more tolerant are they of cheating on taxes relative to benefit fraud. Similarly, the richer they are, the more morally justifiable do they think prostitution is relative to avoiding paying the fare on public transport.

As an additional robustness check, I run regressions separately for each country and question. Although the results differ slightly between countries, the pattern from Table 1 persists. With smoking in public buildings as the main exception, income coefficients

Table 2. Moral values regressions (differences)

	European countries		All countries	
	Short	Long	Short	Long
Benefit fraud – cheating on taxes	**5.07*****	**2.69*****	**4.22*****	**2.50*****
	9.39	3.76	11.08	5.01
	32/32799	32/24045	65/75628	55/52293
Avoiding fare – prostitution	**3.64*****	0.96	**6.85*****	**4.07*****
	3.66	0.76	14.26	6.38
	13/14819	13/11045	47/58769	36/39026

The table reports income coefficients multiplied by 100 with the difference in moral value between two different goods as dependent variables, absolute t values and the number of countries/observations used in estimating the coefficient. Controls in the short specification are country dummies, age and sex. The long specification also controls for marital status, educational level, employment status, occupation, size of home town and number of children. *=10%, **=5% and ***=1% significance level.

are more often significant and positive than significant and negative for inferior goods, and the other way round for normal goods.[23]

The above analysis shows that the relationship between income and moral values observed in the data is consistent with Proposition 1. Can these findings be explained in some other way than by the theoretical model in this paper? The obvious candidate is that some variable correlated with both income and moral values have been omitted in the regressions. However, in order for such an omitted variable to rationalize the empirical findings above, the variable must be correlated differently with moral values depending on whether the good is normal or not. It is hard to see what kind of omitted variable this would be. A potential alternative interpretation of the empirical findings could be peer group effects – if the poor mainly socialize with the poor, they might adjust their moral values to each other. But this alone cannot explain the empirical pattern, since it does not provide an account for why the poor should be more tolerant toward, for example, benefit fraud in the first place.

As the theoretical model predicts *changes* in moral values for a given individual, rather than *levels* across individuals, the ideal test should involve measuring people's moral values before and after an exogenous change in income or prices. Naturally, it is hard to find such data in the field, and the cleanest test of the model probably requires experimental methods. Cognitive dissonance has long been studied experimentally by psychologists (several of these experiments are discussed by Aronson 2003). One such

[23] Except for the smoking question, this holds true both in the short and long regression with two exceptions – the long specification for the question regarding soft drugs and the short specification regarding tax evasion show the opposite pattern. The results from these regressions are available on request from the author.

study that is relatively close to the setting discussed here is the experiment on school children by Mills (1958). In the experiment, pupils participated in a classroom contest and were told that the best student would win a prize. The contest was such that experimenters could detect who had cheated. The children were asked both before and after the contest about their attitudes toward cheating. On average, those who had cheated also changed attitude toward thinking cheating was more acceptable, which is consistent with the model.[24]

4. Concluding Remarks

We have seen that if a change in prices or income lead to higher consumption of an immoral good, the consumer becomes more morally tolerant toward that good. This suggests that there may be a "personal price" to pay for higher incomes – higher income not only leads to more consumption, but also to changes in moral values.

A priority for future research is an experimental test of the theory. On the theoretical side it would be interesting to extend the model to include the social environment. Social influences on moral values might for example be modelled as direct peer-group influence on moral values, or by incorporating social pressure in terms of social rewards and punishments.[25] The model can also be incorporated in a general equilibrium framework to study indirect social effects on values and norms through prices.

[24] This effect was statistically significant when compared to those that did not cheat, but not compared to the control group (which didn't participate in a contest).

[25] See Nyborg (2003) for an interesting discussion about the potential effects of such social influence and its impact for public policy in a related context.

Appendix: Comparative Statics

Since the utility function is increasing in both goods it follows that the budget constraint will be binding (there is no "moral cost" of consuming more of the moral good and the consumer will therefore be better off consuming more of that good). In addition, we have assumed that the solution is interior so that we can use the first-order condition of the Langrangian to solve the optimization problem. The Langrangian is

$$\mathcal{L}\left(x_M, x_I, m\right) = u(x_M, x_I) - d(m, x_I) - \delta(m_0 - m)^2 - \lambda_B\left(p_M x_M + p_I x_I - w\right).$$

The first-order conditions are

$$\begin{aligned} w - p_M x_M - p_I x_I &= 0, \\ \frac{\partial u(x_M, x_I)}{\partial x_M} - \lambda_B p_M &= 0, \\ \frac{\partial u(x_M, x_I)}{\partial x_I} - \frac{\partial d(m, x_I)}{\partial x_I} - \lambda_B p_I &= 0, \\ -\frac{\partial d(m, x_I)}{\partial m} + 2\delta(m_0 - m) &= 0. \end{aligned}$$

A sufficient condition for this system to define a unique maximum is that the naturally ordered principal minors of the bordered Hessian alternate in sign (see Varian 1992). This condition also implies the we can apply the Implicit Function Theorem so that the solutions will be locally differentiable with respect to the parameters.

Let u_{XY} and d_{XY} denote the second derivatives of the material utility and dissonance functions w.r.t. to X and Y where 1 denotes the moral good, 2 the immoral good, and 3 the moral value. Substituting the solutions x_M^*, x_I^*, m^* and λ_B as functions of p_M, p_I, w, m_0 and δ into the system of first order conditions and totally differentiating with respect to w gives

$$\begin{bmatrix} 0 & -p_M & -p_I & 0 \\ -p_M & u_{11} & u_{12} & 0 \\ -p_I & u_{21} & u_{22} - d_{22} & -d_{32} \\ 0 & 0 & -d_{23} & -2\delta - d_{33} \end{bmatrix} \begin{bmatrix} d\lambda_B/dw \\ dx_M^*/dw \\ dx_I^*/dw \\ dm^*/dw \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Note that the coefficient matrix is the bordered Hessian of the Lagrangian and that the sufficient condition implies that the determinant is negative. Let the determinant of the bordered Hessian be denoted by $|H|$. Using Cramer's rule we can solve for the derivatives with respect to income:

$$\begin{aligned}
\frac{dm^*}{dw} &= \frac{d_{32}\left(u_{21}p_M - u_{11}p_I\right)}{|H|}, \\
\frac{dx_M^*}{dw} &= \frac{\left(2\delta + d_{33}\right)\left(u_{22}p_M - d_{22}p_M - u_{12}p_I\right) + d_{23}d_{32}p_M}{|H|}, \\
\frac{dx_I^*}{dw} &= \frac{\left(2\delta + d_{33}\right)\left(u_{11}p_I - u_{21}p_M\right)}{|H|}.
\end{aligned}$$

If we substitute the expression for dx_I^*/dw into the expression for dm^*/dw we get

$$\frac{dm^*}{dw} = -\frac{d_{32}}{2\delta + d_{33}}\frac{dx_I^*}{dw}.$$

We have assumed that $d_{32} > 0$, $d_{33} > 0$ and $\delta > 0$. This means that the sign of the derivative depends on whether the immoral good is normal or inferior. If the immoral good is normal, $dx_I^*/dw > 0$, then $dm^*/dw < 0$, while if the immoral good is inferior, then $dm^*/\partial w > 0$. This finishes the proof of Proposition 1.

Applying the same method with respect to p_M and substituting dx_I^*/dw and dx_I^*/dp_M into the expression for dm^*/dp_M gives

$$\begin{aligned}
\frac{dm^*}{dp_M} &= d_{32}\left[\frac{p_M p_I \lambda_B}{|H|} + \frac{x_M^*}{\left(2\delta + d_{33}\right)}\frac{dx_I^*}{dw}\right] \\
&= -\frac{d_{32}}{\left(2\delta + d_{33}\right)}\frac{dx_I^*}{dp_M}.
\end{aligned}$$

Sufficient conditions for optimality implies that $|H|$ is negative. Therefore, if the immoral good is inferior, then moral values are decreasing in p_M. Hence, it need not be the case that the moral value is decreasing in p_M when the immoral good is normal. This is so because a price increase of p_M need not lead to an decrease in consumption of the immoral good. From the second expression, however, we see that moral values are decreasing in p_M whenever $dx_I^*/dp_M > 0$, i.e., when the two goods are gross substitutes, and increasing in p_M whenever they are gross complements.

Finally, we follow the same procedure with respect to p_I and substituting dx_I^*/dw into the expression for dm^*/dp_I we get

$$\frac{dm^*}{dp_I} = d_{32}\left[\frac{dx_I^*}{dw}\frac{1}{\left(2\delta + d_{33}\right)}x_I^* - \frac{p_M^2\lambda_B}{|H|}\right].$$

We see that this derivative will always be positive if the immoral good is normal. This result finishes the proof of Proposition 2 and 3.

As a final remark we can substitute dx_I^*/dw and dx_M^*/dw into the expression for dx_I^*/dp_I and dx_M^*/dp_M to get

$$\begin{aligned}\frac{dx_I^*}{dp_I} &= \left[(2\delta + d_{33})\,\frac{p_M^2\lambda_B}{|H|}\right] - \frac{dx_I^*}{dw}x_I^*,\\ \frac{dx_M^*}{dp_M} &= \left[(2\delta + d_{33})\,\frac{p_I^2\lambda_B}{|H|}\right] - \frac{dx_M^*}{dw}x_M^*.\end{aligned}$$

The first term of these expressions corresponds to the substitution effect and the second part to the income effect. The income effect might be either positive or negative, but the substitution effect is always negative. These expressions are identical to the standard Slutsky equation except that the substitution effect is multiplied with $(2\delta + d_{33})$. This means that we can analyze the problem in terms of substitution and income effects, with the only difference that the substitution and income effects have a different magnitude than in the standard utility maximization problem.

Appendix: Description of Data

The 69 countries in the WVS 1999-2004 wave for which at least one of the moral values questions are available are listed in Table A1. Northern Ireland is included as a separate country. The asterisks indicate the 32 European Values Survey countries where data is available for additional moral values questions.

Table A1. List of countries

Albania
Algeria
Argentina
Austria*
Bangladesh
Belarus*
Belgium*
Bosnia and Herzegov.
Bulgaria*
Canada
Chile
China
Croatia*
Czech Republic*
Denmark*
Egypt
Estonia*
Finland*
France*
Germany*
Great Britain*
Greece*
Hungary*
Iceland*
India
Indonesia
Iran
Ireland*
Israel
Italy*
Japan
Jordan
Kyrgyzstan
Latvia*
Lithuania*
Luxembourg*
Macedonia
Malta*
Mexico
Moldova
Morocco
Netherlands*
Nigeria
Northern Ireland*
Pakistan
Peru
Philippines
Poland*
Portugal*
Puerto Rico
Romania*
Russia*
Serbia and Montenegro
Singapore
Singapore
Slovakia*
Slovenia*
South Africa
South Korea
Spain*
Sweden*
Tanzania
Turkey*
Uganda
Ukraine*
USA
Venezuela
Vietnam
Zimbabwe

The categorical variables used in the short regressions are gender (WVS code: x001) and age (WVS code: x003). In the long regression, dummies are included for employment status, educational attainment, occupation, size of home town and number of children of the respondent. Educational attainment (WVS code: x025) is measured in eight different categories, ranging from inadequately completed elementary education to university education with a degree. The employment categories (WVS code: x028) include full-time, part-time, self-employed, retired, housewife, student, unemployed and other. There are 16 occupational dummies (WVS code: x036), for example skilled manual worker, farmer and professional worker. The size of town dummies (WVS code: x049) contains eight different size brackets for the size of the town where the respondent lives. Finally, the number of children (WVS code: x011) of the respondent are included as dummies.

The wording of the moral values questions are reported in Table A2.

Table A2. Moral values questions

Question	WVS Code	Wording of question
		Please tell me for each of the following statements whether you think it can always be justified, never be justified, or something inbetween:
Benefit fraud	f114	"Claiming government benefits to which you are not entitled"
Avoiding fare	f115	"Avoiding a fare on public transport"
Joyriding	f125	"Taking and driving away a car belonging to someone else (joyriding)"
Smoking in public	f133	"Smoking in public buildings"
Prostitution	f119	"Prostitution"
Taking soft drugs	f126	"Taking the drug marijuana or hashish"
Drink and drive	f130	"Driving under the influence of alcohol"
Overspeeding	f134	"Speeding over the limit in built-up areas"
Paying cash for services	f131	"Paying cash for services to avoid taxes"
Cheating on taxes	f116	"Cheating on taxes if you have a chance"

References

Akerlof, G. A. and Dickens, W. T. (1982), "The Economic Consequences of Cognitive Dissonance", *American Economic Review* 72(3), 307–319.

Andreoni, J., Erard, B. and Feinstein, J. (1998), "Tax Compliance", *Journal of Economic Literature* 36(2), 818–860.

Aronson, E. (2003), *The Social Animal*, 9 edn, Worth Publishers, New York.

Babcock, L. and Loewenstein, G. (1997), "Explaining Bargaining Impasse: The Role of Self-Serving Biases", *Journal of Economic Perspectives* 11(1), 109–126.

Beasley, R. K. and Joslyn, M. R. (2001), "Cognitive Dissonance and Post-Decision Attitude Change in Six Presidential Elections", *Political Psychology* 22(3), 521–540.

Bénabou, R. and Tirole, J. (2003), "Intrinsic and Extrinsic Motivation", *Review of Economic Studies* 70(3), 489–520.

Bénabou, R. and Tirole, J. (2006*a*), "Belief in a Just World and Redistributive Politics", *Quarterly Journal of Economics* 121(2), 699–746.

Bénabou, R. and Tirole, J. (2006*b*), "Incentives and Prosocial Behavior", *American Economic Review* 96, 1652–1678.

Bowles, S. (1998), "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions", *Journal of Economic Literature* 36(1), 75–111.

Brekke, K. A., Kverndokk, S. and Nyborg, K. (2003), "An Economic Model of Moral Motivation", *Journal of Public Economics* 87(9-10), 1967–1983.

Chaloupka, F. and Warner, K. (2000), "The Economics of Smoking", in A. J. Culyer and J. P. Newhouse, eds, *Handbook of Health Economics 1B*, Elsevier, Amsterdam.

Dickens, W. T. (1986), "Crime and Punishment Again: The Economic Approach with a Psychological Twist", *Journal of Public Economics* 30(1), 97–107.

Ellingsen, T. and Johannesson, M. (2008), "Pride and Prejudice: The Human Side of Incentive Theory", *American Economic Review* 98(3), 990–1008.

Festinger, L. (1957), *A Theory of Cognitive Dissonance*, Stanford University Press, Stanford.

Frey, B. S. (1997), *Not Just For the Money: An Economic Theory of Personal Motivation*, Edward Elgar Publishing, Cheltenham & Brookfield.

Haagsma, R. and Koning, N. (2002), “Endogenous mobility-reducing norms”, *Journal of Economic Behavior and Organization* 49, 523–547.

Konow, J. (2000), “Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions”, *American Economic Review* 90(4), 1072–1091.

Lieberman, M. D., Ochsner, K. N., Gilbert, D. T. and Schacter, D. L. (2001), “Do Amnesics Exhibit Cognitive Dissonance Reduction? The Role of Explicit Memory and Attention in Attitude Change”, *Pscyhological Science* 12(2), 135–140.

Lindbeck, A. (1995), “Welfare State Disincentives with Endogenous Habits and Norms”, *Scandinavian Journal of Economics* 97(4), 477–494.

Lindbeck, A. (1997), “Incentives and Social Norms in Household Behavior”, *American Economic Review* 87(2), 370–377.

Mills, J. (1958), “Changes in Moral Attitudes Following Temptation”, *Journal of Personality* 26(4), 517–531.

Mullainathan, S. and Washington, E. (2006), “Sticking with Your Vote: Cognitive Dissonance and Voting”, *NBER Working Paper* 11910.

Nyborg, K. (2003), “The Impact of Public Policy on Social and Moral Norms: Some Examples”, *Journal of Consumer Policy* 26, 259–277.

Oxoby, R. J. (2003), “Attitudes and allocations: status, cognitive dissonance, and the manipulation of attitudes”, *Journal of Economic Behavior and Organization* 52, 365–385.

Oxoby, R. J. (2004), “Cognitive Dissonance, Status and Growth of the Underclass”, *Economic Journal* 114, 727–749.

Rabin, M. (1994), “Cognitive Dissonance and Social Change”, *Journal of Economic Behavior and Organization* 23(2), 177–194.

Rachels, J. (1999), *The Elements of Moral Philosophy*, 3 edn, McGraw-Hill, New York.

Saffer, H. and Chaloupka, F. (1999), "The Demand for Illicit Drugs", *Economic Inquiry* 37(3), 401–411.

Shinar, D., Schechtman, E. and Compton, R. (2001), "Self-reports of Safe Driving Behaviors in Relationship to Sex, Age, Education and Income in the US Adult Driving Population", *Accident Analysis and Prevention* 33(1), 111–116.

Shleifer, A. (2004), "Does Competition Destroy Ethical Behavior?", *American Economic Review* 94(2), 414–418.

Titmuss, R. M. (1970), *The Gift Relationship: From Human Blood to Social Policy*, Allen & Unwin, London.

Van de Ven, J. (2003), "Optimal Subsidies with Rationalizing Agents: Subsidize Enough, but Don't Subsidize Too Much", in *Psychological Sentiments and Economic Behaviour*, PhD Thesis, Tilburg University, chapter 7, pp. 157–177.

Varian, H. R. (1992), *Microeconomic Analysis*, 3 edn, W. W. Norton & Company, New York & London.

PAPER 4

Identity and Redistribution

with Erik Lindqvist

Abstract. This paper models the interaction between individuals' identity choices and redistribution. Both redistributive polices and identity choices are endogenous, and there might be multiple equilibria. The model is applied to ethnicity and social class. In an equilibrium with high taxes, the poor identify as poor and favor high taxes. In an equilibrium with low taxes, at least some of the poor identify with their ethnic group and favor low taxes. The model has two main predictions. First, redistribution is highest when society is ethnically homogenous, but the effect of ethnic diversity on redistribution is not necessarily monotonic. Second, when income inequality is low, an increase in income inequality might induce the poor to identify with their ethnic group and therefore favor lower taxes.

1. Introduction

> The Marxian solidarity between the toilers of all the earth will, indeed, have a long way to go as far as concerns solidarity of the poor white Americans with the toiling Negro. (Myrdal 1944, p. 69)

Both economic theories of redistribution (e.g. Meltzer and Richard 1981) and Marxian theory assume that people's political preferences are determined by their economic position in society. This view is controversial. Conflicts along other dimensions, such as ethnicity, race, religion or gender, may be more important than social class. In particular, it has often been argued that class conflict is rare in societies that are ethnically divided. For example, the racial diversity among the American working class is a recurring theme in the literature on the failure to establish a strong worker's movement in the United States.[1]

We would like to thank Alberto Alesina, Johan Almenberg, George Baker, Kaushik Basu, Milo Bianchi, Tore Ellingsen, Ola Granström, Magnus Johannesson, Glenn C. Loury, Erik Mohlin, Ola Olsson, Torsten Persson, Daniel N. Posner, Per Sonnerby, Björn Tyrefors, Björn Wallace and seminar participants at the EEA Congress 2006, Massachusetts Institute of Technology, Stockholm School of Economics and the Zeuthen Workshop on Behavioral Economics 2005 for valuable comments. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

[1] See, for example, Myrdal (1944), Glazer and Moynihan (1970) and Lipset and Marks (2000).

The view that there are multiple dimensions of political conflict invokes the questions under what circumstances voters come to identify with a certain group and how their identities affect redistributive policies. Social psychology research indicates that people tend to identify with groups that have high status, suggesting that redistribution in turn might affect identity choices.[2] In this paper, we develop a formal framework where both redistribution and identity choices are endogenously determined. We use ethnicity and social class (defined as intervals of the income distribution) as our leading example throughout the paper, but the model is applicable to any situation where there are two potential dimensions of social cleavage.[3]

We view identity as altruism toward a subset of the population. Each agent belongs to an ethnic group and a social class, but chooses to identify with only one of them. Two factors determine the identity choice. First, agents want to identify with groups with high status, which is given by the group's average after-tax income. Second, identification with a group involves a cognitive cost which is determined by the proportion of different types in the group. For example, a person feels more distant to her social class if it consists of many people from other ethnic groups. Identity choices affect preferences for redistribution, and redistribution in turn affect identity choices through its effect on the status of different groups. In equilibrium, we require that voting and identity choices are consistent in the sense that nobody wants to switch identity or vote differently. This equilibrium concept is called *Social Identity Equilibrium* and is due to Shayo (2007). We show that there always exists at least one equilibrium in the general case with any number of ethnic groups and social classes. In addition, we show that in this general setting, but with restrictive assumptions on the income distribution, increasing the size of a small ethnic group, or adding a small ethnic group, might reduce the level of redistribution. This is in line with the empirical finding that ethnic diversity is associated with lower levels of redistribution (see Alesina and La Ferrara 2005 for a survey of this literature).

In order to make the analysis more tractable, we further specify the model so that there are only two income levels (rich and poor) and two ethnic groups (black and white). We assume that the poor are in majority, blacks are in minority and the average income of blacks is the same as for whites. In this setting, the level of redistribution is determined by the identity choices of poor blacks and poor whites. In a more extensive working paper version of this paper (Lindqvist and Östling 2007), we

[2] See the study by Roccas (2003) and references cited therein for empirical evidence that people tend to identify with groups that have high status.

[3] See Posner (2005) for a variety of different examples of two-dimensional social cleavages, and Alesina and La Ferrara (2005*a*) for further motivation why ethnic identitification is likely to be endogenous to economic policy.

also consider the case when blacks are on average poorer than whites, as well as some additional results.

Poor blacks and poor whites choose whether to identify with the poor or their respective ethnic group. Identity choices are determined by a trade-off between the relatively higher status of the ethnic identities, and the potentially lower cognitive distance to the poor identity. Identity choices in turn affect voters' preferences for redistribution. If poor blacks and poor whites identify as poor, they are altruistic toward the poor and vote for a high tax rate. If they instead identify with their respective ethnic groups, they favor low taxes since their altruism is now confined to the relatively richer black and white groups. The implemented tax rate in turn determines the status of the poor, black and white identities through the effect on the average after-tax income of these groups.

Poor whites are most prone to identify as poor and favor high taxes when there are no blacks in society at all. As the number of blacks increases, the cognitive distance to the poor identity grows and poor whites might switch to a white identity and favor lower taxes. This mechanism can explain why social class seems to be more important, and redistribution higher, in ethnically more homogeneous societies (e.g. Scandinavia compared to the US). However, for poor blacks, a higher proportion of blacks also reduces the cognitive distance to the poor identity. Consequently, an increase in the number of blacks might induce poor blacks to identify with the poor and favor more redistribution. The effect of ethnic diversity on redistribution is therefore not necessarily monotonic in our model.

The standard model of redistribution (e.g. Meltzer and Richard 1981) predicts that an increase in pre-tax income inequality, in the sense of a larger distance between the median and average income, leads to more redistribution. In contrast, but in accordance with empirical evidence (e.g. Perotti 1996 and Lind 2005), higher pre-tax income inequality need not imply more redistribution in our model. The reason is that higher income inequality may increase the status difference between ethnic and poor identities, which makes the poor more likely to identify with their ethnic groups and favor low taxes.

A feature of the model is that there is a complementarity between tax rates and identity choices. A higher tax rate increases the status of the poor identity relative to ethnic identities and makes it more likely that poor whites and poor blacks identify with the poor and prefer higher taxes. The complementarity between tax rates and identity choices implies that there might be multiple equilibria. For example, we may have one high tax equilibrium where the poor identify as poor and one low tax equilibrium where they identify with their respective ethnic group.[4]

Our model extends the redistribution model in Shayo (2007), where individuals have the choice between identifying with their social class (rich or poor) and a common nationalist identity. There are two main conceptual differences. First, extending the model with an ethnic dimension, i.e., several ethnic identities rather than a single nationalist identity, implies that more than one group can influence the tax rate. This creates an interdependence between the identity choices of different groups. Second, we define individuals' cognitive distances to different identities as a function of the proportions of different types in the population. This allows us to explicitly study the effects of changes in the demographic composition of society, such as an increase in ethnic diversity. These differences affect the results. For example, Shayo (2007) argues that an increase in ethnic diversity concentrated among the poor reduces redistribution, which is not always the case in our model.

Our approach differs from previous economic theories of ethnic diversity and redistributive policies.[5] First, there are models that expand the policy space with a non-economic issue or targeted transfers. Roemer (1998) studies how an additional non-economic political issue such as religion or race leads to a bundling effect of political policies.[6] A citizen that favors a high degree of redistribution may vote for a political party that advocates a low degree of redistribution if he favors the political party's position on racial issues. Fernández and Levy (2008) instead consider endogenous political parties where the policy space consists of general redistribution and targeted public goods. For intermediate levels of preference diversity they find that the rich might form a winning coalition with special interest groups among the poor to reduce general redistribution. In contrast to these models, we consider a one-dimensional policy space. A second type of explanation, put forward by, e.g., Alesina et al. (1999) and Alesina et al. (2001), concerns a direct effect of ethnic fragmentation on voter preferences for redistribution. In their view, a voter's altruistic motive for

[4] The presence of multiple equilibria suggests that the model may be difficult to test empirically. However, the formally stated results provide predictions given initial identity choices, incomes and population proportions of different groups. Income and population proportions are easily available data, and there are several ways to empirically measure people's identities, for example using survey responses or the probability of homogamy (see Bisin et al. 2006 for a recent example). Given such data, our model provides empirically testable predictions for both the level of redistribution and individuals' identity choices.

[5] More generally, ethnic heterogeneity might of course also influence economic outcomes through other channels than the political system. For example, ethnicity might influence the ease by which people cooperate, act as focal points in coordination games or affect the possibility to enforce social forms through social networks. See Habyarimana et al. (2007) for references and an overview of this literature.

[6] Austen-Smith and Wallerstein (2006) develops a related model of legislative bargaining.

redistribution is confined to people that belong to her own ethnic group. Common to both types of explanations is that voters' political preferences on ethnic issues are exogenous, whereas both preferences and redistribution are determined endogenously in our model.

In the next section of the paper, we develop the model with arbitrarily many ethnic groups and social classes. In Section 3 we specify the black and white model that is used throughout the rest of the paper. The following three sections discuss the implications of the black and white model. Section 7 discusses how "American Exceptionalism" – the difference in redistribution between the United States and Western Europe – can be explained by our model. Section 8 concludes the paper.

2. General Model

In this section we extend the model of redistribution in Shayo (2007) to arbitrarily many ethnic and class identities. The model could equally well be applied to identities along two other dimensions, for example language and religion. We refer to Shayo (2007) for a detailed discussion of how the model relates to the social psychological research on social identity.

Consider a set of N agents, a finite set C of social classes and a finite set E of ethnic groups. We view a social class as a particular interval of the income distribution, i.e., all agents within a certain income interval belongs to the social class corresponding to that interval.[7] Each agent also belongs to an ethnic group.[8] All social classes are represented in every ethnic group, and we refer to a particular combination of class and ethnicity as a type. Agents must choose to identify with either their ethnic group or their social class.[9] Given this identity choice, agents also choose which tax rate to

[7] It is not necessary to define social classes in terms of income intervals – one can think of more complicated mappings that takes educational and cultural aspects into account. Proposition 1 holds also with such alternative interpretations.

[8] We assume that there is an uncontroversial way to determine which ethnic group each agent in the economy belongs to. In practice, this is of course easier said than done. For an axiomatic approach to determination of group membership, see Kasher and Rubinstein (1997).

[9] This assumption raises three related issues. First, why can't an individual identify with her type, i.e., her particular combination of class and ethnicity? We don't allow this since it would be very similar to allowing each agent to identify with herself, which corresponds to the standard model of redistribution. Second, why can't an individual identify with both her ethnic group and her social class? In this setting, an agent "can't have both" since she has to vote for her preferred level of redistribution which forces her to decide on how much to favor either of her two groups. However, it is straightforward to allow for intermediate identification, e.g. 30 percent class identification and 70 percent ethnic identification (see footnote 12). Third, we don't allow agents to identify with a group they don't belong to. Though we could allow agents to identify with any group in society, this aspect is not relevant in contexts where it is very costly to shift ethnic identity (for example, from black to white in the US).

vote for. Simple majority voting selects the winning tax rate and in equilibrium we require that the resulting tax rate is consistent with identity and voting choices.

Each agent in the economy is endowed with pre-tax income $y_i > 0$ and the average income in the population is denoted by y. There is a single proportional tax rate t and tax revenues are redistributed lump-sum.[10] There is a quadratic deadweight loss of taxation equal to $(t^2/2)\, y$.[11] This implies that that the income after taxes and transfers of agent i is $\bar{y}_i = (1-t)\, y_i + (t - t^2/2)\, y$. Similarly, let y_j denote the average pre-tax income of the agents belonging to ethnic group $j \in E$ or social class $j \in C$ so that their average after-tax income (including transfers) is given by $\bar{y}_j = (1-t)\, y_j + (t - t^2/2)\, y$.

Since each agent belongs to one social class and one ethnic group, the average income of these two categories will generally differ. For an agent with low income, the average income in her ethnic group will typically be higher than in her social class, whereas for rich people the social class will typically have a higher average income than the ethnic group. We refer to the category with the higher pre-tax income as the agent's high status identity and the other as the low status identity. The average pre-tax income in the high status versus low status identity is denoted by y_H and y_L where $y_H \geq y_L$. The identity choice consists of choosing l_i to be either zero or one, where $l_i = 1$ means that the agent identifies with the low status group and $l_i = 0$ that she identifies with her high status group.[12]

An agent's utility consists of two parts: material utility arising from after-tax income including transfers and the immaterial utility arising from identification with a group. The utility function is assumed to be additively separable and take the following form:

$$U_i = \bar{y}_i\,(t) + \gamma\left(l_i \bar{y}_L\,(t) + (1-l_i)\, \bar{y}_H\,(t)\right) - \delta\left(l_i d_L + (1-l_i)\, d_H\right), \tag{2.1}$$

where t is the prevailing tax rate and $\bar{y}_H\,(t)$ and $\bar{y}_L\,(t)$ are the after-tax incomes of the two categories the agent belongs to. The first term in the utility function represents direct material benefit of after-tax income, the second term the status of the group the individual identifies with, and the last term the cognitive distance to the same group. The parameters γ and δ are positive so that utility is increasing in status and

[10] We do not allow targeted redistribution. Although this might be a relevant extension, there are two main reasons why we focus on general redistribution (i.e. from rich to poor). First, the empirical literature concerning ethnic diversity and redistribution mainly concerns general redistribution. Second, many democracies have high legal barriers to discriminatory redistributive policies which limits the scope for redistribution targeted to specific ethnic groups.

[11] The results in this section of the paper holds as long as the deadweight loss is strictly convex in t so that unique solutions to agents' voting problems exist and preferences are single-peaked.

[12] We assume that an agent cannot partially identify with a group. However, Proposition 1 is unaffected if the agent is allowed to pick $l_i \in [0,1]$ as long as the specification implies a unique solution $l_i^*\,(t)$ that is non-decreasing in t.

decreasing in cognitive distance. Group status is linearly increasing in the group's after-tax average income and so the higher income group has higher status than the lower income group.

The cognitive distance to a group is higher if there are many people of a different type than oneself in the group.[13] To make this precise, let p_{jk} denote the proportion in the population that belong to social class $j \in C$ and ethnic group $k \in E$. The distance of an agent that belongs to social class j and ethnic group k to his social class j is given by

$$d_{jkj} = d\left(\beta \sum_{h \in E \setminus \{k\}} p_{jh} / \sum_{h \in E} p_{jh}\right),$$

where $d(\cdot)$ is some positive and increasing real-valued function and $\beta > 0$. In other words, the distance to the class identity is increasing in the proportion of people belonging to the same social class that is from a different ethnic group. The parameter β is a measure of ethnic tensions – if β is high, the distance to the class identity is large since the members of a social class come from different ethnic groups. Similarly, an agent that belongs to social class j and ethnic group k has the following distance to his ethnic group k:

$$d_{jkk} = d\left(\alpha \sum_{h \in C \setminus \{j\}} p_{hk} / \sum_{h \in C} p_{hk}\right),$$

where $\alpha > 0$ is a measure of class tensions. Note that cognitive distances do not depend on tax rates – the tax rate only affects the material utility and the relative status of groups. The above specifications imply that the distance to an identity is unaffected by the identity choices of other agents – the distance to a certain identity only depends on characteristics of the population.

The tax rate t is determined by simple majority voting or some other political process that selects the median tax under the assumption of single-peaked preferences. The political process will hence be a mapping Γ from the vector of all tax votes $t^* \in \times_{i \in N}[0,1]$ to a median tax rate $t \in [0,1]$.

Following the definition of social identity equilibrium in Shayo (2007) we require that three conditions must hold in equilibrium:

[13] The assumption that people tend to identify with people similar to themselves fits well with research within evolutionary psychology. People tend to be more altruistic toward kin than nonkin, but, as argued by for example Waldman (1987), the recognition of kin is not perfect and relies on a variety of proximate mechanisms. Similarity in terms of ethnicity or social standing may thus function as perceptual cues that trigger altruistic behavior even when actual kinship bonds are weak.

(1) All individuals vote for their preferred tax rate given the identity choice l_i:

$$t_i^*(l_i) = \underset{t\in[0,1]}{\arg\max}\left\{\bar{y}_i(t) + \gamma\left(l_i\bar{y}_L(t) + (1-l_i)\,\bar{y}_H(t)\right) - \delta\left(l_i d_L + (1-l_i)\,d_H\right)\right\}.$$

(2) All agents choose identity optimally given the prevailing tax rate t:

$$l_i^*(t) = \underset{l_i\in\{0,1\}}{\arg\max}\left\{\bar{y}_i(t) + \gamma\left(l_i\bar{y}_L(t) + (1-l_i)\,\bar{y}_H(t)\right) - \delta\left(l_i d_L + (1-l_i)\,d_H\right)\right\}.$$

(3) The median tax rate is consistent with identity choices and voting behavior of all individuals:

$$\Gamma\left(t^*\left(l^*(t)\right)\right) = t.$$

Note that identity and voting choices are taken separately. The main reason for this assumption is that these are two conceptually different decisions that are likely to be made under different circumstances and at different points in time. The equilibrium concept is aimed to capture the steady state of a dynamic process where people vote for taxes given their identity choices, but may change their identity choice as a new tax rate is implemented.[14] A second reason is that preferences for taxes are single-peaked only for given identity choices. In order to be able to use the median voter theorem we cannot admit agents to switch identity at the same time as they choose tax rates.

First consider the agents' voting choices. The utility function (2.1) is strictly concave in t and we can therefore use the first-order condition to derive the solution to the agent's tax voting decision:

$$t_i^*(l_i) = \max\left\{1 - \frac{y_i + \gamma\left(l_i y_L + (1-l_i)\,y_H\right)}{(1+\gamma)\,y}, 0\right\}. \tag{2.2}$$

Note that this tax rate is non-decreasing in l_i, i.e., the more the agent identifies with the low status identity, the higher is her preferred tax. The reason is that people are altruistic toward the group they identify with. Since the low status group is poorer, an agent favors more redistribution if she identifies with that group. That people tend vote for tax rates in precisely this way is shown in an experiment by Klor and Shayo (2007).

Now consider optimal identity choices. For a given tax rate t, an agent chooses the high status identity, i.e., $l_i = 0$, if[15]

$$\gamma(1-t)(y_H - y_L) > \delta(d_H - d_L). \tag{2.3}$$

[14] Furthermore, we implicitly assume that players aren't forward looking in the sense that they anticipate their own or others future tax and identity choices when making identity and voting decisions.

[15] In the unlikely event that an agent is indifferent between the two identities, we will assume that the agent chooses the low status identity.

It is clear from this condition that $l_i^*(t)$ is non-decreasing in t. In other words, for given cognitive distances, a higher tax rate implies that the low status identity becomes relatively more attractive since redistribution benefits the low status group more. The higher the prevailing tax rate, the more likely it is that people identify with their low status identity, which in turn would imply that they vote for higher tax rates. Since the median tax rate is non-decreasing in the vector of tax rate choices, there is a complementarity between identity choices and the tax rate. This complementarity means that there are potentially many equilibria, but it also allows us to establish that at least one equilibrium exists.

Proposition 1. *There exists at least one social identity equilibrium.*

All proofs are provided in the Appendix.

It is difficult to derive any general comparative statics without further specifying the model. In order to derive results for the effects of an increase in ethnic diversity that are not confounded by income differences between ethnic groups, we first study the simplest possible distribution of income.

Suppose there are only two income levels, $y_R > y_P > 0$, and consequently two social classes, rich (R) and poor (P). The poor are in majority and all ethnic groups have same proportion of poor. From the expression for the most preferred tax rate (2.2) we see that these assumptions imply that the rich prefer zero taxes irrespective of how they identify themselves, and that the poor always prefer positive taxes. The median voter(s) must therefore be poor. From the condition for high status identification (2.3) it is clear that the identity choices of the poor only differ in the distances to the poor identity. The larger the ethnic group a poor individual belongs to, the smaller is the distance to the poor identity and the more likely she is to identify with the poor. There are only two possible equilibrium tax rates in this setting. In the high tax equilibrium, the poor in relatively large ethnic groups identify as poor and they are sufficiently many to create a majority for their preferred tax rate, which is high since they are altruistic toward the poor. In the low tax equilibrium, the poor in relatively small ethnic groups identify with their ethnic groups and are sufficiently many to be pivotal. This tax rate is lower since they are altruistic toward their ethnic groups that contain both rich and poor.[16] Increasing the size of an existing ethnic group, or adding an ethnic group with the same proportion of poor as the already existing ones, implies that all other ethnic groups shrink in relative size, and consequently, that the distances to the poor identity increase. As long as the enlarged or added ethnic group is sufficiently small, the set of parameter values that can support the high tax equilibrium shrinks,

[16] There may of course also be equilibria where all poor identify either as poor or with their ethnic groups.

which is stated in Proposition 2.[17] Proposition 2 and the propositions that follow below provide conditions under which a change in parameters may render the initial equilibrium unfeasible. Hence, we do not consider the possibility that the economy may shift from one equilibrium to another in the presence of multiple equilibria.

Proposition 2. *Let p_j denote the proportion of the population belonging to ethnic group j. If the assumptions in the previous paragraph hold, there is a threshold $\widehat{p}$ such that the poor of ethnic group j identify with the poor if $p_j \geq \widehat{p}$, and with their ethnicity if $p_j < \widehat{p}$. Increasing the size of an existing ethnic group smaller than $\widehat{p}$, or adding a new ethnic group that has the same proportion of poor as the pre-existing population and a size smaller than $\widehat{p}$, increases the proportion of poor that identify with their ethnic group and might lower the equilibrium tax rate.*

In line with Proposition 2, several papers have shown empirically that there is a negative relation between ethnic heterogeneity and redistribution both across countries and between communities within countries. For example, Alesina et al. (2001) found social spending to be lower in countries with a high degree of racial fractionalization; Alesina et al. (1999) found a lower degree of public goods provision in ethnically fragmented metropolitan areas in the US, and Soss et al. (2001) found that when US states were given greater autonomy to set their own welfare policies, states with higher proportion of blacks implemented more punitive welfare regulations. Luttmer (2001) shows that support for welfare spending in the US is higher among people living in areas where the proportion of welfare recipients from their own racial group is high. Similarly, Orr (1976) found a negative correlation between aid to families with dependent children and the proportion of non-white welfare recipients across US states.

A seemingly paradoxical finding that our model can explain is why class voting, i.e., the extent to which voting behavior coincide with social class, seems to be particularly important in Scandinavian countries – which have the lowest income inequality in the world. Our answer is that the Scandinavian countries are relatively ethnically homogeneous, suggesting that the poor identify with their class.[18] In line with this explanation, Nieuwbeerta and Ultee (1999) found a negative correlation between religious and ethnic diversity and the level of class voting.

The model also suggests that members of small ethnic groups tend to identify with their ethnicity, which resonates well with the picture of New York in the 1960s

[17] Although Proposition 2 is stated in terms of a threshold $\widehat{p}$ that is not directly observable, the threshold can be indirectly inferred from initial identity choices (since ethnic groups identify differently depending on whether they are below or above the threshold).

[18] According to Nieuwbeerta and Ultee (1999), class voting was particularly important in the Scandinavian countries at least until the 1980s. Since then class voting has declined, but on the other hand the Scandinavian countries have also become more ethnically heterogenous due to immigration.

described by Glazer and Moynihan (1970).[19] A similar idea has also been used by the authoritarian former leader of Singapore, Lee Kuan Yew, to legitimize Singapore's one-party system:

> In multiracial societies, you don't vote in accordance with your economic interests and social interests, you vote in accordance with race and religion. Supposing I'd run their system [democracy] here: Malays would vote for Muslims, Indians would vote for Indians, the Chinese would vote for Chinese. (Spiegel 2005, p. 23)

Although Lee Kuan Yew may be right that Malays and Indians in Singapore would vote for their own ethnic groups if they were allowed to vote, it is less clear that the Chinese would do so since they constitute roughly three quarters of the population.[20]

The idea that ethnic identification is stronger the smaller is the ethnic group is also in line with the study of ethnic minorities in the UK by Bisin et al. (2006). They find evidence that the higher is the percentage of a person's own ethnic group in the neighborhood, the lower is the degree of ethnic identification and the probability of homogamy. Similarly, Fryer and Torelli (2006) find that the phenomenon of 'acting white' among blacks – interpreted as racial differences in the relationship between academic performance and popularity – is stronger in US schools with few black students.

However, it is also plausible that the relationship between ethnic identification and the size of ethnic groups might be different for very small ethnic groups in ways not captured by our model. For example, Miguel and Posner (2006) find that there is a *negative* relationship between the degree of ethnic identification and ethnic diversity for twelve ethnically diverse African countries.

Proposition 2 requires quite strong assumptions on the distribution of income and size of ethnic groups. We therefore specify a simpler version of the model with only two ethnic groups, which allows us to study the effects of ethnic diversity and income inequality under less restrictive assumptions.

3. Black and White Model

In the remainder of the paper, we consider a simpler model with two social classes, poor (P) and rich (R), and two ethnic groups, black (B) and white (W). This simplification is relevant for the US, where the main ethnic division has traditionally been between the Afro-American population and people of European origin. Based on survey data on self-reported social class, the restriction to two social classes seems relevant for

[19] The notion that members of small ethnic groups identify ethnically is consistent with the empirical evidence in Scheve and Slaughter (2001) showing that immigrants have more favorable attitudes toward immigration, also when income is controlled for.

[20] In all cases, it is a poor argument for not allowing the citizens of Singapore to vote.

the US.[21] The simplified model is also likely to be relevant for other countries – for example ethnic divisions between the native population and non-European immigrants in Europe and between the French and English speaking population of Canada.

We denote the proportion of the four different types in the population by p_{PB}, p_{RB}, p_{PW} and p_{RW} and as before we assume that all four types are represented in the population. In addition, we assume that no type, or sum of two or three types, consists of exactly half the population since this allows us to disregard the possibility of the median falling between two types' preferred tax rate.

All individuals of a certain type are identical – all people in the rich income group have pre-tax income y_R and everybody in the poor group have income y_P satisfying $y_R > y_P > 0$.[22] This specification implies that the status of the ethnic groups is in between the status of the poor and rich groups. In other words, the ethnic identity is the high status identity for poor people, whereas it is the low status identity for rich people.

Actual income distributions are typically skewed so that the median income is less than the average income. Since there are only two income levels in the model, we therefore assume that the poor population is in majority, i.e., $p_{PB} + p_{PW} > 1/2$. Without loss of generality, we also assume that the white population is in majority, i.e., $p_{PW} + p_{RW} > 1/2$. Given these assumptions, we have two different cases. First, if poor whites are in majority, the tax rate is uniquely determined by their identity choice. Second, if poor whites do not constitute a majority of the population, both poor whites and poor blacks could potentially determine the tax rate. We assume that the white and black population have the same average income, i.e., $p_{RW}/p_{PW} = p_{RB}/p_{PB}$. In Lindqvist and Östling (2007), we derive formal results also for the case when the white population is richer on average, i.e., $p_{RW}/p_{PW} > p_{RB}/p_{PB}$. In this version of the paper, we mostly give a verbal account for this case.

The cognitive distance function $d(\cdot)$ is given in Table 1. Note that cognitives distances are linear, which implies that it is costless to identify with a group where everybody is of the same type as oneself, whereas the cost goes to α or β when there are nobody like oneself in that group.

We now turn to determining the equilibria of this model. First, recall from (2.2) that the optimal tax rate of someone belonging to social class j and ethnic group k

[21] For example, in the General Social Surveys 1972-2004 (Davis et al. 2005), 46 percent of respondents classify themselves as working class, whereas 46 percent classify themselves as middle class. Of the remaining 8 percent, 5 percent classify themselves as lower class and 3 percent as upper class.

[22] It would admittedly be more realistic to allow agents with different individual incomes, but that makes it difficult to derive interesting comparative statics results. In the Appendix of Lindqvist and Östling (2007) we nevertheless consider this alternative and show that the result in Proposition 3 is similar in that setting.

TABLE 1. Cognitive distance function

	Poor black	Poor white	Rich black	Rich white
Black	$\alpha\frac{p_{RB}}{p_{PB}+p_{RB}}$		$\alpha\frac{p_{PB}}{p_{PB}+p_{RB}}$	
White		$\alpha\frac{p_{RW}}{p_{PW}+p_{RW}}$		$\alpha\frac{p_{PW}}{p_{PW}+p_{RW}}$
Poor	$\beta\frac{p_{PW}}{p_{PB}+p_{PW}}$	$\beta\frac{p_{PB}}{p_{PB}+p_{PW}}$		
Rich			$\beta\frac{p_{RW}}{p_{RB}+p_{RW}}$	$\beta\frac{p_{RB}}{p_{RB}+p_{RW}}$

that identifies with his social class j is given by:

$$t^*_{jkj} = \max\left\{1 - \frac{y_j}{y}, 0\right\}. \tag{3.1}$$

It is clear that rich people who identify themselves as rich prefer a zero tax rate (since $y_R > y$) and that poor people who identify themselves as poor prefer the tax rate $1 - y_P/y$. Similarly, someone identifying with her ethnic group k prefers the tax rate

$$t^*_{jkk} = \max\left\{1 - \frac{y_j + \gamma y_k}{(1+\gamma)\,y}, 0\right\}. \tag{3.2}$$

The optimal voting choices (3.1) and (3.2) imply that preferred tax rates can be ordered within ethnic groups – for example, rich whites always prefer lower taxes than poor whites. Since we need these results later, we state them formally.

LEMMA 1. *Optimal tax rates always satisfy the following:*
(i) $0 = t^*_{RWR} \leq t^*_{RWW} \leq t^*_{PWW} \leq t^*_{PWP} = 1 - y_P/y$,
(ii) $0 = t^*_{RBR} \leq t^*_{RBB} \leq t^*_{PBB} \leq t^*_{PBP} = 1 - y_P/y$.
In addition, if whites are richer than blacks, then $t^*_{RWW} \leq t^*_{RBB}$ *and* $t^*_{PWW} \leq t^*_{PBB}$.

It should be noted that we cannot say whether the preferred tax rate of poor whites identifying themselves as white (t^*_{PWW}) is higher or lower than the tax preferred by rich blacks identifying themselves as black (t^*_{RBB}), even if we make the additional assumption that whites are richer on average. The reason is that the relation between these two tax rates also depends on the parameter γ, i.e., how much individuals care about the status of the group they identify with.

Lemma 1 also implies that when whites are richer than blacks and all types identify with their ethnic group, blacks prefer the same or higher taxes than whites holding income constant. Fong (2001), Alesina et al. (2001) and Alesina and La Ferrara (2005*b*) show empirically that white people in the US are more negative toward redistribution than Afro-Americans also when personal income is held constant. This suggests that the poor in the US identify along ethnic lines rather than with their social class.

If whites are richer than blacks, the status of the ethnic identity is higher for poor whites than for poor blacks. However, that whites are richer also means that the distance for poor whites to the white identity is larger than the distance for poor blacks to the black identity. It is possible to show that the latter effect dominates and that poor blacks always identify as black if poor whites identify as white. This result is dependent on the linear specification of the distance function and the assumption that the class and ethnic tension parameters are the same for all types, but it is also plausible – if the poor in the majority group favor their ethnic group, then we would probably not expect the poor in the minority group to identify with the poor.

Lemma 2. *If whites have the same or higher average income than blacks and poor whites identify as white, then poor blacks identify as black.*

From Proposition 1 we know that at least one equilibrium exists. Since we have assumed that the poor are in majority, we can show that only the identity choices of the poor matter for the equilibrium tax rate. When whites and blacks are equally rich, there can only be two different tax rates in equilibrium since poor whites and poor blacks prefer the same tax rate when they make the same identity choice. For simplicity, we denote the two possible equilibrium tax rates the poor ($t^*_{PWP} = t^*_{PBP}$) and ethnic ($t^*_{PWW} = t^*_{PBB}$) tax rate.

Lemma 3. *If blacks and whites have the same average income, then there are two feasible equilibrium tax rates:*

(1) *If poor whites are in majority and identify as white, or if poor whites are in minority and poor blacks identify as black, the equilibrium tax rate will be the ethnic tax rate* ($t^*_{PWW} = t^*_{PBB} = 1 - (y_P + \gamma y) / (1 + \gamma) y$).

(2) *If poor whites are in majority and identify as poor, or if poor whites are in minority and poor blacks identify as poor, the equilibrium tax rate will be the poor tax rate* ($t^*_{PBP} = t^*_{PWP} = 1 - y_P / y$).

A feature of the model is that there might be multiple equilibria which implies that an equilibrium can be suboptimal in the sense that each agent of a certain type would reach a higher utility level if the other agents of the same type changed identity and preferred tax rate. However, given the identity choices of the other agents, no agent has an incentive to change identity or vote differently.[23] For example, we might have one high tax equilibrium where poor whites identify as poor and one low tax equilibrium

[23] Since voting and identity choices are made separately, an agent can end up in a suboptimal equilibrium even if he is the only agent in the economy (and cognitive distances are defined so that this is possible). A single agent might prefer to simultaneously switch identity and preferred tax rate, but this is ruled out by the definition of an equilibrium.

where they identify as white. Based on Marxian theory it might be tempting to conclude that the poor are better off in a high tax equilibrium, and that the poor should be made "class conscious" if the low tax equilibrium prevails. However, although a class identity would benefit their material interest, our model allows no such conclusion since agents also get utility from their identity – it may well be the case that the poor's utility is lower in a high tax than in a low tax equilibrium.[24]

Although differences in redistribution can be explained in terms of multiple equilibria for identical parameter values, we now go on to study how the set of potential equilibria changes with the parameters of the model. These results together with some empirical evidence are presented in the following three sections.

4. Ethnic Diversity

The main lesson from Proposition 2 is that ethnic diversity might induce the poor to identify with their ethnic group and therefore favor lower taxes. In this section we will see that this conclusion does not hold universally.

In the black and white model, blacks constitute a minority and we therefore model an increase of ethnic diversity as an increase in the black population. The results differ depending on whether poor whites are in majority or not. When poor whites are in majority, their identity choices alone determine the tax rate. When poor whites are in minority, the identity choices of poor blacks also affect the tax rate (unless poor whites already identify as white).

The effect of ethnic diversity depends both on whether poor whites are in majority and the extent of interethnic income inequality. Proposition 3 focuses on an increase in the proportion of blacks when whites and blacks have the same average income. In Lindqvist and Östling (2007), we consider the case when blacks are poorer on average, and we also study an increase in the number of poor blacks when blacks are poorer. Table 2 summarizes the overall effect on the level of redistribution in these three different cases. As is clear from Table 2, the effect of ethnic diversity is typically not monotonic.

When blacks and whites are equally rich and the proportion of poor and rich blacks increases proportionally, the only effect of an increase in the proportion of blacks on identity choices is to increase the cognitive distance to the class identity for poor whites and decrease it for poor blacks. The relative status of both identities and distance to the ethnic identity is unaffected by changes in ethnic diversity. As the proportion of

[24] More generally, the idea that people may hold dysfunctional identities is often raised in the literature on identity and may be important in order to understand self-destructive behaviors such as "ghetto culture" (see Akerlof and Kranton 2000 for references and further discussion).

TABLE 2. Effects on redistribution of an increase in ethnic diversity

	Poor whites are in majority	Poor whites are in minority
Increase of blacks, no interethnic inequality (Proposition 3)	$-$	$+$
Increase of blacks, interethnic inequality (Prop. 4 in LÖ, 2007)	$-$	$-/+$
Increase of poor blacks, interethnic inequality (Prop. 5 in LÖ, 2007)	$-/+$	$-/+$

blacks increases, poor blacks therefore become more prone to identify as poor whereas poor whites become more prone to identify as white.

PROPOSITION 3. *Suppose blacks and whites have the same average income. If poor whites are in majority, then an increase in the black population implies the following for the equilibrium tax rate:*

(1) *If poor whites initially identify as poor, then poor whites might to switch to the white identity resulting in a lower equilibrium tax rate.*
(2) *If poor whites initially identify as white, nothing happens to identity choices and tax rates.*

If poor whites are in minority, then an increase in the black population implies the following for the equilibrium tax rate:

(1) *If poor whites initially identify as white or both poor whites and poor blacks identify as poor, then the tax rate is unchanged.*
(2) *If poor whites initially identify as poor and poor blacks identify as black, then poor blacks might switch to the poor identity resulting in a higher equilibrium tax rate.*

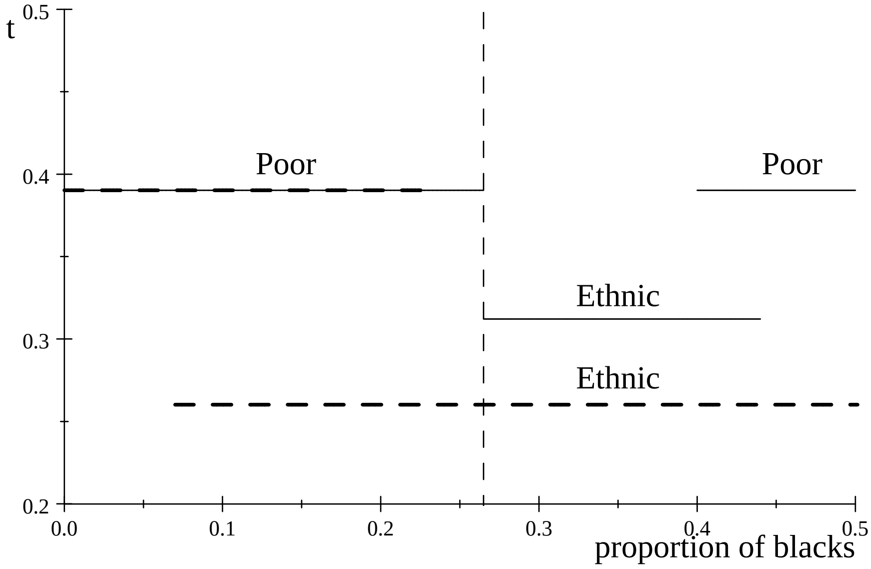
To illustrate the full comparative statics, we consider two parametric examples (illustrated in Figure 1) with different status parameters. The thin dashed vertical line in Figure 1 indicates the proportion of blacks above which poor whites are in minority.

The thick dashed horizontal line in Figure 1 indicates the equilibrium tax rate as a function of the proportion of blacks when agents care relatively much about the status of the group they identify with ($\gamma = 0.5$).[25] The more important is status, the more likely it is that poor blacks and poor whites identify with their respective ethnic groups. In this example, the status parameter is so high that poor blacks always identify themselves as black. Poor whites, on the other hand, identify as poor when society is ethnically homogenous (less than 7 percent blacks) and the higher poor tax rate is the only possible equilibrium. The poor tax rate is an equilibrium also when

[25] The parameters used in this example are $p_P = 0.68$, $y_P = 100$, $y_R = 300$, $\gamma = 0.50$, $\delta = 20$, $\alpha = 4$ and $\beta = 1.3$.

the proportion of blacks is between 7 and 24 percent. However, since poor whites now identify as white at the lower ethnic tax rate, this can also be an equilibrium. When the proportion of blacks is above 24 percent, poor whites identify as white at all tax rates and only the ethnic tax rate is an equilibrium. Hence, poor whites already identify as white at the point when they become a minority (at 27 percent blacks), implying that the tax rate is unaffected by this shift in potential majorities.

FIGURE 1. Increase in ethnic diversity (no interethnic income inequality)



Thick dashed lines: High status ($\gamma = 0.5$). Thin lines: Low status ($\gamma = 0.25$).

The equilibrium tax rate in the second parametric example is indicated by the thin lines in Figure 1. The only difference compared to the previous example is that status is less important ($\gamma = 0.25$ compared to $\gamma = 0.50$), which has two different effects: It makes it more likely that the poor identify with their social class, and it leads to a higher ethnic tax rate. In this example, poor whites always identify as poor. When the proportion of blacks is below 27 percent, poor whites are in majority and since they always identify as poor, the poor tax is implemented. If the proportion of blacks is between 27 and 40 percent, poor whites are not in majority and poor blacks identify as black at both tax rates, so only the ethnic tax rate is an equilibrium. When the proportion of blacks is between 40 and 44 percent, poor blacks identify as poor at the poor tax rate and as black at the ethnic tax rate, implying that both tax rates are equilibria. Finally, as the proportion of blacks is above 44 percent, poor blacks identify as poor at all tax rates and only the poor tax rate can be an equilibrium.

Note that though the effect of ethnic diversity on redistribution was monotonic and negative in the first parametric example, this is not the case in the second example. Instead, redistribution is high when ethnic diversity is either very low or very high. In the intermediate case, there are enough blacks to influence the tax rate, but so few that poor blacks are reluctant to identify with the poor. This provides an explanation for the finding in Dincer and Lambert (2006) that there is a U-shaped relationship between redistribution and ethnic fractionalization and polarization across US states.[26]

Proposition 3 only applies to the case when blacks and whites have the same income. In many societies, the minority population is poorer than the majority group. In this case, increasing the size of the minority group decreases the average income in the population, leading to lower tax rates for given identity choices. We analyze this case formally in Proposition 4 in Lindqvist and Östling (2007) and the results are similar to Proposition 3. One difference, however, is that there are three instead of two potential equilibrium tax rates. Since the black group is poorer than the white group, poor blacks now favor higher taxes when they identify as black than poor whites when they identify as white.

So far, we have assumed that the number of poor blacks and rich blacks increase proportionally, ensuring that the incomes of the black and white groups are held constant. In many cases, for example immigration, it is more reasonable to assume that it is only the proportion of poor blacks that increases. This implies a change in income inequality both within and between the two ethnic groups, which introduces a more intricate interaction between social class and ethnicity. In Proposition 5 in Lindqvist and Östling (2007), we consider an increase in the proportion of blacks among the poor, holding the average income of the population constant. One way to think about this case is as an inflow of poor black immigrants. Such an inflow has counteracting effects on the identity choices of poor blacks. On the one hand, both the status of the black identity and the cognitive distance to the poor identity decrease, which makes it more likely that poor blacks identify as poor and favor high taxes. On the other hand, a higher proportion of poor among blacks also implies that the cognitive distance to the black identity decreases. If there are few black people, the latter effect is stronger and an increase in the proportion of poor blacks might induce poor blacks to identify with the black group and favor less redistribution. Somewhat paradoxically, an increase in

[26] Dincer and Lambert (2006) report that the relationship between fractionalization and redistribution is U-shaped, but do unfortunately not state the sign of the square of their measure of ethnic polarization, just that it is statistically significant in all specifications. However, their graphical evidence strongly suggests that the relationship between ethnic polarization and redistribution is U-shaped. We have contacted the authors in order to clarify this point, but are still awaiting a response. Note also that since we only have two ethnic groups in the model, we cannot distinguish between ethnic fractionalization and polarization.

the proportion of poor blacks may therefore reduce the support for redistribution also among poor blacks.

The results for ethnic diversity may shed some light on the evidence on class voting, i.e., the extent to which social class determines voting behavior (see Nieuwbeerta and Ultee 1999 for references to this sociological literature). The model suggests that immigration of foreign low-skilled people might induce poor whites – and possibly also poor blacks – to identify with their ethnic group and support lower taxes. The inflow of relatively poor immigrants may therefore be part of the explanation for why class voting has declined in Europe during the last decades, as well as why European anti-immigration political parties seem to have gained in popularity.[27] The latter is supported by empirical studies by Knigge (1998) and Golder (2003) showing that the support for anti-immigration parties is indeed increasing in the level of immigration. A competing explanation for the relatively strong support that anti-immigration parties get from the working class is a fear for increased competition in the labor market. However, in contrast to our model, this does not explain why these parties often advocate a low level of redistribution (see for example Betz 1993, Poglia Mileti et al. 2002 and McGann and Kitschelt 2005).[28] In addition, the empirical evidence on the relationship between support for anti-immigration parties and the level of unemployment is ambiguous (see Knigge 1998 and Golder 2003).[29]

5. Income Inequality

Income inequality can mean two different things in this model – the income difference between social classes and the difference in income between ethnic groups. We first analyze the effects of income inequality between rich and poor.

Standard models of income redistribution, e.g. Meltzer and Richard (1981), predict that redistribution increases as a response to an increase in pre-tax income inequality as measured by the distance between the average and median income. When income inequality increases, the poor become poorer compared to the rich which increases their demand for redistribution. In our model there is a counteracting effect since an

[27] Examples of such parties include FPÖ (Austria), Schweizerische Volkspartei (Switzerland), Dansk Folkeparti (Denmark), Vlaams Blok (Belgium), Fremskridtspartiet (Norway) and Front national (France).

[28] The political bundling effect in a two-dimensional policy space demonstrated by Roemer (1998) can explain why a voter could vote for a right-wing party although she favors a high degree of redistribution, but not why anti-immigration parties tend to focus on right-wing economic policies in the first place.

[29] Of course, the absence of any relation between the unemployment rate and anti-immigration sentiments is only an argument against the labor market hypothesis if agents are not perfectly forward-looking regarding the effects of increased immigration, but adjust their beliefs about negative effects of immigration in response to a high level of unemployment.

increase in income inequality increases the status of ethnic identities, which might lead to a shift to ethnic identities and lower tax rates. Since the comparative statics are considerably more complicated if blacks are poorer than whites, without bringing many additional insights, Proposition 4 is only stated for the case when whites and blacks have the same average income.[30]

PROPOSITION 4. *If poor whites are in minority and if whites and blacks have the same average income, then an increase in pre-tax income inequality* $(y_R - y_P)$*, while average income* y *is held constant, implies the following for the equilibrium tax rate:*

(1) *If poor blacks identify as black, the tax rate increases. Furthermore, if in addition income inequality is high* $(y_P/y < (1-\gamma)/2)$*, poor blacks (and possibly poor whites) might switch to the poor identity which increases the tax rate further. If income inequality instead is low* $(y_P/y > (1-\gamma)/2)$*, the identity choices of poor blacks are unchanged.*

(2) *If poor whites and poor blacks initially identify as poor, the tax rate increases and the identity choices of the poor are unchanged if income inequality is high* $(y_P/y < 1/2)$*. If instead income inequality is low* $(y_P/y > 1/2)$*, poor blacks (and possibly poor whites) might switch to ethnic identities which leads to a lower tax rate.*

As can be seen in Proposition 4, the effect of an increase in pre-tax income inequality depends on the initial degree of income inequality. If income inequality is initially high, the tax rate increases so much in response to an increase in pre-tax income inequality that after-tax income inequality decreases, which decreases the relative status of the ethnic identities. On the other hand, if income inequality is initially low, after-tax income inequality increases and the ethnic identities becomes more attractive.

To see why this is the case, note that an increase in income inequality has two counteracting effects on the relative status of ethnic and poor identities. For given tax rates, higher pre-tax income inequality implies that the ethnic identities become more attractive for poor blacks and poor whites. On the other hand, for given identity choices, the tax rate increases as a response to higher inequality, which makes the poor identities more attractive. To see these two effects, note that the status difference between the ethnic and class identities is some population parameter times $(1-t)(y_R - y_P)$.[31]

[30] Proposition 4 is only stated for the case when poor whites are in minority since the case when poor whites are in majority follows directly once it is noted that the identity choices of poor blacks do not affect the equilibrium.

[31] One way to see this is to consider the conditions for ethnic identification for poor blacks and poor whites in the Appendix, i.e., A.1 and A.2.

Differentiating with respect to $y_R - y_P$ gives

$$\frac{\partial (1-t)(y_R - y_P)}{\partial (y_R - y_P)} = (1-t) - \frac{\partial t}{\partial (y_R - y_P)} (y_R - y_P) . \tag{5.1}$$

The first term in this expression is the direct effect of income inequality, whereas the second term is the effect through the increase in the tax rate. Since $\partial t / \partial (y_R - y_P)$ does not depend on $y_R - y_P$, this latter effect is stronger if income inequality is initially high.

The result that an increase in income inequality has ambiguous effects on redistribution fits well with recent empirical evidence showing no clear connection between income inequality and redistribution (see Perotti 1996 and Lind 2005). However, our model is not the first to produce this result. For example, in Corneo and Grüner (2000) the median voter prefers less redistribution as economic inequality increases, since the cost of taxation in terms of lost social prestige relative to the working class increases with economic inequality. The result that an increase in pre-tax income inequality might induce the poor to switch identity and thus favor lower taxes is also present in a slightly different flavor in Shayo (2007).

The model also allows us to study the effect of a change in income differences across ethnic groups. In Proposition 7 in Lindqvist and Östling (2007), we model interethnic income inequality as an increase in the number of poor blacks and a corresponding decrease in the number of poor whites, while the total number of poor and blacks is held constant.[32] The main prediction is that the level of redistribution falls as interethnic income inequality increases if the black minority group is small, but might increase if the black group is large. This results is partly in line with the theoretical and empirical results provided by Lind (2007), who argues that interethnic inequality reduces the support for redistribution.[33] The novel idea behind our result is that higher income inequality between ethnic groups might induce the poor of the majority group to switch to their ethnic identity in order to enjoy the higher status of the ethnic group.

6. Ethnic and Class Hostility

Apart from the size and economic position of minorities, there are probably also differences in hostility between ethnic groups across countries. For example, a country like the US, with its history of slavery and discriminatory laws, arguably has more

[32] Studying interethnic inequality in this way implies that the cognitive distances are affected, whereas these are unaffected by a change in standard income inequality. In a model with more than two income groups, interethnic inequality could instead be analyzed as income changes that wouldn't affect cognitive distances.

[33] Lind (2007) shows this theoretically in a model where people's altruism are targeted towards their own group. He also provides somewhat weak empirial support that between group inequality reduces the support for redistribution (using U.S. panel data from 1969 to 2000).

tense ethnic relationships than, say, Sweden. In terms of the model, ethnic tension imply that β is high and distances to class identities are large. In turn, it is more likely that individuals identify with their ethnic group, which implies lower taxes and less redistribution.

The model also allows us to study exogenous variation in hostility between social classes. If there are sufficiently weak tensions between rich and poor, i.e., if α is low, both poor whites and poor blacks identify with their ethnic group and in equilibrium the tax is low. Stronger class tension increases the distances to ethnic identities and for sufficiently high α we can be sure that poor blacks and poor whites identify as poor and the tax rate is the highest possible.

Although class hostility as captured by α might vary between countries, beliefs about the causes of poverty may also be important.[34] Alesina et al. (2001) and Fong (2001) show empirically that the belief that poverty is caused by laziness and not bad luck is a strong predictor of negative attitudes toward redistribution.[35] Arguably, if the poor believe that the rich have worked hard for their higher incomes, they are less likely to feel aversion toward the rich. Conversely, the rich would feel more sympathetic toward the poor if they thought that poverty was caused by bad luck instead of laziness. Therefore, such beliefs are not captured very well by the class conflict parameter α. However, if we reinterpret α as the strength of the belief that poverty is caused by laziness, a high α implies that the rich feel more distant to their ethnic identity and that the poor closer to their ethnic identity. To incorporate this in the model, we can replace α by $1/\alpha$ in the distance functions to the ethnic identity for poor blacks and poor whites. In this case, a high α tend to push the rich toward class identification, whereas the poor are pushed toward ethnic identification. This provides a simple argument for why beliefs about the causes of poverty may matter for redistribution. Strong beliefs that poverty is caused by laziness make it more likely that the poor identify with their ethnic group, which in turn imply low taxes and little redistribution (compared to the case with class identification). Such beliefs are of course likely to directly affect preferences for redistribution, but the possibility of identity shifts demonstrates an extra channel through which those beliefs may lead to lower redistribution.

[34] See for example Piketty (1995), Alesina and Angeletos (2005) and Bénabou and Tirole (2006) for models where such beliefs are endogenously determined.

[35] Gilens (1999, p. 172–173) develops a similar argument: "The belief that black Americans lack commitment to the work ethic is central to whites' opposition to welfare. But it appears that this race-based opposition does not primarily reflect either a general racial animosity or an effort to defend whites' concrete group interests. Rather, the racial component of white opposition to welfare seems to reflect the most important nonracial basis of welfare opposition: the perception that welfare recipients are undeserving."

7. American Exceptionalism

Why is redistribution so much lower in the US compared to Western Europe? In terms of our model, the US population with European origin is represented by whites, whereas the Afro-American population is represented by blacks.[36] In Western Europe, white refers to the native population and black refers to non-European immigrants.

First, pre-tax income inequality is higher in the US than in Western Europe. On the one hand, the tax preferred by the poor is increasing in income inequality for given identities. On the other hand, if income inequality in the US and Western Europe is lower than the threshold in Proposition 4, then the higher level of income inequality in the US will make the poor more likely to identify with their ethnic group. Hence, the effect on redistribution from the higher pre-tax income inequality in the US is ambiguous.

Second, the higher degree of ethnic diversity in the US may imply that poor whites in the US are more likely to identify as white and favor a low level of redistribution (under the conditions given in Proposition 5 of Lindqvist and Östling 2007). Similarly, to the extent that interethnic income inequality is higher in the US, Proposition 7 in Lindqvist and Östling (2007) shows that this might be an additional force in the same direction. Moreover, the preferred tax rate of poor whites when they identify themselves as white is decreasing in the affluence of whites. Hence, to the extent that whites in the US are more wealthy than their counterparts in Europe, poor whites in the US who identify as white favor a lower tax rate than poor whites in Europe identifying as white, holding everything else constant.

Third, Americans are much more prone than Europeans to believe that the poor are lazy rather than unlucky. In addition, the US has historically a more troubled racial relationship than most European countries. Both of these differences suggest that the poor whites should be more likely to identify as white in the US.

These different explanations do not yield an unambiguous prediction, but, given the argument above, it indeed seems plausible that poor white Americans should be more likely to identify as white and favor low taxes than poor white Europeans. It should be noted, however, that the difference in redistribution between the US and Western Europe could also be rationalized in terms of multiple equilibria.[37] Even if the US and Western Europe were identical in terms fundamentals (i.e., parameter values), it could

[36] There are of course many other ethnic groups in the US, but we focus on blacks and whites. This has also been the focus in the literature on racial issues in the US, with Myrdal (1944) as the classic reference. Loury (2002) provides a more recent account on racial stigmatization in the US.

[37] Several other economists have also argued that differences in the level of redistribution across countries can be understood in terms of multiple equilibria. See Alesina and Angeletis (2005) and Bénabou and Tirole (2006) for two recent contributions.

be the case that poor whites in Europe identify as poor simply because redistribution is relatively high – if redistribution had been at US levels they would have switched to ethnic identities and supported lower taxes.

We are not the first to raise the argument that ethnic diversity is important in explaining differences in redistribution between the US and Europe. Shayo (2007) argues that a high degree of ethnic diversity concentrated to the poorer segments of society induces the poor to identify with their nation instead of their class, thereby reducing the support for redistribution. Alesina et al. (2001) claim that differences in beliefs about the poor and ethnic heterogeneity explains the comparably low level of redistribution in the US through its impact on altruism. However, since altruism is itself an exogenous parameter in their theoretical framework, they do not explicitly model how these factors explain altruism. Moreover, Alesina et al. (2001) consider altruism to be nondiscriminatory across groups, whereas altruism in our model is only directed at a particular subgroup of the population. Lind (2007) studies such directed altruism, but unlike our approach the decision to sympathize with a particular subgroup is not endogenous in his model.

8. Concluding Remarks

The model presented in this paper treats social categories as given despite the fact that such social constructs typically are not unchanged in the longer term (see Alesina and La Ferrara 2005*a* and Posner 2005 for discussion and examples). An interesting and challenging task for future research would therefore be to make social identities, and not only social identification, endogenous. Relatedly, groups may have incentives to manipulate the identity choices of others. For example, the rich might try to reduce the level of redistribution by convincing the poor to identify with their ethnic group. This could be done by directly influencing their identity choice through propaganda, or, more indirectly, by trying to create new ethnic categories.

Appendix: Proofs

A.1. Proof of Proposition 1. We know that $l_i^*(t)$ is non-decreasing in t and that $t_i^*(l_i)$ is non-decreasing in l_i. Consequently, $t_i^*(l_i^*(t))$ is non-decreasing in t, which implies that the median tax is non-decreasing in t. Equilibrium tax rates are given by the fixed points of $\Gamma(\times_{i \in N} t_i^*(l_i^*(t)))$. Note that this is a non-decreasing mapping $\Gamma(t) : [0,1] \rightarrow [0,1]$. This mapping will typically not be continuous, but since it is a non-decreasing mapping from the unit interval to the unit interval, Tarski's fixed point theorem implies that there is at least one fixed point of $\Gamma(t)$ (see Theorem M.I.3 in Mas-Colell et al. 1995).

A.2. Proof of Proposition 2. It is clear from (2.2) that the poor always prefer positive tax rates, the rich prefer zero taxes, and that preferred tax rates for given identity choices is unchanged. Since the poor are in majority, the median voter(s) is poor. The condition for ethnic identification (2.3) for the poor only differs in the cognitive distances to the poor identity. Letting p_j denote the proportion of the population belonging to ethnic group j, we can re-write the distance to the poor identity for a poor person in this ethnic group as

$$d_{PjP} = d(\beta(1 - p_j)).$$

In other words, the higher p_j is, the lower is the distance to the poor identity and the more likely is identification with the poor. This implies that there is a $\widehat{p}$ such that the poor in ethnic groups larger than $\widehat{p}$ identify with the poor, and the poor in smaller ethnic groups identify with their ethnic group. It may also be the case that all ethnic groups are above or below this threshold. Increasing the size of one ethnic group or adding a new ethnic group implies that the other ethnic groups shrink in size, which might induce them to shift to an ethnic identity.

Suppose first that the high tax rate prevails in which sufficiently many of the poor identify as poor. Increasing the size of an ethnic group or adding a new ethnic group could then lead some of the poor to switch to ethnic identities, which may result in the low equilibrium tax rate. (The new low equilibrium tax rate may in turn induce the poor in other ethnic groups to switch to ethnic identities.)

Now suppose instead the sufficiently many poor initially identified with their ethnic groups so that the low tax rate preferred by poor identifying with their ethnic group prevails. Enlarging or adding one ethnic group may then induce some of the poor in the other groups to switch to ethnic identities. However, we want to rule out that the enlarged or new group isn't so large that the poor in that group identify with the poor. This cannot happen if the group is smaller than $\widehat{p}$.

A.3. Proof of Lemma 1. The result follows directly from (3.1) and (3.2) once it is noted that $y_P < y_W < y_R$ and $y_P < y_B < y_R$ for the first part, and $y_W > y_B$ for the second.

A.4. Identity Inequalities: Black and White model. The condition (2.3) for high status identification can be rewritten as follows for poor blacks and poor whites:

$$PB : (1-t)\frac{p_{RB}}{p_{PB}+p_{RB}}(y_R - y_P) > \frac{\delta}{\gamma}\left(\alpha\frac{p_{RB}}{p_{PB}+p_{RB}} - \beta\frac{p_{PW}}{p_{PB}+p_{PW}}\right), \tag{A.1}$$

$$PW : (1-t)\frac{p_{RW}}{p_{PW}+p_{RW}}(y_R - y_P) > \frac{\delta}{\gamma}\left(\alpha\frac{p_{RW}}{p_{PW}+p_{RW}} - \beta\frac{p_{PB}}{p_{PB}+p_{PW}}\right). \tag{A.2}$$

For several of the proofs it is useful to rewrite the two inequalities (A.1) and (A.2) as follows by dividing by $p_{RB}/(p_{PB}+p_{RB})$ and $p_{RW}/(p_{PW}+p_{RW})$, respectively:

$$PB : \frac{p_{PW}}{p_{PB}+p_{PW}}\frac{p_{PB}+p_{RB}}{p_{RB}} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)(y_R - y_P)\right), \tag{A.3}$$

$$PW : \frac{p_{PB}}{p_{PB}+p_{PW}}\frac{p_{PW}+p_{RW}}{p_{RW}} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)(y_R - y_P)\right). \tag{A.4}$$

A.5. Proof of Lemma 2. We want show that the left hand side of (A.3) is larger than the left hand side of (A.4). Rearranging this condition we get

$$p_{PW}\frac{p_{RW}}{p_{PW}+p_{RW}} > p_{PB}\frac{p_{RB}}{p_{PB}+p_{RB}}.$$

If white and black have the same average income, we only need to show that $p_{PW} > p_{PB}$. This follows from the fact that whites are in majority and that blacks and whites have the same average income (to see this, divide $p_{RW}+p_{PW} > p_{PB}+p_{RB}$ by p_{PB} and substitute $p_{RB}/p_{PB} = p_{RW}/p_{PW}$). Now suppose that whites are on average richer than blacks. Rewriting the above condition, we get

$$\frac{p_{PW}}{p_{PW}+p_{RW}}\frac{p_{RW}}{p_{PW}+p_{RW}}(p_{PW}+p_{RW}) > \frac{p_{PB}}{p_{PB}+p_{RB}}\frac{p_{RB}}{p_{PB}+p_{RB}}(p_{PB}+p_{RB}).$$

We know that $p_{PW}+p_{RW} > p_{PB}+p_{RB}$ so it is sufficient to show that

$$\frac{p_{PW}}{p_{PW}+p_{RW}}\frac{p_{RW}}{p_{PW}+p_{RW}} > \frac{p_{PB}}{p_{PB}+p_{RB}}\frac{p_{RB}}{p_{PB}+p_{RB}}.$$

Using that $p_{RW}/(p_{PW}+p_{RW}) = 1 - p_{PW}/(p_{PW}+p_{RW})$ (and similarly for blacks) we can rewrite this condition as

$$\left(\frac{p_{PB}}{p_{PB}+p_{RB}}\right)^2 - \left(\frac{p_{PW}}{p_{PW}+p_{RW}}\right)^2 > \frac{p_{PB}}{p_{PB}+p_{RB}} - \frac{p_{PW}}{p_{PW}+p_{RW}}.$$

Since whites are richer than blacks, both the left and right hand sides of this expression are positive. The left hand side can be rewritten as

$$\left(\frac{p_{PB}}{p_{PB}+p_{RB}}+\frac{p_{PW}}{p_{PW}+p_{RW}}\right)\left(\frac{p_{PB}}{p_{PB}+p_{RB}}-\frac{p_{PW}}{p_{PW}+p_{RW}}\right)>\left(\frac{p_{PB}}{p_{PB}+p_{RB}}-\frac{p_{PW}}{p_{PW}+p_{RW}}\right).$$

Since blacks are on average poorer, the right hand side is positive and we can therefore divide both sides by the expression on the right hand side so that the condition simplifies to

$$\left(\frac{p_{PB}}{p_{PB}+p_{RB}}+\frac{p_{PW}}{p_{PW}+p_{RW}}\right)>1.$$

Since $p_{PB}/\left(p_{PB}+p_{RB}\right)>p_{PW}/\left(p_{PW}+p_{RW}\right)$, we can write $p_{PB}/\left(p_{PB}+p_{RB}\right)=p_{PW}/\left(p_{PW}+p_{RW}\right)+\varepsilon$ for some $\varepsilon>0$. Hence, we need to show that $p_{PW}/\left(p_{PW}+p_{RW}\right)>\left(1-\varepsilon\right)/2$. The assumption that poor are in majority implies

$$\frac{p_{PB}}{p_{PB}+p_{RB}}\left(p_{PB}+p_{RB}\right)+\frac{p_{PW}}{p_{PW}+p_{RW}}\left(p_{PW}+p_{RW}\right)>\frac{1}{2}.$$

Substituting $p_{PB}/\left(p_{PB}+p_{RB}\right)$ we can write this condition as

$$\left(\frac{p_{PW}}{p_{PW}+p_{RW}}+\varepsilon\right)\left(p_{PB}+p_{RB}\right)+\frac{p_{PW}}{p_{PW}+p_{RW}}\left(1-p_{PB}-p_{RB}\right)>\frac{1}{2},$$

which can be further rewritten as

$$\frac{p_{PW}}{p_{PW}+p_{RW}}>\frac{1-2\left(p_{PB}+p_{RB}\right)\varepsilon}{2}.$$

Since $\left(p_{PB}+p_{RB}\right)<1/2$, we have $\left(1-2\left(p_{PB}+p_{RB}\right)\varepsilon\right)/2>\left(1-\varepsilon\right)/2$, and we have therefore shown what we needed to show.

A.6. Proof of Lemma 3. When whites and blacks are equally rich, we can see from (3.1) and (3.2) that the rich will always prefer zero taxes irrespective of how they identify themselves. Poor identifying with their ethnic identities will prefer the tax rate $1-\left(y_P+\gamma y\right)/\left(1+\gamma\right)y$ whereas poor identifying with their class will prefer the tax rate $1-y_P/y$. Since the poor are in majority, the median tax rate will be either of these two tax rates. The remainder of the result follows directly from the assumptions that the poor are in majority, blacks are in minority and poor whites are in minority.

A.7. Proof of Proposition 3. Let p_B denote the proportion of poor and p_P the proportion of rich. Since whites and blacks are equally rich, $p_{PB}=p_Pp_B$, $p_{RB}=\left(1-p_P\right)p_B$, $p_{PW}=p_P\left(1-p_B\right)$ and $p_{RW}=\left(1-p_P\right)\left(1-p_B\right)$. Using these relations the two conditions for ethnic identification (A.3) and (A.4) can be rewritten as

$$PB : \frac{1-p_B}{1-p_P} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)(y_R - y_P)\right),$$

$$PW : \frac{p_B}{1-p_P} > \frac{1}{\beta}\left(\alpha - \frac{\gamma}{\delta}(1-t)(y_R - y_P)\right).$$

Since whites and blacks both have the average income y and the average income of both ethnic groups is unchanged, the median tax rates for given identities and the status of different identities remain unchanged.

First suppose that poor whites initially identify as white so that the tax rate is $t^*_{PWW} = t^*_{PBB}$ (according to Lemma 3). An increase of p_B implies that the left hand side of the identity choice inequality for poor whites increases which in turn implies that identity choices and hence the tax rate will remain unchanged. This is true both irrespective of whether poor whites are in majority or not.

Now suppose instead that poor whites are in minority and that both poor blacks and poor whites identify as poor so that the equilibrium tax rate is $t^*_{PWP} = t^*_{PBP}$. An increase in p_B will then decrease the left hand side of the identity choice inequality for poor blacks which implies that they will not change their identity. Since the right hand side is the same for both poor blacks and poor whites and $p_B < 1/2$, poor whites will never identify as white unless the poor black identify as black and so the tax rate remains unchanged.

Finally, suppose that poor whites are in minority, poor whites identify as poor and poor blacks identify as black so that the tax rate is $t^*_{PWW} = t^*_{PBB}$. Increasing p_B might then induce the poor blacks to switch to the poor identity so that the tax rate will be $t^*_{PWP} = t^*_{PBP}$. Alternatively, poor whites may switch to the white identity, but that would leave the tax rate unaffected (recall that Lemma 2 implies that not both black and white can switch identities in this case). If instead poor whites are in majority and initially identify as poor, the tax rate is $t^*_{PWP} = t^*_{PBP}$ and an increase in p_B might induce them to shift to the white identity, resulting in the low tax rate $t^*_{PWW} = t^*_{PBB}$.

A.8. Proof of Proposition 4. As income inequality changes, the only thing that changes in the conditions for ethnic identification (A.1) and (A.2) is the term $(1-t)(y_R - y_P)$ which reflects the relative status of the ethnic identity over the poor identity. Since blacks and whites are equally rich, there are only two tax rates in equilibrium: $t^*_{PWW} = t^*_{PBB} = 1 - (y_P + \gamma y)/(1+\gamma)y$ and $t^*_{PWP} = t^*_{PBP} = 1 - y_P/y$. Clearly, both these tax rates increase with income inequality for given identity choices. Since both the tax rate and income inequality increase, the effect on relative status $(1-t)(y_R - y_P)$ is ambiguous.

Since we keep average income constant, i.e., $\partial y/\partial\left(y_R - y_P\right) = 0$, it must hold that

$$\frac{\partial y_P}{\partial\left(y_R - y_P\right)} = -\frac{1-p_P}{p_P}\frac{\partial y_R}{\partial\left(y_R - y_P\right)}.$$

It is the case that $\partial\left(y_R - y_P\right)/\partial\left(y_R - y_P\right) = 1$ and this implies that

$$\frac{\partial y_R}{\partial\left(y_R - y_P\right)} - \frac{\partial y_P}{\partial\left(y_R - y_P\right)} = 1.$$

Combining these two observations we get

$$\frac{\partial y_P}{\partial\left(y_R - y_P\right)} = -\left(1-p_P\right) \text{ and } \frac{\partial y_R}{\partial\left(y_R - y_P\right)} = p_P.$$

Now consider the case when poor blacks identify as black (and perhaps poor whites identify as white). For given identity choices, the tax rate increases. If poor blacks switch to the poor identity, then the tax rate increases further. To see if this can happen, note that the effect on relative status of higher income inequality is given by

$$\begin{aligned}\frac{\partial\left(1-t^*_{PBB}\right)\left(y_R - y_P\right)}{\partial\left(y_R - y_P\right)} &= \left(1-t^*_{PBB}\right) - \frac{\partial t^*_{PBB}}{\partial\left(y_R - y_P\right)}\left(y_R - y_P\right) \\ &= \frac{y_P + \gamma y}{\left(1+\gamma\right)y} - \frac{1-p_P}{\left(1+\gamma\right)y}\left(y_R - y_P\right).\end{aligned}$$

Rearranging the latter expression shows that relative status is increasing in income inequality if and only if $y_P/y > \left(1-\gamma\right)/2$. If this condition is satisfied, then the identity choice of poor blacks remain unchanged. Otherwise, poor blacks might switch to the poor identity. From Lemma 2 we also know that if poor blacks switch to the poor identity, then poor whites will switch to the poor identity as well (unless they already identified as poor)

Now consider the case when both poor blacks and poor whites identify as poor. In this case, the effect on relative status is given by

$$\begin{aligned}\frac{\partial\left(1-t^*_{PBP}\right)\left(y_R - y_P\right)}{\partial\left(y_R - y_P\right)} &= \left(1-t^*_{PBP}\right) - \frac{\partial t^*_{PBP}}{\partial\left(y_R - y_P\right)}\left(y_R - y_P\right) \\ &= \frac{y_P}{y} - \frac{1}{y}\left(1-p_P\right)\left(y_R - y_P\right).\end{aligned}$$

This is increasing if $y_P/y > 1/2$. Note that if this condition holds, then relative status at the ethnic tax rate is increasing for all values of γ. If $y_P/y > 1/2$, then poor blacks (and possibly poor whites) might switch to ethnic identities, implying that the net effect on the tax rate is ambiguous. However, if $y_P/y < 1/2$, identity choices remain unchanged and the tax rate will increase.

References

Akerlof, G. A. and Kranton, R. E. (2000), "Economics and Identity", *Quarterly Journal of Economics* 115(3), 715–753.

Alesina, A. and Angeletos, G.-M. (2005), "Fairness and Redistribution", *American Economic Review* 95(4), 960–980.

Alesina, A., Baqir, R. and Easterly, W. (1999), "Public Goods and Ethnic Divisions", *Quarterly Journal of Economics* 115(4), 1167–1199.

Alesina, A., Glaeser, E. and Sacerdote, B. (2001), "Why Doesn't the United States Have a European-Style Welfare State?", *Brookings Paper on Economics Activity* 2001(2), 187–254.

Alesina, A. and La Ferrara, E. (2005*a*), "Ethnic Diversity and Economic Performance", *Journal of Economic Literature* 43(3), 721–761.

Alesina, A. and La Ferrara, E. (2005*b*), "Preferences for Redistribution in the Land of Opportunities", *Journal of Public Economics* 89(5-6), 897–931.

Austen-Smith, D. and Wallerstein, M. (2006), "Redistribution and Affirmative Action", *Journal of Public Economics* 90, 1789–1823.

Bénabou, R. and Tirole, J. (2006), "Belief in a Just World and Redistributive Politics", *Quarterly Journal of Economics* 121(2), 699–746.

Betz, H.-G. (1993), "The New Politics of Resentment: Radical Right-Wing Populist Parties in Western Europe", *Comparative Politics* 25(4), 413–427.

Bisin, A., Patacchini, E., Verdier, T. and Zenou, Y. (2006), "'Bend It Like Beckham': Identity, Socialization and Assimilation", CEPR Discussion Paper No. 5662.

Corneo, G. and Grüner, H. P. (2000), "Social Limits to Redistribution", *American Economic Review* 90(6), 1491–1507.

Davis, J. A., Smith, T. W. and Marsden, P. V. (2005), *General Social Surveys 1972-2004*, ICPSR04295-v1, 2005-09-02, Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

Dincer, O. and Lambert, P. J. (2006), “Taking care of your own: Ethnic and religious heterogeneity and income inequality”, *University of Oregon Economics Department Working Paper*.

Fernández, R. and Levy, G. (2008), “Diversity and Redistribution”, *Journal of Public Economics*, forthcoming.

Fong, C. (2001), “Social Preferences, Self-Interest, and the Demand for Redistribution”, *Journal of Public Economics* 82(2), 225–246.

Fryer, R. G. and Torelli, P. (2006), “An Empirical Analysis of ‘Acting White’”, mimeo, Harvard University.

Gilens, M. (1999), *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*, University of Chicago Press, Chicago.

Glazer, N. and Moynihan, D. P. (1970), *Beyond the Melting Pot: The Negroes, Puerto Ricans, Jews, Italians, and Irish of New York City*, 2 edn, MIT Press, Boston.

Golder, M. (2003), “Explaining Variation in the Success of Extreme Right Parties in Western Europe”, *Comparative Political Studies* 36(4), 432–466.

Habyarimana, J., Humphreys, M., Posner, D. N. and Weinstein, J. (2007), “Why Does Ethnic Diversity Undermine Public Goods Provision? An Experimental Approach”, *American Political Science Review* 101, 709–725.

Kasher, A. and Rubinstein, A. (1997), “On the Question ‘Who is a J?’: A Social Choice Approach”, *Logique et Analyse* 160, 385–395.

Klor, E. F. and Shayo, M. (2007), “Social Identity and Preferences over Redistribution”, Mimeo, Hebrew University.

Knigge, P. (1998), “The Ecological Correlates of Right-wing Extremism in Western Europe”, *European Journal of Political Research* 34(2), 249–279.

Lind, J. T. (2005), “Why is there so little redistribution?”, *Nordic Journal of Political Economy* 31, 111–125.

Lind, J. T. (2007), “Fractionalization and the size of goverment”, *Journal of Public Economics* 91, 51–76.

Lindqvist, E. and Östling, R. (2007), "Identity and Redistribution", *SSE/EFI Working Paper Series in Economics and Finance* No 659.

Lipset, S. and Marks, G. (2000), *It didn't happen here: Why socialism failed in the United States*, W. W. Norton & Co, New York.

Loury, G. (2002), *The Anatomy of Racial Inequality*, Harvard University Press, Cambridge.

Luttmer, E. F. P. (2001), "Group Loyalty and the Taste for Redistribution", *Journal of Political Economy* 109(3), 500–528.

Mas-Colell, A., Whinston, M. D. and Green, J. R. (1995), *Microeconomic Theory*, Oxford University Press, New York.

McGann, A. J. and Kitschelt, H. (2005), "The Radical Right in the Alps", *Party Politics* 11(2), 147–171.

Meltzer, A. H. M. and Richard, S. F. (1981), "A Rational Theory of the Size of Government", *Journal of Political Economy* 89(5), 914–927.

Miguel, E. and Posner, D. N. (2006), "Sources of Ethnic Identification in Africa", mimeo.

Myrdal, G. (1944), *An American Dilemma. The Negro Problem and Modern Democracy*, Harper & Brothers Publishers, New York and London.

Nieuwbeerta, P. and Ultee, W. (1999), "Class Voting in Western Industrialized Countries, 1945-1990: Systematizing and Testing Explanations", *European Journal of Political Research* 35, 123–160.

Orr, L. L. (1976), "Income Transfers as a Public Good: An Application to AFDC", *American Economic Review* 66(3), 359–371.

Perotti, R. (1996), "Growth, Income Distribution and Democracy: What the Data Say", *Journal of Economic Growth* 1(2), 149–187.

Piketty, T. (1995), "Social Mobility and Redistributive Politics", *Quarterly Journal of Economics* 110(3), 551–584.

Poglia Mileti, F., Tondolo, R., Plomb, F., Schultheis, F., Meyer, M.-H., Hentges, G., Flecker, J. and Mairhuber, I. (2002), “Modern Sirens and their Populist Songs: A European Literature Review on Changes in Working Life and the Rise of Right-wing Populism”, Report from the SIREN project [http://www.siren.at].

Posner, D. N. (2005), *Institutions and Ethnic Politics in Africa*, Cambridge University Press, New York.

Roccas, S. (2003), “The Effects of Status on Identification with Multiple Groups”, *European Journal of Social Psychology* 33, 351–366.

Roemer, J. E. (1998), “Why the poor do not expropriate the rich: an old argument in new garb”, *Journal of Public Economics* 70, 399–424.

Scheve, K. F. and Slaughter, M. J. (2001), “Labor Market Competition and Individual Preferences over Immigration Policy”, *Review of Economics and Statistics* 83(1), 133–145.

Shayo, M. (2007), “A Theory of Social Identity with an Application to Redistribution”, mimeo, Hebrew University.

Soss, J., Schram, S., Vartanian, T. and O’Brien, E. (2001), “Setting the Terms of Relief: Political Explanations for State Policy Choices in the Devolution Revolution”, *American Journal of Political Science* 45(2), 378–395.

Spiegel (2005), “It’s stupid to be afraid”, *Spiegel Special International Edition* 2005(7), 21–23.

Waldman, B. (1987), “Mechanisms of kin recognition”, *Journal of Theoretical Biology*, 128, 159–185.

PAPER 5

Political Polarization and the Size of Government

with Erik Lindqvist

Abstract. We study the effect of political polarization on government spending and redistribution using the dispersion of self-reported political preferences as our measure of polarization. Politically polarized countries have lower levels of redistribution and government consumption. The relationship between political polarization and the size of government is stronger in democratic countries, indicating that the effect goes through the political system. The results are robust to a large set of control variables, including GDP per capita and income inequality.

1. Introduction

In the canonical model of redistribution (e.g., Meltzer and Richard, 1981), the preferences of the median voter determines the size of government. This view can be challengened on both theoretical and empirical grounds. In richer politico-economic models, polarization of voter preferences leads to policies that involve both lower spending on public goods (Alesina et al, 1999) and a lower degree of redistribution (Fernandez and Levy, 2008). There is substantial empirical evidence (e.g., Gerber and Lewis, 2004) that politicians implement policies further away from the median voter the more heterogeneous is the electorate, implying that voters in polarized electorates may rationally prefer a smaller government. In this paper we perform the first direct test of the link between polarization of political preferences and the size of government.

We derive country-level measures of polarization of political preferences from responses to multiple choice questions about economic policy and find that these measures are strong predictors of the size of government. Countries that are more polarized have governments that both consume and redistribute less. The implied relationship

We would like to thank Alberto Alesina, Martin Bech Holte, Roberta Dessi, Mikael Elinder, Tore Ellingsen, Justina A. V. Fischer, Christina Fong, Ola Granström, Magnus Johannesson, Henrik Jordahl, Ethan Kaplan, Torsten Persson, Daniel N. Posner, Paul Seabright, Jean Tirole, Björn Tyrefors and seminar participants at the ENTER Jamboree 2007 in Mannheim, the Research Institute for Industrial Economics (Stockholm), the 2006 Ratio Colloquium for Young Social Scientists, Stockholm School of Economics and Uppsala University for encouragement and valuable comments. Both authors gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Foundation.

is strong. For example, consider one of our measures of political polarization based on attitudes toward private ownership of business. Controlling for the mean value of preferences, a one standard deviation increase in this measure predicts a decrease in transfers and subsidies as share of GDP by 2 percentage points. We also find that the relationship between polarization and government size is stronger in democratic countries, which is consistent with theoretical explanations that focus on the political system as the link between polarization and the size of government. For democratic countries, the relationship between polarization and government size is robust to controlling for a large set of variables that have been found by the previous literature to predict the size of government. The results are also robust for using other measures of polarization than the standard deviation.

The primary concern with a causal interpretation of our results is that political polarization might be endogenous with respect to the size of government. For example, government spending might reduce pre-tax income inequality which reduces distributional conflict. Though endogeneity cannot be ruled out completely, the results for democratic countries remain robust when we control for income inequality. Moreover, polarization of political preferences does not seem to be the result of divisive economic policies, an unwillingness to give extreme answers in certain countries, or higher uncertainty about how to answer survey questions in developing countries.

One related finding is worth mentioning. Though the effect of political polarization is robust to controlling for ethnic fractionalization, the estimated effect of ethnic fractionalization on redistribution is sensitive to political polarization. This indicates that political polarization is one of the channels through which ethnic fractionalization affects government size, which is consistent with the view put forward by Alesina et al. (1999), Bandiera and Levy (2007) and Fernández and Levy (2008).

We describe the related theoretical and empirical literature in the next section. In Section 3, we discuss the data and how we measure polarization. We test the effect of political polarization on the size of government and investigate a number of alternative explanations in Section 4. Section 5 focuses on polarization in other domains than economic policy and Section 6 concludes the paper. Data sources and definitions of variables are provided in the Appendix. All empirical results not provided in the text are available online in a Supplementary Appendix.[1]

[1] The Supplementary Appendix is available at: http://swopec.hhs.se/hastef/abs/hastef0628.htm.

2. Related Literature

2.1. Political mechanisms. Economic theory and empirical work in political economics suggest a number of mechanisms through which polarization may affect redistribution and the provision of public goods.

Alesina et al. (1999) develop a model of two-stage voting where the size of the budget is decided before its composition. In the model, citizens first vote for the amount to spend on a public good and then on the type of public good to provide. As voters in the first stage anticipate the outcome of the second stage, support for spending on the public good is decreasing in the dispersion of preferences over its type.

Fernández and Levy (2008) study how preference diversity among the poor affects their ability to extract resources from the rich in a model of endogenous party formation. In their model, each taste is represented by an interest group. As taste diversity increases from a low level, redistribution becomes more and more tilted toward special interest groups and general redistribution to the poor goes down. However, at a certain threshold, all special interest coalitions break down and general redistribution to the poor increases.

Alesina et al. (1999) and Fernández and Levy (2008) both focus on the effects of preference diversity through the political system in democratic countries. Whereas the prediction of Alesina et al. (1999) is unambigous, Fernández and Levy's (2008) model predict a U-shaped relationship between preference heterogeneity and general redistribution, but also an inverse U-shaped relationship between polarization and redistribution targeted to specific groups.

Another mechanism by which political polarization may affect the size of government is through legislator shirking. Politicians facing a polarized electorate may have more discretion how to spend public funds (for example because it is more difficult for voters to form interest groups that monitor politicians), implying that voters rationally prefer a smaller government. There is substantial empirical evidence that politicians deviate more from the opinions of the median voter in polarized electorates. Kalt and Zupan (1990) ran a model of senators' voting records on a standard set of control variables. They found that the absolute value of the residuals from this model were lower in states that voted in a more consistently liberal or conservative way in the 1972 presidential election. Goff and Grier (1993) found that same-state senators had more different voting records the higher was heterogeneity with respect to the distribution of income, ethnicity and workforce composition.[2] Bailey and Brady (1998) found that

[2] As pointed out by Goff and Grier (1993), the differences in same-state senators voting records need not reflect legislator shirking. Another potential explanation for this result is that senators appeal to different constituencies.

the relationship between constituency characteristics and senator's voting records on free trade was less predictable for states that scored high on measures of socioeconomic and cultural diversity. Gerber and Lewis (2004) calculated a measure of district heterogeneity in voter preferences from voting records of binary legislative proposals in California. They found that legislators were more constrained by the preferences of the median voter in homogeneous districts.

There are a few empirical studies that consider the effect of electorate heterogeneity on the size of government, but they all use indirect measures of political preferences.[3] In particular, ethnic diversity has been shown to be associated with lower levels of redistribution and smaller government (Alesina et al, 1999; Alesina et al., 2001; Luttmer, 2001, and Vigdor, 2004).[4] To the best of our knowledge, our paper provides the first investigation of the effect of political polarization on the size of government using a measure of polarization based on self-reported opinions.[5]

2.2. Other mechanisms. Polarization might affect government size in at least two other ways than the political mechanisms suggested above.

First, polarization of voters could have a direct effect on political preferences, which in turn affect government size. For example, research in social psychology have shown that attitudinal similarity is a strong predictor of altruism, attraction and friendship, see Newcomb (1961); Byrne (1961, 1971); Suedfeld et al. (1972); Batson et al. (1981); McGrath (1984); Feren et al. (1988) and Chen and Kendrick (2002). Polarization of preferences might therefore lead to lower levels of altruism, which in turn decreases the preference for redistribution and public goods provision.[6] Relatedly, research in experimental economics has shown that heterogeneity in preferences leads to lower provision of public goods (Ledyard, 1995).

[3] Several studies have focused on polarization of political parties and economic outcomes. See, for example, Cukierman et al. (1992); Svensson (1998); Frye (2002) and Alt and Dreyer Lassen (2006).

[4] There are two types of explanations for the negative association between size of government and ethnic diversity put forward in the previous literature. One argument is that ethnic diversity is related to preference heterogeneity (Alesina et al, 1999; Bandiera and Levy, 2007; Fernandez and Levy, 2008). The second argument is that people are less altruistic toward people from different ethnic groups.

[5] We are not the first, however, to study the dispersion of responses to multiple choice questions at the country level. DiMaggio et al. (1996) use survey data to study changes in dispersion and polarization of attitudes over time. Au (1999) and Au and Cheung (2004) study how variation in job characteristics affects certain social outcomes, such as job satisfaction. There is also a literature within social psychology on the determinants of value consensus (e.g. Shalom and Sagie, 2000).

[6] The relationship between attitudinal similarity and altruism can be rationalized in terms of evolutionary psychology, see Hamilton (1964); Tesser (1993); Olson et al. (2001) and Park and Schaller (2005).

Second, polarization could increase transaction costs and have a negative effect on national income, which in turn may affect government size. Gradstein and Justman (2002) theoretically study the relationship between social cohesion and economic growth. They argue that schooling choices of parents affect future productivity since large social distance between transacting agents reduces productivity. This argument is supported by research in management studies which has shown that groups where members hold similar values have fewer intragroup conflicts and better performance (Jehn et al., 1997; Jehn et al., 1999; Jehn and Mannix, 2001), more conversation exchanges (Oetzel, 1998), and a higher degree of social integration (Harrison et al., 1998).

2.3. Determinants. As this paper studies the effect of political polarization on the size of government, we are concerned that political polarization is itself determined by the size of government. In particular, the standard economic theory of redistribution (e.g. Meltzer and Richard 1981) assumes that voters' preferences for redistribution merely reflect their relative position in the distribution of income. As a result, political polarization on issues of redistribution is assumed to be increasing in income inequality before taxes. Hence, policies that affect the distribution of human capital in the population – such as public schooling for the poor – may have long-term effects on the level of political polarization. Another form of reverse causality occurs if large governments invest more in policies that foster homogeneity. It is not clear, however, exactly what these policies are. For example, there is no indication in our data that political polarization is lower in countries with censorship and lack of a free press.[7]

There are also sources for variation in political polarization which are exogenous with respect to the size of government. For example, Bernhardt et al. (2008) develop a model where media firms maximize profits by catering to partisan audiences by suppressing information that their audiences do not want to hear. As each individual voter has a negible probability of affecting the electoral outcome, citizens' care more about the entertainment value of news than its informational content. Hence, the consumption value of biased news may be higher than for unbiased news. This could give rise to electoral inefficiencies even if voters update rationally from the information they get. That media bias actually affects voting behavior has been shown empirically by DellaVigna and Kaplan (2007) and Gerber et al. (2006).

Political polarization could also be caused by a divergence in beliefs about the effects of different policies rather than by a conflict of interest or partisanship. For example,

[7] The partial correlations between our measures of polarization in economic policy (presented in the next section) and Reporters without borders (2002, 2003) index of press freedom ranges from .18 and .33, where a higher score indicates less freedom.

Dixit and Weibull (2007) show that if people have heterogeneous priors, agents interpret signals about the effectiveness of various policies differently and political polarization result after the realization of bad outcomes. Whereas the setup in Dixit and Weibull (2007) implies that people's beliefs will converge in the long run, Acemoglu et al. (2007) develop a related model where polarization of beliefs might be persistent. As attitudes are partly heritable, political polarization may also be determined by the genetic variation within a society.[8]

3. Data and Measurement

We calculate our measures of polarization from responses to multiple-choice questions in the World Values Survey (WVS) 1999-2002 (European Values Study Group and World Values Association, 2004).[9] The WVS is based on face-to-face interviews with between 417 and 6,025 respondents in 81 different countries. For most countries, the WVS contains data from about 1,000 respondents. We excluded some countries (Bosnia and Hercegovina, Montenegro, Northern Ireland, Puerto Rico and Serbia) since they were not included as separate entities in the other data sources and some other countries (Armenia, Azerbaijan, Belarus and Moldova) because government size data was unavailable. For each question in the WVS, we calculated the mean and standard deviation of the responses for each country. Thereafter, we excluded all variables with binary responses, fewer than 55 observations or a correlation between the country mean and standard deviation with an absolute value of 0.5 or more. We also excluded constructed indexes, questions about personal characteristics (income, age, etc.) and questions for which the alternatives have no natural ordering. The 43 variables that met these criteria were then further classified into the following five categories depending of their theme of inquiry:

1. Economic policy (4 variables).
2. Confidence in governmental and non-governmental institutions (9 variables).
3. Attitudes toward democracy as a political system (9 variables).
4. Other political questions (12 variables).
5. Personal matters (9 variables).[10]

[8] See Tesser (1993); Olson et al. (2001) and Park and Schaller (2005) for evidence that attitudes are heritable.

[9] Data for some of the countries in this data set is from the 1994-1998 wave of the WVS. To calculate changes in polarization, we use the cumulative WVS 1981-2004 (European Values Study Group and World Values Association, 2006).

[10] Most questions in the *Personal matters* category refers to the respondent's own behavior.

We focus on economic policy for two reasons. First, unlike the questions about confidence in various institutions and personal matters, the questions about economic policy measure political preferences. Second, economic policy is arguably an important dimension of politics in most countries. However, we discuss the main results for other question categories in Section 5.

Broadly speaking, the economic policy questions measure various economic aspects of left and right on a 1 to 10 scale. This means that standard deviations vary between 0 and 4.5, where 0 corresponds to the case where everybody gives the same answer and 4.5 to the case where half of the population says "1" and the other half "10". Table 1 displays the wording of each question.

Table 1. Economic policy questions

Equality (e035)	How would you place your views on this scale? 1 means you agree completely with the statement "Incomes should be made more equal"; 10 means you agree completely with the statement "We need larger income differences as incentives".
Private (e036)	How would you place your views on this scale? 1 means you agree completely with the statement "Private ownership of business should be increased"; 10 means you agree completely with the statement "Government ownership of business and industry should be increased".
Government (e037)	How would you place your views on this scale? 1 means you agree completely with with the statement "People should take more responsibility to provide for themselves"; 10 means you agree completely with the statement "The government should take more responsibility".
Competition (e039)	How would you place your views on this scale? 1 means you agree completely with the statement "Competition is good. It stimulates people to work hard and develop new ideas"; 10 means you agree completely with the statement "Competition is harmful. It brings out the worst in people".

The questions have been slightly abbreviated.

Note that the Equality, Private and Government questions all refer to the status quo in each country. The questions measure different aspects of economic policy, but the standard deviations calculated from these questions are nevertheless strongly correlated. As shown in Table 2, the lowest correlation is 0.69 and the highest 0.88. The average responses to these questions are correlated to much lower extent – the correlation varies from 0.05 to 0.37. This suggests that it is much less of a problem that questions are country-specific when focusing on the dispersion of responses than mean values.

We focus on the standard deviation as our measure of polarization, which raises a couple of issues related to the measurement of polarization.

First, the standard deviation is only one of many plausible measures of polarization. Unfortunately, we lack clear guidance from theory as to which measure is most

TABLE 2. Pairwise correlations of standard deviations

	Equality	Private	Government	Competition
Equality	1			
Private	0.75	1		
Government	0.73	0.88	1	
Competition	0.77	0.72	0.69	1

appropriate. Two of the theories mentioned above suggest particular measures of polarization. The model by Alesina et al. (1999) predicts that public goods provision depends on the median deviation from the median response. This is a measure of dispersion as much as polarization, just like the standard deviation. However, we don't calculate the median distance to the median since it will be a very crude measure given that there are only a few discrete responses to the questions in the WVS. Fernández and Levy (2008) calls for a measure of heterogeneity based on the probability that two randomly matched individuals in the population hold the same opinion. Yet this measure does not resonate well with multiple-choice questions as it treats "4" and "5" on a 1 to 10 scale as two groups as distinct as "1" and "10". To check the robustness of our results, we consider nine other measures of dispersion and polarization. We calculate the polarization measure suggested by Esteban and Ray (1994) for three different parameter values and the kurtosis (see DiMaggio et al. 1996 for further discussion about kurtosis as a measure of bimodality). Esteban and Ray's (1994) measure of polarization includes a parameter α which loosely speaking measures the extent of sensitivity to polarization – the higher α is, the more the measure departs from the Gini coefficient. In order to satisfy their axioms, α has to be between zero and approximately 1.6. We calculate their measure when α is 0.5, 1.0 and 1.5. We also calculate the variance, average absolute deviation, mean difference and the Gini coefficient.[11] Table 3 displays the pair-wise correlations between the standard deviation and the other measures. We use the Government question to calculate these correlations, but the results are very similar for the other economic policy questions. The correlations between the standard deviation and the polarization measures from Esteban and Ray (1994) depend entirely on the parameter α. The correlation is lower the larger is this parameter. The main

[11] Average absolute deviation is defined as

$$\frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|$$

and the mean difference as

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i-x_j|\,.$$

results of the paper are very similar regardless of whether we use the standard deviation or any of the other measures, except for kurtosis which has a weaker relationship with government size.

TABLE 3. Correlations with other measures (Government)

Measure	Corr.	Measure	Corr.
Variance	0.996	Average absolute deviation	0.992
Polarization 0.5	0.961	Mean difference	0.981
Polarization 1.0	0.827	Kurtosis	−0.736
Polarization 1.5	0.689	Gini	0.414

Second, all measures of dispersion and polarization we consider treat the ordinal scale of responses to multiple-choice questions as an interval scale, i.e., we assume that the difference between an answer of 1 and 2 is the same as the difference between 5 and 6. Mouw and Sobel (2001) demonstrate that it is possible to measure dispersion without this assumption, but we believe that the cost in terms of difficulties in interpreting the results using their measure is greater than the benefits of using a well-known and simple measure like the standard deviation.

Third, as the standard deviation may be correlated with the mean value of responses, we control for the mean in all regressions. We have also excluded all questions where the absolute value of the correlation between mean and standard deviation is above 0.5. The correlations between mean and standard deviation are low for the economic policy questions – the exception being the question regarding private ownership of business with a correlation of 0.44.[12] However, the mean value need not perfectly reflect the true mean of preferences if responses are centered on either end of the scale (e.g., 1 or 10 on a 1 to 10 scale). For example, consider two countries with continuously normally distributed underlying distributions of preferences with the same mean. If the mean is above 5.5, the country with the highest standard deviation of the underlying distribution will have a lower observable mean because a larger share of respondents have their answers censored at 10. Hence, the measured standard deviation might be informative about the true mean of preferences even if we control for the measured mean. A related concern is that countries with mean values closer to one end point or the other will appear to have a lower standard deviation, since end point censoring reduces variability. We control for this by including the absolute deviation from 5.5 as an additional control variable. This test is also done to control for "false consensus", see Section 4.3.

[12] The correlations between mean value and standard deviation for the other polarization measures are: Equality, -0.15; Government, 0.11 and Competition, 0.23.

Fourth, if there is a nonlinear relationship between responses at the individual level and the dependent variable, the variance will have a direct effect on the dependent variable. Since we have no data on individual-level counterparts to the aggregate-level outcomes we consider in this paper, we are not able to test for such nonlinearities and this caveat should be kept in mind when interpreting our results.[13]

Finally, answers to survey questions may not only capture underlying differences in political preferences. For example, stated answers may partly reflect a misunderstanding of the question or a social norm not to deviate from consensus. The answers could also reflect the partisanship of politics in a given country rather than fundamental differences in ideology. This could lead to omitted variable bias or reverse causality from government size to the *measured* level of political polarization. We discuss these issues in more detail in Section 4.2 and 4.3.

What characterizes countries with a high or low degree of political polarization? Table 4 lists the ten countries with the highest and lowest standard deviation in the question about government interventions (Government). Perhaps surprisingly, Pakistan is the country with the lowest level of political polarization. This is not a peculiarity of this particular question – as shown in Table A1 in the Appendix, Pakistan has a very low standard deviation for the other economic policy questions as well. However, Pakistan is also among the countries with the lowest response rates. In the case of the Government question, 37 percent of respondents in Pakistan said they didn't know or gave no answer at all.[14] The other countries on the list are less surprising with three Scandinavian countries among the ten most cohesive and five Latin American countries among the most polarized.[15]

In order to get an idea how stable political polarization is over time, we consider the intertemporal correlations using data from previous waves of the WVS (1989-1993 and 1994-1998).[16] As shown in Table 5, the correlations between the old polarization measures and those from WVS 1999-2002 vary between 0.52 and 0.83, suggesting that the degree of political polarization is relatively stable over time.[17]

[13] The problem of nonlinearities at the individual level and aggregate outcomes is discussed by Deaton (2003) in the context of health and income inequality.

[14] The report from the person responsible for collecting WVS data in Pakistan does not reveal anything particular except that certain regions of the country couldn't be included in the survey for political and security reasons (for example close to the Afghan border). The data from Pakistan may thus not be fully representative although this is probably not a very serious problem.

[15] It should be kept in mind, however, that we cannot readily compare polarization across continents since we only have data from 72 countries. In addition, although WVS contains many developing countries there is a tendency that larger and more developed countries are more likely to be included.

[16] There is also a 1981-1984 wave of the WVS, but none of the four economic policy questions were used in the first wave.

[17] Another indication that political polarization is relatively stable over time is that the degree of political polarization in East and West Germany in 1990, the year of the reunification, are remarkably

Table 4. Countries with lowest and highest level of political polarization

Rank	Lowest	Government	Rank	Highest	Government
1	Pakistan	1.92	63	Turkey	3.27
2	Israel	1.99	64	Zimbabwe	3.32
3	Netherlands	2.11	65	Tanzania	3.32
4	Denmark	2.15	66	India	3.43
5	Sweden	2.22	67	Bangladesh	3.43
6	South Korea	2.27	68	Brazil	3.44
7	Norway	2.34	69	Venezuela	3.46
8	Great Britain	2.39	70	Dominican Republic	3.52
9	Taiwan	2.42	71	Mexico	3.55
10	Estonia	2.42	72	El Salvador	3.70

To test if changes in measured polarization partly reflect measurement error, we regress change in polarization during the 1990's on polarization in 1990. If polarization is measured with error, countries with a high measured level of polarization in 1990 should have a expected decrease in polarization during the 1990's. The coefficient has a negative sign in all regressions and is statistically significant (at the ten percent level) in three out of four regressions.

Table 5. Intertemporal correlations

	1989-1993		1994-1998	
	Correlation coefficient	Obs.	Correlation coefficient	Obs.
Equality	.52	30	.75	31
Private	.83	28	.75	29
Government	.56	38	.75	35
Competition	.64	37	.76	33

See the Supplementary Appendix for further details about how the correlations were calculated.

We now turn to our measures of government size and control variables. We focus on two different measures of government size: general government consumption as a fraction of GDP and transfers and subsidies as a fraction of GDP. The indicators of government size were obtained from Gwartney & Lawson (2005) and refer to year 2000 for most of the countries. The basic control variables used are the same as La Porta et al (1999), although we have used data from World Bank (2005) for GDP per capita and

similar. Ranking all countries by the degree of polarization in 1990, the rank of East and West Germany are 13 and 16 for Equality, 11 and 4 for Private, 29 and 11 for Government and 13 and 9 for Competition.

the ethnic fractionalization measure from Alesina et al (2003). The sources for these and the other control variables are described in Table A2 in the Appendix. We use the democracy index from the Polity IV project (Marshall & Jaggers, 2002) to classify countries as weak and strong democracies.

Table 6 shows the partial correlations between the standard deviations and the set of control variables controlling for mean responses. Political polarization is moderately correlated with ethnic fractionalization. Countries with French legal origin are more polarized, whereas countries with German or Scandinavian legal origin are more cohesive. Socialist and British legal origin are not significantly correlated with polarization. The proportion of Catholics is positively correlated with polarization, whereas countries with many Protestants are more cohesive. In line with the conjecture by Fukuyama (1995), political polarization is strongly negatively correlated with trust.[18] We also check the correlations between polarization and another measure of social capital – the proportion of respondents who participate in at least one civic organization (cf. Putnam, 1993, 2000). This correlation is negative, implying that fewer people participate in polarized countries, but quite weak.[19] There is a strong positive correlation between polarization and income inequality as measured by the Gini index. Polarization is negatively correlated with logarithm of GDP per capita as well as openess to trade. The electoral rule used in the country is uncorrelated with polarization, but countries with presidential regimes are more polarized. Polarization is strongly correlated with distance from the equator – countries close to the equator are more polarized. Some polarization measures are moderately correlated with both land area and size of the population – larger countries (in terms of area and population) are more polarized.[20] We find no significant correlations between population density or mountainous terrain and polarization. Polarization is negatively correlated with the proportion in the population between 15 and 64 and above 65, implying that countries with many children (below 15 years of age) are more polarized. Finally, we find that political polarization is highest in Latin America and Sub-Saharan Africa, and lowest in North America, Europe and Oceania.[21]

[18] There are only two possible answers to the trust question ("Most people can be trusted" and "Need to be very careful") and a higher value implies lower trust. Hence, we cannot calculate a measure of dispersion based on the trust question. This question is used by Knack and Keefer (1997) and other papers that empirically study trust and social capital.

[19] The modest correlation between participation in civic organizations and political polarization masks a strong negative correlation between polarization and participation for European countries.

[20] Our finding that polarization is positively correlated with country size supports the assumption in Alesina and Spolaore (1997) that large nations have more heterogeneous and diverse populations.

[21] We classify Pakistan and Iran as part of the Middle East. The exact classification is available from the authors upon request.

TABLE 6. Partial correlations with control variables

	Equality	Private	Government	Competition
Basic control variables				
Ethnic fractionalization$^+$	.23*	.35***	.28**	.43***
Legal origin: Socialist$^+$	.09	.11	.01	.00
Legal origin: French$^+$	.13	.09	.29**	.43***
Legal origin: German$^+$	−.20	−.27**	−.20*	−.32**
Legal origin: Scandinavian$^+$	−.18	−.25*	−.30**	−.37***
Legal origin: British$^+$	−.02	.11	−.04	.03
Religion: Catholic$^+$	.31**	.07	.25**	.34***
Religion: Muslim$^+$	−.14	.18	.09	.20
Religion: Other$^+$	−.06	−.03	−.10	−.14
Religion: Protestant$^+$	−.25**	−.32**	−.36***	−.45***
Log of GDP per capita$^+$	−.44***	−.60***	−.59***	−.62***
Social capital and inequality				
Average trust	.54***	.60***	.47***	.57***
Participation in civic organizations	−.15	−.03	−.16	−.26*
Gini index	.53***	.46***	.56***	.65***
Political variables				
Openess to trade	−.37**	−.24	−.25*	−.35**
Majoritarian rule	−.14	−.02	.00	.00
Presidential regime	.42***	.26*	.43***	.56***
Federal political structure	.33**	.22	.30**	.28**
Geography and demography				
Absolute distance from equator	−.32**	−.41***	−.56***	−.52***
Percent mountainous terrain	.05	.02	.14	.16
Log of area (sq. km)	.23*	.05	.26**	.32**
Log of total population	.18	.00	.30**	.32**
Log of population density	−.12	−.07	−.05	−.09
Proportion 15-64 years	−.29**	−.42***	−.35***	−.50***
Proportion 64- years	−.34***	−.52***	−.57***	−.61***
Regional dummies				
Sub-Saharan Africa	.35***	.48***	.22*	.23*
Middle East and North Africa	.25**	.06	−.13	.14
Latin America and the Caribbean	.41***	.22*	.48***	.61***
Asia	−.10	−.19	.09	−.13
Europe, North America, Oceania	−.26**	−.34***	−.48***	−.50***

Partial correlations with control for mean responses. Variables marked with a plus sign belongs to the set of controls used in the long regression reported in Table 7. * = 10 percent, ** = 5 percent, and *** = 1 percent significance level.

4. Results

In this section, we analyze whether polarization in preferences for economic policy has predictive power for the size of government.

Let y_i denote one of the two measures of government size (spending or redistribution) in country i. For each measure y, we run the regression

$$y_i = \alpha + \beta Polarization_i + \mathbf{X}_i \boldsymbol{\gamma} + \varepsilon_i, \tag{1}$$

where $Polarization_i$ is a measure of polarization in country i and $\mathbf{X}_i$ is a vector of control variables measured at the country level.[22]

There are three potential problems that could bias the estimate of β. First, there might be an omitted variable bias if variables which are correlated with both polarization and government size are not included in X_i. Second, $Polarization_i$ might be endogenous to government size. Third, measurement error in political polarization could bias our estimate of β toward zero. Our econometric analysis proceeds as follows. We first report the results from our basic specifications. We then perform sensitivity analysis controlling for the variables listed in Table 6. Finally, we discuss the potential endogeneity problems due to income inequality, partisan politics and issues of survey data in more detail.

In the basic specifications, we run two different specifications for each dependent variable and measure of polarization. In the "short" regression, we include the mean response, controls for legal origin (Socialist, French, German and Scandinavian), share of religious denominations (Muslim, Catholic and Protestant) and ethnic fractionalization. We include these variables as they are arguably exogenous with respect to political polarization and likely to affect the size of government. In the "long" regression we also include the logarithm of GDP per capita.[23] We include GDP per capita as a control variable as political polarization and size of government might be increasing in national income. Yet including GDP per capita implies that we are controlling for one mechanism by which political polarization may affect size of government. Moreover, GDP per capita may in itself be endogenous with respect to size of government.

The results from the regression with the main specification and the standard deviation as our measure of polarization are reported in Table 7.

[22] We have considered using an index combining all economic policy questions instead of reporting the results for the four questions separately. We have abstained from doing so for three reasons. First, there would be fewer observations for this index than for any single question. Second, the results for an index would be less straightforward to interpret. Third, comparing the results for four related questions is a sensible first robustness check.

[23] Our "short" specification is identical to one of the specifications in La Porta et al. (1999) with the exception that we use a different measure of ethnic fractionalization.

Table 7. Political polarization and government size

Dependent variable	Equality		Private		Government		Competition	
	Short	Long	Short	Long	Short	Long	Short	Long
All countries								
Transfers & subsidies	−5.40**	−1.84	−4.27**	−3.37*	−9.65***	−6.67**	−9.45***	−6.70**
	(−2.09)	(−.76)	(−2.35)	(−1.82)	(−3.65)	(−2.57)	(−3.09)	(−2.16)
Consumption	−6.15**	−.87	−4.62**	−1.69	−9.40***	−4.94*	−10.94***	−6.74**
	(−2.46)	(−.33)	(−2.19)	(−.78)	(−3.88)	(−1.90)	(−4.12)	(−2.37)
Weak democracies								
Transfers & subsidies	3.56*	3.97*	1.05	1.57	3.79	4.43*	3.52	4.94
	(1.87)	(1.96)	(.31)	(.49)	(1.51)	(1.75)	(1.13)	(1.27)
Consumption	.30	3.08	−1.05	1.58	−2.90	−.21	−5.35	−1.84
	(.08)	(1.02)	(−.27)	(.46)	(−.67)	(−.06)	(−.97)	(−.37)
Strong democracies								
Transfers & subsidies	−11.15***	−7.01	−13.56**	−17.53**	−15.85***	−14.52***	−15.06***	−13.91***
	(−3.44)	(−1.47)	(−2.39)	(−2.39)	(−6.60)	(−2.93)	(−2.97)	(−2.87)
Consumption	−12.15***	−9.29	−8.09	−18.60**	−11.96***	−11.55***	−16.76***	−15.83***
	(−5.11)	(−1.57)	(−1.59)	(−2.24)	(−5.62)	(−2.87)	(−4.46)	(−2.97)

The table reports coefficients and heteroskedasticity robust t statistics for the standard deviation of each question for two different specifications. The left column for each question includes controls for the mean response, legal origin, religious denomination and ethnic fractionalization, whereas the right column also include the logarithm of GDP per capita. One asterisk denotes 10 percent significance level, two asterisks denotes 5 percent significance level and three asterisks denotes 1 percent significance level.

The first rows of Table 7 show that polarized countries have smaller governments both in terms of spending and redistribution. In the short specification, the coefficients are both economically and statistically significant. Including the logarithm of GDP per capita in the regression reduces the estimated coefficients, but they are still negative for all polarization measures. For the Private, Government and Competition questions, the estimated coefficients are reduced by roughly 30 to 40 percent when GDP is included, but they remain statistically significant in most cases.

The mechanisms by which polarization among the electorate may affect the size of government – the budget decision process, legislator shirking or a larger scope for coalition formation – all hinge on the assumption that politics is democratic. In contrast, many of the potential endogeneity problems, like the effect of government size on pre-tax income inequality, should not depend on democratic development. A simple test of a causal effect is thus to see if the relationship between polarization and size of government is stronger in more democratic countries.[24]

We classify the 33 countries with a democracy score of 9 to 10 as “strong” democracies and the 35 countries with a score of 0 to 8 as “weak” democracies. We chose this particular cutoff in order to get roughly half of the countries in each group. Table 7 reports the main regressions when the sample is restricted to weak and strong democracies. The democracy index is not available for four countries, but we classify three of these countries as strong democracies.[25] The coefficients vary in sign for weak democracies. The estimated coefficients suggest that polarization might have a positive effect on redistribution in weak democracies, whereas the effect on spending is ambigous. However, when Pakistan is excluded from the sample, none of the coefficients for weak democracies are statistically significant. When restricting the sample to strong democracies, however, the estimated coefficients for government size are negative, large and robust to the inclusion of the logarithm of GDP. The results are very similar when we use more inclusive definitions of “strong” democracies.

To test for heterogenous effects more formally we include an interaction term between political polarization and strong democracy together with a strong democracy dummy in regression (1). The common effect of polarization is most often negative,

[24] The test of heterogeneous effects with respect to democracy is only indicative. It is possible, for example, that government spending in democracies foster homogeneity whereas less democratic countries spend public funds on projects that have no effect on social cohesion.

[25] The countries for which data is missing are Iceland, Luxembourg, Malta and Peru. However, Iceland, Luxembourg and Malta receive the highest possible score on the political and civil rights indices in 2000 published by Freedom House (2005) and it therefore seems uncontroversial to classify these as strong democracies. The democracy index is not available for Peru in 2000 because the country was “in transition” with a very low score on the democracy index prior to 2000 and a high score thereafter.

but sometimes positive. The interaction effect is negative for all questions, though not always statistically significant. These results indicate that democratic countries have a stronger relationship between political polarization and government size. That the interaction effect isn't always significant is partly due to the loss of statistical power from decomposing one effect into two different, additive effects (i.e., the common effect of polarization and the specific effect of polarization in democracies).

An alternative explanation for the stronger relationship between polarization and size of government in strong democracies is that the measurement error variance is larger in weak democracies, exacerbating attenuation bias. However, classical measurement error cannot explain why the estimated coefficients with respect to redistribution are positive in weak democracies while negative in strong democracies.

The results in Table 7 are robust to controlling for demographic and geographic factors. Separately including percent of mountainous terrain, the logarithm of country area, absolute distance to equator, the logarithm of total population, population density, proportion of the population between 15 and 64 or the proportion of the population above 65 does not qualitatively affect the results.[26] The coefficients for strong democracies are particularly robust, and often strengthened when these control variables are included. The size and significance of the coefficients estimated on the full sample are reduced when we include regional dummies, whereas the results for strong democracies are strengthened somewhat.

We also control for a number of variables that has been found to explain the size of government in previous studies: openness to international trade (Rodrik 1998); whether the country has a federal political structure; whether the country has a presidential regime and whether the lower house is elected under majoritarian rule (Persson and Tabellini, 2003).[27] Separately including these variables in regression (1) does not change the estimated coefficients significantly. This is also the case when we control for the average level of trust and participation in civic organizations.

In the following sections we consider the endogeneity problems mentioned above in more detail, and a number of alternative explanations and potentially confounding factors for the relationship between polarization of political preferences and government size.

4.1. Income inequality. If government consumption and redistribution reduces (pre-tax) income inequality, political polarization may be endogenous in regression (1) unless inequality is controlled for. The most straightforward way control for income

[26] Alesina and Wacziarg (1998) show that smaller countries have a larger share of public consumption of GDP.

[27] We define openness to trade as exports plus imports over GDP.

inequality is to include the Gini coefficient in the main regression. There are two potential problems with this approach. First, as the Gini coefficient is highly correlated with political polarization and polarization is likely to be measured with error, attenuation bias is exacerbated by including the Gini index in the regression. The second and more important problem is that the Gini index is based on consumption data for developing countries, whereas income net of employer taxes is used for developed countries (Luxembourg Income Study). Consequently, the Gini index is endogenous with respect to redistribution, thereby spuriously reducing the estimated effect of political polarization. Keeping these caveats in mind, Table A3 reports the same regressions as in Table 7 when the Gini coefficient is included in both the short and long specification. The estimated coefficients decrease when the Gini coefficient is included in the regressions with all countries included. However, when the sample is restricted to strong democracies, the estimated relationship between polarization and government size are highly robust to the inclusion of the Gini coefficient.

Another way to control for income inequality is to create an "income-adjusted" measure of political polarization. For each country j we run the individual question answers on the respondents' income.[28] For example, the regression in the case of the Competition question is

$$Competition_{ij} = \alpha_j + \beta_{1j} Income_{ij} + \varepsilon_{ij},$$

where all variables refer to respondent i. We then calculate an "income-adjusted opinion" for each respondent by subtracting the difference between the predicted opinion and the mean predicted opinion from the actual answer. In the case of Competition, we calculate the income-adjusted opinion as

$$Adj.Competition_{ij} = Competition_{ij} - \widehat{\beta}_{1j} \left(Income_{ij} - \overline{Income_j} \right),$$

where $\overline{Income_j}$ is the average income in country j. Finally, we calculate a new polarization measure for each country using $Adj.Competition$. The income-adjusted polarization measures are very similar to the ordinary polarization measures. All correlation coefficients between the standard and income-adjusted polarization measures are above 0.99. Not surprisingly, re-running the regressions of government performance using the income-adjusted standard deviations yields very similar results.[29] The reason for the small differences between the income-adjusted and ordinary polarization measures is

[28] Note that the 1 to 10 scale of income is specific to each country.

[29] As educational attainment might be a better proxy for permanent income than current income, we also calculate adjusted standard deviations where we include educational attainment (on a 1 to 3 scale) as a regressor in the first stage above. This adjusted measure of polarization is also highly correlated with our standard measure with correlation coefficients of 0.99 for the economic policy questions.

found in the low explanatory power of stated income for individual preferences. A potential explanation for this result is that income is measured with error, giving rise to attenuation bias. However, if permanent income is truly an important determinant of political opinions, we would expect individuals' responses to the different economic policy questions to be strongly correlated within countries. In fact, most of the correlations are weak, indicating that the economic policy questions are not easily divided on a single left to right political scale.

4.2. Partisan economic policies. Since the economic policy questions often relate to the current situation in a particular country, political polarization may be a direct effect of the partisanship of policy. For example, suppose that the government in a certain country redistributes income from group X to group Y. If people in this country think of "redistribution" as redistribution from X to Y, then measured preferences for "redistribution" may be polarized, even if preferences are homogenous regarding some other redistribution scheme (say from the rich to the poor). If countries in which the government pursues partisan policies also spend and redistribute less, there will be a spurious correlation between political polarization and government size. In effect, political polarization as measured by the survey questions could be endogenous with respect to the size of government. However, as shown above, the results for economic and government performance are similar for all four economic policy questions. For political polarization to be a direct consequence of partisan economic policies, the partisan policies must thus shape opinions on all these questions. If divisive policies cause polarization, we would therefore expect the responses to the economic policy questions to be strongly correlated at the individual level in countries where polarization is high. In reality, we have the opposite case. The correlation between individual-level responses is much higher in countries with a low level of political polarization.[30]

4.3. Survey data. A potential problem in using survey data is that people in some countries are uncertain about their political preferences. For example, people in developing countries may be more likely to randomize their responses, giving rise to a spurious correlation between political polarization and poverty.[31] This could also be the case if uncertain respondents choose particular focal values.[32] Part of this potential problem is solved by controlling for GDP per capita, but the size of the public sector

[30] In addition, the view that political polarization is a consequence of divisive policies does not fit well with the result that the effect of political polarization is stronger in democracies.

[31] This correlation is not necessarily positive as uniform randomization on a 1 to 10 scale implies a standard deviation of approximately 2.63, far from the theoretical maximum of 4.5 (when half of the respondents answer 1 and the other half 10).

[32] If uncertain respondents have the same focal value, we underestimate the true polarization of preferences. If uncertain respondents instead randomize between extreme values, we overestimate it.

could affect measured polarization also when economic prosperity is held constant. For example, the ability of the relatively poor to understand the survey question may hinge on public investment in schooling. We perform three different tests to check whether political polarization may be due to the inability of certain groups to understand the survey questions.

First, we include the country response rate as a control variable in regression (1). The idea is that in countries where many people are uncertain about their preferences, more people will also state that they "don't know" what they think or not answer the question at all. The estimated coefficients in Table 7 are robust to this test.

Second, we calculate the country-level correlations between individuals' responses to the economic policy questions. For each question, we then calculate the country average of the absolute value of the correlations with the other questions. This measure captures the extent to which question responses follow a certain pattern, and we expect it to be lower the more uncertain people are about their political preferences. Political polarization remains a robust predictor of government size though somewhat fewer coefficients are statistically significant when we include this measure in regression (1).

Third, as Pakistan is the country with the lowest response rates, we run the regressions with Pakistan excluded from the data set. This leads to somewhat stronger results for the regression with all countries.

A related problem with survey data is that people might not want to deviate too much from the opinions of others. A respondent who wants to minimize the maximum deviation from other respondents will answer five or six provided he does not know the others' responses. If so, countries where people are concerned with consensus would have their responses centred around 5.5. Hence, we include the absolute deviation from 5.5 as a control variable in regression (1) as a rough way of controlling for "false consensus". This is also a rough way of controlling for a correlation between the standard deviation and the true mean of preferences due to censoring of the data, as discussed in Section 3. The results in this robustness check are almost exactly the same as in the basic specifications. Another indication that censoring is unlikely to be quantitatively important is that there is no systematic relationship between the mean values (in terms of left or right) and government size in the basic specification of regression (1).

4.4. Ethnic fractionalization. To check the sensitivity of political polarization with respect to ethnic fractionalization, we re-run the regressions excluding ethnic fractionalization as a regressor, which leaves the estimated coefficients largely unchanged.

If political polarization is the main channel by which ethnic fractionalization affects government size (as suggested by for example Alesina et al., 1999), we would expect the coefficient on ethnic fractionalization to be sensitive to the inclusion of political polarization. Estimating regression (1) without the mean and standard deviation, shows that ethnic fractionalization has a statistically significant negative effect on redistribution in the short regression, but not in the long regression. The effect of ethnic fractionalization on government spending is statistically insignificant in both specifications. Including the mean and standard deviation reduces the effect of ethnic fractionalization on redistribution in the short regression, in particular when the Government or Competition questions are used as control variables.

5. Other Questions

We also run regression (1) using the other question categories as measures of political polarization.

Comparing the results for the economic policy questions with those for the other question categories gives us an indication of which mechanisms are more important: If the politico-economic explanations are true, we should expect a stronger relationship between polarization and size of government for the questions of economic policy than for questions that do not reflect political preferences. If the relationship between polarization and size of government is instead due to less altruism in polarized societies, there is little reason to expect substantial differences between question categories.

This distinction is clear-cut when we compare the results for the economic policy question with polarization in terms of *Personal matters*, but less so for the other question categories. Though clearly related to the political situation in a country, citizens' confidence in governmental and non-governmental institutions is not in itself a measure of political preferences. Similarly, it is not obvious from a theoretical viewpoint that polarized attitudes toward democracy as a political system have an effect on political outcomes – even though people disagree over the merits of a certain political system, they may still agree on economic policy. The category "other political questions" contains questions concerning, among other things, the justifiability of divorce, willingness to pay for preventing pollution of the environment and immigrant policy. The importance of these questions are likely to vary widely depending on the political context.

As it turns out, polarization in personal matters and attitudes toward democracy as a political system are almost unrelated to the size of government. As there are nine questions for each of these two categories and we run twelve specifications for each question, we should expect about eleven coefficients per category to be statistically significant at the ten percent level even if there does not exist a relationship between polarization on these issues and size of government. For personal matters, there are six statistically significant coefficients of either sign compared to eight statistically significant coefficients of either sign for attitudes toward democracy. The relationship is somewhat stronger for the other political questions with 24 statistically significant coefficients with a negative sign and 10 with a positive sign out of 144 regressions. The results for polarization of confidence in institutions are similar to those for economic policy for government consumption, but there is no relationship with redistribution.

6. Conclusion

This paper has shown that politically polarized countries have smaller public sectors. This relationship is robust to a large set of control variables and the use of other measures of polarization than the standard deviation. Moreover, we found no evidence that the correlation between political polarization and government size is a direct consequence of income inequality, divisive economic policies, false consensus, or an inability to understand the survey questions in developing countries.

A priority for future research is to better identify the underlying causal mechanism. In order to do this, it would probably be useful to use panel data on political polarization, but it would also be interesting to examine if political polarization can explain variation in redistribution and spending within countries. A definite test of the causal mechanism, however, requires some kind of exogenous variation in political polarization that does not have an independent effect on the size of government. Unfortunately, it is very difficult to come up with an instrument that only affects government size through political polarization.[33]

The relationship between political polarization and government size is relevant for policy. If political polarization has a causal effect on the size of government, then policies that foster homogeneity might increase redistribution and public goods provision. One such policy instrument is the educational system to the extent that it plays a role in shaping the values of future citizens (as suggested by Gradstein and Justman, 2002).

[33] One potential instrument could be a randomized propaganda campaign, although it is probably difficult to find such campaigns that are both exogenous and successful enough to change people's political preferences.

Appendix

Table A1. Political polarization in economic policy

Country	Equality	Private	Government	Competition
Albania	2.47269	2.42767	2.77046	2.02929
Algeria	2.57152	3.19376	3.01380	
Argentina	3.25348	3.07496	3.15056	3.18853
Australia	2.55512	2.29370	2.61804	2.11603
Austria	2.56862	2.10130	2.56659	2.01197
Bangladesh	2.85097	3.35134	3.43077	2.47480
Belgium	2.92396		2.72886	2.60430
Brazil	3.44113	3.19661	3.44391	2.92267
Bulgaria	3.00508		2.85846	2.46537
Canada	2.66328	2.21450	2.56175	2.35627
Chile	2.96206	2.80484	2.67257	2.82391
China	3.11140	2.86740	3.21962	2.18632
Colombia	2.96834	3.15352	3.13940	
Croatia	2.96147	2.98273	3.11691	2.43001
Czech Republic	2.78061	2.53557	2.56544	2.20023
Denmark			2.14797	2.23760
Dominican Republic	2.86784	3.17444	3.52405	2.99318
Egypt	2.05136	2.87055	2.77806	
El Salvador	3.39994	3.44250	3.70209	3.20272
Estonia	2.40385	2.44452	2.42079	2.34325
Finland	2.58286	2.08597	2.45625	2.23000
France	2.97815	2.21928	2.51053	2.69906
Georgia	2.64499	3.05580	2.74663	2.41603
Germany		2.28356	2.70390	2.18147
Great Britain	2.54724	2.18622	2.39326	2.15212
Greece			2.58134	2.53479
Hungary			2.83871	2.54501
Iceland	2.84525	2.11932	2.64113	1.85494
India	3.58164	3.32503	3.42869	2.93711
Indonesia	2.32837	2.63504	3.10194	
Iran	2.39829	2.61761	2.69548	
Ireland	2.74860	2.27580	2.54011	2.28960
Israel	2.31262		1.99117	
Italy	2.72860	2.21390	2.67469	2.48730
Japan	2.20095	1.85110	2.58949	2.06766
Jordan	2.73477	2.95101	2.81872	
Latvia			2.66536	2.28656
Lithuania	3.07459	2.89135	2.82974	2.69752
Luxembourg	2.60595		2.46838	2.48415
Malta			2.75868	2.08569
Mexico	3.60650	3.30571	3.54808	3.23273
Morocco	3.09866	3.50752	3.26359	2.21190

Country	Equality	Private	Government	Competition
Netherlands	2.02512	1.85953	2.11136	2.04367
New Zealand	2.62809	2.26910	2.70147	2.26087
Nigeria	2.87664		2.86610	
Norway	2.26377	1.90572	2.34372	1.90249
Pakistan	2.15079	1.64685	1.91684	
Peru	2.80644	2.86111	3.22385	2.76792
Philippines	2.71600	2.67870	2.91533	2.49010
Poland	3.18268	2.83358	2.61729	2.76091
Portugal		2.35647	2.72656	2.65784
South Korea	2.74708	2.38578	2.27107	2.23890
Macedonia	3.02271	2.87731	2.87864	2.53609
Romania	3.04189	3.23263	3.23421	2.31758
Russia	3.00578	2.76346	2.90544	2.69169
Singapore	2.30463	2.44428	2.67522	2.06676
Slovakia			2.61098	2.23223
Slovenia	2.68866		2.65498	2.27948
South Africa	3.11986	3.06017	3.11434	2.53226
Spain	2.85794	2.49406	2.50493	2.35026
Sweden			2.22051	1.92397
Switzerland	3.07271	2.38122	2.71851	2.35609
Taiwan	2.35119	2.15162	2.41842	2.03901
Turkey	3.24765	3.29784	3.26722	3.09322
Uganda	3.18213	3.01569	3.10382	2.31025
Ukraine	2.98101	2.97537	2.99808	2.97123
Tanzania	3.83358	3.83577	3.31942	3.13652
USA	2.56661	2.23922	2.69689	2.39604
Uruguay	3.25013	2.71755	2.86099	2.86888
Venezuela	3.44249	3.30915	3.46030	3.01533
Vietnam	3.06162	2.86592	2.94052	2.64291
Zimbabwe	3.43612	3.43149	3.31763	2.62255

TABLE A2. Dependent and control variables

Variable name	Description and source	Obs.
Transfers and subsidies	Government transfers and subsidies as a percentage of GDP. Data from 2000, except for Georgia, Macedonia and Vietnam (data from 2003). Source: Gwartney and Lawson (2005)	65
Government consumption	General government consumption expenditures as a percentage of GDP. Data from 2000, except for Georgia, Macedonia and Vietnam (data from 2003). Source: Gwartney and Lawson (2005).	72
Ethnic fractionalization	Reflects the probability that two randomly selected individuals belongs to the same ethnic group. This is calculated as one minus the sum of squared shares of each group and therefore takes values between 0 and 1. Source: Alesina et al. (2003).	72
Legal origin	Identifies the legal origin of the Company Law or Commercial Code of each country. There are five possible origins: (1) English Common Law; (2) French Commercial Code; (3) German Commercial Code; (4) Scandinavian Commercial Code; and (5) Socialist or Communist laws. Source: La Porta et al. (1999).	72
Religion	Identifies the percentage of the population of each country that belonged to the three most widely spread religions in the world in 1980. For countries of recent formation, the data is available for 1990-95. The numbers are in percent (scale from 0 to 100). The three religions identified here are: (1) Roman Catholic; (2) Protestant; and (3) Muslim. The residual is called "other religions". Data for protestants in Lithuania was missing and has been set to 1.9 percent which is the figure reported in CIA World Fact Book 2005. Source: La Porta et al. (1999) and Central Intelligence Agency (2005).	72
GDP per capita	GDP per capita in US dollars (constant 2000). Source: World Bank (2005).	71
Trust	Average (binary) response from World Values Survey 1999-2001 to the question whether people can generally be trusted. Source: European Values Study Group and World Values Association (2004).	72

Variable name	Description and source	Obs.
Participation civic org.	Percentage of population answering “yes” to at least one of the questions of participation	52
Gini index	Estimates of the Gini index based on primary household survey data obtained from government statistical agencies and World Bank country departments. Data for high-income economies are from the Luxembourg Income Study database. Data refers to various years between 1990 and 2004 and the observation closest to year 2000 have been used. Source: World Bank (2005).	69
Openess to trade	Sum of exports and imports as a share of GDP (in the 1990s). Source: Persson & Tabellini (2003).	56
Majoritarian rule	Dummy equal to 1 if lower house elected under plurality rule (during 1990s), 0 otherwise. Source: Persson & Tabellini (2003).	56
Presidential rule	Dummy variable equal to 1 in presidential regimes (during 1990s), 0 otherwise. Source: Persson & Tabellini (2003).	56
Federal structure	Dummy equal to 1 if the country has a federal political structure (in the 1990s), 0 otherwise. Source: Persson & Tabellini (2003).	55
Distance to equator	The absolute value of the latitude of the country, scaled to take values between 0 and 1. Source: La Porta et al. (1999).	72
Percent mountainous terrain	We obtain this variable from the data provided by Montalvo and Reynal-Querol (2005). They in turn base this variable on work by geographer A. J. Gerard for the World Bank’s	57
Area	“Economics of Civil war, Crime, and Violence” project Area measured in square kilometers. . Data refers to year 2000. Source: World Bank (2005)	70
Total population	Data refers to year 2000. Source: World Bank (2005).	71
Population density	People per square kilometer. Data refers to year 2000. Source: World Bank (2005).	71
Population age 15-64	Proportion of population in age 15 to 64. Data refers to year 2000. Source: World Bank (2005)	71
Population age 65-	Proportion of population 65 years or older. Data refers to year 2000. Source: World Bank (2005)	71

Table A3. Political polarization and government size (controlling for income inequality)

Dependent variable	Equality		Private		Government		Competition	
	Short	Long	Short	Long	Short	Long	Short	Long
All countries								
Transfers & subsidies	−.6	1.39	−1.87	−.63	−6.25**	−3.4	−4.9	−2.63
	(−.2)	(.59)	(−.96)	(−.3)	(−2.39)	(−1.44)	(−1.54)	(−.79)
Consumption	−3.19	.03	−3.56	−.78	−8.11***	−4.46	−7.62**	−4.64
	(−1.05)	(.01)	(−1.27)	(−.29)	(−2.67)	(−1.52)	(−2.19)	(−1.28)
Weak democracies								
Transfers & subsidies	4.7**	5.43**	1.59	2.44	3.93	4.77*	3.14	5.2
	(2.21)	(2.48)	(.41)	(.67)	(1.49)	(1.85)	(.73)	(.98)
Consumption	−.22	2.24	−1.93	1.04	−3.42	−.92	−4.35	−1.23
	(−.05)	(.56)	(−.41)	(.24)	(−.72)	(−.22)	(−.63)	(−.18)
Strong democracies								
Transfers & subsidies	−10.21**	−3.8	−23.39**	−18.59**	−13.78***	−11.18**	−14.41***	−12.49**
	(−2.96)	(−.88)	(−2.89)	(−2.26)	(−6.48)	(−2.54)	(−3.02)	(−2.51)
Consumption	−11.31***	−9.77	−5.96	−15.95	−11.9***	−12.81**	−16.65***	−15.85**
	(−3.56)	(−1.38)	(−.82)	(−1.7)	(−5.07)	(−2.6)	(−2.85)	(−2.45)

The table reports coefficients and heteroskedasticity robust t statistics for the standard deviation of each question for two different specifications. The left column for each question includes controls for the mean response, legal origin, religious denomination, ethnic fractionalization and Gini, whereas the right column also include the logarithm of GDP per capita. One asterisk denotes 10 percent significance level, two asterisks denotes 5 percent significance level and three asterisks denotes 1 percent significance level.

References

Acemoglu, D., Chernozhukov V. and Yildiz, M. (2007), “Learning and Disagreement in an Uncertain World”, unpublished manuscript, Massachusetts Institute of Technology.

Alesina, A., Baqir, R. and Easterly, W. (1999), “Public Goods and Ethnic Divisions”, *Quarterly Journal of Economics* 115(4), 1167–1199.

Alesina, A., Glaeser, E. and Sacerdote, B. (2001), “Why Doesn’t the United States Have a European-Style Welfare State?”, *Brookings Paper on Economics Activity* 2001(2), 187–254.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2003), “Fractionalization”, *Journal of Economic Growth* 8(2), 155–194.

Alesina, A. and Spolaore (1997), “On the Number and Size of Nations”, *Quarterly Journal of Economics,* 112(4), 1027-1056.

Alesina, A. and Wacziarg, R. (1998), “Openness, Country Size and Government”, *Journal of Public Economics*, 69(3), 305-321

Alt, J. E. and Dreyer Lassen, D. (2006), “Transparency, Political Polarization, and Political Budget Cycles in OECD Countries”, *American Journal of Political Science* 50(3), 530-550.

Au, K. (1999), “Intra-Cultural Variation: Evidence and Implications for International Business”, *Journal of International Business Studies* 30(4), 799–812.

Au, K. and Cheung. M. (2004), “Intra-cultural Variation and Job Autonomy in 42 Countries”, *Organization Studies* 25(8), 1339–1362.

Bailey, M. and Brady, D. (1998), “Heterogeneity and Representation: The Senate and Free Trade”, *American Journal of Political Science* 42(2), 524-544.

Bandiera, O. and Levy, G. (2007), “The Diminishing Effect of Democracy in Diverse Societies”, Manuscript.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T. and Birch, K. (1981), “Is Empathic Emotion a Source of Altruistic Motivation?”, *Journal of Personality and Social Psychology* 40(2), 290-302.

Bernhardt, D. Krasa, S. and Polborn, M. (2008), “Political polarization and the electoral effects of media bias”, *Journal of Public Economics*, 92(5-6), 1092-1104.

Byrne, D. (1961), “Interpersonal Attraction and Attitude Similarity”, *Journal of Abnormal Social Psychology* 62, 713-715.

Byrne, D. (1971), *The Attraction Paradigm*, New York: Free Press.

Central Intelligence Agency (2005), *World Fact Book 2005*, Washington, DC.

Chen, F. and Kenrick, D. (2002), “Repulsion or Attraction? Group Membership and Assumed Attitude Similarity”, *Journal of Personality and Social Psychology* 83(1), 111-125.

Cukierman, A., Edwards, S. and Tabellini, G. (1992), “Seignorage and Political Instability”, *American Economic Review* 82(3), 537-555.

Deaton, A. (2003), “Health, Inequality and Economic Development”, *Journal of Economic Literature* 41(1), 113-158.

DellaVigna, S. and Kaplan, E. (2007), “The Fox News Effect: Media Bias and Voting”, *Quarterly Journal of Economics*, 122(3), 1187-1234.

DiMaggio, P., Evans, J. and Bryson, B. (1996), “Have Americans’ Social Attitudes Become More Polarized?”, *American Journal of Sociology* 102(3), 690—755.

Dixit, A. and Weibull, J. (2007), “Political Polarization”, *Proceedings of the National Academy of Sciences* 104, 7351-7356.

Esteban, J-M. and Ray (1994), “On the Measurement of Polarization”, *Econometrica* 62(4), 819-851.

European Values Study Group and World Values Association (2004), *European and World Values Surveys Integrated Data File 1999-2002*, RELEASE I, 2nd ISCPSR version.

European Values Study Group and World Values Association (2006), *European and World Values Surveys Four-wave Integrated Data File 1981-2004, version 20060423*, downloaded from www.worldvaluessurvey.org.

Feren, D., Carroll, S. and Olian, J. (1988), "Effects of Managerial Performance and Attitudinal Similarity on Interpersonal Attraction", *Basic and Applied Social Psychology* 9(1), 33-44.

Fernández, R. and Levy, G. (2008), "Diversity and Redistribution", *Journal of Public Economics*, forthcoming.

Freedom House (2005), *Freedom in the World Country Rankings 1973-2005*, Washington, DC. Data retrieved from www.freedomhouse.org.

Frye, T. (2002), "The Perils of Polarization: Economic Performance in the Postcommunist World", *World Politics* 54(3), 308-337.

Fukuyama, F. (1995), *Trust: The social virtues and the creation of prosperity*, Free Press, New York.

Gerber, E. and Lewis, J. (2004), "Beyond the Median: Voter Preferences, District Heterogeneity, and Representation", *Journal of Political Economy* 112(6), 1364-1383.

Gerber, A., Karlan, D. S. and Bergan, D. (2006), "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions", *Yale Economic Applications and Policy Discussion Paper* No. 12.

Gradstein, M. and Justman, M. (2002), "Education, Social Cohesion, and Economic Growth", *American Economic Review*, 92(4), 1192-1204.

Goff, B. and Grier, K. (1993), "On the (Mis)measurement of Legislator Ideology and Shirking", *Public Choice* 76, 5-20.

Gwartney, J. and Lawson, R. (2005), *Economic Freedom of the World: 2005 Annual Report*, Fraser Institute, Vancouver. Data retrieved from www.freetheworld.com.

Hamilton, W. D. (1964), "The Genetic Evolution of Social Behaviour", *Journal of Theoretical Biology* 7, 1-52.

Harrison, D., Price, K. and Bell, M. (1998), "Beyond Relational Demography: Time and the Effects of Surface- and Deep-level Diversity on Work Group Cohesion", *Academy of Management Journal*, 41(1), 96-107.

Jehn, K., Chadwick, C. and Thatcher, S. (1997), "To Agree or Not To Agree: The Effects of Value Congruence, Individual Demographic Dissimilarity, and Conflict on

Workgroup Outcomes", *The International Journal of Conflict Management* Vol. 8, No. 4 (October), 287-305.

Jehn, K., Northcraft, G. and Neale, M. (1999), "Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in Workgroups", *Administrative Science Quarterly*, 44, 741-763.

Jehn, K. and Mannix, E. (2001), "The Dynamic Nature of Conflict: A Longitudinal Study of Intragroup Conflict and Group Performance", *Academy of Management Journal*, Vol. 44, No. 2, 238-251.

Kalt, J. and Zupan, M. (1990), "The Apparent Ideological Behavior of Legislators: testing for Principal-Agent Slack in Political Institutions", *Journal of Law & Economics* 33, 103-131.

Knack, S. and Keefer, P. (1997), "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation", *Quarterly Journal of Economics* 112(4), 1251–1288.

La Porta, R., Lopez de Silanes, F., Shleifer, A. and Vishny, R. (1999), "The Quality of Government", *Journal of Law, Economics and Organization* 15(1), 222–279.

Ledyard, J. O. (1995). "Public Goods: A Survey of Experimental Research", Chapter in *Handbook of Experimental Economics.* Eds.: J. Kagel and A. E. Roth. Princeton: Princeton University Press, pp. 111-194.

Luttmer, E. F. P. (2001), "Group Loyalty and the Taste for Redistribution", *Journal of Political Economy* 109(3), 500–528.

Marshall, M. G. and Jaggers K. (2002), Political Regime Characteristics and Transitions 1800-2002, Polity IV Project, University of Maryland. Data retrieved from http://www.cidcm.umd.edu/inscr/polity/.

McGrath, J. E. (1984), *Groups: Interaction and Process*, Englewood Cliffs, NJ: Prentice-Hall.

Meltzer, A. H. and Richard, S. F. (1981), "A Rational Theory of the Size of Government", *Journal of Political Economy* 89(5), 914-927.

Montalvo, J, G. and Reynal-Querol, M, (2005), "Ethnic Polarization, Potential Conflict, and Civil Wars", *American Economic Review* 95(3), 796-816.

Mouw, T. and Sobel, M. E. (2001), “Culture Wars and Opinion Polarization: The Case of Abortion”, *American Journal of Sociology* 106(4), 913—943.

Newcomb, T. M. (1961), *The Acquaintance Process*, New York: Holt, Rinehart & Winston.

Oetzel, J. (1998), “Explaining Individual Communication Processes in Homogeneous and Heterogeneous Groups Through Individualism-Collectivism and Self-Construal”, *Human Communication Research*, 25(2), 202-224.

Olson, J. M., Vernon, P. A. and Jang, K. L. (2001), “The Heritability of Attitudes: A Study of Twins”, *Journal of Personality and Social Psychology* 80(6), 845-860.

Park, J. and Schaller, M. (2005), “Does Attitude Similarity Serve As a Heuristic Cue for Kinship? Evidence of an Implicit Cognitive Association”, *Evolution and Human Behavior* 26, 158-170.

Persson, T. and Tabellini, G. (2003), *The Economic Effects of Constitutions*, The MIT Press, Cambridge, MA.

Putnam, R. (1993), *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton University Press, NJ.

Putnam, R. (2000), *Bowling Alone: The Collapse and Revival of American Community*, Simon and Schuster, New York, NY.

Reporters without borders (2002), *Worldwide Press Freedom Index 2002.* Data retrieved from www.rsf.org.

Reporters without borders (2003), *Worldwide Press Freedom Index 2003.* Data retrieved from www.rsf.org.

Rodrik, D. (1998), “Why Do More Open Economies Have Bigger Governments?”, *Journal of Political Economy*, 106(5), 997-1032.

Shalom, S. and Sagie, G. (2000), “Value Consensus and Importance”, *Journal of Cross-Cultural Psychology* 31(4), 465–497.

Suedfeld, P., Bochner, S. and Wnek, D. (1972), “Helper-Sufferer Similarity and a Specific Request for Help: Bystander Intervention During a Peace Demonstration”, *Journal of Applied Social Psychology* 2 (1), 17–23.

Svensson, J. (1998), “Investment, Property Rights and Political Instability: Theory and Evidence”, *European Economic Review* 42(7), 1317-1341.

Tesser, A. (1993), “The Importance of heritability in Psychological Research: The Case of Attitudes”, *Psychological Review* 100(1), 129-142.

Vigdor, J. L. (2004) “Community Composition and Collective Action: Analyzing Initial Mail Response to the 2000 Census.” *Review of Economics and Statistics* 86(1), 303–312.

World Bank (2005), *World Development Indicators 2005*, Washington, DC. Data retrieved from www.worldbank.org.