Procedural Justice, Social Norms and Conflict:

Human Behavior in Resource Allocation

EFI THE ECONOMIC RESEARCH INSTITUTE

EFI Mission

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution that works independently of economic, political and sectional interests. It conducts theoretical and empirical research in the management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics. At EFI, the researchers select their projects based on the need for theoretical or practical development of a research domain, on their methodological interests, and on the generality of a problem.

Research Organization

The research activities at the Institute are organized into 20 Research Centres. Centre Directors are professors at the Stockholm School of Economics.

EFI Research Centre:	*Centre Director:*
Management and Organization (A)	Sven-Erik Sjöstrand
Centre for Entrepreneurship and Business Creation (E)	Carin Holmquist
Public Management (F)	Nils Brunsson
Information Management (I)	Mats Lundeberg
Centre for People and Organization (PMO)	Andreas Werr
Centre for Innovation and Operations Management (T)	Pär Åhlström
Centre for Media and Economic Psychology (P)	Richard Wahlund
Centre for Consumer Marketing (CCM)	Magnus Söderlund
Centre for Information and Communication Research (CIC)	Per Andersson
Marketing, Distribution and Industry Dynamics (D)	Björn Axelsson
Centre for Strategy and Competitiveness (CSC)	Örjan Sölvell
Accounting and Managerial Finance (B)	Johnny Lind
Centre for Fin. Analysis and Managerial Econ. in Accounting (BFAC)	Kenth Skogsvik
Finance (FI)	Clas Bergström
Centre for Health Economics (CHE)	Magnus Johannesson
International Economics and Geography (IEG)	Mats Lundahl
Economics (S)	Paul Segerstrom
Economic Statistics (ES)	Anders Westlund
Centre for Business Law (RV)	Johnny Herre
Centre for Tax Law (SR)	Bertil Wiman

Chair of the Board: Professor Carin Holmquist
Director: Associate Professor Filip Wijkström

Address

EFI, Box 6501, SE-113 83 Stockholm, Sweden • Website: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

To Dominic, my best friend and brother

To my mother, for allowing me, from an early age, to grow into the person I wanted to be

And to Aduke, my true love – without you this would not have been possible

"I am only one; but still I am one.
I cannot do everything, but still I can do something.
I will not refuse to do the something I can do."
— Helen Keller

"Do or do not... there is no try."
— Yoda

Table of Contents

Introduction

Resource Allocation. The three chapters in this book all study the allocation of scarce resources among competing needs and wants. These allocation decisions are meaningful to the welfare of individuals and our society, and this focus lies at the core of the word "economics" in its ancient form: *οἰκονομία*[1]*,* more or less meaning "the wise management and allocation of resources." Within this realm, the chapters of this book span ideas from various epochs of human thought, testing ideas from Aristotle, Locke and de Tocqueville, and the very latest theories in current publications and also work done by the most recent Nobel Laureate, Elinor Ostrom.

Procedural Justice, Social Norms and Conflict. While all the chapters are unified under the sub-title *Human Behavior in Resource Allocation*, they each focus on a distinct element of the main title *Procedural Justice, Social Norms and Conflict.*

The first chapter, on *Procedural Justice,* provides one possible explanation for the large difference between American and Western European tax and redistribution levels.

The second chapter, on *Social Norms,* investigates two potential reasons why solutions to social dilemmas such as insurance systems, tax compliance, effort-shirk decisions at work, and contribution to irrigation systems can persist over time without being destroyed by the negative forces of free-riding.

In the last chapter, on *Conflict*, the focus is shifted to bilateral bargaining and the reasons for conflict and impasses. Whether they manifest as strikes, job resignations, or trade embargoes, these failures of the negotiation process create tremendous loss of social welfare.

Human Behavior. During my early undergraduate studies I felt that economics, as a science, was in need of development. From everyday experience, I was convinced that mankind is much more than the perfectly rational maximizer of own material wellbeing that is captured in the notion of 'Homo Economicus'. When I first was introduced to the wave of thought-

[1] From the two Greek words *oikos* or "house" and *nomos* or "law, rule, custom."

provoking research that integrates insights on social preferences, cognitive and evolutionary psychology, and sociology into economic theory, I knew I had discovered the area in which I wanted to do research.

The two first chapters of this book fit soundly into this field, behavioral economics. However the third chapter is a summary of the first phase in a two-part project; behavioral economics will play a larger role in the second part of the project. For this reason, I have chosen to present this chapter last although it likely is the strongest research piece in this book.

Empirical data. Another common element among these papers is their shared basis in observations of real human behavior in the lab. The total set of empirical data consists of:

- 204 M.B.A. students and 96 M.Sc. students from Harvard University, the Stockholm School of Economics, the Royal Institute of Technology, Stockholm University and Karolinska Institutet
- 5 experiments
- 21 experimental sessions
- 2 520 observations

Procedural Justice:
Chapter 1 - Luck, Effort and Redistribution

This chapter attempts to answer questions such as: Why is redistribution so much lower in the U.S. than in Western Europe, where tax rates are more progressive and the overall size of government is about 50 percent larger? Why is this the case, when pre-tax inequality is higher in the US than in Europe?

I argue that part of the answers to these questions could perhaps be found in Aristotle's theory of proportionality and John Locke's concept of natural law/desert, which both would suggest that people perceive income differences to be fair when driven by personal effort, but unfair when due to forces such as luck. If people have fairness preferences not just for outcomes, but also with regards to *how* those outcomes are reached—procedural justice concerns—then beliefs about what determines income and success in a society could have a large impact on the level of redistribution in that society.

Aristotle's proportionality and Locke's desert have been developed in procedural justice and equity theory, which are important among philosophers, sociologists, psychologists and political scientists. These concepts have rarely been tested experimentally with incentive compatible real stakes, and have so far received little focus in the economics literature focused more on outcomes than the processes that lead to the outcomes.

I find experimentally that preferences for procedural justice are, in fact, strong and that these effects have a large impact on participant's preferences for redistribution. In line with the theories of proportionality and natural law/desert, people act as if they perceive income differences to be fair when driven by personal effort, but unfair when due to external forces such as luck. Interestingly, income differences due to ability appear to be as fair as those due to effort.

When the only exogenous difference is the role of luck versus effort, people from all regions make relatively similar decisions – higher redistribution when income is decided by luck, lower when based on effort.

The above is interesting in relation to the fact that the Americans (Europeans) in the study have real world ideological views directionally in line with the lower (higher) level of redistribution and taxes in the American (European) society.

Together the findings suggest that one potential reason for differing redistribution levels in the US and Europe, could be differences in citizens' beliefs about the relative importance of effort and luck in determining success in their society.

While outside the scope of the study, in the conclusion of this chapter I discuss the fact that if luck actually plays a smaller role in determining individual prosperity in Western Europe than in the US (as several studies suggest), it could be the case that one or both of the two populations has biased beliefs. The two very different redistribution levels in the US and Europe could then both be rational given people's beliefs; however *given biases in beliefs, one or both of the regions' tax and redistribution level decisions could nonetheless be significantly suboptimal* relative to the true underlying preferences of the population(s).

Social Norms:
Chapter 2 - Participation and Peers in Social Dilemmas

In this second chapter I build on the notion from empirical data, experimental research and everyday knowledge that people, when in contexts of social dilemmas, are strongly motivated by concerns additional to those of material payoffs.

I look at two potential reasons for why solutions to social dilemmas such as insurance systems, tax compliance, effort-shirk decisions at work, contribution to irrigation systems and usage levels of offshore fisheries can persist over time without being destroyed by the negative forces of free-riding. The two questions I study are (1) whether individuals are more willing to abide by rules that they have participated in creating or voted over and (2) whether people are more willing to abide by rules if a larger fraction of those around them also abide, than when a lower proportion does.

In contrast to other literature, voting over or having a higher degree of participation towards, in this case, an insurance system does not have any positive impact on the levels of free-riding. However, previous trends in the extent to which peers misuse or do not misuse the system are found to have a strong effect on future levels of free-riding - given the same economic incentive, people are more likely to lie when they believe that others will lie.

In this chapter I also find indicative support for the so called false consensus effect - that people who engage in a given behavior believe that the behavior is more prevalent than those who do not. This is well discussed in psychology, but the concept does not (yet) play a large role in economics.

Conflict:
Chapter 3 – Commitment and Impasses in Negotiation

Whether they manifest as strikes, job resignations, or trade embargoes, impasses in negotiation are Pareto-inefficient and create tremendous social welfare losses. Hence, it is important to understand why impasses occur.

Yet so far, the existing literature primarily emphasizes only three reasons why impasses occur: due to incomplete information, bounded rationality, or fairness concerns. But, as Schelling argued more than 50 years ago, one significant reason for conflict in negotiation could be that people make aggressive commitments in order to increase their share of the pie, although they know that such commitments can lead to impasses. This is at odds with the general literature's requirement of incomplete information or irrationality for there to be impasses.

While versions of Schelling's argument have appeared in formal theory, in Crawford (1982) and Ellingsen & Miettinen (2008), I am not aware of any experimental or empirical studies on any of the three theories.

Thus I decided to study games of bilateral bargaining where negotiators can make commitments attempts in the lab. I find strong support for the conjectures of Shelling and predictions based on Crawford. Some support, albeit weaker, exists for the primary predictions based on Ellingsen & Miettinen; support for their secondary predictions is strong.

In sum, the results suggest that ex ante, conflict (negotiators making demands that could lead to impasse) is more likely to occur where commitment devices are weak. In these domains I frequently observe outcomes in which one party takes most of the pie and the other receives little. Although the reasons are distinct, uneven outcomes are also common in domains of strong commitment devices.

These results run counter to the typical predictions in most formal theory that more symmetric "compromise outcomes" will be realized, and as Ellingsen & Miettinen discuss, this can perhaps help explain the infrequency of side payments in real life agreements

Finally we find that the level of welfare destruction is much higher than most formal theory would predict. Welfare destruction is around 60% of what my predictions based on Crawford suggests, but only 15-30% of what than my predictions based on Ellingsen & Miettinen theories suggest.

These results may be relevant for understanding complex, multiparty international negotiations such as the Israeli-Palestinian peace process or attempts to create environmental and trade agreements like Kyoto 2 and a Doha agreement. At the same time, these insights apply to everyday, one-on-one negotiations between children and parents, or workers and employers, or between two firms.

—

Now that you have a full overview of why this book is entitled *Procedural Justice, Social Norms and Conflict* and how the component chapters relate to *Human Behavior in Resource Allocation,* you should feel to read the chapters in whichever order you desire.

Acknowledgements

I have conducted most of this work while juggling other full-time commitments. As such, the value of the support, critique, and comments I have received throughout my research has been particularly important. While I probably cannot do justice in expressing the tremendous gratitude I feel to the key people I am grateful to, it is important for me that I at least try.

First, I would like to thank my main supervisor (and possibly future co-author) Magnus Johannesson. Without the many insightful discussions challenging my ideas, and the many rounds of feedback on my draft papers, this thesis would not be close to what it has finally become. Magnus, without your belief that my lack of time was not the same as a lack of commitment and motivation, I would likely not have been allowed to conduct this research on a part-time basis. Without your beliefs and subsequent recommendations of my potential and ability, I would probably not have had the opportunity to spend time at the *Department of Economics at Harvard University* and the *Harvard Business School*, to attend classes at *MIT*, or to be a participant in the ENABLE *Summer Institute on Economics and Psychology* and the *Seventh Russell Sage Summer Institute in Behavioral Economics* – all of these experiences significantly increased the quality of my research. Thank you so much for believing in me!

Papers underlying two chapters in this book were also co-supervised by professors at Harvard: Deepak Malhotra (Luck, Effort and Redistribution) and Guhan Subramanian (Commitment and Impasses in Negotiation). Their challenging of my ideas, feedback, and critique was extremely valuable.

Tore Ellingsen provided very important feedback and input throughout the initial, middle and final stages of the work on Commitment and Impasses in Negotiation. Fort this I am very thankful.

An earlier study on the topic in Participation and Peers in Social Dilemmas was part of my undergraduate thesis, supervised by Professor Martin Dufwenberg (then at *Stockholm University*, now Director of *The Institute for Behavioral Economics* at *the University of Arizona*). Martin, I remain thankful for your supervision. I also am indebted to you for enabling me to become a teaching assistant while still early in undergraduate studies and for supporting my application to begin the Ph.D. program one year earlier than normal.

Two colleagues and friends, Erik Lindqvist and Robert Östling have provided especially valuable input and feedback on all three of my papers. Both also provided enormous help when I was preparing for exams, taking course work, and at the same time often was working full-time outside of academia.

I am also grateful to Victor Corzo, my co-author on the study preceding Participation and Peers in Social Dilemmas, both for letting me further develop our initially shared idea by myself, and for providing valuable feedback on the new study I undertook.

For inviting me to spend time at the *Department of Economics* at *Harvard University*, I am grateful to David Laibson and Alvin Roth.

I am also grateful to the following people for providing valuable comments and assistance:

- Topi Miettinen, Geir Asheim, Damian Migueles, Jesper Roine, Chris Dawes, James Fowler, Lakshmi Iyer, Rafael Di Tella,
- Seminar participants at the *Stockholm School of Economics* (in 2006 and 2009),
- Seminar participants at the *Swedish Research Institute of Industrial Economics,*
- Seminar participants at the *Institute for International Economic Studies at Stockholm University*
- Several course members in Max Bazerman's *Behavioral Approaches to Decision Making and Negotiations Seminar* at Harvard,
- Matthew Rabin, Colin Camerer, Ernst Fehr and other participants at the *Seventh Russell Sage Summer Institute in Behavioral Economics*,
- Participants at the *European Network for the Advancement of Behavioral Economics' Summer Institute on Economics and Psychology* in Munich.

I thank Anneli Sandbladh and Lilian Öberg and of course Carin Blanksvärd and Ritva Kiviharju for outstanding and friendly administrative support.

I also enjoyed and am thankful for haven been given the opportunity to teach seminars in Richard Friberg's micro economics course.[2]

[2] For which I was voted *3rd place winner* in the *2005 Teacher of the Year Competition* alongside tenured professors.

A large debt is also owed to my doctoral classmates: Erika Färnstrand Damsgaard, Ola Melander, Björn Tyrefors, David Cesarini, Milo Bianchi, Anna Breman, Martin Mossberg, Roine Vestman, Anna Dreber Almenberg, Ola Granström, Kristin Magnusson, Per Sonnerby, Anders Åkerman and especially Martin Bech Holte.

Family and friends have in many ways supported me in this effort, not least in agreeing to be test pilots of my five different experiments: my mother Ylva Eriksson, brother Dominic Eriksson Giwa, and the Thelwell/Richardson family and my friends Christian Sommer, Mikael Grynblat, Frida Hobroh, Emilio Dauvin, Lina and Johan Zander, Jonathan Kipowsky, Åsa Bjering, Jonas Klang, Martin Winiarsky, Matthias Rubin de Lima, Eric Thunander, and more than twenty friends from Harvard Business School helped out.

I could not have executed the experimental sessions without the co-experimental leadership of Soledad Piñero Misa, Lauryn Hale, Carl-Richard Häggman and Brooke Jenkins. I also express gratitude to the exactly 300 experiment participants.

I would also like to thank Marcus Dahlgren and DRAGNET (www.dragnet.se) for excellent programming support.

Financial support from the following institutions is appreciatively acknowledged:

- the *Jan Wallander and Tom Hedelius Foundation*,
- the *Dr Marcus Wallenberg Foundation*,
- the *Swedish Finance Political Research Institute* and
- the *European Network for the Advancement of Behavioural Economics* (ENABLE)

Although she entered later than most of the people mentioned above, the level of feedback, quality of suggestions and never-ending support I have received from Aduke Thelwell over the past year has been invaluable. Aduke, the way I accelerated the work on this thesis would not have been possible without you.

New York, October 2009

Sebastian

Chapter 1

Luck, Effort and Redistribution[1]

"The labour of his body, and the work of his hands, we may say, are properly his"

John Locke

We find experimentally that preferences for procedural justice – for how an end outcome is reached, not just what the end outcome is – can be strong.

This has a large impact on participant's preferences for redistribution. In line with Aristotle's theory of *proportionality* and John Locke's concept of *natural law/desert*, people act as if they perceive income differences to be fair when driven by personal effort, but unfair when due to external forces such as luck. Income differences due to ability appear to be as fair as those due to effort.

When the only exogenous difference is the role of luck versus effort, people from all regions make relatively similar decisions – higher redistribution when income is decided by luck, lower when based on effort.

The above is interesting in relation to that the Americans (Europeans) in the study have real world ideological views directionally in line with the lower (higher) level of redistribution and taxes in the American (European) society.

Together the findings suggest that one potential driver of the different redistribution levels between the US and Europe could be differences in beliefs about the relative importance of effort and luck in determining success in society.

[1] I am grateful to Magnus Johannesson, Aduke Thelwell, Deepak Malhotra Martin Bech Holte for invaluable help and critique. I thank Colin Camerer, Ernst Fehr and participants at the Seventh Russell Sage Summer Institute in Behavioral Economics as well as Erik Lindqvist, Jesper Roine, Robert Östling, Chris Dawes, James Fowler, Lakshmi Iyer, Rafael Di Tella and also seminar participants at the Institute for International Economic Studies at Stockholm University for valuable comments.

Part of this project was carried out when I was a visiting doctoral student at the Department of Economics at Harvard University and later when I was a graduate student at the Harvard Business School.

Financial support from the Jan Wallander and Tom Hedelius Foundation, the Dr Marcus Wallenberg Foundation and the Swedish Finance Political Research Institute is appreciatively acknowledged.

Finally, I thank Lauryn Hale for valuable help in executing the experimental sessions and Marcus Dahlgren and DRAGNET (www.dragnet.se) for excellent programming support.

I – Introduction

Why citizens in democratic countries support or oppose redistribution is still inadequately understood. For instance, one riddle is why redistribution is so much lower in the U.S. than in (Western) Europe, where tax rates are more progressive and the overall size of government is about 50 percent larger. Why is this the case, when pre-tax inequality is higher in the US than in Europe? The pre-tax Gini coefficient is above 40 in the US while around 30 in Europe.[1]

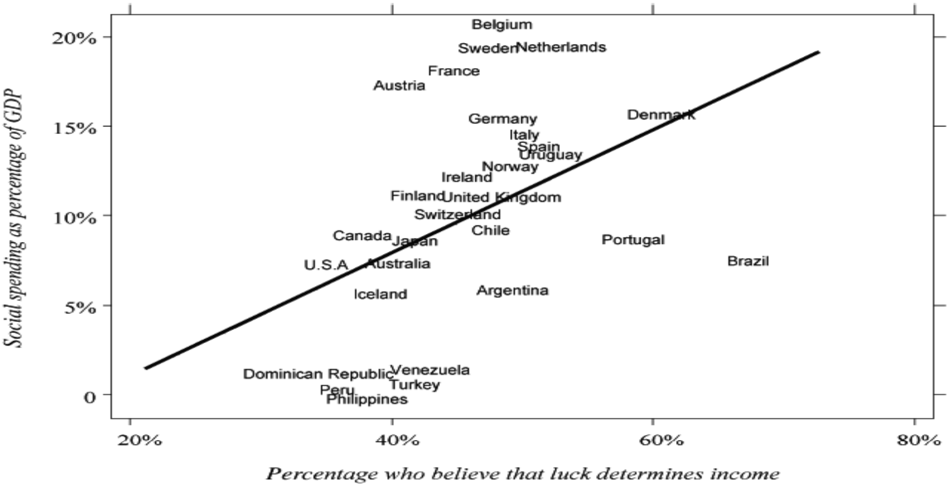
Part of the answer could perhaps be found in Aristotle's theory of *proportionality* and John Locke's concept of *natural law/desert*, which both would suggest that people perceive income differences to be fair when driven by personal effort, but unfair when due to forces such as luck.

If people have fairness preferences not *just for outcomes*, but also with regards to *how those outcomes are reached*—procedural justice concerns—then *beliefs* about what determines income and success in a society could have a large impact on the level of redistribution in that society.

One indication that this could be the case, is the strong correlation between how much a country spends on social programs and its citizen's beliefs about whether luck/connections or effort/ability determine income (see *Figure 1*).

[1] The paradox of high (low) pre-tax inequality together with low (high) redistribution in the U.S. (Europe) is not in agreement with either Mirrlees' (1971) social insurance paradigm or Meltzer-Richard's (1981) redistribution paradigm.

Figure 1 (Reproduced from Alesina & Angeletos [2005])



As argued in Fong (1999) and Alesina & Angeletos (2005), these correlations are robust to controlling for countries' Gini coefficients, per capita GDPs, continent dummies and political variables, among other things.[2] However, as discussed by Fong, the correlations could be spurious and there is little evidence of causality.

Note that the *x*-axis above is a measure of people's *beliefs* about what determines income and success in a society, not a measure of what actually does determine these outcomes. Many argue that today luck, family and connections play a smaller role in Western European and especially Scandinavian countries than in the US (see for instance Erikson & Goldthorpe [1992], Arrow, Bowles & Durlauf [2000] and Jäntti et al [2006]). In this context, the question of whether Europeans and Americans have

[2] Similar patterns to the above and support for the view that people see inequality as more acceptable when it is due to hard work (rather than luck or connections), can also be found in the Gallup pole used in Fong (2001), the General Social Survey, the International Social Survey and the International Social Justice Project. Jencks & Tach (2006) report that a majority of people in both high tax/redistribution countries, such as Germany and Japan,

different procedural justice concerns is important, since such differences could be a driver of different redistribution regimes, even if the beliefs discussed above are biased[3].

Aristotle's proportionality and Locke's desert have been developed in procedural justice and equity theory, which are important among philosophers, sociologists, psychologists and political scientists[4]. However these concepts have barely been tested experimentally with incentive compatible real stakes, and have so far received much less focus in the economics literature, which focuses more on outcomes than the processes that lead to the outcomes[5].

We propose an experimental design in which we exogenously vary subjects' beliefs about the extent to which luck versus effort versus ability determine an initial ("pre-tax") income. We then, in an incentive compatible way and behind a "veil of ignorance", let subjects decide on the level of redistribution they want. We aim to shed light on any *preferences for procedural justice* and the potential *causal effect* they could have on redistribution levels.

Four experimental sessions of the experiment were executed at the Harvard Business School during April and May 2009, generating a total of 528 observations.

---

and in low tax/redistribution countries, such as the U.S., agree with the statement that "[inequality] is fair but only if there are equal opportunities."

[3] See also Footnote 6 below.

[4] See for instance Aristotle (1925), Locke (1978), Rawls (1971), Nozick (1974), Buchanan (1986), Leventhal & Michaels (1971), Dworkin (2000), Heider (1958), Rotter (1966), Weiner & Kukla (1970), Homans (1958), Adams (1965), Walster, Walster & Bercheid (1973) and Schokkaert & Overlaet (1989), some of which we will revisit in the section of related theory and research.

[5] See for instance Konow's (2003) positive analysis of justice theories. Exceptions are the attempts by Hoffman & Spitzer (1985) and Burrows & Loomes (1994) which will be discussed below.

We find strong support that procedural justice concerns can have a large and highly significant impact on people's preferences for redistribution. The redistribution rates chosen by subjects are twice as high when initial income is decided by luck than when due to effort. The subjects also behave as if income differences due to ability are as fair as those due to effort.

We find no support for Americans in the study having stronger desert preferences than Europeans. When the only exogenous difference is the role of luck versus effort, people from all regions make similar decisions – higher redistribution when income is decided by luck, lower redistribution when based on effort and/or ability.

The above is interesting in light of the fact that the Americans (Europeans) in the study have real world ideological views that mirror the lower (higher) level of redistribution and taxes in the American (European) society.

Together our evidence suggests that the vast difference in redistribution levels between the US and Europe could partially be due to differences in beliefs about the relative importance of effort and luck in determining a person's income and success in the two societies.

Our results provide support for the relevance and importance of incorporating procedural justice concerns in formal theory and models such as the one presented in Piketty (1995) and Alesina & Angeletos (2005).[6]

---

[6] Alesina & Angeletos (2005) argue that the vastly different tax-rates and redistribution levels in the US and in Europe could both be correct equilibria. However, there is also an alternative story – which if true, suggests very large welfare losses for at least one of the two societies.

Given that luck might actually play a smaller role in Western Europe than in the US, it could be the case that one or both of the two populations has biased beliefs. The two very different equilibrium redistribution levels in the US and Europe could then both be rational given people's beliefs, but given the biases in beliefs, one or both of the regions' tax and redistribution regimes could nonetheless be significantly suboptimal in relation to the true underlying preferences of the population(s).

The rest of this paper is organized as follows. In *Section II* we summarize the relevant theories, discuss previous empirical results and formulate our main hypotheses. *Section III* goes through the experimental procedure and design. *Section IV* presents our results. Finally in *Section V*, we conclude and discuss some ideas for future research.

Appendix i contains key parts of experimental instructions and the exit survey, *Appendix ii* and *Appendix iii* provide additional description of the data and regression analysis.

II – Theory, Predictions and Previous Evidence

At least as early as Aristotle with his theory of *proportionality* and John Locke with his concept of *natural law/desert*, philosophers have proposed forms of earnings-based justice. These and subsequent theories have been united by the supposition that fair distributive outcomes must depend on individuals' actions. Aristotle is often seen as the forefather of equity theory and Locke as the forefather of natural law/desert theory (Konow 2003).

Desert theories

The essence of Lockean Desert theory is the notion that, under natural law, an individual deserves and is (only) entitled to any allotment of resources that have been amassed or multiplied through that individual's own efforts, so called 'earned' or 'deserved' desert.

> *"The labour of his body, and the work of his hands,*
> *we may say, are properly his."*
>
> John Locke (1690),
> The Second Treatise of Civil Government,
> Chapter 5, Section 27

Some desert theories suggest broader definitions of what determines fair outcomes. Buchanan (1986) looks beyond effort to identify luck, choice and birth as three other factors which also determine the allocation of economic income across populations. He agrees with Locke on the fairness of resources acquired by effort, and extends this notion of fairness to rewards from luck and choice also. In Buchanan's view, birth is the only driver of success which clashes with common views of justice (pp 129-30).[7]

---

[7] Nozick (1975) goes several steps further to focus on the principle of justice in acquisition, or how resources came to be initially controlled. In his view, the multiplication or increase of

At the other end of the spectrum, we have the social psychology theory called *attribution theory* (Heider [1958], Rotter [1966] and Weiner & Kukla [1970]), in which people hold others accountable *only* for the factors in determining outcomes that an individual can control. This usually means that individuals will be evaluated based on their effort, but not on birth or luck.

In a similar spirit, Rawls (1971) argues that success should be deserved or earned as a reward for an individual's actions, but only to the extent that the individual deserves the conditions that made such action possible. In this line of reasoning, individuals do not deserve all the rewards from their natural ability, since ability is determined before actions occur and cannot be truly deserved.

Equity theory

In the fourth century B.C., Aristotle formulated a theory of justice based on proportionality. This became the root of equity theory of sociologists and social psychologists such as Homans (1958), Adams (1965) and Walster, Walster, & Berscheid (1973). In equity theory, as in Aristotle's *Nicomachean Ethics*, the proposition is usually expressed as:

$$\frac{O_1}{I_1} = \frac{O_2}{I_2}$$

where 1 and 2 are two persons, O stands for the fair outcomes/consequences to receive and I stands for the inputs/contributions (Konow 2003). The key suppositions of this theory are that people are not only self-interested, but are also motivated to create perceived equity by rewarding (penalizing) others for just (unjust) behaviors in social interactions.

economic resources is only unfair if those resources where originally the rewards of fraud, robbery, enslavement or forced exclusion from competitive trade.

In this articulation of the theory, the generality explains everything and hence nothing, and despite a long history of having wide currency, the predictions of equity theory have rarely been tested with experimental rigor (Konow 2003).

Accountability principles

There is an area of possible overlap between equity theory and attribution theory, and this space has been fertile ground for a growing number of social scientists. As one example, equity theorists Leventhal and Michaels (1971) proposed a theory in which they narrowed the set of inputs relevant to fair outcomes and distinguished them based on the degree of control an individual could have over these inputs. In these approaches, input factors are categorized into discretionary factors that an individual can influence (for example, effort), and exogenous variables that are externally determined (such as luck and ability).

Predictions

Based on the procedural justice theories discussed above, we formulate the following predictions that we aim to test in the lab (some of them exclude each other):

Strong predictions on fairness of luck and effort

$P^{\text{Luck should not yield desert}}$	Preferences for redistribution are positively correlated to the extent that *luck* determines income
$P^{\text{Effort based desert}}$	Preferences for redistribution are negatively correlated to the extent that *effort* determines income

Weaker predictions on fairness of ability

$P^{\text{Ability should not yield desert}}$	Preferences for redistribution are negatively correlated to the extent that *ability* determines income
$P^{\text{Weak ability based desert}}$	Preferences for redistribution are negatively correlated to the extent that *ability* determines income, but to a lesser extent than when effort determines income
$P^{\text{Strong ability based desert}}$	Preferences for redistribution are negatively correlated to the extent that *ability* determines income, to roughly the same extent that they are when effort determines income

Ancillary predictions

In addition, we are interested in gaining further understanding of the extent to which the differences in redistributive preferences between Americans and Europeans (discussed in the Introduction) could be driven by differences in beliefs and/or differences in desert preferences.

If people in the two regions have the same underlying preferences – for example, that inequality due to effort is fair, but inequality due to luck is unfair – and only differ in beliefs about the extent to which luck versus effort determine income we can make the following prediction:

When the only thing that is varied in the controlled lab setting is the extent to which luck/effort/ability determines income, Americans and Europeans should make the same redistribution decisions.

Previous Evidence

At a general level, fairness and justice concepts are used to improve theoretical and empirical analysis in many fields of economics[8]. However, little empirical or experimental evidence exists of how strong people's potential concerns for procedural justice are, and the scant evidence that does exist is inconclusive.

The evidence on people's concerns for procedural justice is mostly from other social sciences and relies often on vignettes[9] and attitude surveys[10]. However, such studies do not provide good behavior-based measures of preferences, since they are not based on individuals' willingness to act when stakes are involved. Also, in the case of survey data, it is often difficult to ensure causality behind patterns of correlations (see for example, the above discussion of problems regarding the econometric work by Fong [1999] and Alesina & Angeletos [2005]).

Good examples of the value of using experimental economics to test other theories of distributive justice exist in which stakes are involved (for example, Frohlich et al. [1987a,b,] Frohlich & Oppenheimer [1990], Lissowski et al. [1991], and Beckman and Smith [1995]). These experiments indicate the strength of experimental designs based on veil of ignorance settings. However, the studies also use group decision-making mechanisms that limit the power of the designs, since they do not fully allow the observer to study preferences at the level of the individual.

---

[8] See for example Rabin (1993), Güth & Tietz (1990), Kahneman, Knetsch & Thaler (1986), Akerloff & Yellen (1990) and Fehr & Schmidt (1999).

[9] See for instance Leventhal & Michaels (1971), Kluegel & Smith (1986) and Overlaet (1991).

[10] See for instance, Schokkaert & Lagrou (1983), Ordóñez & Mellers (1993) and Farwell & Weiner (1996). See also Footnote 2 for further sources.

The studies that come closest to testing potential differences in redistributive preferences when income is determined by chance versus by effort are the studies by Hoffman and Spitzer (1985), Burrows & Loomes (1994), Clark (1998), Rutström & Williams (2000) and Ball et al (2001).

Both Hoffman & Spitzer (1985) and Burrows & Loomes (1994) study the effects of random versus earned initial entitlement allocations on individual choices. Their predictions are based on Locke's desert theory discussed above, and they hypothesize that earned entitlements will be treated as more fair than randomly allocated ones.

Hoffman & Spitzer's (1985) design is based on a two-person bargaining situation where one agent either earns the role of "controller" through *performance* on a task or is assigned the role through a *random* chance mechanism. The controller can unilaterally invoke and decide the no-deal outcome (given constraints). The task used to induce effort is a hash-mark game. Hoffman & Spitzer interact these effort/luck treatment variables with "moral authority" and "no moral authority": the difference between whether the instructions and experiment design specifically emphasize that the person assigned the role as "controller" has earned it.

While Hoffman & Spitzer find strong and significant differences between "moral authority" versus "no moral authority", the picture is less clear with regard to the random versus effort treatments. There is no significant difference between the pure random and pure effort treatments. However, conducting ancillary analysis and pooling data from both moral authority and no moral authority treatments, Hoffman & Spitzer are able to provide weakly significant (at the 10% level) support for their Lockean prediction.

While the design by Hoffman & Spitzer is good for studying the moral authority predictions, the same design is in many ways problematic for testing their Lockean prediction. While we will mention some of the problems below, Burrows & Loomes (1994, p. 206-208) provide an excellent overview of several other misgivings they have with Hoffman & Spitzer's design.

Burrows & Loomes (1994), while trying to address shortcomings by Hoffman & Spitzer (1985), nevertheless design an experiment that in many ways is very similar. They again use two-person face-to-face bargaining over trades that can create mutual benefits, but where one person has been assigned a roll as "controller". To control for between subject effects the participants take part in both a treatment in which the role of controller is assigned randomly (Stage 1), and then a second treatment where the role of controller is assigned through "effort", determined by performance on a word search task (Stage 2). Burrows & Loomes' results do not confirm Hoffman & Spitzer's conclusions about the weak effects of Lockean desert preferences. However, they propose a modified theory of two-part desert.

Burrows & Loomes find that when entitlements are earned, the share of final outcomes yielding equal resource splits goes down versus when they are random, however the proportion of equal splits of the gains from trade also increases. Based on this finding, they coin the concept of a "two-part Lockean desert" that derives not only from the effort that produces the initial entitlements, but also from exerting effort in the bargaining process.

It is problematic that the designs in Hoffman & Spitzer (1985) and Burrows & Loomes (1994) both are based on two-person, face-to-face communication, in asymmetric bargaining situations where one person is given both the initial income entitlement and the decision-making power for no-deal outcomes. As we know from the general fairness literature, people are more inclined to making equal splits of final joint payoffs when groups are small, negotiations are face-to-face and payoffs are made in public. Their designs hence pose problems of both interaction effects with other drivers on non-selfish behavior and interpreting the robustness of any results to more anonymous situations with more social distance.

In Burrows & Loomes the effort treatment is always played by subjects that already have played in the random treatment in Phase 2. Therefore, the effects could be due to learning and not a difference between the treatments, as discussed by Burrows & Loomes themselves.

Additionally, these studies share one more characteristic that is problematic. The time it takes to assign both the initial income entitlement and the decision-making power differs between the effort and chance treatments. This means that any potential effects very well could be driven by the so called endowment effects (see for example Kahneman, Knetsch & Thaler [1990]).

Ball et al. (2001) provide interesting results that go against the results of Hoffman & Spitzer and Burrows & Loomes, but note that their study was primarily designed to study status effects, not procedural justice. While they have one treatment based on randomization and one based on performance on a quiz ("effort"), these are not treated separately from potential interaction with and crowding out from status[11].

[11] The experimental leader goes to "great lengths" to convince subjects of the status difference, including holding award ceremonies where everyone claps for the awardees, providing preferential treatment throughout the experiment, and designating symbolic names and badges of stars.

Keeping in mind that the experiment was not designed to test for desert effects, but for status, Ball et al. make the following conclusion with regards to their random versus quiz treatments (p. 177):
"The effect of status was slightly stronger when it was randomly allocated than when it was awarded … contradicts Hoffman & Spitzer (1985) who show that an earned advantageous role made subjects more willing to exploit their opponents, than a randomly assigned advantageous role".
However, Ball et al. conjecture that this effect was due to the fact that the quiz asked the subjects questions that "we expected them not to know the answers to" and that the subjects "likely considered the test to be unfair." This could naturally crowd out any desert preferences[12].

Outside the domain of two-person bargaining experiments, Clark (1998) finds some suggestive support that members of a group are less likely to redistribute from the "rich" to the "poor" when initial income is assigned based on performance on a quiz, than when initial income is randomly assigned. While Clark picks up this pattern in two distinct ways, it is never significant at conventional levels (with values $P = 0.31$ and $P = 0.49$). Note however, that these results were ancillary and the experiments were not primarily designed to test for desert preferences.

Rutström & Williams (2000) put together a study where they try to distinguish between not just effort/ability versus chance but also between effort and productivity (ability). However, their study is primarily a test of whether individuals have preferences over distribution that differ from own-payoff maximization. Their dominant result is that high income individuals

[12] Such perceived unfairness might even go against Nozick's very limited principle of justice in acquisition, due to the unfair nature of the competition. See Footnote 7.

chose no redistribution and low income individuals chose perfect equality. With regards to earnings-based preferences, they conclude that they can see some weak indication, but not to an extent sufficient enough to reject self-interest.

Given the timing of making redistribution decisions which are made ex post, the study has problems with the following confounds: (a) preferences to respect property rights on the entitlements/endowment effects, as discussed above and (b) self-serving biases (see, for example, Babcock et al [1995]) and pretension of fairness/ self-deception (see, for example, Ellingsen & Johannesson [2005]).

III - Experimental Procedure and Design

Procedure

The experiment was conducted at the Harvard Business School, where four sessions were carried out in April and May 2009. A total of 88 subjects were recruited from the first and second year classes of the MBA program, generating 528 observations across the 6 treatments that each subject participated in. No one participated more than once in the experiment. See Table 1a for demographic data.

Table 1a - Summary Statistics - Demographics

	Gender		Age			Undergraduate Education				
	Female	Male	Lower Qtr	Med- ian	Upper Qtr	Busi-ness	Econ	Social Sci/ Human	Engin./ Natural Sci	Other
Experiment 1	26%	71%	26 yrs	27 yrs	28 yrs	22%	24%	14%	39%	2%
Experiment 2										

	Primary Region Lived In			Avg. Time in Region		Year at HBS	
	US	Europe	Other	US	Europe	1st year	2nd year
Experiment 1	57%	22%	22%	15.4 yrs	6.1 yrs	61%	38%
Experiment 2							

	Pre-MBA Job											
	Consul-ting	Other Services	High tech/ Communi-cation	Venture Capital/ Private	Invest-ment Banking	Invest-ment Mgmt	Other Financial Services	Manu-facturing	Non-Profit	Cons-umer Prod-	Health-care/ Biotech	Military
	1	2	3	4	5	6	7	8	9	10	11	12
Experiment 1	36%	5%	10%	9%	2%	2%	5%	6%	8%	3%	2%	11%
Experiment 2												

Each of the four experimental sessions lasted for about 80 minutes. After the experiment, subject earnings were paid in a separate location, using anonymous identification numbers, by people other than the experiment leader. Each subject earned on average USD 23.6, the maximum was USD 42 and the minimum USD 8.

To make sure everyone would easily understand how to use the experimental software interface, it was done as a web interface. (The software can be found at www.giwa.com/econ-experiments/).

Throughout the experiment, the subjects were given instructions both verbally and in written form. At small steps, the computer took care of guiding the subjects forward without any particular verbal support. At one stage, written instructions were handed out on paper at which time an oral presentation was given. To guarantee understanding, on key occasions each subject had to pass a comprehensive battery of control questions to be allowed by the computer to continue to the next step. While these questions were quite complex and often involved several computations, the median number of attempts needed to get the questions right was 1, and the average number of attempts was 1.2. We take this, in concert with exit interviews we conducted on participants in the pilots (from the same subject pool), as evidence that the participating MBA students had a strong understanding of the instructions.

The participants were asked to raise any questions by raising their hands, and general discussion was avoided by answering questions individually.[13] At the end of each session, all participants where given information on how to register to receive a draft version of this paper.

Design

Our design benefits from specifically being targeted at studying procedural justice as opposed to the studies by Clark (1998) and Ball et al. (2001), for which it is ancillary, or as opposed to Hoffman & Spitzer (1985), who interact it with their concept of moral authority.

---

[13] No questions were asked (nor would have been answered) about the nature or aim of the study.

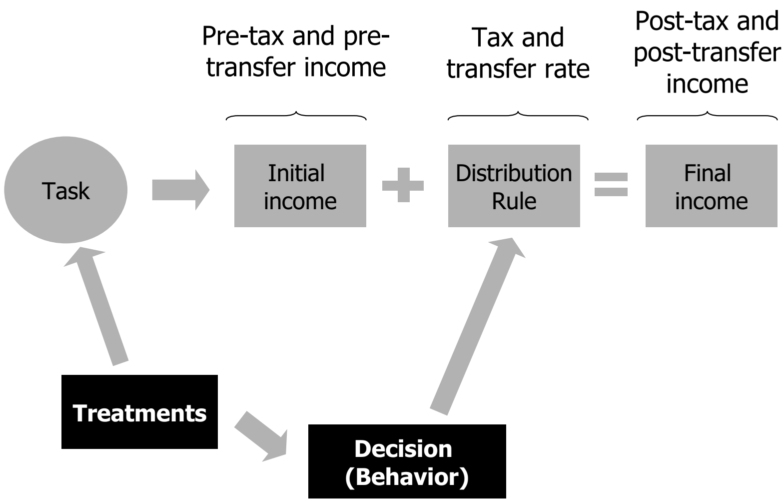
Improving on Rutström & Williams (2000), and in contrast to the other studies discussed, we distinguish between effort and ability measures, and also elicit to what extent the participants believe that the effort (ability) measure does measure effort (ability). Previous studies use problematic measures of effort where they use performance as a proxy for effort. Most of the performance measures seem more related to ability than effort (mastery of trivia knowledge, word puzzles), or even related to chance (a hash-mark game during a short time frame[14]).

Other benefits of our design are:

- It avoids problems of desert preferences interacting with fairness concerns, such as those driven by face-to-face bilateral bargaining where one player has both the entitlement and the decision-making power (and responsibility)
- We avoid problems of the endowment effect driving results, since the task performed in our study is exactly the same for each of our six treatments (and hence chance and effort/ability treatments take the same amount of time)
- We randomize subjects to treatments (in fact we randomize each person to all treatments within each session)
- We circumvent problems of limited number of observations: with 528 observations we have 4-5 times – and in some cases, 10 times – as many observations as those of other studies.

[14] In the hash-mark game the person who (randomly) gets to start has a first mover advantage and he/she can guarantee a victory through backward induction or experience. On the other hand, a person who has no experience of the game before (and does not figure out the backward induction solution), will quite likely feel that victory is largely due to randomness, since they might have no clue which decisions to make.

Our basic experimental design is a follows:



While the initial and final income and the distribution rule are analogous to pre- and post- tax income and a tax rate, we refrained from using such language to avoid potential framing effects.

The design has three main phases:

- **Phase 1** entails a decision being made, behind a veil of ignorance, over what distribution rule to apply to the initial income outcome of the task in Phase 2.

  Here we randomly induce our six different treatments: 100% effort, ability or luck and the 50/50% combinations luck/effort, luck/ability and effort/ability.

- In **Phase 2,** subjects individually conduct a task for approximately 30 minutes for which measures of effort, ability and/or luck determine the initial income outcome.

- In **Phase 3**, the distribution rule from Phase 1 is applied to the initial income outcome from Phase 2 to reach the final income which then is paid out.

Augmenting the design in Rutström & Williams (2000), we use the Tower of Hanoi puzzle as our task, since it enables us to create measurements of ability on the one hand and effort on the other. Luck is introduced by a random chance mechanism.

Before entering Phase 1 the subjects are informed about how each phase will work. The subjects are not informed about the nature of the task, but are told that the same task has been used in other academic work because its nature allows an observer to separately measure and distinguish various dimensions of people's performance on the task. The subjects are also told that the task has been piloted and calibrated for nearly a month, on fellow MBA students at Harvard.

The subjects are informed that there are several methods of allocating points and therefore several ways initial income can be awarded in Phase 2, but they are not told what the methods are. To avoid strategic behavior, subjects are told that they will not find out which of the methods of allocating points will be used for their group, until they are in Phase 3.

The three phases

Phase 1 – Decision over Distribution rule

The subjects are randomly assigned to groups of four. They are then asked to choose the distribution rule they want to apply in Phase 3 to the initial outcome of Phase 2.

We employ the *Random Dictator Rule*, an incentive compatible mechanism under which each person has the same chance of dictating the outcome of the game and there are no strategic voting considerations[15].

Each person is asked to cast a vote for the distribution rule they would like if a specific method of assigning points were to be randomly selected for their group. This process is then repeated successively, until votes for all six treatments have been cast. See *Appendix i* for the exact formulation and format used to elicit these votes.

The order in which they are faced with the different treatments is randomized, and statistical tests to check for potential order effects indicate that data for all six treatments per person can be used.[16]

At this point, we also elicit the subjects' beliefs with regards to (a) their own level of effort and ability on the task to come, and (b) the extent to which they perceive that the methods used will be able to measure the dimensions of effort and ability[17].

Contrary to Rutström & Williams (2000), who use four discrete redistribution rules, we employ a continuous variable to gain more statistical power[18]. This variable, D, is the percent of each person's initial income that will be collected, and evenly distributed to each person in the group in order to reach the final income distribution. The choice space is 0, 1, 2... 100%. In order to ensure participants can easily understand what various distributions

---

[15] In order to show how the Random Dictator Rule works and demonstrate that the best strategy under it is to vote honestly according to one's true preferences, a demonstration of the mechanism was conducted in all but one of the sessions using five different types of snacks (of which the participants got to eat the chosen type).

[16] One strength of this design is that if there had turned out to be any order effects, we could have confined the analysis to only the round(s) before such order effects kicked in. See *Section IV - Results* for the statistical analysis.

[17] See *Appendix i* for the exact formulation/procedure.

rules would imply, the participants are provided a calculator in the software; for any distribution rule 1-100%, the calculator instantaneously shows the final income share for each of the four group members.[19]

Phase 2 – Conducting the Task

The subjects now find out what the selected task is, and after a short practice run[20], they each perform this task for approximately 30 minutes. The computer software measures and collects all the data needed to generate our measures of effort exertion and ability on the task. The task determines how total group income will be divided to reach initial income[21]:

Rank on task	**Initial share of total group income**
Highest scoring person	40%
Second highest scoring person	30%
Third highest scoring person	20%
Fourth highest scoring person	10%

Total group income is determined by aggregate performance of the four group members on the task compared to other groups in the session.

---

18 Since our entire experiment is computerized, we can do this, while in Rutström & Williams (2000) it would have required too many complex real-time calculations to be feasible during the non-computerized parts of their experiment.

19 In the paper instructions that are handed out, the participants also have a table showing the results for each of the four group members under example distributions rules 0, 20, 40, 60, 80 and 100%.

20 Additionally, in order to even the playing field, a quick check is made to ensure no one is using an external mouse.

21 We chose to use income categories, to ensure comparability between this and a potential future study that will investigate the effects of distribution rules on effort and performance. As discussed by Rutström & Williams (2000), income categories enable distribution rule decisions to be compared across treatments and groups, even if total income varies with aggregate production in the group.

Phase 3

The participants are now asked to answer a brief battery of survey questions based on the 2005 World Values Survey to get a gauge of their ideological views. Thereafter, we also collect demographic data.

At this point, subjects find out which randomized treatment they are in, which of the group's four distribution rules has been chosen, and therefore what the distribution rule in use for their group actually is. They then find out which initial income category they are in, as well as the production and resulting total income of the group. Lastly, they learn what their individual final income is after the distribution rule is applied.

The methodology behind the effort and ability measurements, and the rationale for using this methodology, are also explained.

With their anonymous and randomly self-chosen ID numbers, subjects can now go to another room and collect their payments from payment clerks, who are individuals other than the experiment leader.

The task and how we measure effort and ability

The Tower of Hanoi

In phase 2 the subjects find out that the task they will be asked to perform is a computerized version of the *Tower of Hanoi* puzzle[22]. In the puzzle there are X pegs and Y disks. Each disk is of a different size and at the start of the puzzle the disks are arranged as a pyramid on the leftmost peg, the "start peg". The objective is to move all disks to the rightmost peg, the "destination peg", so that they form a new pyramid. The rules are

a) only one disk may be moved at a time and
b) a larger disk may never be placed on a smaller one.

Our primary configuration of the puzzle entails 3 pegs and 5 disks, creating a relatively complex problem that requires a minimum of 31 moves to solve.

Measures of effort and ability

In order for the different treatments to be distinct and for the experimental results to be powerful, it is enough that the subjects *believe* that we have credible and fair measures of effort and ability on the task, *not that we actually have* such measures. However, in order to avoid entering a zone of deception, the task and measurements described below were tested during two smaller and one larger pilot. In the exit interviews of these pilots, a vast majority of the participants felt that the measures reasonably and validly measure and distinguish between ability and effort[23].

The ability measure is quite simple: how many/how fast the subjects are able to solve puzzles during a limited time of about five minutes. The fact that the time is short ensures that any rewards from effort are kept to a bare minimum. The ability score is computed by comparing how fast and how many puzzles a subject completes compared to everyone else in the same experimental session.

For the effort measure the subjects are asked to solve as many puzzles as possible during approximately 25 minutes. The duration makes

---

[22] Due to French mathematician Édouard Lucas' in 1883. Puzzle sometimes called the Towers of Brahma. We thank Romuald Zylla (http://wipos.p.lodz.pl/zylla) for letting us use and adapt his java script.

[23] And while we got feedback on many other aspects on the experiment from several participants in the real sessions, no feedback with regards to the suitability of the measures was voiced by the otherwise very vocal subject pool.

effort an important factor[24]. However, to arrive at a measurement of the effort exerted, we need to filter out any effects of ability. We do this by using the ability measure from above as a control.

For the effort score, each person is grouped with the three people who had the closest scores on the ability measure. Comparing the number of puzzles that are solved within a control group provides a gauge of effort when ability has been roughly filtered away. Also, in order to minimize any large effects of learning, the first five minutes of the long session are excluded from the measure.

Both the effort and ability scores are normalized on a scale between 1 and 100, so that effort points can be compared and combined with ability points for the cases when the total points are based on a combination of ability and effort.

[24] From the pilot study, we learned that participants found the task tiring after about 10 minutes.

IV - Results

Clear support for procedural justice concerns

We find strong support that procedural justice concerns can have a large and highly significant impact on people's preferences for redistribution.

Fig. 1a - Effort & Ability vs. Chance

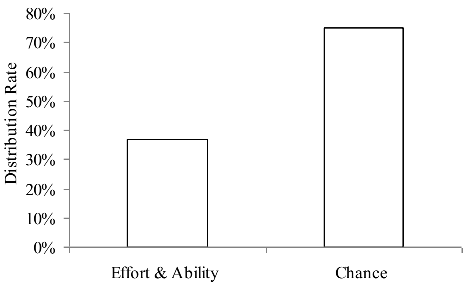


Fig. 1b - Effort & Ability versus Chance (filtered for Round = 1)

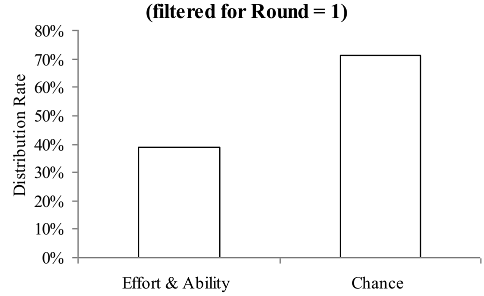


Fig. 1c - Effort & Ability vs. Chance (filtered for Beliefs)

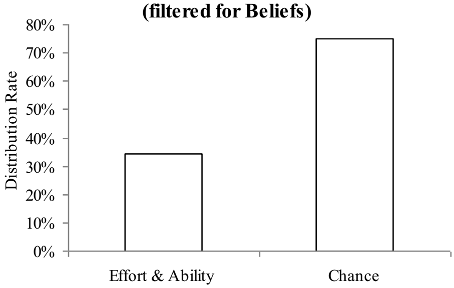
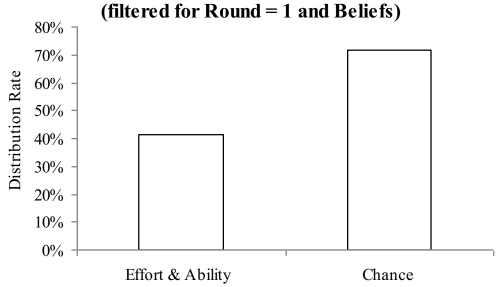


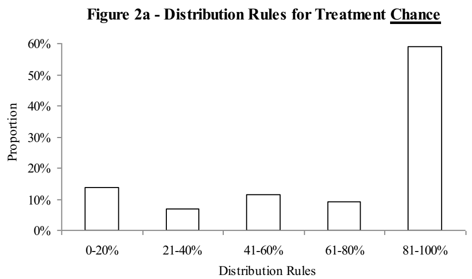
Fig. 1d - Effort & Ability versus Chance (filtered for Round = 1 and Beliefs)

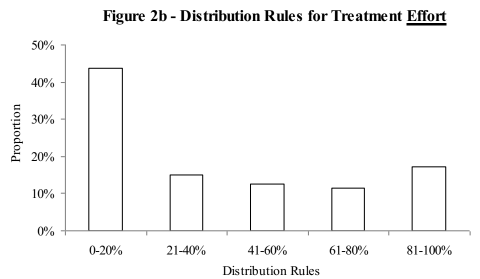


From *Figure 1a* it is clear that the distribution rates chosen by subjects are about twice as high (75%) when initial income is based on chance than when based on effort or ability or a combination of both (37%).

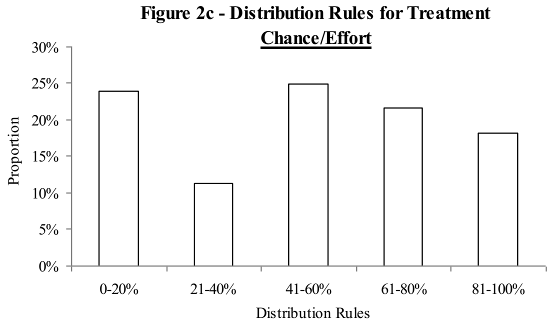
Figure 1c only includes data from the first treatment that subjects met and provides an indication that the results are robust for order effects. Figure 1b provides an early indication of robustness to controlling for successfully having induced the effort vs. ability vs. luck treatments. We will revisit these robustness checks more thoroughly later and also see that the results above are significant at the 0.1% level.

Taking a deeper look at the data with regards to distribution levels in the effort and chance treatments we can see that, while there is some noise, the patterns in *Figure 2a* and *2b* are basically opposites.



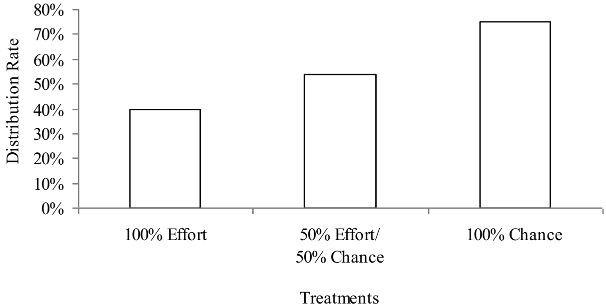


From the histogram in Figure 2c, we can see that, in the treatment with a convex combination of 50% luck and 50% effort, the distribution rates lie in between the pure effort and chance treatments.



In Figure 3 we can see that the average distribution rate appears to be roughly a linear function of chance. These differences are significant at the 0.1% level.

Figure 3 - Pure Effort, Effort/Chance and Pure Chance

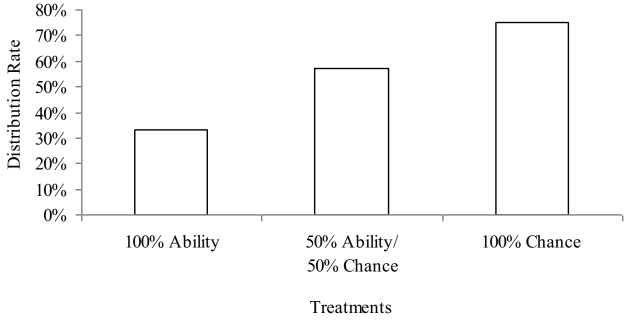


Why does our study pick up such strong evidence of preferences for effort-based Lockean desert while others studies have not? One reason could be that our effort measurement is perceived to be a better measure of *effort* than previous studies which have used performance measures based on ability (or even chance) as proxies for effort. This is a significant distinction if people do not have preferences for ability-based desert.

However, from *Figure 4*, we see that the subjects behave as if income differences due to ability are as fair as those due to effort. (In *Appendix ii* we also show the histograms for the ability treatments - which are very similar to the ones shown for the effort treatments above). The difference between ability and ability/chance, and between ability/chance and chance are both significant at the 0.1% level.

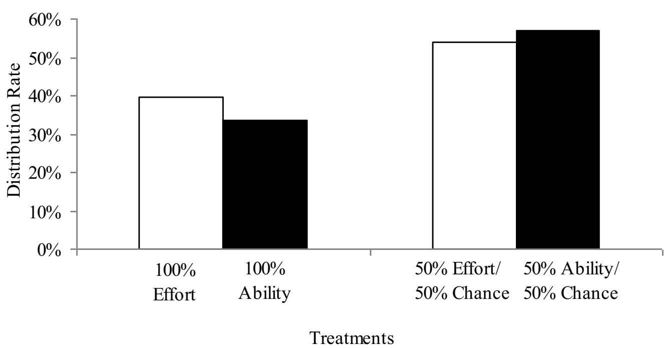
That previous studies have not been able to pick up these strong effects could likely be due to the problems with their designs and choices of tasks as discussed in sections II and IV above.

Fig. 4 - Pure Ability, Ability/Chance and Pure Chance



Comparing *Figure 3* and *Figure 4,* it appears as if the average distribution rate might actually be lower in ability treatments than in effort treatments. But note also that in the 50/50% combinations with luck, the reverse might be the case - the

Figure 5 - Effort versus Ability



distribution rate in the *effort*/chance treatment is lower than in the *ability*/chance treatment.

Let us now turn to a more formal analysis of the results to get a better understanding of their statistical significance and their robustness.

Table 2 - Main Regression Results and Robustness Analysis

	OLS with clustering†	OLS with order dummies‡ and clustering†	R.E. (z-stat and chi2 stats)	Tobit with lower and upper bound at 0 and 100% respectively	R.E. filtered for beliefs (z-stat and chi2 stats)	OLS with clustering† controlled and adjusted for (over)confidence†† (All stats = F-stats)
	Distribution	**Distribution**	**Distribution**	**Distribution**	**Distribution**	**Distribution**
Constant	0.75	0.74	0.75	1.04	0.75	0.75
(t-stat)	*(20.06)****	*(15.51)****	*(19.89)****	*(12.83)****	*(16.69)****	*(19.96)****
Effort	-0.36	-0.36	-0.36	-0.74	-0.37	-0.35
(t-stat)	*(-6.62)****	*(-6.66)****	*(-8.86)****	*(-6.65)****	*(-7.37)****	*(37.98)****
Ability	-0.42	-0.42	-0.42	-0.86	-0.45	-0.41
(t-stat)	*(-7.43)****	*(-7.54)****	*(-10.41)****	*(-7.58)****	*(-8.86)****	*(42.94)****
Effort/Chance	*-0.21*	*-0.22*	*-0.21*	*-0.50*	*-0.19*	*-0.22*
(t-stat)	*(-5.26)****	*(-5.31)****	*(-5.31)****	*(-4.59)****	*(-3.77)****	*(24.86)****
Ability/Chance	*-0.18*	*-0.18*	*-0.18*	*-0.44*	*-0.19*	*-0.16*
(t-stat)	*(-4.66)****	*(-4.66)****	*(-4.53)****	*(-4.08)****	*(-3.66)****	*(15.22)****
Effort/Ability	*-0.38*	*-0.39*	*-0.39*	*-0.75*	*-0.41*	*-0.37*
(t-stat)	*(-8.20)****	*(-8.20)****	*(-9.56)****	*(-6.82)****	*(-7.99)****	*(57.48)****
Ordereffect dummies		*Yes*				
Confidence interaction dummies						*Yes*
Effort = Effort/Chance?	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected
(F-stat)	*(17.05)****	*(16.89)****	*(12.74)****	*(5.46)***	*(12.95)****	*(12.98)****
Ability = Ability/Chance?	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected
(F-stat)	*(35.80)****	*(38.01)****	*(34.74)****	*(15.64)****	*(27.02)****	*(32.89)****
Effort = Ability?	Borderline	Borderline	Not rejected	Not rejected	Not rejected	No rejected
(F-stat)	*(3.18)**	*(3.24)**	*(2.38)*	*(1.21)*	*(2.22)*	*(2.4)*
Effort/Chance = Ability/Chance	Not rejected	Not rejected	Not rejected	Not rejected	Not rejected	Not rejected
(F-stat)	*(0.91)*	*(0.90)*	*(0.61)*	*(0.31)*	*(0.01)*	*(1.72)*
Effort = Effort/Ability	Not rejected	Not rejected	Not rejected	Not rejected	Not rejected	Not rejected
(F-stat)	*(0.54)*	*(0.48)*	*(0.48)*	*(0.01)*	*(0.39)*	*(0.44)*
Ability = Effort/Ability	Not rejected	Not rejected	Not rejected	Not rejected	Not rejected	Not rejected
(F-stat)	*(0.90)*	*(0.97)*	*(0.72)*	*(0.99)*	*(0.75)*	*(0.69)*
Average (Round 1 + 2) = = Average (Round 5 + 6)		Not rejected				
(F-stat)		*(0.00)*				
Round 1 = Average (Round 5 + 6)		Not rejected				
(F-stat)		*(0.01)*				
Order Effects DO NOT Jointly Improve Model		Not rejected				
(F-stat)		*(0.46)*				
(Critical Value at 95% Conf.)		*(2.21)*				

* = 10% level, ** = 5% level, *** = 1% level

† Clustered around participants; Robust standard errors

‡ No order dummy is individually significant (P values are between .35 and .95). Testing for order effects on the individual treatments does not yield significant order effects either (Joint tests that the model IS NOT improved by adding order dummies cannot be rejected and 24 out of the 25 effects are insignificant).

†† Evaluated at the 50th percentile (no over/under confidence). For instance, the change from the Chance to the Effort treatment is calculated:
Effort-dummy-coefficient + Effort-Confidence-Interaction-dummy-coefficient * 50.0
where 50.0 adjust the 44.55 overconfidence for the Effort-Confidence measure.

Note: Dummies for type such as gender or year in MBA program are non-significant and do not change any of the regressions.

In all of the regressions in *Table 2* the dependent variable is the distribution rate and dummies are used for all treatments except the chance treatment.

Hence, each dummy indicates the percentage point difference in distribution rate compared to the chance treatment.

Before discussing the differences between the regressions, we can note the following: All of the treatments are significantly different from the chance treatment at the 0.1% level, in all regressions. The differences between the pure effort and pure ability treatments and their convex combinations with luck are also highly significant. However, we cannot reject that the distribution rates in effort and effort/ability or the ability and effort/ability treatments are equal.

Before returning to the question of whether the distribution rates in the effort and ability treatments are different, let us first turn to the differences in the regressions.

The first regression "OLS with clustering" takes into account the fact that each participant generates six distinct observations and creates robust standard errors through clustering. The second regression "OLS with order dummies and clustering" shows (together with its footnote) that there are no significant order effects, that is that our design, based on increasing the sample size but letting participants meet all of the treatments in a randomized order, worked. This is confirmed again though the panel regressions that we have run; see the third regression.

The Tobit regression shows that the results are significant when one also takes into account the fact that the data is, to a large extent, grouped around the end points: 0% and 100% distribution.

Regression (5) in Table 2, "R.E. filtered for beliefs", shows the results when we exclude the participants who potentially may have doubted that the ability and effort treatments would be based on good and fair measurements.[25] Note that, if anything, the effect of not being fully able to induce treatments in the minds of the subjects should simply create noise – and decrease the differences between treatments – not increase differences or produce spurious results. While the filter excludes about 1/3rd of the sample, the results from this regression are almost identical to the other regressions.

Let us now turn to the last regression "OLS with clustering controlled and adjusted for effects of (over)confidence", which adjusts for potential (over)confidence.

Could larger (over)confidence be driving the results?

As mentioned in *Section II*, directly after the participants vote on the distribution rules[26], we elicit the top percentile in which they estimate that their performance would fall, if based on each of (1) effort, (2) ability and (3) chance.

The average estimate if no overconfidence were to exist would be the 50th percentile, which is the percentile ranking for the random chance

---

[25] We say "potentially" because the filter is intentionally conservative and is probably overly strict in categorizing subjects as "doubters" and then excluding them from the data set. The filter is based on responses to questions regarding the extent to which participants believed that the measures will capture what they are supposed to measure (See *Appendix i*).

We think it quite likely the question could, in and of itself, have induced doubt. It is also possible that people prefer to appear skeptical – even if in fact they were not skeptical at the time they were making their votes. (Recall that the questions were asked AFTER the participants made their distribution rule decisions). Two observations support this view.

A. It is clear from people's open-ended answers to questions about why they chose the distribution rates that they chose, that a large portion of the people mechanically filtered out by our procedure did in fact talk about the effort and ability treatments in a way that clearly sets them apart from each other and from luck.
B. During open-ended interviews with pilot participants, people hardly voiced any skepticism about the validity of the measures when given the opportunity.

[26] And before they learn anything more about the task, i.e. the information they have is exactly the same as when they just recently made their votes on distribution rules (our independent variables).

outcome of 50.1. However, across all participants the average estimated percentile ranking is lower, at 43.0 for ability, and at 44.6 for effort[27]. Both averages are significantly lower than 50 at the 0.1% level, suggesting that some level of overconfidence exists in these two domains. Could overconfidence be the driver of average distribution being lower in effort and ability than in chance treatments, and/or could differences in people's confidence be a key driver of distribution rule choices?

Confidence about effort and ability do not have significant explanatory power over distributions levels on their own (P = 0.16 and 0.25 respectively), but their point estimates both go in the direction that theory would predict, and we proceeded to use them as controls. The results in column (6) of *Table 2* above, "OLS with clustering controlled and adjusted for effects of (over)confidence", are based on a regression where we have added confidence interaction dummies to our standard regression and then evaluated the regression at the 50th percentile for both effort and ability confidence.

As theory would predict, we can see that the effect of the effort and ability treatments are smaller than before, but only slightly. Compared to the other regressions, the "desert effect" of effort at -35% is only about one percentage point smaller. The adjusted desert effect of the pure ability treatment is now -41%, or roughly one percentage point less than in the other regressions. In this adjusted regression we can NOT reject that the effects of the effort and ability treatments are the same (P = 0.13).

[27] The more confident the lower percentile. E.g. 25th percentile means top 25%.

Differences and similarities among Europeans and Americans

While we have presented the main results of this study above, below we will conduct some ancillary analysis to compare Europeans and Americans. Since the European sample size is quite small (18 subjects and 108 observations), we also do the analysis below on Americans versus Non-Americans (37 subjects and 220 observation), in which we will see that the overall picture roughly parallels that of Americans versus Europeans. (The sample size of Americans is 49 subjects and 294 observations).

From Table 3 below, we can see that the Americans in the sample differ in ideology compared to non-Americans (see the upper part of the table) and Europeans (see the lower part) in the same direction as in the real world[28].

Questions 3 to 7 are replicated from the World Value Survey.[29] The ideology measure in (1) is a simple average of Questions 3 to 7, while the measure in (2) is created through factor rotation.

[28] Using Questions 3 to 6 to create a proxy of preferences for government/redistribution in the real world we see a similar pattern to that discussed by Fong (2001) and Alesina & Angeletos (2005) in *Figure 1* in the *Introduction*. That is, we find a strong correlation between preferences for government/redistribution and beliefs about whether luck/connections or effort/ability determine income and success in society (significant at 0.1% level).

[29] See Appendix for exact formulations.

Table 3 - Differences in "Ideology"

Americans participants versus participants from all other regions

	(1) Ideology Based on straight average	(2) Ideology Factor rotated (varimax)	(3) Equal vs Differences in Income	(4) Government v.s Private Private Ownership	(5) Government vs. People Responsible	(6) Competion bad vs. good	(7) Income Determined by Luck vs. Effort
Constant	6.67	-0.19	5.49	7.22	4.78	8.30	7.57
(t-stat)	*(79.95)****	*(-3.56)****	*(37.21)****	*(60.16)****	*(30.51)****	*(87.91)****	*(55.84)****
US-dummy	**0.48**	**0.34**	**0.20**	**0.58**	**0.79**	**0.51**	**0.31**
(t-stat)	*(4.35)****	*(4.70)****	*(1.03)*	*(3.67)****	*(3.82)****	*(4.09)****	*(1.74)**

American versus European participants

	(8) Ideology Based on straight average	(9) Ideology Factor rotated (varimax)	(10) Equal vs Differences in Income	(11) Government v.s Private Private Ownership	(12) Government vs. People Responsible	(13) Competion bad vs. good	(14) Income Determined by Luck vs. Effort
Constant	6.30	-0.37	4.25	7.50	3.92	8.50	7.33
(t-stat)	*(17.10)****	*(-1.53)*	*(7.04)****	*(16.13)****	*(5.26)****	*(20.38)****	*(12.60)****
US-dummy	**0.85**	**0.51**	**1.44**	**0.30**	**1.66**	**0.31**	**0.55**
(t-stat)	*(2.26)***	*(2.09)***	*(2.33)***	*(0.63)*	*(2.18)***	*(0.72)*	*(0.92)*

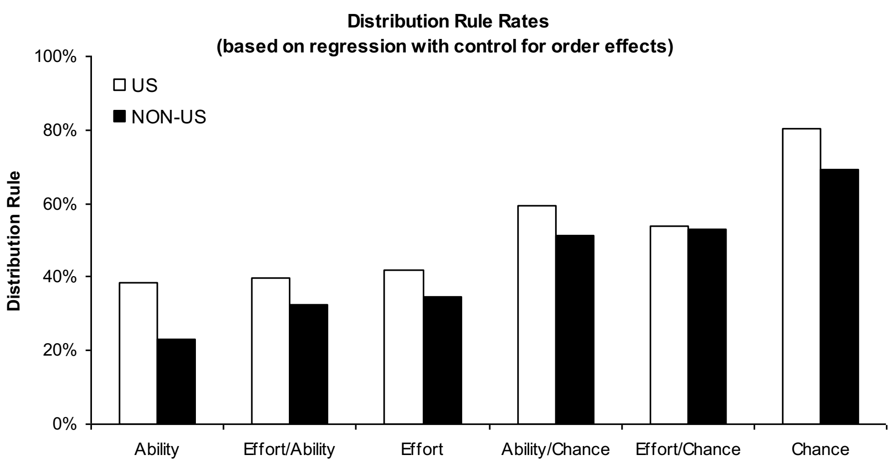
* = 10% level, ** = 5% level, *** = 1% level

Note: The more towards the right (left) a participant's views are, the higher (lower) the variable is. As in the World Value Survey half of the questions were asked in the opposite fashion (were "right" was towards the "left" in the format of the question), but those variables are transformed here.

Let us now turn to the question of whether Americans differ from Europeans / Non-Americans with regard to their preferences for procedural justice. To gain an understanding of this, we run regressions with region and region-treatment dummies (see *Appendix iii*) which underlie *Figures 6* and *7*.

From Figure 6 we can see the average treatment effects per treatment for Americans and Non-Americans.

Figure 6



We do not find statistical support for Americans having stronger desert preferences – "i.e. larger jump in the distribution level in the chance treatment to the distribution levels in each of the other treatments" - than Non-Americans:

We cannot reject the joint hypotheses that adding region and region/treatment-interaction dummies DOES NOT improve the regression model. Also, no differences between Americans and non-Americans in the treatment effects are significant (p-values in the 0.70's). That is, we cannot say that any jump from the chance treatment to each of the other treatments is different for Americans than for Non-Americans.

If anything, the point estimates of the levels suggest that the Americans in the study actually want more redistribution overall than Non-Americans when in the experiment and behind the Veil of Ignorance. This, despite the Americans in the sample having more right leaning views in the world outside of the lab than the Non-Americans have (as shown in *Table 3*). However, only the difference in the level of distribution in the ability-treatment between Americans and Non-Americans is significant, and then only at the 10% level.

The pattern for Europeans is somewhat similar, but – due to small sample size – more noisy. Again, we cannot reject the joint hypotheses that adding region and region-treatment interaction dummies DOES NOT improve the regression model.

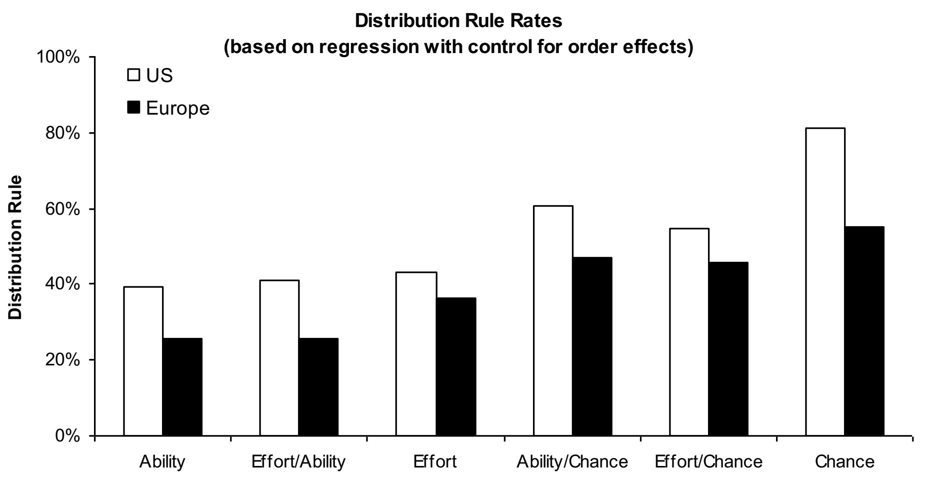
However, if we nevertheless look at the difference between Americans and Europeans with respect to individual *desert* point estimates – i.e. in the jump from chance to each of the other treatment we see the following: Two of the five desert effects – "Chance to Pure Effort" and

"Chance to Effort/Chance" - are now borderline of having significantly different distribution level jumps between Americans and Europeans. We cannot reject at the 5% level, but can reject at the 10% level that they each are equal for Americans and Europeans.

It appears that the driver of this result is that Americans opt for more redistribution in the chance treatment than Europeans. The difference between the level of distribution in the chance treatment for Americans and Europeans is significant at the 5% level.

As can bee seen in Figure 7, on a level basis, the point estimates suggest that the Europeans opted for less redistribution in the five other treatments also. However, running ancillary tests we see that only one of these is statically significant. The level of distribution in pure effort/ability is significantly lower for the Europeans than for the Americans at 10% level.

Figure 7



The raw data underlying the regressions in Figure 6 and 7 is displayed in *Appendix iii*.

V – Summary and Conclusions

We find strong support for people having fairness preferences not *just for outcomes*, but also with regard to *how those outcomes are reached* – procedural justice concerns. These preferences have a large and highly significant impact on participant's preferences for redistribution. The redistribution rates chosen by subjects are twice as high when initial income is decided by luck than when due to effort and/or ability.

While the sub-samples are small and cannot be seen as representative for the general populations of America and Europe we find some tentative results that would be interesting to study further and on a more representative sample: We find suggestive support for Americans in the experiment not having stronger desert preferences than Europeans. When the only exogenous difference is the role of luck versus effort versus ability, people from all regions make relatively similar decisions – higher redistribution when income is decided by luck, lower redistribution when based on effort and/or ability. This is interesting in light of the fact that the Americans (Europeans) in the study have real world ideological views that mirror the direction of the lower (higher) level of redistribution and taxes in the American (European) society.

Together our findings suggests that one reason for the vast difference in redistribution levels between the US and Europe potentially could be due to differences in *beliefs* about the relative importance of effort and luck in determining income and success in the two societies. Is appears as if the more people believe that luck determines income and success, the more redistribution they prefer while the more they believe that effort determines income and success, the lower redistribution they prefer. In this sense our results support the *causality* in the stories in Fong (2001), Alesina & Angeletos (2005) and Jencks & Tach (2006) discussed in the introduction.

The results beg several questions. How do people trade-off procedural justice concerns with other concerns, such as preferences for dynamic efficiency (for example, in a case where taxes/redistribution create deadweight losses)? How generalizeable are our indicative results that there are no differences in Europeans' and Americans' procedural justice concerns, or more specifically their desert preferences? Is the result that people find success due to ability to be as fair as success due to effort a robust one?

If higher redistribution rates in Europe versus the U.S. are in fact due to Europeans attributing a larger portion of success to luck and Americans attributing a larger portion to effort, as our results provide some suggestive support for, it is important to understand whether those beliefs about the drivers of success are correct or biased.[30]

Today in the 21st century, income mobility could in fact be higher in Western European countries (and especially Scandinavian countries), with broader access to quality schooling, free university education and other social benefits, than in the U.S. (see for example Arrow, Bowles & Durlauf [2000]). The picture looks similar for movement among classes and occupations (see for example Erikson & Goldthorpe [1992]).

---

[30] The notion that social mobility can shape political attitudes, and hence policies, dates back to at least de Tocqueville (1835), who proposed that the difference in attitudes towards redistribution between Europe and the U.S. could be explained by different beliefs about mobility rates. See also Lipset and Bendix (1959) and Lipset (1966, 1977, 1992) for arguments suggesting that differences in redistribution between Europe and the U.S. could be due to differences in beliefs about mobility rates, rather than differences in actual mobility rates.

The two, very different, equilibrium redistribution levels in the U.S. and Europe could both be rational given people's beliefs. However, due to biases in those beliefs, one or both of the regions' tax and redistribution regimes could nonetheless be significantly suboptimal in relation to the true underlying preferences of the population(s).

That people can have different beliefs and that "societies remain in different redistributive equilibrium, although the underlying structural parameters of mobility are essentially the same"[31], are features common to the models of Piketty (1995) and Alesina & Angeletos (2005). There are also several studies that suggest that people have biased estimates of their chances of being rich (see for example, DipRete [2007] on how Americans overestimate their chances of being rich).

There are, of course, several alternative explanations for regional redistribution disparities (one possible could be differing preferences regarding efficiency versus equity tradeoffs across the two regions). However, even the mere risk that one or both regions have potentially gotten the crucial decision of where to set the level of taxes and redistribution wrong relative to the true preferences of their citizens suggests that further research is warranted in this domain.

[31] Piketty (1995, p 554) further conjectures, that this is "particularly likely if a country exhibited, at some time in the past, a significantly different experiment of social mobility before joining the 'common' pattern." As Alesina & Angeletos (2005, p 974) observe, in Europe "[t]he 'invisible hand' frequently favored the lucky and privileged rather than the talented and hardworking." Europeans thus came to want aggressive redistributive policies and other forms of government intervention. Alesina & Angeletos (2005, p 974) continue: "In the 'land of opportunity,' on the other hand, the perception was that those who were wealthy and successful had 'made it' on their own." Americans have thus historically chosen low redistribution and limited regulation. Thus further understanding of the role of history and culture in shaping beliefs and norms is potentially very important in this context.

Appendix i – Instructions, Questions and Exit Survey

1. The Main Instructions (paper instructions that were handed out)

2. How votes for the distribution rules were elicited (from software)

3. World Values Survey questions that we replicated

4. How the belief gauge of whether treatments were induced was created

5. For the rest of the instructions, questions and exit survey please visit www.giwa.com/econ-experiments/

An Experiment with Three Phases - Main Instructions

This experiment involves three phases. The key characteristics of each phase are:

- **In Phase 1**, you will randomly be assigned to a group of four participants. You will then be asked to vote for the type of *distribution rule* that you would like to have applied to the money that members of your group receive in Phase 2. The distribution rule will determine how much each person is finally paid in Phase 3.
- **In Phase 2** you will individually conduct a task where points will be assigned. These points will determine the initial money that each group member receives.
- **In Phase 3,** you will find out how points were assigned in Phase 2 (6 different potential ways) and which distribution rule was selected. The distribution rule is applied to the money received in Phase 2, leading to a final payoff for each group member (in real cash).

You will meet the phases in chronological order, but let us learn more about Phase 2 first.

Phase 2

A task
In Phase 2 of the experiment, each participant in your group will individually be asked to perform a task.

Determinant of total group income
The size of your group's total initial income will depend on how many solutions to the problem your group reaches in total. The group with the lowest number of solutions in the room will receive a total group income of 80 dollars whereas the group with the highest number will receive 105 dollars. (Any ties will be broken with help of a lottery).

Member's initial share of group income
The members of the group will receive initial shares of the total group income as follows:

Table 1. Rank by number of points	Initial share of group income
Highest scoring person	40%
Second highest scoring person	30%
Third highest scoring person	20%
Fourth highest scoring person	10%

For instance, the person with the least points, in the group with the least points, would have an initial income of 10% * 80 dollars = 8 dollars. The person with the most points, in the group with most points, would similarly receive 40% * 105 = 42 dollars.

A brief note on the task
The specific task has been chosen because it has been used in other academic studies and because its nature allows us to measure various features of performance. The measurements have also been calibrated on and with the help of about 30 HBS MBA students during April.

Phase 1

A group of four
You will here randomly be assigned to a group of four, including yourself.

A vote for your group's distribution rule
You will then be asked to vote for the distribution rule you would like to have applied to the initial income outcomes of Phase 2 to reach the final distribution in Phase 3.

The distribution rule, D, is the % of each person's initial income that will be collected and then evenly distributed to each person in your group to reach the final income distribution.

The table below shows example final income shares for some different rules. When you get to the point in the experiment were you are asked to vote for D you will be provided a calculator that will show you what the final shares would be for any potential D.

Table 2	Final Share (Phase 3) of total group income for person with:			
Distribution Rule "D"	Highest score	Second highest score	Third highest score	Fourth highest score
100%	25%	25%	25%	25%
80%	28%	26%	24%	22%
60%	31%	27%	23%	19%
40%	34%	28%	22%	16%
20%	37%	29%	21%	13%
0%	40%	30%	20%	10%

Choice of distribution rule
How will your group choose the distribution rule? We are going to use a voting rule that determines the outcome through a lottery. Here is how it works. Each person will vote for the distribution rule he or she prefers on their computer. The software will then randomly select a vote among the entries of all participants in your group. As there are four individuals in your group, there is a 1 in 4 chance your vote will be chosen. The chosen vote determines the rule that will be used to determine everyone's final payoffs in your group.

A note on your optimal voting strategy
You should vote for whichever distribution rule (D) you most prefer. If your vote is chosen, then your preferred distribution will take effect. If your vote is not chosen, it does not affect the outcome in any way.

Different configurations of the task and different votes
There are six possible ways of assigning points in Phase 2/6 versions of the task. Each group will eventually get one of the versions to work through. As you in Phase 1, do not yet know which version of the task you will get, you will be asked to vote for your preferred distribution rule for six different versions of the task. In other words, you (and the others in your group) will be asked for their votes on D with regards to each of the six versions.

A chance mechanism will decide which of the six versions will actually be used for your group in Phase 2.

Phase 3
In this phase you will find out which version of the task, i.e. how the points from Phase 2 were assigned, and which of the distribution rules that was selected for your group –
<u>you and the others in your group will NOT know this information when in Phase 2</u>.

You will then find out your individual number of points, the total number of solutions found by your group and which initial income category you were placed in during Phase 2. You will then see what your final income is after the distribution rule from Phase 1 has been applied.

After the experiment, you can go to another room and use your anonymous ID number to collect your payments from payment clerks. The payment clerks will be individuals other than the experiment leaders. Before this you will be asked exit questions.

A vote for the distribution rule for the first potential task version

In this version of the task, points in Phase 2 will be assigned based

[**the one of the six treatment conditions that was randomly chosen**:
- 100% on the effort measure
- 100% on the ability measure
- 100% on luck (randomized)
- 50% on the effort measure and 50% on luck
- 50% on the ability measure and 50% on luck
- 50% on the effort measure and 50% on the ability measure]

Recall that the points from Phase 2 will determine the initial income, as shown in Table 1 in your paper instructions.

However, the final income (the income that will be paid out in real money after the experiment) will depend on your group's distribution rule.

See Table 2 in your paper instructions for some sample distribution rules. **We also encourage you to use this calculator to see what outcomes different distribution rules imply:**

Input

Distribution rule =	D

Final income distribution

	Share of total group income to receive as final income
Highest scoring person	X
Second highest scoring person	Y
Third highest scoring person	Z
Fourth highest scoring person	W

Please vote for the distribution rule you would prefer for the case that this version of the task is chosen.

Recall, in this version of the task, points in Phase 2 are assigned based on [the one of the six treatments]

My vote for distribution rule for this configuration is:	

Continue =>

[Pop up box for the screen above]

You are just about to cast the following vote for the case where points in Phase 2 would be assigned based on [the one of the six treatment conditions]

Your vote

Distribution rule =	V

Final income distribution

	Share of total group income to receive as final income
Highest scoring person	X
Second highest scoring person	Y
Third highest scoring person	Z
Fourth highest scoring person	W

Is this the vote you want to cast?

[Yes] **[No, I want to change]**

World Values Survey questions that we replicated

Note that while every other question has 10 as the most left leaning, and every other has 10 as the most right leaning, in our analysis we transform them so that 10 always is the most right leaning.

How would you place your views on the following questions on the following scale?

- 1 means you agree completely with the statement on the left;
- 10 means you agree completely with the statement on the right; and
- if your views fall somewhere in between, you can choose any number in between.

We need larger income differences as incentives for individual effort — Incomes should be made more equal

1 2 3 4 5 6 7 8 9 10

Private ownership of business and industry should be increased — Government ownership of business and industry should be increased

1 2 3 4 5 6 7 8 9 10

The government should take more responsibility to ensure that everyone is provided for — People should take more responsibility to provide for themselves

1 2 3 4 5 6 7 8 9 10

Competition is good. It stimulates people to work hard and develop new ideas — Competition is harmful. It brings out the worst in people

1 2 3 4 5 6 7 8 9 10

In the long run, hard work usually brings a better life — Hard work doesn't generally bring success—it's more a matter of luck and connection

1 2 3 4 5 6 7 8 9 10

How the belief gauge of whether treatments were induced was created

A) In some of the sessions the following two questions were asked

"To what extent do you think you will agree that the method for measuring ability on the task does measure ability on the task when we later explain how the measurement works?

To a high degree
To a relatively high degree
To a moderately high degree
To a moderately low degree
To a relatively low degree
To a low degree"

"To what extent do you think you will agree that the method for measuring effort exertion on the task does measure effort exertion on the task when we later explain how the measurement works?

Same six choices as above"

In these session the belief gauge threw out any who answered low, relatively low or moderately low on EITHER (or both) of the two questions above.

B) In some of the sessions the following question was asked instead

"Which of the following would you say *best* describes your state of mind while you were making your votes for the various distribution rules?

i. I simply thought about the ability measure as ability and the effort measure as effort, without giving it that much thought when making my votes for the distribution rules

ii. I gave a the measures a lot of thought and concluded that while the measures might not be perfect, they would still be a form of measurement for what they were designed to measure. As such, I thought about the ability measure as ability and the effort measure as effort when making my votes

iii. I did not believe that the ability measure would be better at measuring ability than the effort measure would be at measuring ability (and vice versa) and hence did not treat the ability measure and the effort measure different, but as the same thing when making my votes

Important note: It is possible that the very process of being asked to answer and hence concretely think about this question has changed they way you perceive the measures. Therefore, please do your best to answer the questions based on your actual state of mind while you were making your votes on the last six screens. Do not to base your answers on the way you perceive the measures right now, after thinking concretely about them."

In these session the belief gauge threw out anyone who indicated that they did not believe that the ability measure would be better at measuring effort than the effort measure and vice versa (option 3).

Appendix ii – Summary of Data

Figure 1a - Distribution Rules for Treatment Effort

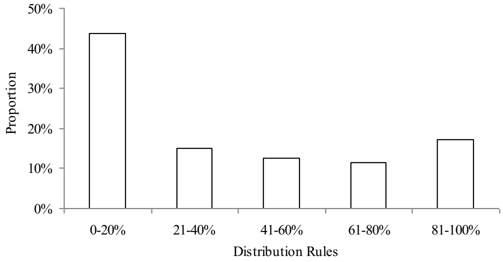


Figure 1b - Distribution Rules for Treatment Effort (Filtered for Beliefs = 1)

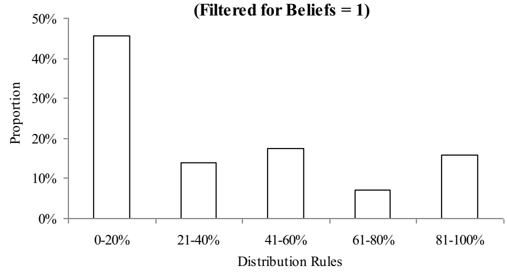


Figure 2a - Distribution Rules for Treatment Ability

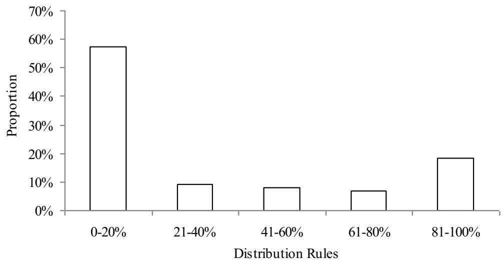


Figure 2b - Distribution Rules for Treatment Ability (Filtered for Beliefs = 1)

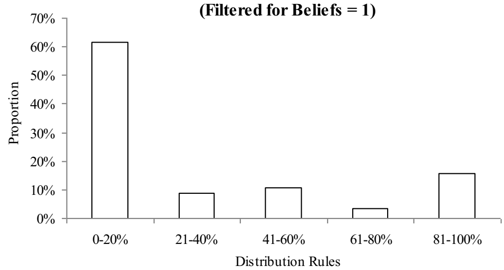


Figure 3a - Distribution Rules for Treatment Chance

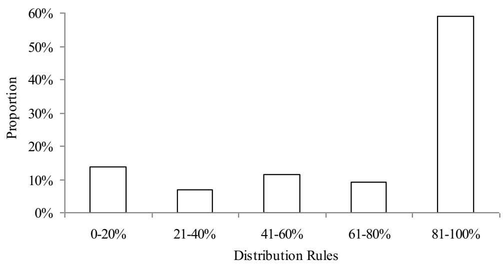


Figure 3b - Distribution Rules for Treatment Chance (Filtered for Beliefs = 1)

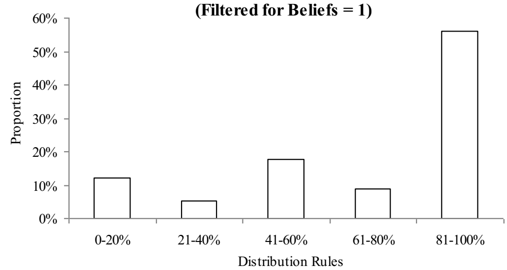


Figure 0a - Distribution Rules for All Treatments

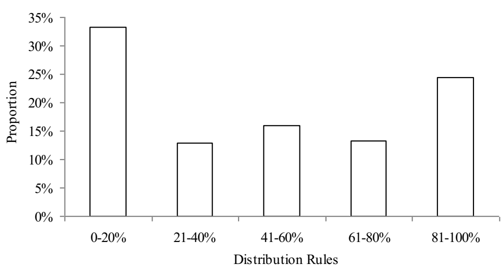
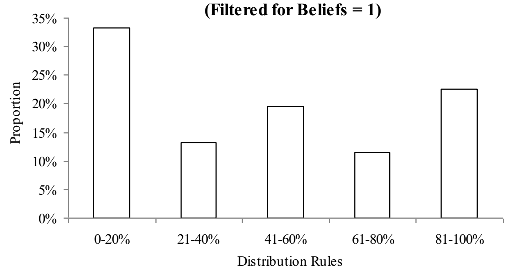


Figure 0b - Distribution Rules for All Treatments (Filtered for Beliefs = 1)



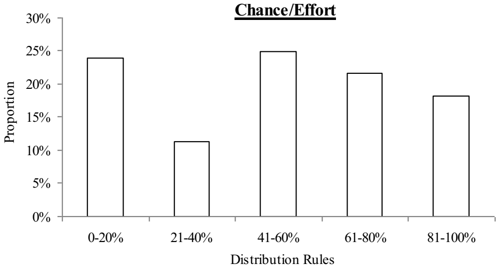


Figure 4a - Distribution Rules for Treatment Chance/Effort

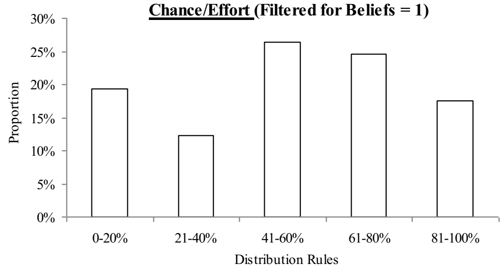


Figure 4b - Distribution Rules for Treatment Chance/Effort (Filtered for Beliefs = 1)

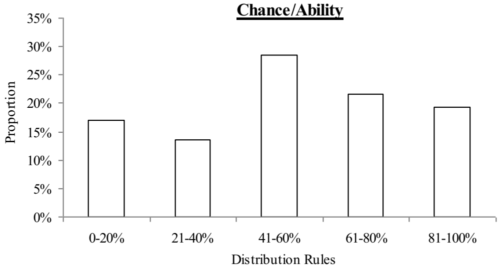


Figure 5a - Distribution Rules for Treatment Chance/Ability

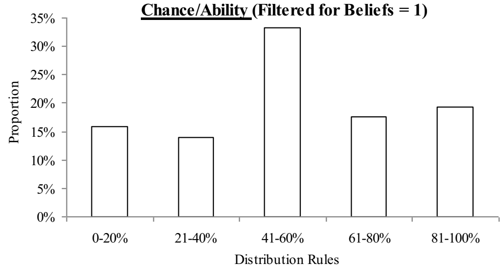


Figure 5b - Distribution Rules for Treatment Chance/Ability (Filtered for Beliefs = 1)

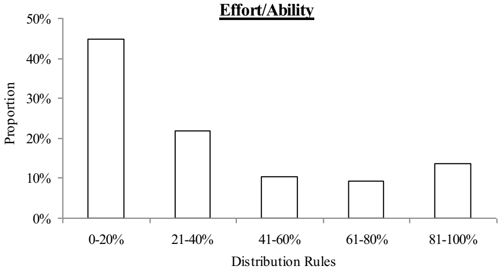


Figure 6a - Distribution Rules for Treatment Effort/Ability

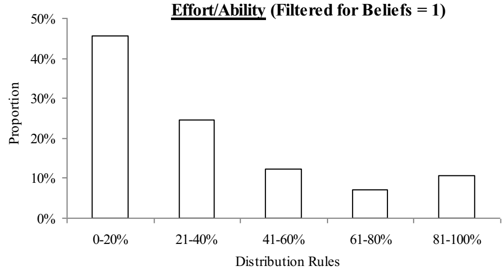


Figure 6b - Distribution Rules for Treatment Effort/Ability (Filtered for Beliefs = 1)

Appendix iii – Regressions and Raw Data underlying Figures 6 and 7

Table A iii - Regional Analysis

	US vs. Non-US	US vs. Europe (Note: Small Sample)	
	OLS with order dummies and clustering†	OLS with order dummies and clustering†	
	Distribution	**Distribution**	**Comment**
Effort	-0.35	-0.19	These 5 dummies indicate the "Desert Effect"
(t-stat)	*(-4.27)****	*(-2.23)***	- the **difference** in distribution level in the chance
Ability	-0.46	-0.29	treatment and each of the other treatments -
(t-stat)	*(-5.71)****	*(-2.98)****	for Non-Americans in first column,
Effort/Chance	*-0.16*	*-0.10*	and for Europeans in second column
(t-stat)	*(-2.89)****	*(-1.35)*	
Ability/Chance	*-0.18*	*-0.08*	
(t-stat)	*(-3.53)****	*(-1.12)*	
Effort/Ability	*-0.37*	*-0.30*	
(t-stat)	*(-5.57)****	*(-3.07)****	
Constant	0.69	0.55	Redistribution **level** for Non-Americans
(t-stat)	*(10.22)****	*(1536)****	and Europeans respectively in the Chance treatment
Dummies			
US-Effort-Interaction	-0.04	-0.19	These six dummies indicate the difference in
(t-stat)	*(-0.36)*	*(-1.71)**	in "Desert Effect" for Americans, compared to
US-Ability-Interaction	0.05	-0.12	Non-Americans in first column and
(t-stat)	*(0.41)*	*(0.97)*	Europeans in second column
US-Effort/Chance-Interaction	*-0.10*	*-0.17*	Again, by "Desert Effect" we mean the
(t-stat)	*(-1.25)*	*(-1.79)**	difference in distribution level in the chance
US-Ability/Chance-Interaction	*-0.03*	*-0.12*	treatment and each of the other treatments
(t-stat)	*(0.34)*	*(-1.35)*	
US-Effort/Ability-Interaction	*-0.04*	*-0.10*	
(t-stat)	*(-0.38)*	*(-0.88)*	
US	0.11	0.26	**Additional** redistribution in the Chance
(t-stat)	*(1.42)*	*(2.30)***	treatment for the Americans
Collection of regional dummies DO NOT Jointly Improve Model	Not rejected	Not rejected	We cannot say that adding the collection
(F-stat)	*(1.23)*	*(1.46)*	of regional dummies improves the model
(Critical Value at 95% Conf.)	*(2.10)*	*(2.10)*	
Regional interaction Effects DO NOT Jointly Improve Model	Not rejected	Not rejected	Same conclusion if we only look at the
(F-stat)	*(0.86)*	*(1.05)*	interaction dummies
(Critical Value at 95% Conf.)	*(2.21)*	*(2.21)*	

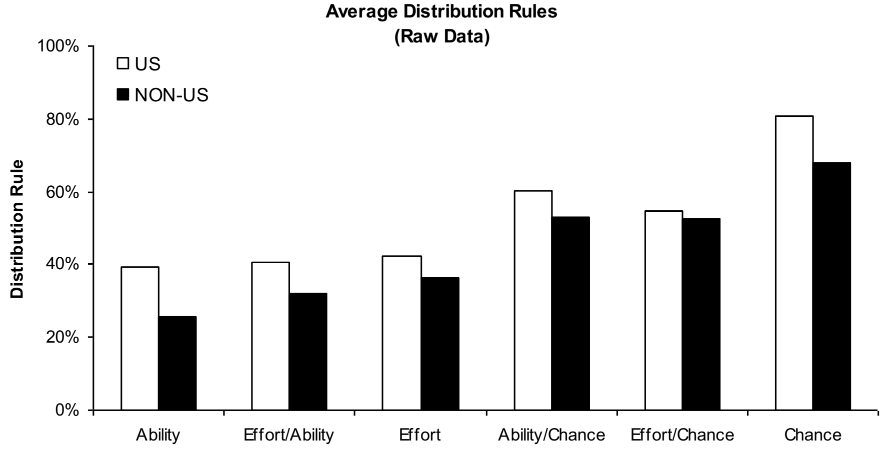
* = 10% level, ** = 5% level, *** = 1% level, † Clustered around participants; Robust standard errors. Controls for order effects included

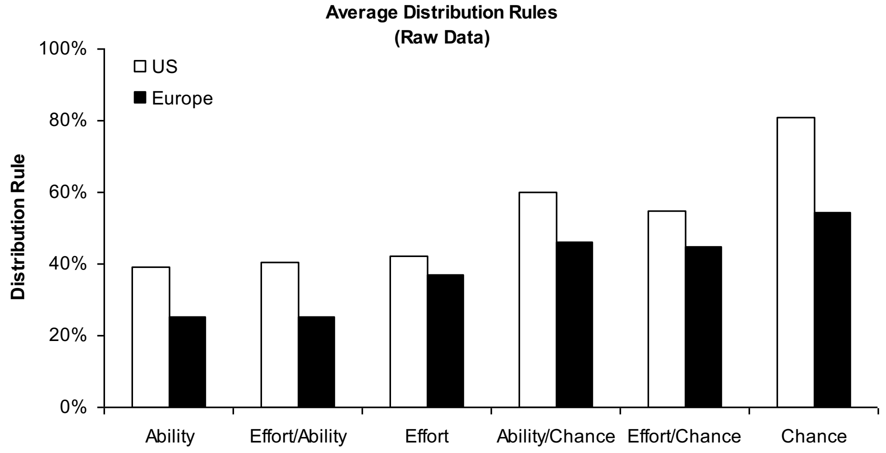
Levels

The graphs and ancillary tests are based on point estimates of the levels that are calculated as follows:

Level	Non-Americans / Europeans	Americans
Chance	Constant	Constant + US-dummy
Effort	Constant + Effort dummy	Constant + US-dummy + + Effort-dummy + US-Effort-Interaction dummy
Effort/Chance	Constant + Effort/Chance dummy	Constant + US-dummy + + Effort/Chance-dummy + US-Effort/Chance-Interaction dummy
Ability, Ability/Chance, Effort/Ability	Analogous to last two above	Analogous to last two above

Summary of raw data underlying regressions for Figures 6 and 7





References

Adams, J. Stacy. 1965. "Inequality in Social Exchange" In *Advances in Experimental Social Psychology, Vol. 2,* ed. Leonard Berkowitz, 267–99. NY, London: Academic Press Inc.

Akerlof, George A. & Janet Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *Quarterly Journal of Economics,* 105(2): 255–83.

Alesina , Alberto & George-Marios Angeletos. 2005. "Fairness and Redistribution: US versus Europe." *American Economic Review*, 95(4): 960-98.

Aristotle. 1925. *The Oxford translation of Aristotle: Vol. IX. The Nichomachean Ethics*, ed. WD Ross. London: Oxford University Press.

Arrow, Kenneth, Samuel Bowles & Steven Durlauf, eds**.** 2000. *Meritocracy and Economic Inequality.* Princeton, NJ: Princeton University Press.

Babcock, Linda, George Loewenstein, Samuel Issacharoff & Colin Camerer. 1995. "Biased Judgments of Fairness in Bargaining." *American Economic Review,* 85: 1337–1343.

Ball, Sheryl, Catherine Eckel, Philip J. Grossman & William Zame. 2001. "Status In Markets." *The Quarterly Journal of Economics*, 116(1): 161-188.

Beckman, S. R. & J. W. Smith. 1995. "Efficiency, Equity And Democracy: Experimental Evidence On Okun's Leaky Bucket." *CRESP Working Paper No. 9509.* Department of Economics, University of Colorado, Denver, CO.

Buchanan, James M. 1986. *Liberty, Market and State: Political Economy in the 1980s.* NYU. Press, Columbia University Press.

Burrows, P. & G. Loomes. 1994. "The impact of fairness on bargaining behavior." *Empirical Economics,* 19: 201–221.

Clark, Jeremy. 1998. "Fairness in Public Good Provision: An Investigation of Preferences for Equality and Proportionality." *Canadian Journal of Economics*, 31(3): 708–29.

DiPrete, Thomas A. 2007. "Is this a Great Country? Upward mobility and the chance for riches in contemporary America." *Research in Social Stratification and Mobility*, 25(1): 89-95.

Dworkin, Ronald. 1981. "What Is Equality? Part 2: Equality of Resources." *Philosophy Public Affairs,* 10(1): 283–345.

Ellingsen, Tore & Magnus Johannesson. 2005. "Sunk Costs and Fairness in Incomplete Information Bargaining." *Games and Economic Behavior,* 50: 155-177.

Erikson, R. & J. H. Goldthorpe. 1992. *The Constant Flux - A Study of Class Mobility in Industrial Societies.* Oxford: The Clarendon Press.

Farwell, Lisa & Bernard Weiner. 1996. "Self-Perceptions of Fairness in Individual and Group Contexts." *Personality and Social Psychology Bulletin,* 22(9): 867–991.

Fehr, Ernst & Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics,* 114(3): 817–68.

Fong, Christina. 2001. "Social Preferences, Self-Interest, and the Demand for Redistribution." *Journal of Public Economics*,82(2): 225–46.

Frohlich, N. & J. A. Oppenheimer. 1990. "Choosing justice in experimental democracies with production." *American Political Science Review*, 80 (2): 461–477.

Frohlich, N., J. A.Oppenheimer & C. Eavey. 1987a. "Laboratory Results On Rawls' Distributive Justice." *British Journal of Political Science*, 17: 1–21.

Frohlich, N. & J.A.Oppenheimer & C. Eavey. 1987b. "Choices Of Principles Of Distributive Justice In Experimental Groups." *American Journal of Political Science*, 31: 606–636.

Güth, Werner & Reinhard Tietz. 1990. "Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results." *Journal of Economic Psychology,* 11: 417–49.

Heider, Fritz. 1958. *The Psychology of Interpersonal Relations.* NY: John Wiley & Sons.

Hoffman, E. & M. L. Spitzer. 1985. "Entitlements, rights, and fairness: an experimental examination of subjects' concepts of distributive justice.*" Journal of Legal Studies*, 14(2): 259–297.

Homans, George C. 1958. "Social Behavior as Exchange." *American Journal of Sociology*, 63: 597–606.

Jäntti, Markus, Bernt Bratsberg, Knut Roed, Odbdjoern Raaum, Robin Naylor, Eva Österbacka, Anders Björklund & Tor Eriksson. 2006. "American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries.*" IZA DP no. 1938.*

Jencks, Christopher & Laura Tach. 2006. "Would Equal Opportunity Mean More Mobility?" In *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, eds. Stephen Morgan and David Grusky. CA: Stanford University Press.

Kahneman, Daniel, Jack L. Knetsch & Richard H. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 98(6): 1325-4.

Kluegel, James R. & Eliot R. Smith. 1986. *Beliefs About Inequality: Americans' Views of What Is and What Ought To Be.* NY: Aldine De Gruyter.

Konow, James. 1996. "A Positive Theory of Economic Fairness." *Journal of Economic Behavior & Organization*, 31(1): 13–35.

Konow, James. 2003. "Which is the Fairest One of All? A Positive Analysis of Justice Theories." *Journal of Economic Literature*, XLI: 1188–1239.

Leventhal, Gerald S. & James W. Michaels. 1971. "Locus of Cause and Equity Motivation as Determinants of Reward Allocation." *Journal of Personality and Social Psychology*, 17(3): 229–35.

Lipset, S. M. 1966. "Elections: the Expression of the Democratic Class Struggle" In *Class, Status and Power*, eds. R. Bendix and S. M. Lipset.

Lipset, S. M. 1977. "Why No Socialism in the US" In *Sources of Contemporary Radicalism*, eds. S. Bialer and S. Blutzar.

Lipset, S. M. 1992. "Foreword: The Political Consequences of Social Mobility." In *Social Mobility and Political Attitudes: Comparative Perspectives*, ed. F. Turner.

Lipset, S. M. & R. Bendix. 1959. *Social Mobility in Industrial Society.* University of California Press.

Lissowski, G. & T. Tyszka & W.Okrasa. 1991. "Principles of distributive justice." *Journal of Conflict Resolution*, 35: 98–119.

Locke, John. 1978. *Second Treatise of Civil Government.* W.B. Eerdmans, Grand Rapids.

Nozick, Robert. 1974. *Anarchy, State, and Utopia.* NY: Basic Books.

Ordóñez, Lisa D. & Barbara A. Mellers. 1993. "Trade-Offs in Fairness and Preference Judgments." In *Psychological Perspectives on Justice*, eds. Barbara A. Mellers and Jonathan Baron. Cambridge U. Press, 138–54.

Overlaet, Bert. 1991. "Merit Criteria as Justification for Differences in Earnings." *Journal of Economic Psychology,* 12(4): 689–706.

Piketty, Thomas. 1995. "Social Mobility and Redistributive Politics." *Quarterly Journal of Economics*, 110(3): 551–84.

Rabin, M. 1993. "Incorporating fairness into game theory and economics." *American Economic Review*, 83(5): 1281–1302.

Rawls, J. 1971. *A Theory of Justice.* Harvard University Press, Cambridge, MA.

Rotter, Julian B. 1966. "Generalized Expectancies for Internal versus External Control of Reinforcement." *Psychological Monographs,* 80(1): 609.

Rutström, Elisabet E. & Melonie B. Williams. 2000. "Entitlements And Fairness: An Experimental Study Of Distributive Preferences. " *Journal of Economic Behavior and Organization*, 43(1): 75-89.

Schokkaert, Erik & Leo Lagrou. 1983. "An Empirical Approach to Distributive Justice." *Journal of Public Economics*, 21:33–52.

Schokkaert, E. and Overlaet, B. 1989. Moral Intuitions and Economic Models of Distributive Justice. *Social Choice and Welfare,* 6:19-31.

Tocqueville, Alexis De. 1935. " Democracy in America Vol II." New York.

Walster, Elaine, G. William Walster & Ellen Berscheid. 1973. "New Directions in Equity Research." *Journal of Personality and Social Psychology*, 25(2):151–76.

Weiner, Bernard & Andy Kukla. 1970. "An Attributional Analysis of Achievement Motivation." *Journal of Personality and Social Psychology*, 15(1): 1–20.

Chapter 2

Participation and Peers in Social Dilemmas[1]

Aspects that might influence the degree of free-riding in social dilemma situations are studied by conducting two social insurance experiments in the lab. Evidence is found that people are motivated by concerns additional to those of material payoffs and that the concerns are connected to the concept of reciprocity.

In contrast to other literature, voting over or having a higher degree of participation towards, in this case, an insurance system does not have any positive impact on the levels of free-riding.

Previous trends in the extent to which peers misuse or do not misuse the system are found to have an effect on future levels of free-riding. Given the same economic incentive, people are more likely to lie when they believe that others will lie.

Support is also found for the false consensus effect.

[1] I am grateful to Magnus Johannesson, Victor Corzo and Martin Dufwenberg, without whose help and involvement this project would not have happened. I thank Robert Östling, Erik Lindqvist and seminar participants at the Stockholm School of Economics, as well as David Laibson, Matthew Rabin, Colin Camerer, Ernst Fehr and other participants at the Seventh Russell Sage Summer Institute in Behavioral Economics and also participants at the European Network for the Advancement of Behavioral Economics' Summer Institute on Economics and Psychology in Munich.

Part of this project was carried out when I was a visiting doctoral student at the Department of Economics at Harvard University and later when I was a graduate student at the Harvard Business School.

Financial support from the Jan Wallander and Tom Hedelius Foundation, the Dr Marcus Wallenberg Foundation, the Swedish Finance Political Research Institute and the Network for the Advancement of Behavioural Economics (ENABLE) is appreciatively acknowledged.

Finally, I thank Soledad Piñero Misa and Carl-Richard Häggman for excellent help in executing the experimental sessions.

I - Introduction

Solutions to social dilemmas such as insurance systems, tax compliance, effort-shirk decisions at work, contribution to irrigation systems and usage levels of inshore fisheries have persisted over time without being destroyed by the negative forces of free-riding. Empirical data, experimental research and everyday knowledge show that people, when in contexts of social dilemmas, are strongly motivated by concerns additional to those of material payoffs. It is commonly believed that social norms to some extent mitigate free-riding (Lindbeck et al [2003]).

Further knowledge of these concerns are important in order to understand how the welfare state has survived so far and to better comprehend the challenges it faces with respect to increasing levels of cheating and public debate over whether peoples' values of morality and honesty have declined in recent times.

This paper looks at two questions that could prove to be important in this context: (1) whether individuals are more willing to abide by rules that they have participated in creating or voted over (for example, Ostrom [1990, 2000] and Frey [1997]) and (2) whether people are more willing to abide by rules if a larger fraction of those around them also abide, than when a lower proportion does (Lindbeck et al [1999]). I also evaluate the potential occurrence of the false-consensus effect due to Ross, Greene & House (1977).

Much work has been done on how social norms and reciprocity have an impact on moral hazard problems. However, as far as I know, no experimental work has been done on the extent to which a person's degree of involvement in or attachment towards an insurance system affects their levels of free-riding. Perhaps different levels of involvement lead to different levels of honesty and different formations of social norms, in turn leading to different levels of free-riding and thus overall insurance costs (see Ostrom [1990, 2000], Bowles [1998], Feld & Tyran [2002] for examples where participation in decision-making leads to more cooperation). A related hypothesis is that it might be the case that younger generations feel lower levels of involvement towards the social security systems than older generations. Older generations might have a feeling of being part of the decision-making that led to the welfare state, leading them to more strongly internalize social norms against free-riding on others than younger generations do[1].

For this purpose an experiment is designed and conducted to capture the impact brought on by differing levels of involvement, generated in one context with a high degree of involvement and one with a low degree of involvement towards an otherwise identical insurance game. This is done by using two different treatment conditions, "Choice-of-system"-treatment and "Given-system"-treatment. In contrast to other literature, no positive impact on the level of free-riding is found when participants have higher participation in devising the rules of the insurance system. On the contrary, there is a non-significant indication that being given the opportunity to choose the system might actually lead to increased free-riding.

Some people argue that welfare states are, or are on their way to being, caught in vicious downward spirals[2]: When people hear from media and politicians that others are misusing social insurance systems, they tend to be more likely to misuse the systems themselves, and so a downward spiral starts that might be hard to curb. Experimental evidence is presented that indicates that people are more likely to lie and cheat when they believe that others will lie or cheat, but are more likely to be honest when they think that others will be honest. This is inline with the conjecture that the larger the

[1] This could be in line with Ostrom's (2000) transmission failures from one generation to the next.

[2] See for instance Lindbeck (2002).

proportion of a group or society that adhere to a social norm, the more strongly the norm is internalized by individuals.

Although the paper focuses on games of social insurance, the results are also relevant for other types of social dilemmas and problems of the commons (tax compliance, effort-shirk decisions at work, contribution to irrigation systems and usage levels of grazing lands) and many aspects of the welfare state in general.

The rest of this paper is organized as follows. In Section II, related literature and relevant theory is visited. Section III goes through the experimental procedure. Section IV presents design and results for the participation experiment. Section V goes through design and results for the peers experiment. In Section VI, I conclude and discuss some findings relating to the false consensus effect.[3]

[3] Appendix i and ii contain mathematical derivations and experimental instructions.

II - Related literature and theory

In a Quarterly Journal of Economics paper, Lindbeck et al (1999) analyze the interplay between social norms and economic incentives in the context of the modern welfare state, with focus on people's decisions about work and benefits. They assume that it is a social norm to not live on transfers from others. A paper, by Stutzer and Lalive (2004), based on data from Switzerland presents empirical evidence suggesting the existence of a strong norm of self-sufficiency and living off one's own work, and suggesting that this norm has a substantial effect on unemployed people's behavior. In a theoretical paper about social norms and moral hazard, Dufwenberg & Lundholm (2001) study the effects of social rewards (in addition to pecuniary rewards) on social insurance. Even though their results show that social rewards have an effect on economic behavior, "the theory is silent on why an erosion of social reward might occur" (Dufwenberg & Lundholm [2001, page 16]).

This study is an attempt to shed some light on two factors that might be important in this context; firstly, the impact of having the opportunity to vote over, for instance an insurance system, and secondly the role of an individual's expectations about how others will behave given identical pecuniary incentives.

Evolutionary theory and empirical research support the assumption that humans have an inherited inclination to learn social norms (Pinker [1994]). But which norms we learn and the power we let them have over us, vary from situation to situation (Ostrom [2000]). For instance, experimental research has consistently shown that communication in the form of "cheap talk" can substantially increase the level of cooperation between individuals in games of social dilemmas (see for example, Sally [1995] or Dawes, McTavish and Shaklee [1977]). The most compelling reasons that communication increases levels of co-operation include: driving a sense of mutual commitment, facilitating increased trust, reinforcing or even creating social norms, and strengthening of group identity (Ostrom [1998]).

From management theory, we know that participation and involvement in decision making are believed to have positive impacts on cooperation levels in groups and corporations through the same channels as communication – by driving a sense of mutual commitment, facilitating increased trust, reinforcing or even creating social norms and strengthening group identity.

Ostrom (1990, 2000) discusses design principles that characterize well functioning common pool regimes. Her third principle states the importance of individuals being able to participate in making and modifying the rules of a system or regime. This participation creates a stronger sense of fairness, which in turn leads to smaller problems of moral hazard. Similarly, Bowles (1998) discusses that individuals are more willing to abide by rules that they participated in creating, because such rules meet a kind of shared concept of fairness.

Evidence from the lab indicates that rules that are externally imposed can crowd-out cooperative behavior (Frey [1994]). Bardhan (2001) provides a field study from India that links different levels of perceived involvement/participation in rule creation to different levels of perceived fairness of rules, and subsequently different levels of rule violations.

Feld and Tyran (2002) show that when taxpayers can directly influence tax laws and rates, tax evasion appears to be lower. Alm, Jackson and McKee (1993) provide additional experimental evidence for this assertion, while Pommerehne and Weck-Hannemann (1996) and Frey (1997) highlight supporting field evidence.

From the above, a natural hypothesis can be formulated: If people feel that they were part of the decision making process over the rules in a social insurance system, they may to a higher extent internalize a norm to avoid living unnecessarily on transfers. This could in turn lead them to be less likely to free-ride than if they did not feel that they were a part of the decision making progress. The key purpose of the first experiment is to test this hypothesis.

In the model of Lindbeck et al (1999), the social norm of living off one's own work is assumed to be endogenous, in the sense that the more people adhere to the norm, the stronger the norm is internalized. If this is the case, welfare states could face the risk of being caught in vicious downward spirals. If people increasingly hear from media and politicians that others are misusing social insurance systems, they might tend to be more likely to misuse the systems themselves. The driving force behind this could be the triggering of a kind of negative reciprocity, and/or dampening of what Charness and Dufwenberg (2006) call guilt aversion. It could also be related to the cost of lying changing, if the perceived relative consequence of lying differs in cases when a high versus a low fraction of others lie (Gneezy [2005]).

From the above, the following hypothesis can be formed: In games of social insurance, given the same economic incentives, people are more likely to lie and cheat when they believe that others will lie or cheat; but they are more likely to be honest when they think that others will be honest. The second experiment in this paper aims to shed some light on this hypothesis.

III - Experimental procedure

The experiments on voting/participation ("Experiment 1") and on the impact of different perceived proportions of lying/not lying ("Experiment 2") were conducted in direct succession, in effect creating one big experiment ("The Experiment"). However, none of the subjects knew about the opportunity for them to take part in Experiment 2 until Experiment 1 was finished.

Two parallel sessions of each of the two experiments were carried out, on three different occasions between January and April 2006. All of these (six plus six) experimental sessions where conducted in the computer labs at the Stockholm School of Economics, Sweden.

A total of 96 subjects were recruited from the undergraduate programs in economics, business, medicine and engineering at the Stockholm School of Economics, Stockholm University, Karolinska Insititutet and the Royal Institute of Technology. No one participated more than once in each experiment.

The experiment lasted for between 60 and 75 minutes, with Experiment 1 constituting roughly half of the time, Experiment 2 comprising one quarter of the time and the initial welcome, instructions and the exit survey accounting for the last quarter. Subject earnings were paid after the experiment in a different place, and in an anonymous way, by people other than the experiment leaders. Each subject earned on average SEK 42 (USD 5.7) for Experiment 1, and SEK 41 (USD 5.5) for Experiment 2 in addition to the guaranteed show-up fee of SEK 50 (approximately USD 6.7), leading to average total participant earnings of SEK 133 (USD 17.8). [4]

[4] Exchange rate of SEK 7.5 per USD assumed.

Table 1a - Summary statistics: Demographics

Gender		Education				Years of study		
Female	Male	Stockholm School of Econoimics	Stockholm University	Karolinska Insitutet	Royal Institute of Technology	<1.5	1.5 - 3	>3
39%	61%	36%	29%	21%	13%	35%	38%	26%

Table 1b - Summary statistics: Payoffs

	Experiment 1	Experiment 2	Exp 1 + Exp 2	Show up fee	Totel total
	SEK (USD)	SEK (USD)	SEK (USD)	SEK (USD)	SEK (USD)
Mean	42 (6)	41 (5)	83 (11)	50 (7)	133 (18)
Min	24 (3)	0 (0)	43 (6)	N.a.	93 (12)
Max	56 (7)	50 (7)	99 (13)	N.a.	149 (20)
Median	44 (6)	40 (5)	85 (11)	N.a.	135 (18)

In order to enable communication and voting while preserving anonymity, it was natural to execute both experiments in a computer lab[5]. This also enables the relatively complex re-matching of subjects into new groups at four different occasions in Experiment 2, to be done automatically and without any loss of time. Another important aspect of using a computer lab setting is that it helps preserve some social distance between subjects, hence, leading to a slightly better representation of reality. Furthermore, using a computerized system enables me to guarantee each subject's anonymity with regard to the experimental leaders.

To make sure everyone would easily understand how to use the computer program, a simple Internet browser interface was created.[6] Throughout both experiments the subjects were given instructions both verbally and via the computer. At smaller steps, the computer took care of guiding the subjects forward without any particular verbal back-up. At two

[5] In a non-computerized pilot experiment to this study (Corzo & Giwa [2002]), we received several comments that choosing report/defect would have been easier in a fully anonymous context. Also, the particularly high level of non-defection (80 percent) was likely an effect of the failure to not only guarantee hidden *action,* but to also guarantee hidden *identity*.

[6] The computer interface was programmed in cooperation with people at Mindglowing.

stages in Experiment 1 and at one stage in Experiment 2, written instructions were handed out on paper. To guarantee understanding at key occasions, each subject had to pass a comprehensive battery of control questions to be allowed by the computer to continue to the next step. The participants were asked to raise any questions by raising their hands. A general discussion was avoided by taking and answering the questions individually.[7]

To make the instructions concrete and easy to take in, both experiments were framed within the context of a health insurance system.

At the end of Experiment 2, all participants where given information about how to register to receive draft and final versions of this paper.

[7] The questions answered were to clarify procedure and instructions. No questions about the nature or the aim of the study were answered until after both experiments had been completed.

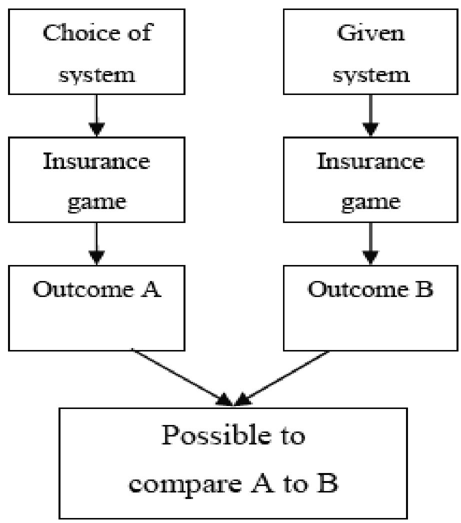
IV - Experiment 1

Design

Two treatment, three stage design

Recall that the purpose of Experiment 1 is to shed some light on whether people free-ride less in a social dilemma context if they feel that they are part of the decision making process over the system's rules. A two treatment, three stage design is created in which subjects play under double anonymity within groups of four. Subjects are randomly divvied into either a "Choice of system" treatment or a "Given system" treatment. When the subjects thereafter have to make a cooperation/defect decision in the health insurance system, it becomes possible to compare the impact of the two treatments on the likelihood of defection (see Figure 1).

Fig 1. The Overall Design



The three main stages of the experiment can be summarized as follows:

1. Each group has up to 10 minutes to discuss and decide – via an anonymous, chat-based communication interface – on an insurance level that will, with a 50% probability, govern the insurance game to be played subsequently

2. By a randomization device, half of the groups play in their chosen insurance system, while half of the groups are given an exogenously decided insurance system to play in

3. Each person decides individually and anonymously whether or not to cooperate or defect in the social dilemma game, that is, to misuse or not to misuse the insurance system in the case that she is "healthy"

Payoffs and game structure

Before the first stage, each subject was given information about all of the separate stages.[8] Each subject started with an initial 1 000 points. There was a 5/6th probability that the subject would turn out to be "healthy" and earn an additional 2 000 points as a wage, but with 1/6th probability the subject would be "sick" and earn zero in additional points. Subjects were told that the monetary value of points was marginally decreasing, and that points would be exchanged for real money after the experiment at the "exchange rate":

Number of points from experiment	0	1000	2000	3000	4000	5000	6000
SEK in hand after experiment	0	28	38	47	54	60	65

Each subject was told that her group of four would have an insurance system, and that given the decreasing marginal monetary value of points, such an insurance system could improve aggregate welfare.

The insurance level would be x % of the 2 000 points salary, where x % would be between 10% and 100%. Each group member would pay an equal share of the cost of the system up to a maximum capped at 1 000

[8] The information was conveyed to the subjects in layman's terms, rather than in the language used here.

points per subject (thus given the initial endowment, no subject could hence earn a negative amount of points).

A person who found herself in the "sick" condition would receive 0 points in salary plus:

1. *x* % of 2 000 points, if the system had enough funds to cover all claims, or
2. if the system did not have enough funds to cover all claims, an equal share of the entire fund (up to 100%, if only one person was claiming to be sick).

The subjects were also informed that there was no way to control whether a person was "sick" or "healthy" and that a healthy person could potentially lie, allege to be sick, and thereby claim insurance on top of her salary.

This main game is therefore a multi-person prisoner's dilemma, with the sub-game perfect, individually rational and dominant strategy being to free-ride. However, this free-riding decreases the aggregate welfare of the group, which would be maximized by cooperation and non-free-riding[9].

[9] See Appendix i for a formal derivation.

Collective-choice process

The procedure used for the collective-choice process was highly stylized. Each group was given ten minutes to reach a decision on the insurance level, x (between 10% and 100%) by unanimity.[10] In line with Walker et al (2000), this process intentionally lacks many features from real voting situations (political parties, agenda setting agents, face-to-face communication, etc) in order to be able to draw clear inferences about the effects of the voting itself.

Strategy method

For methodological reasons, at Stage 3 all participants where asked to decide what they would do if they were healthy, before finding out their state of health. This approach is called the strategy method, and enables the recording of behavior of players, even in cases where they do not reach the information set in question, in our case, the state of "healthy".[11]

Distinguishing a potential effect of participation from that of communication

One challenge when designing the experiment, was to find a good way to isolate the effects of decision making and participation from the effects of communication and learning. The chosen design enables such isolation, by letting subjects in both treatments communicate and make a decision over an insurance system that they will only keep with a 50% probability. Thereafter the subject's groups are, by a randomization device, either allowed to keep

[10] To ensure that groups made their decisions on time, participants were told that not reaching a decision before the time ran out would result in the group not being allowed to play the insurance game, and hence only receiving the show up fee. Luckily, all groups reached decisions within the time limit.

[11] This approach goes back to 1967 and Selten, see for example, Dufwenberg & Gneezy, (2000).

the insurance system they decided on, or given an insurance system by the computer.

Each group in the "Given-system" treatment got (without knowing it) an insurance level that duplicated the insurance level of a "Choice-of-system" treatment group. Hence, on an aggregate level, the "Given-system" treatment exactly mirrors the insurance levels of the "Choice-of-system" treatment, with respect to the payoff structure of the main game.

The only thing that differs between the treatments is whether subjects play in a system that their own group decided upon or in a system that was given externally. Therefore this design nicely isolates the impact of deciding on a system on the choice to free-ride or not.

Results

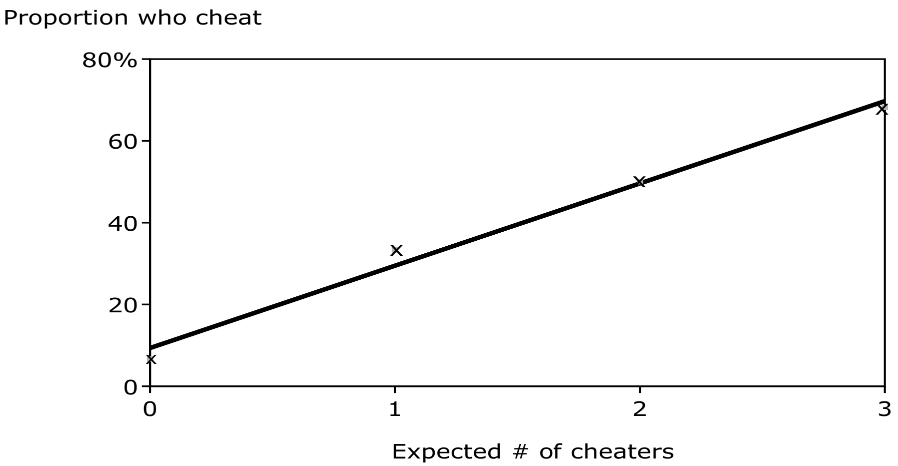
As expected, a prediction based on all subjects being purely motivated by pecuniary payoffs, and playing the sub-game perfect and dominant strategy "cheat", is strongly rejected. But 39% of the participants did chose defection. The overall defection rate is roughly in line with results found in other studies, and significantly different from 100%.

Table 2 - Overall Co-operation Rate

Experiment 1	Sally [1995] Meta analysis	
Co-operation rate	Co-operation rate	Standard deviation
61%	47%	24%

Evidence that some subjects behave according to a form of reciprocity can also be found in that peoples' own behavior and their expectations of others behavior are strongly correlated. A regression of the aggregate proportion of

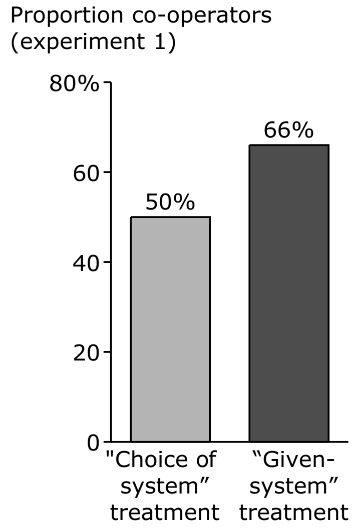
Figure 2



people's decisions whether or not to cheat, against their aggregate (self-reported) expectations of how many others in their respective groups would cheat, is presented in Figure 2. Of course, when interpreting this, it is important to bear in mind that causality may run in both directions (something that will be discussed more in the conclusions of the paper).

A probit regression on dummy variables yields a similar result.

A key finding is that, in contrast to the conjecture and hypothesis based on the literature presented in section II, no positive impact is found on the levels of free-riding when there are higher levels of participation in selecting the insurance system.



OLS and probit regressions of a dummy variable coded 'one' for cooperation and 'zero' for defection, against a dummy that is set to 'one' for "Choice-of-system" and 'zero' for "Given-system" treatment, both yield point estimates that suggest choosing the system may actually reduce the probability to cooperate by 16%. However, this is not significant at conventional levels (p = 0.12 in both cases). The conclusion is identical when non-parametric analysis is conducted. A two-sided chi-2 frequency sample test yields a p-value of 0.12, at the same time that the contingency coefficient (correlations measure between 0 and a theoretical ceiling, which is almost always below 1) is 0.16.[12]

Table 3 - Size and significance of treatment effect (experiment 1)

	OLS	Probit	Non-parametric
	Co-operate	Co-operate	Co-operate
Constant	.5		
(p-value from t-stat)	*(.000)*		
Given-system dummy	.159		
(p-value from t-stat)	*(.119)*		
Given-system dummy (dF/dx for disctrete change)		.159	
(p-value from z-stat)		*(.116)*	
Contigency coefficient			0.158
(p-value from chi-2 frequence sample test)			*(.116)*

From the above, one might ask whether a weak indication is found that being given the opportunity to choose a system might actually lead to increased levels of free-riding. We therefore conduct further regression analysis. Given that males in some studies have been found to free-ride to a

[12] See Seigel (1956) for more information on this non-parametric test.

higher extent than females do, and that males were not allocated by the randomization in a totally identical proportion across both treatments, an additional regression was run with a male dummy as control. Again, the result is found that playing in a system where one's own group had chosen the insurance level, did not increase the probability of cooperation. In this regression the Given-System dummy coefficient is 17% and is significant at the 10%-level. However, when dummies for other back-ground variables are also included, the effect is no-longer significant.

Table 4 - Treatment when using other controls

	Co-operate	Co-operate	Co-operate
Constant	.58	.68	1.0
(t-stat)	*(6.4)*	(5.8)	(6.4)
Given-system dummy	.17	.15	.11
(t-stat)	*(1.7)**	*(1.4)*	(1.1)
Male dummy	- .14	- .15	-15
(t-stat)	*(- 1.3)*	*(- 1.4)*	*(- 1.4)*
SSE dummy		- .19	- .11
(t-stat)		*(-1.5)*	*(- .9)*
KI dummy		- .13	- .9
(t-stat)		*(- .9)*	*(- .6)*
KTH dummy		.15	.17
(t-stat)		*(- .9)*	*(1.0)*
Expect 1 dummy			-.30
(t-stat)			*(- 2.0)***
Expect 2 dummy			-0.41
(t-stat)			*(- 2.8)***
Expect 3 dummy			-0.56
(t-stat)			*(- 3.5)***

A look at proportions indicate that participants in the "Given-system" treatment cheat less than those in the "Choice of system" treatment, regardless of whether they were given a lower insurance rate than their group chose, or a higher rate than their group chose. The proportion of non-cheaters in the "Choice of system" treatment is 50% overall, while it is 65% in the "Given-system" treatment when the insurance rate was higher, vs 67% when the insurance rate was lower. There is no statistical difference between the proportion of non-cheaters in the case where the given-system insurance rate is higher and the case when it is lower. For regression analysis see Table 5 below.

Table 5 - Given treatment divided in two sub-treatments

	Co-operate	Co-operate
Constant	.58	1.0
(t-stat)	*(6.4)***	*(6.4)***
Smaller-system dummy	.19	.12
(t-stat)	*(1.5)*	*(1.0)*
Larger-system dummy	.15	.10
(t-stat)	*(1.2)*	*(.8)*
Male dummy		-15
(t-stat)		*(- 1.6)*
SSE dummy		- .11
(t-stat)		*(- .9)*
KI dummy		- .09
(t-stat)		*(- .6)*
KTH dummy		.17
(t-stat)		*(1.0)*
Expect 1 dummy		-30
(t-stat)		*(- 2.0)**
Expect 2 dummy		-40
(t-stat)		*(- 2.8)***
Expect 3 dummy		-55
(t-stat)		*(- 3.4)***

V - Experiment 2

Design

After playing the first experiment all subjects where invited (and accepted) to participate in Experiment 2. To simplify instructions, the context of this experiment resembles that of Experiment 1. Subjects were informed that by randomization they would be grouped in new groups of four, and that, as before, full anonymity would hold. They learned that the probability of being sick was again 1 in 6 and being healthy was 5 in 6. As before, there would be a health insurance system and its potential costs would be financed equally by each group member.

However, unlike before there was no endowment in the second experiment. If healthy, a wage of SEK 40 would be earned, and if sick, a wage of zero would be earned. The insurance level was now set to 100% of the wage of SEK40 , that is, a person who turned out to be sick would receive SEK 40 in sick insurance, minus 1/4th of the total costs of the insurance system.

Since there was no way to verify whether a person was sick or not, the hazard of a healthy person claiming to be sick and receiving sick benefits existed. Although there was no diminishing marginal value of points, an efficiency loss for claiming sick benefits when healthy was modeled into the payoff structure. If healthy, but claiming to be sick, a person would – instead of earning 80 (40 + 40) minus 1/4th of the insurance cost – only earn 60 (40 + 40 – 20) minus 1/4th of the insurance cost. It was highlighted that such an efficiency loss would not impact a person who got insurance if they really were sick.

The game described above is a social dilemma, in the sense that the individually maximizing, sub-game perfect and dominant strategy to cheat, is opposed to the group welfare maximizing strategy to cooperate.

The subjects were told that they would play the above game between three and five times, each time in a new randomized group, and that they would receive the payoff from one of the games, chosen at random, in real money at the end of the experiment. This enables us to view each game as a one-shot interaction.

The computer system then grouped people at random, but in such a way that each player would over the course of the four games play against groups where zero people, one person, two persons, and all the other three persons had claimed to be sick in the first experiment. The order in which each person met these four games was also randomized.

Note that the marginal value of cheating is constant at SEK 20 no matter how many other people choose to cheat. Informing the subjects before each new game about the number of other group members who had claimed to be sick in the first experiment, created four randomized treatment conditions. The difference between the four treatments nicely isolates the effect of the induced expectations about how many group members would cheat.

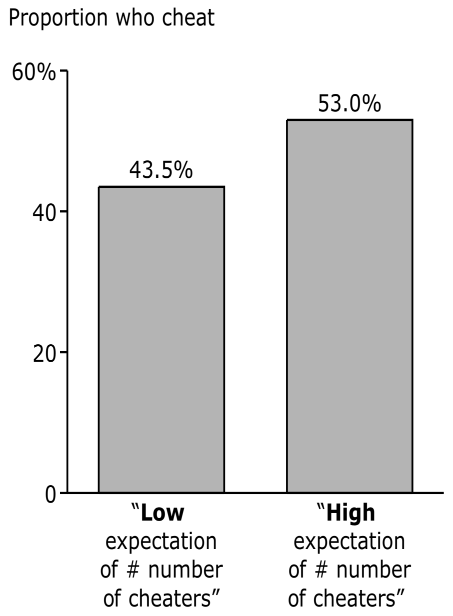
Results

In order to shed some light on the hypothesis that people abide more by rules if a larger fraction of others around them abide than when a lower proportion do, the four treatments in Experiment 2 can be pooled into two:

1. "Low expectation of # number of cheaters" - the treatments where 0 or 1 group member(s) claimed to be sick in Experiment 1
2. "High expectation of # number of cheaters" - the treatments where 2 or 3 group members claimed to be sick

From *Figure 3* it can be seen that cheating in the "high expectations" treatment occurs 10 percentage points or 22 percent more often than in the "low expectations" treatment. This difference is significant at the 5 percent level with a one sided test (p = 0.026). This implies that the null hypothesis, that there is no difference in the extent to which people abide by rules when a larger fraction of others around them abide versus when a lower proportion abide, can be rejected in favor of our alternative hypothesis that such a difference does exist.

Figure 3

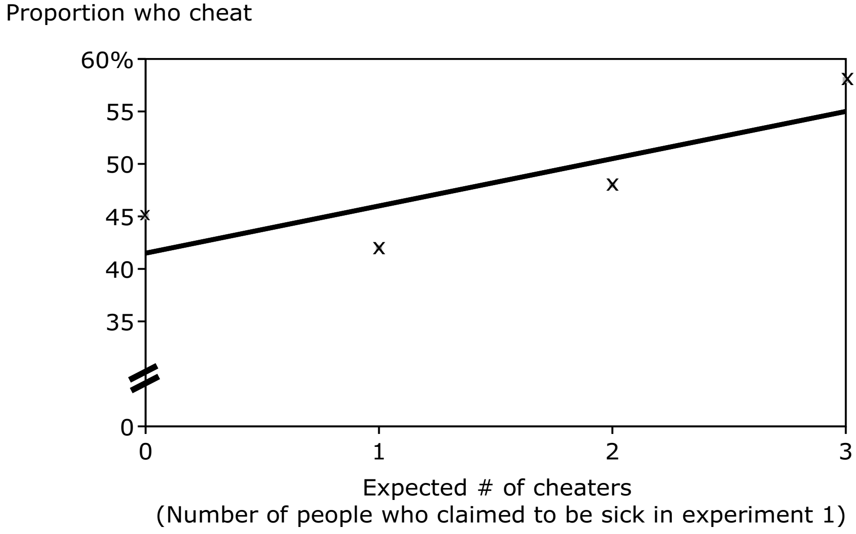


Looking at the data in more detail, by examining all four treatment conditions (information about whether 0, 1, 2 or 3 people claimed to be sick in experiment 3) the picture in Figure 4 emerges. Visual examination indicates that the higher the expected number of cheaters, the more likely a subject is to cheat herself. The difference in the observed proportions cheating between the 0 and 3 treatments and between the 1 and 3 treatments are significant at conventional levels. The difference between treatments 0 and 3 is 13 percentage points and the p-value is 0.030. The difference between treatments 1 and 3 is 16 percentage points with a p-value of 0.011. All other pair-wise comparisons are insignificant, including the difference between the treatment 0 and 1.[13]

---

[13] This "kink" also existed in a pilot study to this experiment. Although I believe that it is due to random sampling errors, it is possible that it is actually showing something that is real. One plausible explanation goes as follows: Some people might have an aversion to "bringing

Figure 4



The visual pattern in Figure 4 is confirmed by running different configurations of fixed effects, random effects and probit regressions (see Table 6 below). In all of these regressions, the null hypothesis, that there is no difference in how much people abide by rules when a larger fraction of others around them abide versus than when a lower proportion do, can be rejected in favor of our alternative hypothesis that such a difference does exist. For instance in the fixed effect regression, a person is, on average, 14 percentage points or 33% more likely to cooperate when facing a group in which none of the others in the group had misused the system in the first

down" the insurance system. They might reason that "If no-one else will misuse the system, I should be able to free-ride without causing any substantial harm. But if one person will misuse the system, then my misuse might create a large burden on the system (and in a dynamic setting, trigger a downwards spiral), which I do not want to be responsible for. However, if two or three people will misuse the system, then at least one or two of them are free-riding, so the system has already failed and I am not going to be a sucker, so I will hence free-ride."

experiment, than when facing a group where all three peers had misused the system.

While the point estimates indicate a potential kink in the pattern (between Info0 and Info1), the kink is insignificant in all regressions.

The second, third and fourth regressions include dummy variables for the order in which the treatment was played. Although, there is a slight indication that people cheat more in later round, the relationship is not statistically significant.

Table 6 - Panel and Probit Regressions on Experiment 2

	Co-operate (F.E.)	Co-operate (R.E.)	Co-operate (R.E.)	Co-operate (Probit)
Constant	0.42	0.48	0.41	
(z-stat)	*(10.5)***	*(6.3)***	*(3.88)***	
Info0 dummy	0.14	0.13	0.13	0.14
(z-stat)	*(2.4)***	*(2.4)***	*(2.4)***	*(2.9)**
Info1 dummy	0.17	0.17	0.17	0.17
(z-stat)	*(3.0)***	*(3.0)***	*(3.0)***	*(2.4)***
Info2 dummy	0.10	0.10	0.10	0.11
(z-stat)	*(1.9)**	*(1.8)**	*(1.8)**	*(1.4)*
Period1 dummy		0.07	0.07	0.07
(z-stat)		*(1.2)*	*(1.2)*	*(0.9)*
Period2 dummy		0.09	0.09	0.10
(z-stat)		*(1.6)*	*(1.6)*	*(1.2)*
Period3 dummy		0.02	0.02	0.02
(z-stat)		*(0.3)*	*(0.3)*	*(0.2)*
Male dummy			-0.03	-0.04
(z-stat)			*(-.05)*	*(-.07)*
SSE dummy			-0.17	-0.19
(z-stat)			*(-1.9)**	*(-2.8)***
KI dummy			-0.21	-0.23
(z-stat)			*(-2.1)***	*(-3.1)***
KTH dummy			-12	-13
(z-stat)			*(-1.0)*	*(-1.4)*
Given system in Exp 1			-0.10	-0.11
(z-stat)			*(-1.4)*	*(-2.0)***
Co-operation in Exp 1			0.27	0.28
(z-stat)			*(3.7)***	*(5.1)***

A quantification of the size of the peer effect can be found in the two regressions below. Given that the potential kink seen in Figure 4 turned out to be insignificant in all the regressions and tests performed to evaluate it, an indicator variable summarizing the treatment information is created. The treatment variable is simply assigned the values 0, 1, 2 or 3 for the treatments where 0, 1, 2 or 3 of the peers had used the system in the first experiment.

In all configurations of the regressions, the picture is the same: increasing the expectation of misuse by one incremental person increases the likelihood to cheat by approximately 5 percentage points or 11% (see Table 7).

Table 7 - Treatment when using other controls

	Co-operate (R.E.)	Co-operate (Probit)
Constant	0.58	
(z-stat)	*(5.5)***	
Treatment (0, 1, 2, 3)	-0.046	-0.051
(z-stat)	*(-2.6)***	*(-2.1)***
Period1 dummy	0.07	0.08
(z-stat)	*(1.3)*	*(1.0)*
Period2 dummy	0.09	0.10
(z-stat)	*(1.5)*	*(1.3)*
Period3 dummy	0.02	0.02
(z-stat)	*(0.3)*	*(0.2)*
Male dummy	-0.03	-0.04
(z-stat)	*(-.05)*	*(-.07)*
SSE dummy	-0.17	-0.19
(z-stat)	*(-1.9)**	*(-2.8)***
KI dummy	-0.21	-0.23
(z-stat)	*(-2.1)***	*(-3.1)***
KTH dummy	-12	-13
(z-stat)	*(-1.0)*	*(-1.4)*
Given system in Exp 1	-0.10	-0.11
(z-stat)	*(-1.4)*	*(-2.0)***
Co-operation in Exp 1	0.27	0.28
(z-stat)	*(3.7)***	*(5.1)***

VI – Summary and conclusions

Both experiments provide clear evidence that people are motivated by concerns additional to those of material payoffs. These concerns are shown to be strongly connected to the concept of reciprocity.

Interestingly, in contrast to other literature, Experiment 1 indicates that voting over or having a higher degree of participation in the devising of an insurance system does not have any positive impact on the levels of free-riding. Further research is needed to better understand this. It is possible that participation has two opposing effects:

A) creating shared beliefs of fairness (thus decreasing free-riding), and
B) inducing negative reciprocity due to conflict during decision making (thus increasing free-riding).

Depending on the relative strength of the two effects, the idea "Participation decreases free-riding in games of moral hazard" must perhaps be complemented with, or even substituted by, an idea such as "Outside authority is a solution to intra-group conflict".

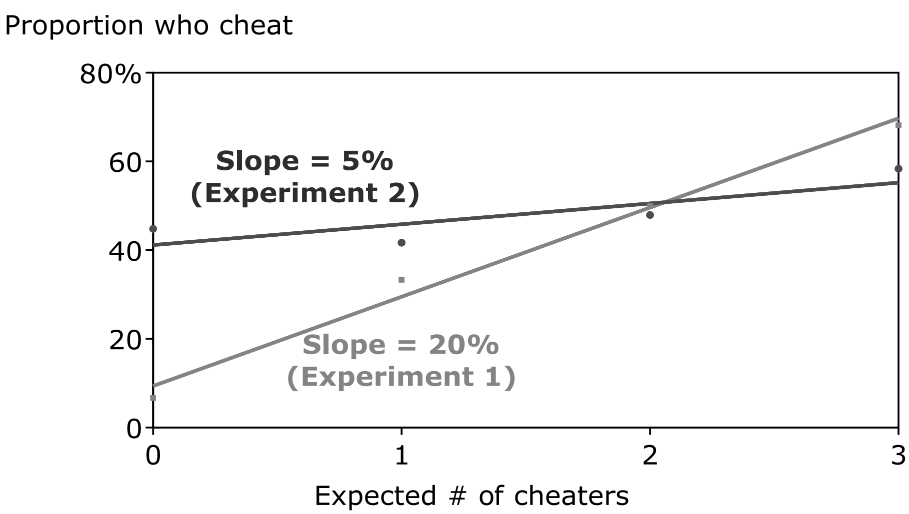
From Experiment 2 it is clear that previous trends in peers misusing or not misusing the system have an effect on future levels of free-riding. People are more likely to lie and cheat when they believe that others will lie or cheat. The effect is of an economically significant magnitude. Increasing the expectation about the number of cheaters by approximately one person increased the likelihood of cheating by 5 percentage points or by 11%. It would be interesting to replicate this effect to further calibrate its magnitude and to investigate how the effect potentially varies in different contexts.

Although outside the original scope of this study, indicative support is also found for the false consensus effect. The false consensus effect refers to the finding that people who engage in a given behavior believe that the behavior is more prevalent than those who do not (Ross, Greene and House [1977]). Using an intra-experiment comparison provides an opportunity to evaluate this potential effect.

In Figure 2 in Experiment 1, we saw a relationship between the expected number of cheaters and the proportion who decided to cheat. Though not exactly comparable, Figure 4 in Experiment 2 also shows the relationship between the expected number of cheaters and the proportion who decided to cheat – but now the expected number of cheaters is given largely exogenously (by the information in each treatment regarding exactly how group members had used the system in the first experiment).

The large difference between the slopes of these two graphs, 20% in Experiment 1 versus 5% in Experiment 2, strongly suggests that the false consensus effect was present and strong in the first experiment.

Figure 5



Appendix i– Formal Formulation of the Insurance Game in Experiment 1

Types: $\theta_i \in \{healthy, sick\} := \{h, s\}$

Messages depends on types: $\sigma_i(\theta_i) \in \{\hat{h}, \hat{s}\}$.

Payoffs: Let P_0 be the initial endowment. Further, P_1 is the wage for a healthy worker, C is the social insurance contribution per worker, and P_I is the ex ante determined social insurance transfer to the sick. We have $C \leq P_0$. Let ι_i^s be an indicator variable taking the value 1 if player i reports sick. Any feasible insurance system must satisfy

$$P_I' \sum_{i=1}^{4} \iota_i^s = 4C \leq 4P_0.$$

The maximum transfer is

$$P_I^{\max} = \frac{4P_0}{\sum_{i=1}^{4} \iota_i^s}$$

Let $\bar{P}_I(\sigma, \theta)$ be the actual transfer for a given strategy profile σ and type profile θ. We have

$$\bar{P}_I(\sigma, \theta) = \min\{P_I, P_I^{\max}(\sigma, \theta)\}.$$

Finally, let $C(\sigma, \theta)$ be the social insurance contribution for a given strategy profile σ and type profile θ.

Given a strategy profile σ_{-i} of the other players, the expected payoffs to a healthy player i is

$$\begin{aligned} \pi(\hat{s}; h, \sigma_{-i}) &= E_\theta \left[\sqrt{P_0 + P_1 - C(\hat{s}, \sigma_{-i}, \theta) + \bar{P}_I(\sigma, \theta)} \right] \\ \pi(\hat{h}; h, \sigma_{-i}) &= E_\theta \left[\sqrt{P_0 + P_1 - C(\hat{h}, \sigma_{-i}, \theta)} \right] \end{aligned}$$

A sufficient condition for a healthy player to always report sick, is that

$$-C(\hat{s}, \sigma_{-i}, \theta) + \bar{P}_I(\sigma, \theta) \geq -C(\hat{h}, \sigma_{-i}, \theta)$$

for all type profiles θ. Can rewrite this as

$$\bar{P}_I(\sigma, \theta) \geq C(\hat{s}, \sigma_{-i}, \theta) - C(\hat{h}, \sigma_{-i}, \theta).$$

In other words, the transfer must be greater than the increase in contribution due to player i also reporting sick. It is clear that this condition always will be satsified.

Case 1: If the feasibility condition always is slack (P_I is sufficiently low), C will not increase if i also reports sick, and since $P_I \geq 0$, this condition is satisfied.

Case 2: If the feasibility condition is binding when i does not report sick, this implies that C cannot increase further, and since $\bar{P}_I$ then must be positive, the condition is satisfied.

The final case to check is when the feasibility condition is slack with i not reporting sick, but binding if i reports sick.

Case 3: In this case we have (let n others report sick):

$$\begin{aligned} P_I n &= 4C \\ P_I(n+1) &\geq 4P_0 \\ \bar{P}_I &= \frac{4P_0}{n+1} \\ &\Rightarrow \Delta C = P_0 - C = P_0 - \frac{P_I n}{4}. \end{aligned}$$

We need

$$\begin{aligned} \bar{P}_I &= \frac{4P_0}{n+1} \geq P_0 - \frac{P_I n}{4} = \Delta C \\ &\Updownarrow \\ \frac{4P_0}{n+1} &\geq P_0 - \frac{P_I n}{4} \\ &\Updownarrow \\ P_0\left(\frac{4}{n+1} - 1\right) &\geq -\frac{P_I n}{4}, \end{aligned}$$

which is always satisfied, as $\frac{4}{n+1} \geq 1$. Thus, it is always a weakly dominant strategy for a healthy player to report sick (strictly if $P_I > 0$).

Appendix ii – Key Parts of the Experiment Instructions

Paper Instructions for Experiment 1 (Page 1)

You will soon participate in an experiment where the aim is to study economic decision making. The experiment will take about 45-75 minutes.

It is important that you read and understand these instructions carefully. Raise your hand and one of the leaders of the experiment will come and answer any questions you might have individually.

Your decisions will be anonymous in relation to the other participants.

Apart from the show-up compensation of SEK 50, you will also receive points, which will depend on your decisions, the decisions of others and chance.

To model decreasing value of the points, the points will be exchanged for real money at the end of the experiment, according to the decreasing rate of exchange, which you have got on your other paper[14].

Extract from the "exchange rate"

No. of points from experiment	0	1000	2000	3000	4000	5000	6000
SEK in hand after experiment	0	28	38	47	54	60	65

Above you can see that the amount of SEK per incremental point becomes less the more points a person gains. For example, for the first 3 000 points, one would get SEK 47, while for the next 3 000 points, that is the points 3 001–6 000, one would only get SEK 18 (65-47).

When everything is completed, the sum of the compensation from the experiment and the show-up compensation of SEK 50 will be paid to

[14] The participants also received a more granular version of this table.

you in another room by persons other than those leading the experiment.

It is important that you take into account that the rate of exchange is decreasing when you later make decisions. Therefore make sure, when it feels appropriate, to refer to the table with the exchange rate between points and SEK that you have on paper.

At some points during the experiment you may have to wait. We ask you to be patient at those occasions.

When you click on "Continue" on your computer, you will be asked to answer some control questions. Therefore, do not continue until you feel that you have understood the above information.

Paper Instructions for Experiment 1 (Page 2)

<u>Groups of 4 participants</u>
In this experiment you will randomly be paired together with three other participants.

<u>1000 points to begin with</u>
Each person automatically starts with 1000 points.

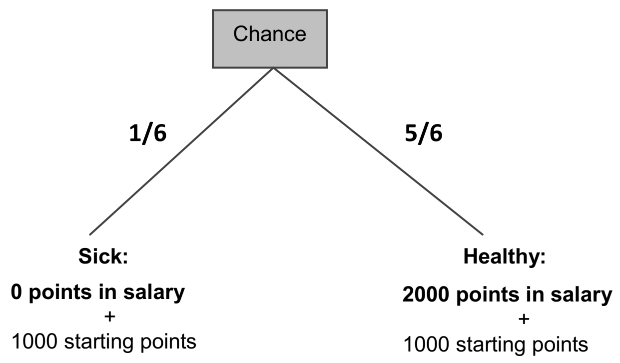
<u>If sick - which occurs with a 1/6th probability – a person gets 0 points in salary</u>

A participant will

- with a probability 1/6 end up in the state of "sick" and earn 0 points salary

and

- with a probability of 5/6 to end up in the state of "healthy" and earn 2000 points salary



<u>Insurance System</u>

Each group will have an insurance system that works in the following way:

- Each participant commits to contribute between 0 and 1000 points to an insurance fund. This means that each person (sick as well as healthy) will pay 1/4 of the insurance costs up to the potential maximum that is capped at 1000 points each.

- The system has a compensation level of x% of the 2000 point salary (x will be between 10% and 100%; more on x later).

- If a person ends up in the state "sick," he/she gets 0 points salary + compensation from the fund as follows:

1. If the fund has enough points, the person or persons who reported being in the state "sick" receive the compensation x% of 2000.

2. If the points in the fund are not enough to cover all claims on the fund, the person or persons who potentially have reported being in the state "sick" will equally share the points that are in the fund.

- Since it is not possible for group members to observe the state in which other persons are, is it possible for a person who is in the state of "healthy" to lie and report that he/she is in a state of "sick". If a person in "healthy" state does this he/she will get his/her salary plus compensation from the fund

When you click "Continue" on your computer you will be asked to answer some control questions. Therefore do not continue before you feel that you understand the above information.

Other Key Extracts from Instructions/Decision Making Phases of the Software (Experiment 1)

Decision on insurance level

Together with the three other participants in your group, you are now going to decide what the level of compensation shall be in your insurance system.

With 50% probability you will play in your own system, that is the system you discussed and decided on together

and

With 50 % probability you will play in a system that the experiment leader has stipulated for you

You will have ten minutes to discuss and debate before you will make your decision.

In order to get the decision registered, everybody has to agree and state the same level of compensation.

If contrary to expectations, you cannot agree, you will get 0 points and therefore SEK 0 for this experiment.

Continue >>

Discussion and decision

Here you have the possibility to communicate with the other players in your group. Your name is player [and then one of 1, 2, 3 or 4].

All discussion takes place with the help of written messages. For technical reasons, it is better that you write whole sentences instead of only single words before you click on "Send".

When you feel that you are beginning to reach an agreement, you can try to make the decision by choosing a level in the box at the bottom and then clicking on "OK".

Everybody has to choose the same level of compensation in order to get a decision registered. You should therefore consider not choosing any level of compensation until you agree in the chat.

Information about the decision
Your group has decided on a level of compensation of X%.

When all groups have finished making their decision it will - through a randomization process - be decided if you and your group are allowed to keep the insurance system you have discussed and decided on together.

With a probability of 50% you will be allowed to play in your own system, that is, the system you discussed and decided on together, and with a probability of 50% you will instead play in the system that the leader of
the experiment has stipulated for you.

Continue >>

Please wait.

Waiting for all groups to finish making their decisions.

[And then either]

Your group is allowed to keep the system you agreed on.

Through the 50-50% randomization process it is decided that your group's insurance system will be the one that you discussed and agreed on together.

The level of compensation is therefore as you decided x%.

[or]

Your group will not keep the system you decided on.

Through the 50-50% randomization process it is decided that your group's insurance system will be one in which the experiment leader has stipulated the level of compensation.

The level of compensation is therefore NOT x% that you decided, but y% as the experiment leader stipulated.

Decision of action

As you earlier were informed, the probability of a person ending up in the state “sick” is 1/6 and the probability of ending up in the state “healthy” is 5/6.

In the case that you end up in state "sick", your state will automatically be stated as "sick".

What is your choice for the potential case were chance decides that your state is "healthy"?

- (●) Report truly that my state is "healthy"
- ( ) Lie and report that my state is “sick”

Paper Instructions for Experiment 2

Random new groups of four, still anonymous

You will, as earlier, be exactly four persons in each group. These new groups will be put together using chance. Your decisions are still completely anonymous in relation to other participants.

Same probabilities, but new payoffs

The probabilities are exactly the same as before, 1/6 to end up in the state of "sick" and 5/6 to end up in the state of "healthy". What is new is that if you turn out to be "healthy", you will now get SEK 40 salary and if you turn out to be "sick", you get SEK 0 salary.

No decreasing "exchange rate"

There is no decreasing exchange rate of points for SEK here - exactly the amount of SEK you gain in this experiment, you will later be paid for participating in this experiment. As there is no decreasing exchange rate, every incremental crown is in other words worth the same.

Each individual pays 1/4th of any costs

As before, the insurance is financed by both those potentially reporting "healthy" and those potentially reporting "sick". Each individual will consequently always pay 1/4th of any potential costs for the insurance, whether he/she is "healthy" or "sick".

Insurance level is 100%

The level of insurance is in this experiment 100%. That is, a person who is "sick" and reports him/herself as sick will get 100% of his/her SEK 40 salary in compensation (minus a fourth for the costs of the system).

Not possible to verify whether stated states are true states

It is not possible to verify with certainty if a person really is "sick", and therefore a person that is in the state of "healthy" can potentially lie and report that he/she is "sick".

Efficiency loss of lying

Unlike earlier, we have in this experiment incorporated some costs and problems with potentially reporting oneself as "sick" when one is not. (For example there are risks, one has to take precautionary actions, and so on). The money pay-off for being “healthy” and working, and at the same time lying and asking for insurance compensation, is set to SEK 60 - that is, there is a loss of efficiency of SEK 20 (40 + 40 -20 = 60).

No efficiency loss if insurance is used by a person in state “sick”

There is no loss of efficiency if insurance money is paid to any person(s) that really are in the state "sick".

References

Alm, J., B.R. Jackson & M. McKee. 1993. "Fiscal Exchange, Collective Decision Institutions and Tax Compliance." *Journal of Economic Behavior and Organization*, 22: 285-303.

Bardan, P. 2001. "Water Community: An Empirical Analysis of Cooperation on Irrigation in South India" In *Community and Market in Economic Development,* ed. M. Aoki and Y. Hayami.

Bowles, S. 1998. "Endogenous preferences: The Cultural Consequences of Markets and Other Economic Institutions." *Journal of Economics Literature*, 36: 75-111.

Charness, G. & M. Dufwenberg. 2006. "Promises & Partnership." *Econometrica,* 74: 1579-1601.

Corzo, V. & S. Giwa. 2002. "Health insurance and moral hazard – an experiment." Bachelor Thesis, Stockholm University.

Dawes, McTavish & Shaklee. 1977. "Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation." *Journal of Personality and Social Psychology*, 35: 1-11.

Dufwenberg, M. & U. Gneezy. 2000. "Measuring Beliefs in an Experimental Lost Wallet Game." *Academic Press.*

Dufwenberg, M. & M. Lundholm. 2001. "Social Norms and Moral Hazard." *The Economic Journal,* 111(473): 506-525.

Feld, L.P. & J-R. Tyran. 2002. "Tax Evasion and Voting: An Experimental Analysis." *Kyklos,* 55.

Frey, B. S. 1994. "How Intrinsic Motivation is Crowded Out and In.", Rationality and Society, 6, 334-52

Frey, B. S. 1997. "A Constitution for Knaves Crowds Out Civic Virtues." *Economic Journal*, 107: 1043-1053.

Gneezy, U. 2005. "Deception: The role of consequences." *American Economic Review*, 384-394.

Lindbeck, A. et al. 1999. "Social Norms and Economic Incentives in the Welfare State." *Quarterly Journal of Economics.*

Lindbeck, A. 2002. "Sverige är inne i en ond cirkel." *Dagens Nyheter*, DN Debatt - October 13.

Lindbeck, A. et al. 2003. "Social Norms and Welfare State Dynamics." *Journal of the European Economic Association* , 1(2-3): 533-542.

Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Ann Arbor, MI: University of Michigan Press.

Ostrom, E . 1998. "A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997." *American Political Science Review*, 92(1): 1-22.

Ostrom, E . 2000. "Collective Action and the Evolution of Social Norms." *Journal of Economic Perspectives*, Vol 14, No 2, 137-158

Pinker, S. 1994. *The Language Instinct.* New York: W. Morrow

Pommerehne, W.W. & H. Weck-Hannemann. 1996. "Tax Rates, Tax Administration and Income Tax Evation in Swizerland." *Public Choice,* 88: 161-170.

Ross, L., Greene, D. & P. House. 1977. "The False Consensus Effect: An Egocentric Bias in Social Perception and Attribution Processes", *Journal of Experimental Social Psychology*, 13: 279-301.

Sally, D. 1995. "Conversation and Cooperation in Social Dilemmas - A Meta-Analysis of Experiments from 1958 to 1992", *Rationality and Society*, vol 7

Seigel, S. 1956. *Nonparametric statistics for the behavioral sciences.* New York: McGraw Hill.

Stutzer, A. & R. Lalive. 2004. "The Role of Social Work Norms in Job Searching and Subjective Well-Being." *Journal of the European Economic Association*, vol. 2(4), 696-719, 06.

Walker, J. M., R. Gardner & A. Herr & E. Ostrom. 2000. "Collective Choice in the Commons: Experimental Results on Proposed Allocation Rules and Votes"*The Economic Journal,* 110: 212-234.

Chapter 3

Commitment and Impasses in Negotiation[1]

We show experimentally that, as Schelling argued more than 50 years ago, one significant reason for conflict in negotiation can be that people make aggressive commitments to increase their share of the pie, although they know that such commitments can lead to impasses. This is at odds with the general literature's requirement of incomplete information or irrationality for there to be impasses.

By studying bilateral bargaining where negotiators can, at a cost *C*, make commitments that succeed with a probability *P*, we find strong support for the predictions based on Crawford (1982). Some support, albeit weaker, exists for the primary predictions based on Ellingsen & Miettinen (2008); support for their secondary predictions is stronger.

The results suggest that *ex ante,* conflict (negotiators making demands that could lead to impasse) is more likely to occur where commitment devices are weak. However because commitment attempts often fail in this domain the *ex post* frequency of impasses is lower than for stronger commitment devices. Contrary to the predictions in most formal theory, in weak commitment device domains we frequently observe outcomes in which one party takes most of the pie and the other receives little. Although the driver is different, uneven outcomes are also common in domains of strong commitment devices.

Welfare destruction is higher than what most formal theory predicts, but lower than what Crawford and E&M predict. The ex post Pareto inefficiency is 60% and 15-30% of the rates predicted by Crawford and E&M respectively.

[1] I am grateful to Magnus Johannesson, Aduke Thelwell, Guhan Subramanian and Robert Östling without whose support, critique, and encouragement this project would not have happened. I thank Tore Ellingsen, Erik Lindqvist, Topi Miettinen, Geir Asheim and seminar participants at the Stockholm School of Economics, at the Swedish Research Institute of Industrial Economics and several of the course members in Max Bazerman's *Behavioral Approaches to Decision Making and Negotiations* Seminar at Harvard for valuable input.

Part of this project was carried out when I was a visiting doctoral student at the Department of Economics at Harvard University and later when I was a graduate student at the Harvard Business School. Financial support from the Jan Wallander and Tom Hedelius Foundation, the Dr Marcus Wallenberg Foundation and the Swedish Finance Political Research Institute is appreciatively acknowledged.

Finally, I thank Lauryn Hale and Brooke Jenkins for their help in executing the experimental sessions, and Marcus Dahlgren at DRAGNET (www.Dragnet.se) for invaluable programming support.

I – Introduction

Whether they manifest as strikes, job resignations, or trade embargoes, impasses in negotiation are Pareto-inefficient and create tremendous loss of social welfare. Hence, understanding why impasses occur is important. Yet so far, the existing literature primarily emphasizes only three reasons why impasses occur: due to incomplete information, bounded rationality or fairness concerns.[1]

However, more than 50 years ago, Schelling (1956, 1960 and 1966) proposed a fourth reason for conflict and impasses in negotiations which has received only scant attention in the literature. Schelling conjectured that despite the likelihood of impasse, two rational parties with complete information may both attempt to commit themselves to an aggressive bargaining stance, in the belief that doing so can allow them to extract surplus from an uncommitted opponent.

While versions of Schelling's argument have appeared in formal theory in Crawford (1982) and Ellingsen & Miettinen (2008), we are not aware of any experimental or empirical studies on any of the three theories.[2]

We therefore designed and conducted two experiments at the Harvard Business School in order to test the predictions that we formulate based on the theory of Schelling, Crawford, and Ellingsen & Miettinen. Our design enables us to study games of bilateral bargaining in which negotiators can, at a cost C, make commitment attempts that succeed with a probability P. The first experiment tests a broad set of predictions, in one-shot negotiations without feedback. The second experiment tests the robustness of the two main results from the first experiment by letting participants play

---

[1] See discussion in Ellingsen & Miettinen (2008) regarding "irrationality" and incomplete information as drivers; see Falk, Fehr and Fischbacher (2003) for a discussion of fairness concerns.

[2] For instance, an overview paper on bargaining experiments, "Reasons for Conflict" by Falk, Fehr & Fischbacher (2003) does not make any reference to similar concepts.

18 one-shot negotiations, each with the opportunity for feedback and learning.

In line with Schelling (1956, 1960 and 1966), we find that people make aggressive commitments in order to increase their share of the pie, even though they know that such commitments can lead to impasse.

We also find support for the prediction based on Crawford (1982) that attempting aggressive demands[3] is the dominant strategy in domains where commitment devices are relatively weak (that is, where $P < 0.5$). In addition, we find support for the prediction based on Crawford that efficient, fair and non-aggressive demand combinations[4] exist when commitment devices are strong (when $P > 0.5$).

We find support for the secondary predications based on Ellingsen & Miettinen (2008), that mixes of either remaining fully flexible or making aggressive demands dominate when either (i) commitment is certain (that is, $P = 1$) or (ii) when the cost of attempting commitment is low, but not very low. Although some support does exist, the balance of the evidence goes against Ellingsen & Miettinen's main prediction that aggressive commitment attempts should dominate even in regions where commitment devices are strong.

Combined, our results suggest that that ex ante conflict in bargaining – people making demands that could turn out to be incompatible with each other – is more likely to occur in situations where commitment devices are inexpensive to use, but not too credible. In these domains of weak commitment devices, we also observe frequent outcomes in which one party takes most of the pie and the other receives little. Although the logic is different, uneven outcomes are also common in domains of strong

[3] Aggressive demand = Incompatible demand = A demand for more than 50% of the pie, a demand such that if the other negotiator makes the same demand, both demands would sum to more than 100% of the pie.

commitment devices. These results run counter to the typical predictions in most formal theory that more symmetric "compromise outcomes" will be realized.

Welfare destruction (of sometimes more than 20% of the pie) is higher than the level most formal theory predicts, but lower than the level Crawford and E&M predict. The ex post Pareto inefficiency is 60% and 15-30% of the rates predicted by Crawford and E&M respectively.

The rest of this paper is organized as follows. In Section II we summarize the relevant theories of Schelling, Crawford, and Ellingsen & Miettinen, and formulate key predictions based on the theories. Section III goes through the experimental procedure. Section IV first presents the design of Experiment 1 and then the results of Experiment 1. Section V does the same for Experiment 2. Finally in section VI, we conclude and discuss some ideas for future research. [5]

---

[4] Non-aggressive demand = "Compatible demand" = A demand for 50% or less of the pie, such that if the other negotiator makes the same demand, both demands would sum to 100% or less of the pie.

[5] The Appendix contains key parts of the experimental instructions.

II – Theory and Hypotheses

More than 50 years ago, Schelling (1956, 1960 and 1966) proposed a reason for conflict and impasses in negotiation that has not received key emphasis in the literature. The first part of his argument is that "the power to constrain an adversary may depend on the power to bind oneself … to burn bridges behind one may suffice to undo an opponent" (Schelling [1956, p 282]). However, the key conjecture is that if two opponents pursue a 'burning bridge' strategy at the same time, they "run the risk of establishing an immovable position that goes beyond the ability of the other, and thereby provoke the likelihood of stalemate or breakdown" (Schelling [1956, p 282]). This is in stark contrast with the literature's emphasis on incomplete information, irrational negotiators and fairness concerns as the drivers of bargaining impasses.

In Schelling's reasoning, despite the likelihood of conflict, two rational parties with complete information may both attempt to commit themselves to an aggressive bargaining stance, in the belief that doing so can allow them to extract surplus from an uncommitted opponent.

While versions of Schelling's argument have appeared in formal theory through Crawford (1982) and Ellingsen & Miettinen (2008), we are not aware of any experimental or empirical studies on the theory. For instance, a overview paper on Bargaining experiments, "Reasons for Conflict" by Falk, Fehr and Fischbacher (2003), does not make any references at all to rational aggressiveness being a potential driver of conflict, nor does it mention any concepts similar to those used in Schelling, Crawford and Ellingsen & Miettinen's theories.

It is therefore natural and potentially fruitful to formulate predictions based on the Schelling, Crawford and Ellingsen & Miettinen's theories that we can take to the lab. Our first prediction is:

P$^{\text{Schelling 1}}$ *Aggressive, incompatible commitments can be a key driver of conflict in simple bilateral bargaining games*

Although Schelling does not build a formal model based on his argument, he does provide an intuition that attempts to make incompatible commitments (and hence create impasses) should occur more often when commitments are weaker in the sense that the commitment devises are less credible (Schelling 1966, Ch 3). The intuition here is that a negotiator in an environment characterized by a low probability of making a successful commitment attempt, knows there is a low probability that her opponent's potential commitment attempt will succeed at the same time as hers, and consequently feels a stronger temptation to attempt an aggressive commitment.

P$^{\text{Schelling 2a}}$ *Attempts to make incompatible commitments occur more often the weaker the commitment devices are*

Based on Schelling's arguments, Crawford (1982) produces a formal model in which negotiators can attempt to commit at the same time and independently choose their stances. In his leading version of the model, there are two Phases. In Phase 1, each player decides either to commit to some demand or to wait and remain flexible. In Phase 2, any commitments (independently) succeed with probability P and fail with probability $(1-P)$[6]. A player who successfully achieves commitment cannot revoke her demand while a player who is flexible best-responds. The cost of attempting commitment is zero.

In line with Schelling's intuition, if the cost of going back on commitments is weak, that is in domains of weak commitment devices, the

[6] We chose to discuss Crawford's model with a similar notation to that used in Ellingsen & Miettinen (2008) instead of Crawford's original notation, to ensure consistency throughout the paper. Also this is the way the experiment was formulated for the participants.

unique equilibrium involves both players attempting to make incompatible commitments. The condition for this is that $P < 0.5$.

P[Crawford 1] *When the cost of commitment is zero and commitment devices are weak (P < 0.5), the unique equilibrium involves incompatible commitments from both players*

The strong feature of this prediction is that it only relies on strategic dominance arguments as solution concept, and the model actually reduces to a Prisoner's Dilemma.

However, in environments where credible commitment devices are available and commitment attempts hence become binding with a high likelihood, the model predicts that several equilibria exist, including fair compromise and efficient equilibria.[7]

P[Crawford 2] *When the cost of commitment is zero and commitment devices are strong (P > 0.5), an efficient and equitable equilibrium exists where both players commit to half the pie*

In a recent American Economic Review article, Ellingsen & Miettinen (2008) make the objection that Crawford's theory primarily explains how impasses can occur in domains where commitments are weak and *P* is low. They argue that "negotiators often have access to quite efficient commitment technologies" and "[s]ince both negotiators prefer strong commitments to weak ones, real negotiations may take place with values of [*P*] that are high" (Ellingsen & Miettinen [2008, p 1630]).

---

[7] With regards to his conjecture that conflict should be more common for weak commitment devices, Crawford (1982, p 635) says "[There] is not support for the common belief that making commitment more difficult (in the sense of lowering the probability of success) makes impasse less likely. The reason is that the "common-sense" argument used to reach this conclusion considers only the partial effect of lowering the success probability; it ignores the resulting changes in attempted commitment positions, whose effects generally go the other way and could easily swamp the partial effect."

Ellingsen & Miettinen then demonstrate in two different ways, that contrary to Crawford's prediction, incompatible commitments will also dominate when *P* is high.

In one of their approaches, they retain all Crawford's assumptions and reach their result by imposing the stronger solution concept, that negotiators use strategies that are iteratively weakly undominated.

P$^{\textbf{Ellingsen\&Miettinen 1}}$ *When the cost of commitment is zero, and for all types of commitment devices (all positive P), the unique equilibrium involves incompatible commitments from both players*[8]

Ellingsen & Miettinen's other approach is more powerful in that it relies only on iterated strict dominance as solution concept. In it they perturb Crawford's leading model by introducing a positive cost of commitment. They distinguish between cases where the cost is 'very low' and cases were the cost is 'low, but not very low.' The main result is that for environments with strong commitment devices (as well as in environments with weak ones, as in Crawford) the unique equilibrium involves both negotiators making incompatible commitment attempts as the cost approaches zero.

While we will not reproduce Ellingsen & Miettinen's proof of why efficient and fair outcomes fail to exist, they state their logic as follows: Consider the case in which both negotiators commit to take exactly half the surplus. If it is costly to make such a commitment, a negotiator is better off being flexible, as this yields the same share of the surplus. At the same time, it cannot be an equilibrium that both

[8] See Asheim & Perea (2009a, b) for a clarification of the solution concept underlying this prediction. These working papers provide both support and critique of Ellingsen & Miettinen (2008) by extending the solution concepts to Proper Equilibrium, Proper Rationalizability and Elimination of Weakly Dominated Strategies with Only Two Rounds of Iteration.

negotiators remain flexible, since one negotiator is then better off by making an aggressive commitment.

The main prediction here is:

P[Ellingsen&Miettinen 2a] *When the cost of commitment is very low and commitment is uncertain (for all $P < 1$) the unique equilibrium involves incompatible commitments from both players*

This perturbed model also provides the following two predictions:

P[Ellingsen&Miettinen 2b] *When the cost of commitment is low, but not very low, and commitment is uncertain (for all $P < 1$) only "aggressive commitments" or "remaining flexible" should be played in equilibrium*

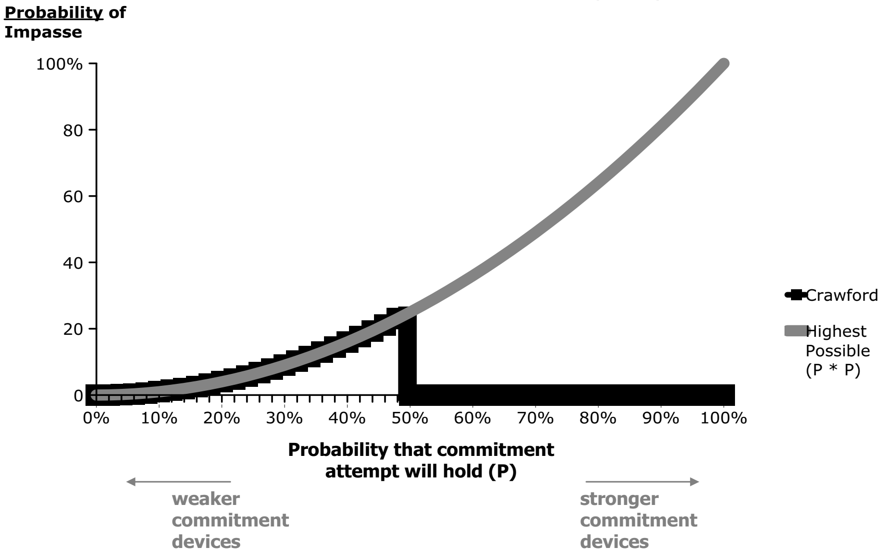
P[Ellingsen&Miettinen 2c] *When the cost of commitment is very low or low and commitment is certain ($P = 1$) only "aggressive commitments" or "remaining flexible" should be played in equilibrium*

These predictions by Schellig, Crawford and Ellingsen & Miettinen defy traditional bargaining theory. In most formal models of bargaining under complete and symmetric information, rational negotiators never end up in an impasse. In this sense, all three theories have in common the view that rational negotiators can end up in conflict because of a struggle to commit to aggressive demands, even though they know that such commitments can lead to an impasse.

However, the predictions of Schelling, Crawford and Ellingsen & Miettinen are also different: both because they are not fully overlapping, and because even where they do overlap, they provide opposite predictions in certain key domains. For instance, it is important to resolve the "debate" between Schelling and Crawford on one side, and Ellingsen & Miettinen on the other, about whether aggressive commitments should be a key driver of impasses in environments where commitment devices are strong.

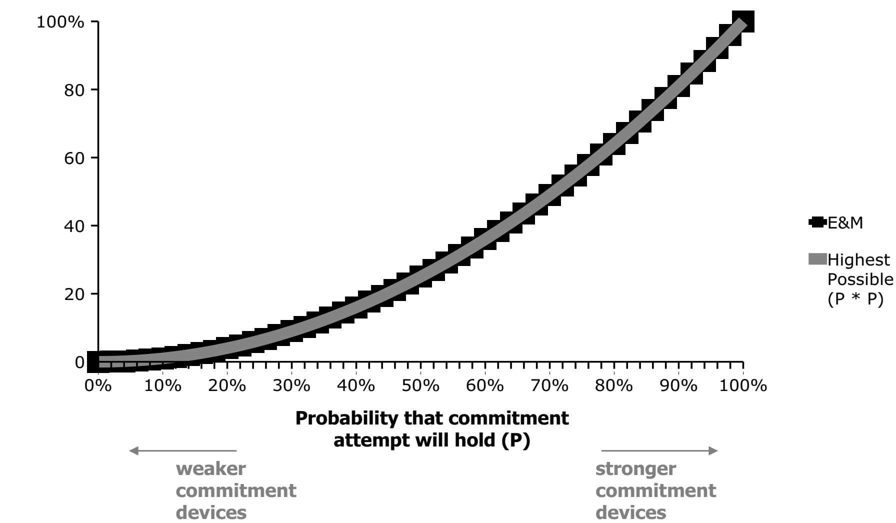
One other important matter to resolve is the extent to which aggressive, incompatible commitments—if they turn out to be common—do in fact create impasses. In Crawford's theory, the probability of impasse in the lowest impasse rate equilibrium[9] is as high as $(0.5)^2 = 0.25$ (Figure 1).

Figure 1 – Lowest Impasse Rate given our *Predictions based* on Crawford (*C* = 0*)*



Ellingsen & Miettinen's models are even more aggressive in that, for $C = 0$ or C very small, they predict impasses with close to full certainty when P is close to 1 (Figure 2).

Figure 2 – Lowest Impasse Rate given our Predictions based on E&M for *C* = 0 or *C* very small



So even if Crawford's model were to be the most accurate of the three, conflict could have a significant Pareto-inefficient welfare destroying effect in negotiations. We can therefore formulate the following final prediction, which is common to all three theories:

P$^{\textbf{welfare destruction}}$ *Aggressive, incompatible commitments lead to significant welfare losses due to frequent conflict and hence impasses*

[9] See Prediction **P**$^{\textbf{Crawford 2}}$ and Footnote 9 above.

III - Experimental procedure

The experiments designed to test the Schelling, Crawford and Ellingsen & Miettinen predictions ("Experiment 1") and the specific follow-on experiment with feedback and learning ("Experiment 2") were both conducted at the Harvard Business School. The three sessions of Experiment 1 were carried out in April 2009 and the two sessions of Experiment 2 were carried out in September 2009.

A total of 120 subjects were recruited from the first and second year classes of the MBA program. See Table 1a for demographic data.

Table 1a - Summary Statistics - Demographics

	Gender		Age			Undergraduate Education				
	Female	Male	Lower Qtr	Med-ian	Upper Qtr	Busin-ess	Econ	Social Sci/ Human	Enginer./ Natural Sci	Other
Experiment 1	28%	71%	24 yrs	27 yrs	28 yrs	28%	18%	24%	25%	6%
Experiment 2	30%	70%	26 yrs	27 yrs	28 yrs	18%	36%	18%	27%	0%

	Pre-MBA Job										
	Consul-ting	Venture Capital/ Private Equity	I-Bank-ing	Invest-ment Mgmt	Other Financial Services	Manu-facturing	Non-Profit	Cons-umer Products	Health-care/ Biotech	Military	Other
Experiment 1	40%	4%	10%	1%	4%	6%	3%	6%	8%	0%	18%
Experiment 2	39%	14%	7%	7%	5%	7%	5%	2%	2%	0%	14%

The experimental session of Experiment 1 lasted for about one hour while the experimental session of Experiment 2 lasted for about 80 minutes. After the experiment, subject earnings were paid in a separate location, using anonymous identification numbers, by people other than the experiment leaders. Each subject earned on average USD 23.6 for Experiment 1 and USD 50.6 for Experiment 2, including the guaranteed show-up fee.

Table 1b - Summary Statistics - Payoffs and Data

	Payoffs						Data	
	Average	Max	Median	Min	Show-up fee	Average + Show up	Subjects	Obser-vations
Experiment 1	13.64	30.00	15.00	-1.40	10	23.64	72	432
Experiment 2	30.57	90.00	29.85	-0.30	10	40.57	44	792

See Table 1b for further summary data on the payoffs. The 72 participants generated a sample of 432 observations in Experiment 1 and the 44 participants produced a sample of 792 observations in Experiment 2.

In order to enable the relatively complex re-matching of subjects into new pairs for each separate configuration of the *P* and *C* parameters, and to preserve anonymity in a double-blind sense, it was natural to execute both experiments in computerized form.

To make sure everyone easily would understand how to use the experimental software interface, it was done as a web interface.[10]

Throughout both experiments, the subjects were given instructions both verbally and via the computer interface. At less significant transitions, the computer took care of guiding the subjects forward without any particular verbal support. At one stage in each experiment, written instructions were handed out on paper, at which time an oral presentation was given. To guarantee understanding on key occasions, each subject had to pass a comprehensive battery of control questions to be allowed by the computer to continue to the next step. While these questions were quite complex and often involved several computations the median number of attempts needed to get the questions right was 1, and the average number of attempts was 1.2. We take this, in concert with exit interviews we conducted

[10] The program can be found at www.giwa.com/econ-experiments/ and was programmed in cooperation with Marcus Dahlgren at DRAGNET (www.dragnet.se).

on participants in the pilots (from the same subject pool), as evidence that the participating MBA students had a strong understanding of the game before they started to play.

The participants were asked to raise any questions by raising their hands, and general discussion was avoided by answering questions individually.[11] To make the instructions concrete and easy to absorb, both games were framed within the context of a two party negotiation. At the end of each experiment, all participants where given information on how to register to receive a draft version of this paper.

[11] The questions answered were solely to clarify procedure/instructions. No questions were asked or would have been answered, about the aim of the study until after both experiments were completed.

IV - Experiment 1: Design and Results

Design

Building on Crawford (1982) and Ellingsen & Miettinen (2008), we use a game of bilateral bargaining in which negotiators can, at a cost *C*, make commitment attempts that succeed with a probability *P*. This enables us to test all of the key predictions we formulated based on Schelling, Crawford and Ellingsen & Miettinen in one simple game, by exogenously varying the treatment variables *C* and *P* in different key configurations.

The game. Both players are given full information about the structure of the game's first and second stages and the exact values of *P* and *C*.

The pie that the negotiation is over is USD 30. To eliminate outcomes that entail negative payoffs, without changing the structure of the game, players receive a show-up fee of 10 dollars (larger than all values of *C*). The potential outcomes, not counting the show-up fee, range from a negative 3.3 dollars to a positive 30 dollars, making this a non-trivial negotiation.

Note the distinction between making a commitment attempt (which is a decision) and being committed or obtaining a binding/successful commitment (which is an outcome).

First stage. Each player *i* chooses simultaneously[12] either to (a) attempt commitment to some demand d_i, where *d* is between 1% and 100%, in increments of 1%, or (b) to wait *w*, and remain flexible.

---

[12] As discussed in Footnote 2 in Ellingsen & Miettinen (2008): Nash (1953), Schelling (1960) and Harsanyi & Selten (1988) and Crawford (1982) argue that "such simultaneous-move bargaining games can capture the dynamic give-and-take of bargaining and the role of expectations in determining outcomes."

Attempting commitment entails a cost C for each player that attempts it.

Second stage. With probability P_i, player I's commitment turns out to be successful and she is hence *committed*, with probability $(1 – P_i)$ the commitment failed and she is *uncommitted*. The same goes for player *j*. Each players' probability, P_i, of successful commitment is drawn from the same P, but P_i and P_j are independent,

Players are informed about both their own and the other player's committed/uncommitted status, but not about whether an opponent who is uncommitted had in fact attempted commitment.

Two uncommitted players: The players are each given half of the surplus. Any player that had attempted to commit also bares the cost, C.

One committed and one uncommitted player: The committed player *i* cannot revoke her demand. The uncommitted player *j* is given the residual share (as if she simply best-responds). Hence player *i* receives $d_i - C$ and Player *j* receives $(1 - d_i)$ and in the case she had attempted commitment also pays C.

Two committed players: If $d_i + d_j = 1$), then player *i* gets $d_i - C$ and player *j* gets $d_j - C$. If $d_i + d_j < 1$, they each get $d_i - C$ and $d_j - C$ and also equally share the remainder $(1- d_i - d_j)$[13]. If the commitments are incompatible (that is, if $d_i + d_j > 1$), then each player gets $–C$ (zero of the surplus while still paying the commitment cost).

[13] Crawford and Ellingsen & Miettinen both hold for a set of sharing rules. The one we have chosen is the simplest one for someone to calculate in their head: "Take the pie minus my share and minus the others share. Now simply split the rest by half". If players were to obtain exactly d_i and d_j, we would simply have a version of the Nash demand game were unclaimed surplus is wasted; this is somewhat unintuitive. The special case considered by Ellingsen (1997) where the full surplus is allocated proportional to relative claims would also have worked, but the computations would be more difficult for a person to do in their head.

Choice of parameters to operationalize our predictions into testable hypotheses. We can operationalize all of the predictions we formulated above based on Crawford and Ellingsen & Miettinen's models by using a relatively parsimonious set of parameters. Certainty trivially has to be $P = 1$, and zero cost is trivially $C = 0$.

We set $P = 0.75$ to represent environments where commitments are quite strong and credible, but not certain. This choice is made since the cut-off for weak commitments is $P = 0.5$, hence $P = 0.75$ is exactly between this cut-off and our certainty condition $P = 1$. For weak commitments, we could use any $P < 0.5$, so we chose $P = 0.25$ in order to again be in the middle of the 0 to 0.5 interval.

When the pie is USD 30, as in our experiment, Ellingsen & Miettinen's "very low cost" implies $0 < C < \$2.8$ in the most restrictive case; so we chose the midpoint $C = \$1.4$. The "low cost" situation for a USD 30 pie implies $\$2.8 < C < \$3.8$ in the most restrictive case, and we opted for the midpoint $C = \$3.3$[14]

In summary, we will use the following parameters to turn the predictions by Schelling, Crawford and Ellingsen & Miettinen into testable hypotheses.

---

[14] The value USD 3.30 works for both values of P because it is lower than USD 3.80, which is the lowest of the two upper bounds for *C* "low." In the case where $P = 0.25$, the upper bound for C to still be "low" is USD 3.80; when $P = 0.75$ it is USD 11.30. The upper bound for *C* to be "very low" is USD 2.8 in both cases. The formal conditions in Ellingsen & Miettinen (2008) are as follows.
For Uncertain Cases (P < 1): A) Upper bound for *C "low":* $C < P *$ Beta * X; B) Upper bound for *C "very low":* $C < P * (1-P) *$ Beta * X
For Certain Cases (P = 1): Upper bound for any *C*: $C <$ Beta * X (that is $C < \$15$ in our case)
where Beta = 50% is the share received if both are flexible and X = USD 30 is the size of the pie.

Table 2 - Parameters to Operationalize Predictions into Testable Hypotheses

Cost of commitment	C-value
zero	0 USD
very low	1.4 USD
low	3.3 USD

Comitment device	P-value
weak	25%
strong	75%
certain	100%

Experimental execution of the game. To increase the power of the study, each subject plays against six different randomized opponents, with each game randomized with respect to *C*-values and *P*-values. No feedback of the outcome is provided until a player has participated in all six treatments. To increase the stakes to a perceived USD 30, we let subjects know that they will be paid for one of the six negotiations, but that they will not find out which one until they have completed all six games.

To be able to control for any potential order effects, we record the order in which each individual meets each of her treatments. The design enables us to get 432 observations out of the 72 participants in this experiment.

Results

In Figure 3 we provide an aggregate overview of the results and in Table 3 we below provide a summary overview of each treatment.

Figure 3 – Summary of All Decisions

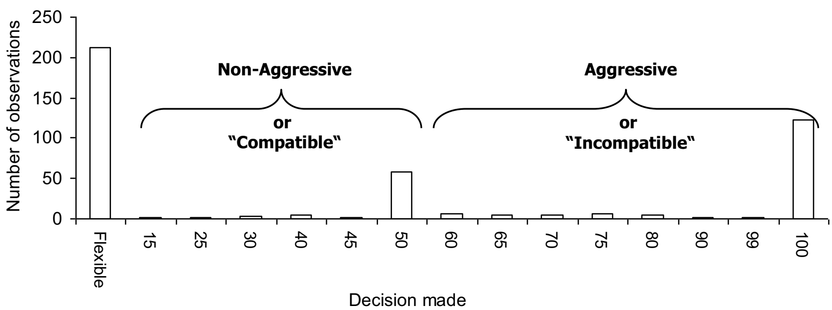


Table 3 Proportion of participans who chose each of the three key types of strategies for each of the different configurations of C and P

Table 3a

All C values		Share of participants who chose: Remain flexible (a)	Non-Aggressive Demand (b)	Aggressive Demand (c)
P = Value	25%	43%	6%	51%
	75%	52%	17%	31%
	100%	53%	24%	23%
	All P	49%	16%	35%

Table 3b

C = 0		Share of participants who chose: Remain flexible	Non-Aggressive Demand	Aggressive Demand
P = Value	25%	24%	10%	67%
	75%	42%	22%	36%
	100%	33%	47%	19%
	All P	31%	22%	47%

Table 3c

C > 0		Share of participants who chose: Remain flexible	Non-Aggressive Demand	Aggressive Demand
P = Value	25%	63%	1%	36%
	75%	55%	17%	28%
	100%	58%	17%	25%
	All P	59%	13%	29%

Table 3d

C = 1.4		Share of participants who chose: Remain flexible	Non-Aggressive Demand	Aggressive Demand
P = Value	25%	56%	3%	42%
	75%	53%	22%	25%
	100%	63%	15%	22%
	All P	58%	14%	28%

Table 3e

C = 3.3		Share of participants who chose: Remain flexible	Non-Aggressive Demand	Aggressive Demand
P = Value	25%	69%	0%	31%
	75%	57%	13%	31%
	100%	53%	19%	28%
	All P	59%	11%	30%

a) No commitment attempts are made.
b) Commitments for 51-100% of the pie. The vast majority in this category are for 100%.
c) Commitments for 1-50% of the pie. The vast majority in this category are for 50%.
Note: The proportions look very similar if we truncate and only display Remain Flexible, Commit to 50% and Commit to 100%. We use both of these sets of definitions for our regressions and explicitly show the almost identical results for the first regressions we report.

Every single value in the right-most column of Table 3 "Aggressive Demand", is significantly different from zero at the 1% level. One could argue that this is enough to support the prediction.

P$^{\textbf{Schelling 1}}$ *Aggressive, incompatible commitments can be a key driver of conflict in simple bilateral bargaining games*

However, a crucial question is if this behavior is based rational choices by the participants to make aggressive commitments in order to increase their share of the pie, although they know that such commitments can lead to a bust negotiation.

As we will see below, patterns in the proportions are systematic and in line with several of our predictions. This provides confidence to not only view the support in a statistical sense, but also in the cautious way required of a behavioral study that hopes to test predictions of a mathematical theory.

Further support is that all plausible forms of "irrational" behavior that we can think of, provide effects that go in the other direction, that is, they should make aggressive demands less likely[15]. That this is the case for any plausible equity concerns is trivial. Additionally, reciprocity could drive a person to commit to up to 50% (with the motivation that good behavior should be rewarded and bad behavior punished), but it should not drive anyone to demand more than 50%.

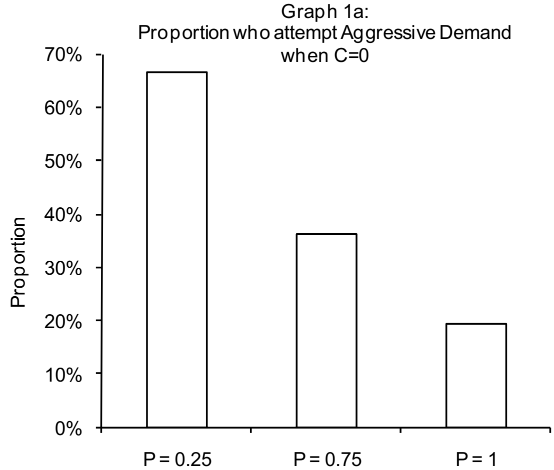
Since making an aggressive demand increases the amount of risk taken, risk-aversion also encourages non-aggressive demands. Only risk-loving preferences could be a partial explanation; however we do not have any reasonable explanations for why risk-loving preferences would explain the systematic patterns in when people opt to make aggressive demands and when they do not.

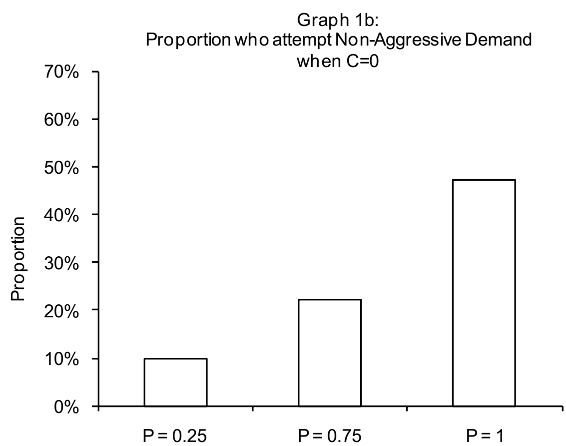
---

15 "Spitefulness" could potentially drive the effect, but we do not see that as plausible.

As seen in *Table 3* above, the frequency of incompatible commitment attempts are very different across different domains. Let us therefore visit the different domains and the sometimes opposing predictions by Crawford and Ellingsen & Miettinen.

Predictions based on Crawford. Clear support for $P^{Crawford\ 1}$ can be seen in the left bar in Graph 1a below. While the bar is not 100% tall, the dominant strategy is clearly aggressive demand. A one-sample test of proportions suggests that the point estimate is 67% and that the 95% confidence interval is 56% to 76%.





Furthermore, clearly the lower *P* is, the higher the proportion of participants who attempt to make aggressive demands. As can been seen below in Table 4, this "slope" is significant, which we take as further support for $P^{\text{Crawford 1}}$.[16]

Table 4 - Regressions on Demand 100% and Demand >50% when C = 0

	(1) Demand 100% (R.E)	(2) Demand 100% (logit)	(3) Demand >50% (R.E)	(4) Demand >50% (logit
Constant	0.15	-1.8	0.2	-1.4
(z-stat)	*(2.1)***	*(-3.8)****	*(2.7)****	*(-3.4)****
P.25-dummy	0.38	2	0.47	2.1
(z-stat)	*(4.8)****	*(3.7)****	*(5.4)****	*(4.3)****
P.75-dummy	0.20	1.3	0.16	0.85
(z-stat)	*(2.4)***	*(1.9)**	*(1.7)**	*(1.6)*
P.25=P.75?	Can Reject	Can Reject	Can Reject	Can Reject
(chi-2)	*(5.1)***	*(4.1)***	*(12.9)***	*(8.7)****

* = 10% level, ** = 5% level, *** = 1% level

We operationalize $P^{\text{Crawford 2}}$ as:

$H^{\text{Crawford 2}}$ *When $C = 0$ and $P > 0.5$ the proportion of negotiators who choose non-aggressive demands is higher than when $C = 0$ and $P < 0.5$*

Support for $H^{\text{Crawford 2}}$ can be seen in Graph 1b above. The low proportion of non-aggressive demands is 10% when $P = 0.25$, but more than doubles for $P = 0.75$ and almost increases five time times for $P = 1$. In Table 5 we report regressions results.

[16] The conclusions here and in the regressions below are robust to running OLS with clustering and robust standard errors, as well as FE panel regressions. Statistical tests show that results are also robust to order effects.

Table 5 - Regressions on Demand Smaller or Equal to 50% when C = 0				
	(5)	(6)	(7)	(8)
	Demand <=50% (R.E)	Demand <=50% (logit)	Demand <=50% (R.E)	Demand <=50% (logit)
Constant	0.86	-2.26	0.47	-0.11
(z-stat)	*(1.8)**	*(-6.20)****	*(-7.6)****	*(-0.3)*
P >= 0.5-dummy	0.18	*1.06*	-	-
(z-stat)	(4.62)***	*(4.09)****		
P.25-dummy	-	-	-0.37	-2.12
(z-stat)			*(-5.2) ****	*(-4.1)****
P.75-dummy	-	-	-0.25	-1.14
(z-stat)			*(-3.4) ****	*(-2.2)***
P.25=P.75?	-	-	Cannot Reject	Borderline
(chi-2)			*(2.5)*	*(3.0)**

* = 10% level, ** = 5% level, *** = 1% level

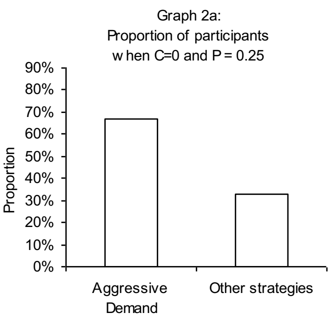
Pooling the $P > 0.5$ treatments and comparing these to the $P = 0.25$ treatment (that is, the $P < 0.5$ treatment), we see that the difference between $P > 0.5$ and $P < 0.5$ is statically different at the 1% level.

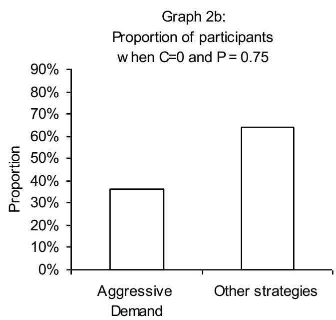
Looking at the individual treatments we see that the proportion when $P = 0.25$ is statistically different from when $P = 1.00$ at the 1% level. However, we cannot reject that the point estimate of 10% when $P = 0.25$ equal to the point estimate of 22% when $P = 75\%$ (except for at the 10% level in the logit regression).

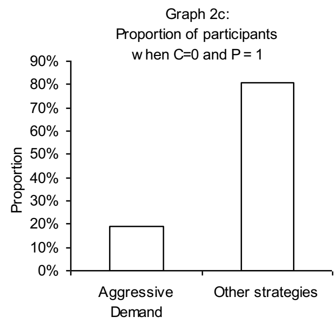
Our first predictions based on Ellingsen & Miettinen. Let us now visit one of the key predictions of Ellingsen & Miettinen for the case when $C = 0$.

$P^{Ellingsen\&Miettinen\ 1a}$ *When the cost of commitment is zero, for all P, the unique equilibrium involves aggressive commitments from both players*

The fact that we found support for $P^{Crawford\ 1}$, implies that we have found support for the part of $P^{Ellingsen\&Miettinen\text{-}1a}$ where $P < 0.5$, that is, in the domain where $P^{Ellingsen\&Miettinen\ 1a}$ is the same as $P^{Crawford\ 1}$. See Graph 2a.







However, in the more interesting cases where $P > 0.5$ and Ellingsen & Miettinen provide an almost opposite prediction to the one we based on Crawford, the support is much weaker for $P^{\text{Ellingsen\&Miettinen 1a}}$. From Table 3b, we can see that the proportion who make aggressive demands when $P = 0.25$, falls by roughly half when $P = 0.75$. A one-sample proportion test suggests that the point estimate is 36% and that the 95% confidence interval is 20% to 52%. From Graph 2b we also see that roughly two thirds of the participants do not make aggressive demands in this domain, but instead pursue other strategies.

The support is even weaker for $P^{\text{Ellingsen\&Miettinen 1a}}$ in the domain where $P = 1$. From Table 3b, we can see that the proportion who make aggressive demands falls further to 19% when $P = 1$. A one-sample proportion test indicates that a 95% confidence interval is from 7% to 32%. From Graph 2b we also see more than 80% of the participants pursue other strategies in this domain.

Note that this prediction, while among the most important ones based on Ellingsen & Miettinen, could be seen as one of the weakest since it relies on survival of iterated weakly undominated strategies as solution concept. What does the picture look like with respect to Ellingsen & Miettinen's potentially more powerful prediction based on survival of iterated strict dominance?

The second set of predictions based on Ellingsen & Miettinen. Let us begin with the case of the remaining ones that is most relevant for real life situations, where commitment is uncertain and commitment costs are very small:

$P^{\text{Ellingsen\&Miettinen 2a}}$ *When the cost of commitment is very low and commitment is uncertain (for all P < 1) the unique equilibrium involves aggressive demands from both players*

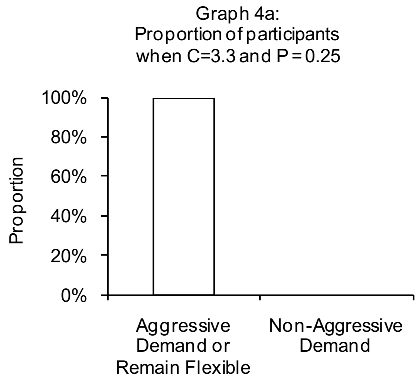
From Graph 3a and Graph 3b, we can see that far from 100% of the participants chose aggressive demands. While aggressive demands are not zero when P = 0.75 they are almost half of that when *P* = 0.25. This is in line with Schelling's conjecture of aggressive demand being more frequent when commitment devices are weaker, and goes against Ellingsen & Miettinen's theory.

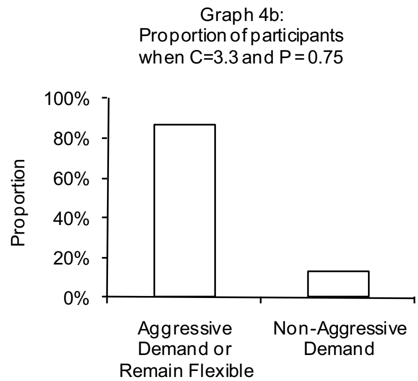
When *P* = 0.25, the proportion who choose aggressive demand is 42%, but is still dominated by the 56% that choose to remain flexible (see Table 3). A one sample proportion test indicates that a 95% confidence interval is from 26% to 58%.

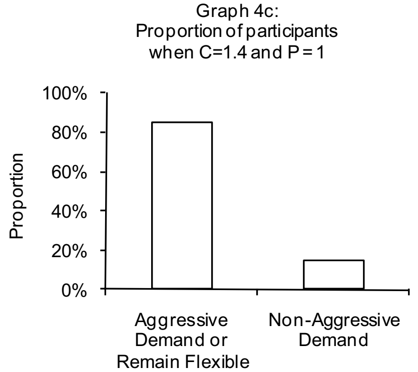
However, for *P* = 0.75 only 25% of the participants make an aggressive demand while 75% chose something else (of which 53 percentage points specifically choose "remain flexible", see Table 3). The 95% confidence interval is from 11% to 39%.

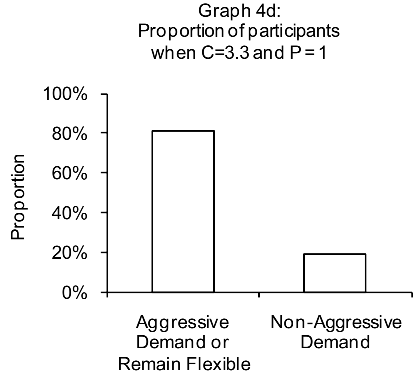
We conclude that also for this prediction where Ellingsen & Miettinen's theory predicts that aggressive demands will dominate in domains of strong commitments devices / high *P* (only) a minority of observations are consistent with their theory.

Strong support for Ellingsen & Miettinen's secondary predictions. For the domains in which Ellingsen & Miettinen predict that either "Aggressive Demands" or "Remaining Flexible" should be played (but never "Non-Aggressive Demands"), we find substantial support. The cases for the domains that $P^{\text{Ellingsen\&Miettinen 2b}}$ makes predictions about are shown in Graphs 4a and 4b, while the cases for the domains of $P^{\text{Ellingsen\&Miettinen 2c}}$ are shown in Graphs 4c and 4d.









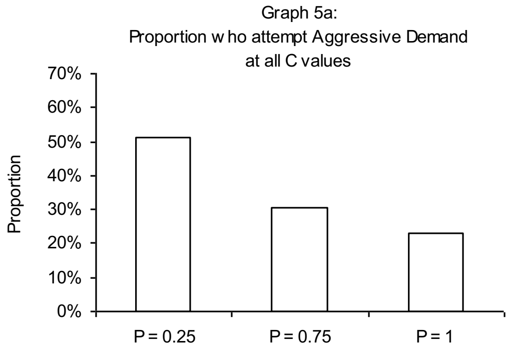
While only the case shown in Graph 4a has the proportion of Non-Aggressive Demand at strictly zero, in the other cases this proportion is close to zero (especially if compared to, for example, the case shown in Graph 1b above). The point estimates [95% confidence intervals] are 13% [6% to 20%] for Graph 4b, 15% [7% to 24] for Graph 4c and 19% [7% to 32%] for Graph 4d.

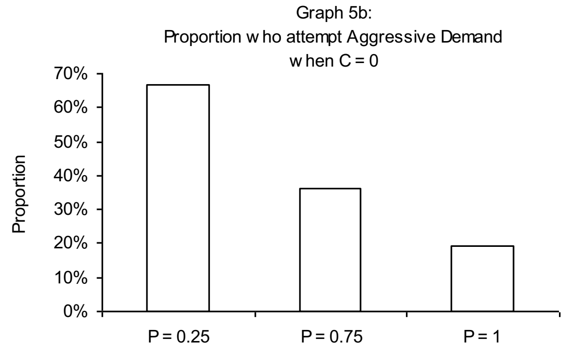
The result is interesting in that it suggests frequent outcomes where one party walks away with most or all of the pie and the other receiving nothing or little. This goes against most formal models in the literature.

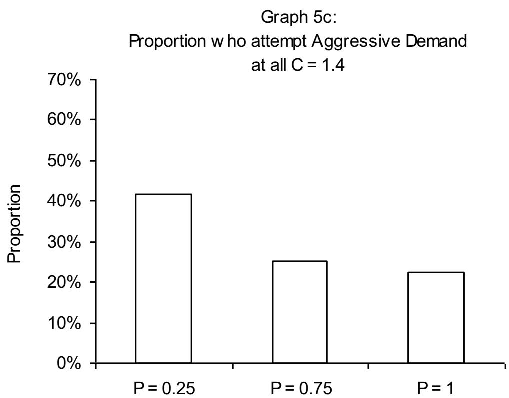
"Schelling's Slope". Let us now revisit the last prediction we made based on Schelling, which we call "Schelling's Slope of Conflict":

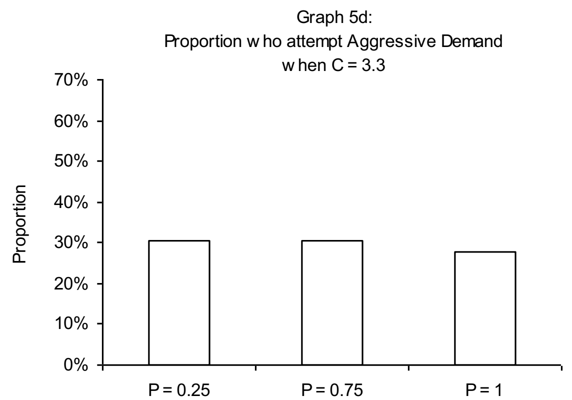
$P^{\text{Schelling 2a}}$	*Attempts to make aggressive demands occur more often, the weaker the commitment devices are*

That we found support for both predictions based on Crawford, implies that we have support for $P^{\text{Schelling 2a}}$ in the domain where $C = 0$. While Schelling did not formalize for what domains of C this conjecture should hold, Graphs 5a–5d show the results for all three cost domains in our experiment, and also results for the aggregate.


Graph 5a:
Proportion w ho attempt Aggressive Demand
at all C values
Proportion
70%
60%
50%
40%
30%
20%
10%
0%
P = 0.25
P = 0.75
P = 1


Graph 5b:
Proportion w ho attempt Aggressive Demand
w hen C = 0
Proportion
70%
60%
50%
40%
30%
20%
10%
0%
P = 0.25
P = 0.75
P = 1


Graph 5c:
Proportion w ho attempt Aggressive Demand
at all C = 1.4
Proportion
70%
60%
50%
40%
30%
20%
10%
0%
P = 0.25
P = 0.75
P = 1


Graph 5d:
Proportion w ho attempt Aggressive Demand
w hen C = 3.3
Proportion
70%
60%
50%
40%
30%
20%
10%
0%
P = 0.25
P = 0.75
P = 1

While there might be some form of slope in the domains where $C > 0$, the pattern is not as strong as when $C = 0$, so we now turn to regression analysis.

Table 6 - Regressions on Demand 100% for all C values

	(9) All C Demand 100% (R.E)	(10) C = 0 Demand 100% (R.E)	(11) C = 1.4 Demand 100% (R.E)	(12) C = 3.3 Demand 100% (R.E)
Constant	0.18	0.15	0.18	0.19
(z-stat)	*(4.2)****	*(2.1)***	*(3.6)****	*(2.7)****
P.25-dummy	0.25	0.38	0.17	0.09
(z-stat)	*(6.0)****	*(4.8)****	*(2.2)***	*(1.0)*
P.75-dummy	0.18	0.20	-0.04	0.07
(z-stat)	*(1.3)*	*(2.4)***	*(-0.5)*	*(0.9)*
P.25=P.75?	No	No	No	Not rejected
(chi-2)	*(22)****	*(5.1)***	*(6.5)***	*(0.8)*

* = 10% level, ** = 5% level, *** = 1% level

From Table 6, we see statistical support for many patterns shown above in the graphs. The regressions show that "Schelling's Slope" is steepest and most significant when $C = 0$.

As mentioned, Schelling does not formalize the relation between this conjectured slope and costs of attempting commitment. If we nevertheless apply his conjecture to both cost domains in Ellingsen & Miettinen's theory where they predict that aggressive demands should exist, we see that Schelling's Slope is, for the most part, significant when $C = 1.4$ – when C is very low – but we cannot conclude that it is significant when $C = 3.3$ – when C is low, but not very low.

Impasses and welfare destruction. Finally, we turn to the question of the welfare destroying effects of conflict. In Table 7a we see that impasses exist and are not immaterial. However, these "raw" numbers should be compared to the highest possible impasse rates in each domain (shown in Table 7b) to take into account that domains of weaker commitment devices also have lower maximum impasse rates (recall Figures 1 and 2 above). This comparison is done in Table 7c.

While raw impasse rates increase as commitment devices become stronger (higher *P*), the more relevant measure of impasse rates are, as a percent of maximal possible impasse rates, the higher the weaker commitment devices are (lower *P*).

Table 7a		**Predicted Impasse Rate Based on Data***			
		C = 0	C = 1.4	C = 3.3	All C values
Prob	**25%**	4%	1%	1%	2%
	75%	16%	10%	10%	11%
	100%	22%	12%	19%	16%
	Average	14%	8%	10%	10%

Table 7b		**Highest Possible Impasse Rate**			
		C = 0	C = 1.4	C = 3.3	All C values
Prob	**25%**	6%	6%	6%	6%
	75%	56%	56%	56%	56%
	100%	100%	100%	100%	100%

Table 7c		**Impasse Rate as % of Highest Possible rate***			
		C = 0	C = 1.4	C = 3.3	All C values
Prob	**25%**	57%	20%	9%	32%
	75%	29%	17%	17%	20%
	100%	22%	12%	19%	16%

* Calculated given the different strategies' respective frequency in the data. [17]

While we above have discussed the predictive power of Crawford's and Ellingsen & Miettinen's theories, we here map the experimental results against their ancillary predictions on impasse rates.

[17] The impasse rates are predictions of the *ex post* likelihood of an impasse, given, for each domain, the proportions of the respective main strategies that are used and the likelihoods that a commitment would succeed and be binding. Impasses can occur when one person makes an Aggressive Demand and the other makes a Non-Aggressive Demand (most common scenario here is 100% + 50% > 100% and hence impasse), or when two Aggressive Demands have been made (most common scenario here is 100% + 100% > 100%).

Table 8a		Crawford predicts (for C = 0)	E&M predicts (for C = 0 & 1.4)	Impasse rate based on data* C = 0	C = 1.4	C = 3.3	All C values
Prob	25%	6%	6%	4%	1%	1%	2%
	75%	0 to 56%	56%	16%	10%	10%	11%
	100%	0 to 100%	0 to 100%	22%	12%	19%	16%

Table 8b		Impass as % of Crawford C = 0	Impass as % of E&M C = 0	C = 1.4
Prob	25%	57%	57%	20%
	75%	n.a.	29%	17%

* Calculated as in Table 7 above, see also the note and footnote that accompany that table.

Table 8a shows what the predictions based on Crawford and E&M imply for absolute impasse rates compared to the level in the data. In Table 8b we show the impasse rates in the experiment, as a percent of the predictions based on Crawford and E&M respectively.

In the domain where E&M have the same prediction as Crawford, the impasse rate is quite high at almost 60% of the prediction. In the domain where E&M have other predictions, the impasse rate is between 15% and 30% of their predictions.

Extrapolating from these results in order to put them in context, we can generate the results in Figure 4 and Figure 5. These patters should of course only be seen as suggestive of what we could expect in general, since they are based on combining the theories with only the few points the experimental data provides predictions for.

Figure 4 – Extrapolation of Experimental Results from *C* = 0 and *P* = 0.25 to the *P* < 0.5 Domain

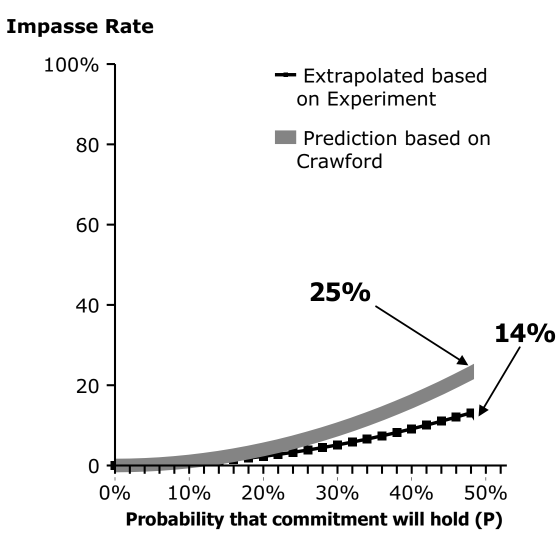
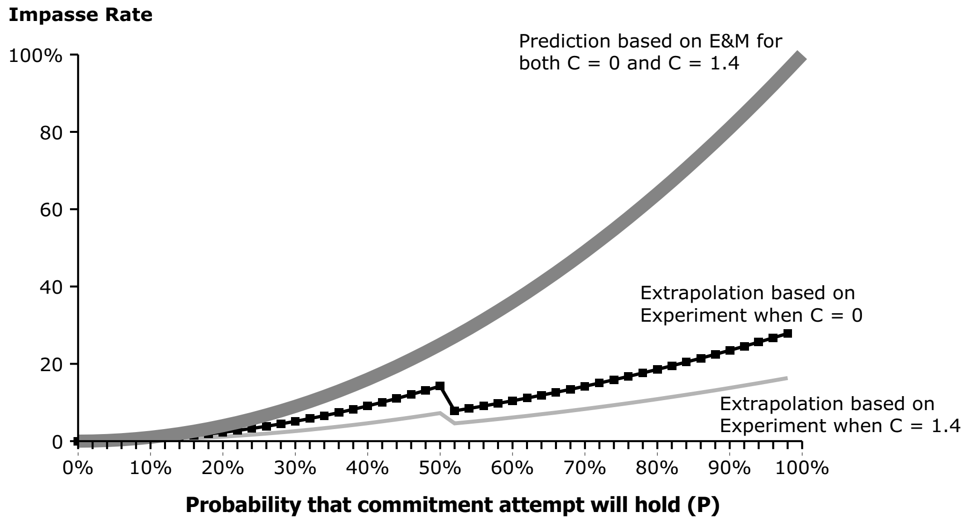


Figure 5 Extrapolation of Experimental Results from when *C* = 0 or 1.4 and *P* = 0.25 or 0.75 to the *P* < 1 Domain



As seen in Tables 7 and 8 above impasses are material across treatments, lending support to:

$\mathbf{P}^{\text{welfare destruction}}$ Aggressive, incompatible commitments lead to significant welfare losses due to frequent conflict and hence impasses

This high level of Pareto-inefficient outcomes in domains where binding contracts can be written goes against what most formal theory predicts should be reached by rational agents under complete information. However, from Table 8 and from Figures 4 and 5 we have also seen that the impasses are not as frequent as the theories of Schelling, Crawford and E&M would predict.

While it appears to be true that conflict (in the sense of attempts to make aggressive demands) is more common when *P* is small, the likelihood of two commitment attempts succeeding at the same time decreases with *P* squared. The net effect is that the absolute level of impasses observed is actually smaller for lower levels of *P* than for higher levels of *P*, lending support for E&M against Schelling's slope and the predictions we based on Crawford.

However, compared to the highest possible level of conflict in each domain, the results provide more support for Crawford than for E&M. Impasse rates are about 60% of the rate based on Crawford and 15-30% of the rates of the main predictions of E&M.

V - Experiment 2: Design and Results

In this second experiment, we test the robustness of the two main results from above, by letting participants play 18 one-shot negotiations, each with the opportunity for feedback and learning.

Design

In this experiment we give participants the opportunity to learn from feedback and conduct many negotiations with the same *C* and *P* parameters. The participants get to negotiate 18 times[18]. Random matching without replacement is used, and participants know that they never will meet a person more than once such that each interaction is, in that sense, a one-shot game. Feedback on the outcome is provided after each negotiation, before new pairs are formed and the next negotiation takes place.

The structure of the game is the same as in Experiment 1.

To induce a sense of high stakes, participants will be awarded the payoff in real dollars for three randomly selected negotiations. The pie at stake in each negotiation will be USD 30. Hence, together with the show-up fee of USD 10, a subject could potentially walk away with as much as USD 100.

We run one treatment to further test the largely opposing predictions, based on the models of Crawford on the one hand, and Ellingsen & Miettinen on the other:

When the cost of commitment is zero and commitment devices are strong (P > 0.5), an efficient and equitable equilibrium exist where both players commit to half

[18] The participants are told that they will negotiate between 15 and 25 times to further avoid any last round effects. This form of random re-matching is in line with Johnson, Camerer, Senc & Rymond (2002).

P[Ellingsen&Miettinen 1a'] *When the cost of commitment is zero and commitment devices are strong ($P > 0.5$), the unique equilibrium involves aggressive demands from both players*

We use parameters $C = \$0$ and $P = 0.75$.

We also run one treatment to further test the strongest prediction of Ellingsen & Miettinen:

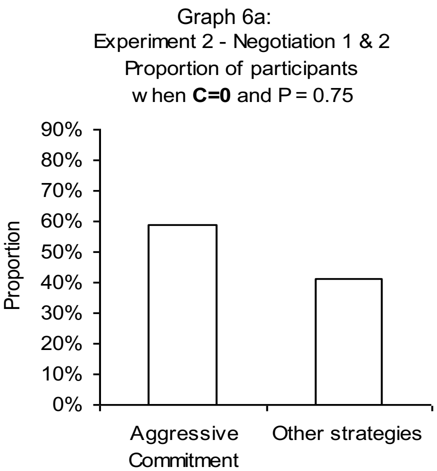
P[Ellingsen&Miettinen 2a] *When the cost of commitment is very low and commitment devices are strong ($P > 0.5$), the unique equilibrium involves aggressive demands from both players*

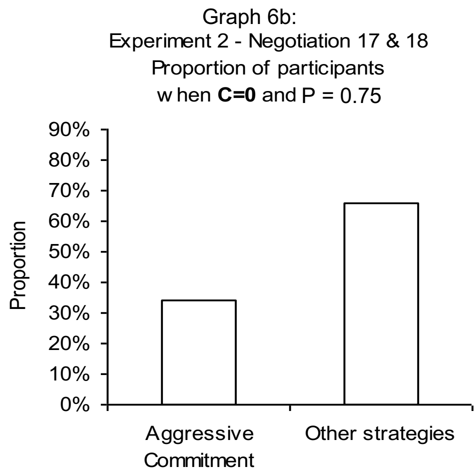
We use parameters $C = \$0.10$ and $P = 0.75$.

While the $C = \$1.4$, that we used for C "very low" in Experiment 1, is far below itsupper bound, which for a pie of \$30 and $P = 0.75$ is $C < \$2.8$, we here opt for an even lower level of C. Since this is a robustness check of the results from Experiment 1 we opt to create the most favorable conditions we can - "C approaching zero" - for testing this E&M prediction. The parameter values were discussed with one of the authors of Ellingsen & Miettinen, and no disagreement on the choice of parameters (or number of rounds) existed.

Results

Treatment 1: *C* = $0 and *P* = 0.75. Enabling learning does not provide more support for the E&M prediction; the share of aggressive demand attempts goes down over time.





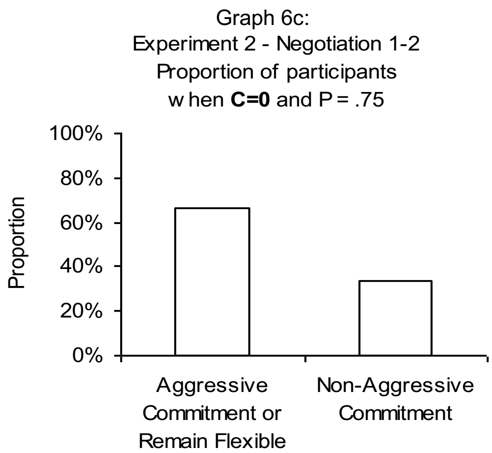
Pooling the first five negotiations, a one-sample proportion test suggests that the point estimate is 55% and that the 95% confidence interval is 46% to 65%. For the last five negotiations, the point estimate is 35% with a 95% confidence interval from 27% to 44%.[19] A two-sample test of proportions

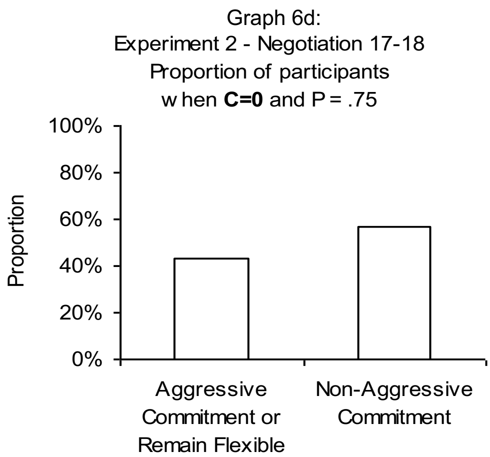
[19] This experiment is conducted on a somewhat different subject pool than Experiment 1 was (half the subject pool from Experiment 1, the second year students, had graduated, and so no subjects from that part of the pool took part in Experiment 2). The change in context and time (end of semester with exams approaching during Experiment 1, versus beginning of semester post-summer return to campus for Experiment 2) may also have lead to different selection effects, regarding who accepted the invitation to participate.

While both experiments are based on one-shot negotiations, playing 5-8 different negotiations with different parameters in Experiment 1 and playing 15-24 negotiations with the same parameters in Experiment 2 might be perceived as different situations by the respective participants. For instance, a participant in Experiment 2 might try out a strategy in the first round because she knows that she will have many more rounds were potential learning from such an early "experiment" can be valuable, whereas a participant in Experiment 1 cannot do this.

If we nevertheless compare the results - since the pie of $30, the expected payoff per round and the parameters are exactly the same as in the equivalent treatment in Experiment 1 - we see that the point estimate of 36% is lower than the level we observe in the first five suggest that the last five rounds are significantly different from the first five rounds at the 1% level[20].

However, learning <u>does</u> appear to provide further support for the prediction we based on Crawford; efficient and equitable equilibria where participants make Non-Aggressive Demands, not only exist, but increase over time.





Pooling the first five negotiations, a one-sample proportion test suggests that the point estimate is 35% and that the 95% confidence interval is 26% to 43%. For the last five negotiations, the point estimate is 54% with a 95% confidence interval from 44% to 63%.[21] A two-sample test of proportions

---

rounds of Experiment 2, but almost identical to the level we observe in the last five rounds of Experiment 2. Note also that the 95% confidence interval of 20% to 52% overlaps with the confidence intervals from both the first and last five rounds of Experiment 2.

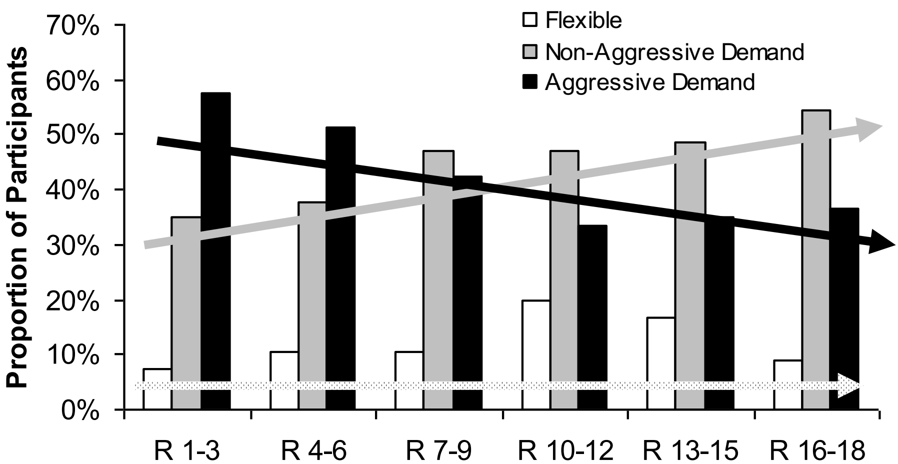
[20] Note that while this test is indicative, it is not strictly correct, since the first and the last five rounds are not independent from each other.

[21] Note that for the many reasons discussed above in Footnote 19 we cannot strictly compare the results for Experiment 2 to those of Experiment 1. If we nevertheless compare the results, we see that the point estimate for Non-Aggressive strategies of 22% is lower than the level we observe in both the first and the last five rounds of Experiment 2. However, the 95% confidence interval of 7% to 36% overlaps with the confidence intervals from the first

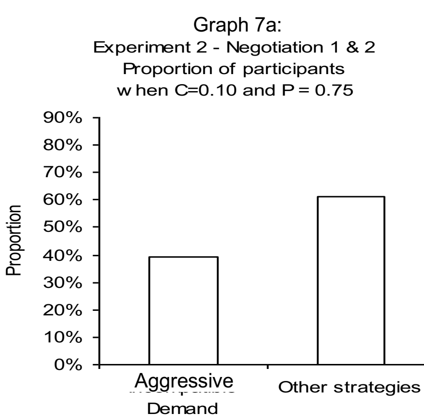
suggests that the last five rounds are significantly different from the first five rounds at the 1% level.[22]

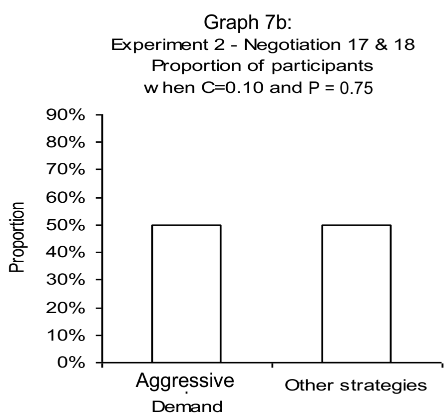
These trends can further been seen in Figure 6:

Figure 6



Treatment 2: C = $0.10 and P = 0.75. Here, the effect of enabling learning is less clear. As can be seen in Graphs 7a and 7b below, the proportion who play aggressive commitment attempts appears to go up over time (although, as we will see below in Figure 7, it starts to go down again towards the end).





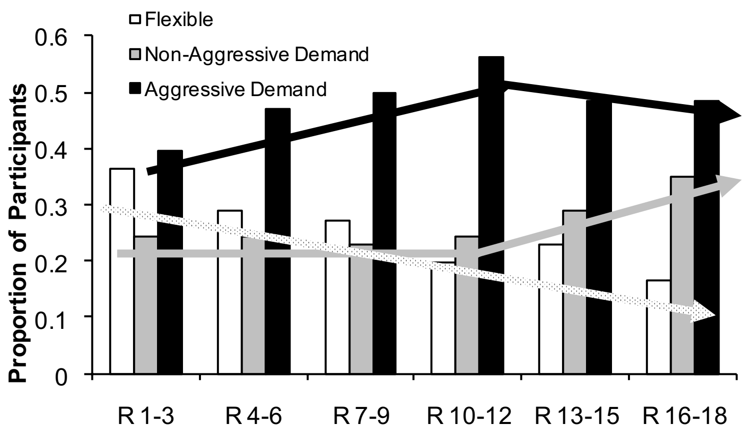
five rounds of Experiment 1. The confidence interval from the last five rounds of Experiment 2 does not overlap with that of Experiment 1.

[22] See Footnote 20.

However, this pattern does not appear to be significant. Pooling the first five negotiations, a one-sample proportion test suggests that the point estimate is 42% and that the 95% confidence interval is 32% to 51%. For the last five negotiations, the point estimate is 48% with a 95% confidence interval from 38% to 58%.[23] A two-sample test of proportions suggests that the first five and the last five rounds might not be significantly different from each other (p-value is 34%).[24]

From Figure 7, we also note that in this treatment, while the point estimates of the proportions of aggressive demands first go up (in rounds 1 to 9), they later appear to go down (in rounds 13 to 18).

Figure 7



---

[23] Note that, for the many reasons discussed above in Footnote 21, we cannot strictly compare the results for Experiment 2 to those of Experiment 1. In this treatment we have also changed how *C* "very low" is parameterized: from C = $1.40 to C = $0.10. If we nevertheless compare the results with regards to this prediction (the case when $P = 0.75$ and *C* is "very low"), we see that the point estimate of the proportion of aggressive strategies of 25% in Experiment 1, is lower than the level we observe in both the first five and the last five rounds of Experiment 2. At least part of this lower result should be expected due to the lower cost of commitment (as seen in Table 3, higher commitment costs always appears to crowd out at least some aggressive commitments). Note also that the 95% confidence interval of 11% to 39% overlaps with the confidence intervals from both the first five and the last five rounds of Experiment 2.

[24] See Footnote 20.

In closing, we would like to comment on the following two part observation:

1. The only difference between the two treatments in Experiment 2, is whether C = \$0 or C = \$0.10, which is a tiny difference in relation to a pie of \$30. All other parameters (structure, strength of commitment device, number of rounds, etc) are exactly the same.[25]

2. The outcomes of the two treatments are nevertheless quite different. [26] The direction of the differences are in line with:
    - the prediction based on Crawford being most valid when $C = 0$, and
    - the prediction based on Ellingsen & Miettinen being more relevant (less non-relevant) when $C = 0.10$ than when $C = 0$, which is inline with the $C > 0$ prediction relying on the weaker solution concept of iterated elimination of strictly dominated strategies, than the $C = 0$ prediction that relies on the stronger concept of iterated elimination of weakly dominated strategies.

---

[25] The participants were also randomized to the two treatments from the same underlying subject pool.

[26] Comparing the last five rounds of each treatment, using two-sample tests of proportions, we see the following:

- The higher share of 54% that opt for Non-Aggressive Demands when $C = 0$, versus the share of 35% that make this choice when $C = 0.10$, is significantly different at the 1% level (using a two-sided test).
- The lower share of 35% that opt for Aggressive Demands when $C = 0$, versus the share of 48% that make this choice when $C = 0.10$, is statistically different at the 10% level with a p-value of 5.6% (again, using a two-sided test).
- There is no statistically significant difference between the 11 and 17% that opt to Remain Flexible (the p-value is 18%).

VI – Summary and Conclusions

Conflict and impasses in negotiation, whether they manifest as strikes, job resignations, or trade embargoes, are Pareto-inefficient and create tremendous loss of social welfare. Hence, a further understanding of why impasses occur is important. In this context we provide evidence that supports a reason for conflict and impasses proposed by Schelling (1956, 1960 and 1966) and formalized by Crawford (1982) and Ellingsen & Miettinen (2008) that so far only has received little attention in the literature. The foundation of these theories is that people can rationally make aggressive demands in order to increase their share of the pie, even though they know that such commitments may lead to impasse.

We find strong support for the prediction based on Crawford (1982) that attempting aggressive demands is the dominant strategy in domains where commitment devices are relatively weak, while efficient, fair and non-aggressive demand combinations are more common when commitment devices are strong.

Support for the main predictions based on Ellingsen & Miettinen (2008) – that aggressive commitments dominate when commitments devices are strong – is weaker, but the theory might be capturing the behavior of a sub-set of the participants: When $C = 0$, we find that 36% in Experiment 1 and an average of 35% in the last five negotiations of Experiment 2 play aggressive demands. When $C > 0$ the corresponding numbers are 25% in Experiment 1 (where $C = 1.40$), and an average of 48% in the last five negotiations of Experiment 2 (where $C = 0.10$).

We find strong support for the secondary predications based on Ellingsen & Miettinen, that mixes of either remaining fully flexible or making aggressive demands dominate when commitment is certain, or when the cost of attempting commitment is low, but not very low.

Combined, the results suggest that *ex ante conflict* (defined as negotiators making demands that could lead to impasse) is more likely to occur where commitment devices are weak. In these domains of weak commitment devices, we also observe frequent outcomes in which one party takes most of the pie and the other receives little. Although the logic is different, uneven outcomes are also common in domains of strong commitment devices. These results run counter to the typical predictions in most formal theory that more symmetric "compromise outcomes" will be realized, and can perhaps, as discussed by Ellingsen & Miettinen (2008), help explain the infrequency of side payments in real life agreements.

In areas of weak commitment devices, where aggressive demands dominate, the uneven outcomes are driven by the high likelihood that only one of the negotiators' commitment attempts will succeed (the predictions by Crawford, and Schelling's Slope, are the drivers here).

In domains where commitment devices are either certain or the cost of attempting commitment is low, but not very low, the uneven outcomes are driven by the predictions based on Ellingsen & Miettinen that equilibria will contain combinations of the two strategies "Aggressive Demand" AND "Remain Flexible".

Welfare destruction (of sometimes more than 20% of the pie) is much higher than the Pareto efficient 0% welfare destruction that most formal theory predicts when binding contracts can be entered. However, the welfare destruction is not as high as Crawford predictions would imply, and far from as high as E&M predictions would imply. The welfare destruction is 60% and 15-30% of the rates predicted by Crawford E&M respectively.

Our study does not investigate the role of fairness concerns, which could prove key to further understanding. From the exit survey in both experiments, it appears as though both equity concerns and preferences for reciprocity are playing a role in decision-making, for at least some portion of the subject population. Further investigation of these and other fairness concerns, and their interaction with the "rational" drivers of conflict suggested by Schelling, Crawford, and Ellingsen & Miettinen could offer important insights.

We conjecture that a model that incorporates fairness concerns into the theory of Ellingsen & Miettinen could predict the experimental data quite well[27]. Similarly, our attempt in Experiment 2 to create an environment where subjects take part in many negotiations with feedback and have ability for learning is only one first step towards gaining insight into the ways that bounded rationality might interact with these theories. One interesting avenue for further research could be to see how the theories could be combined with, for instance, cognitive hierarchy models such as that of Camerer, Ho & Chong (2004).

While we agree with Crawford when he starts his 1982 paper by quoting Sir Arthur Stanley Eddington:

> "It is ... a good rule not to put overmuch confidence in the observational results that are put forward until they are confirmed by theory."

We also believe that the opposite analog - that theory needs to be empirically tested to enable an understanding of the extent to which the theory actually explains human behavior - is equally true.

[27] One approach could be to integrate the Fehr-Schmidt (1999) model with the theory of Ellingsen & Miettinen.

Appendix – Instructions for Experiment 2
(in all material respects identical to Experiment 1)

A two stage, two part negotiation on how to divide 30 dollars

You will soon randomly be paired with one of the participants in the other room with whom you will have a negotiation over how to split a payoff of 30 dollars among the two of you.

The $30 pie is in addition to the $10 show-up fee for attending the experiment.

Please read these instructions carefully. Before the negotiation begins you will answer some control questions to verify that you understand the negotiation structure.

There will be two variables: a cost $*C* in dollars and a probability *P* in %. In the instructions they are called $*C* and *P*. In the negotiation you will get exact values for these variables.

The negotiation will proceed as follows (more detail on each stage will follow)**:**

Phase 1 - Pre-negotiation decision

- You decide whether to remain flexible or to attempt to make a binding commitment to a certain share of the $30
- Your counterpart will have the same decision to make

Phase 2a - Chance decides whether <u>any potential</u> commitments fail or succeed

- If any commitment attempts have been made, the chance mechanism decides whether each commitment attempt succeeds. If commitment attempts are made by both negotiation participants, the success of each will be decided independently.
- Information about whether each party remains flexible or has made a binding commitment is provided

Phase 2b - The Negotiation/Realization of results

- The outcome of each negotiation is decided by a combination of your decision and your counterpart's decision (and potentially also the chance mechanism, when any commitments attempts are made):
  - Compatible claims (summing to less or equal to $30) are paid
  - Incompatible claims lead to a "bust negotiation" and no one receiving any share of $30
  - A person who attempted to commit pays $*C* (whether or not the commitment succeeds)

Phase 1 – Pre-negotiation decision

It is in this phase you make your decision (as does the person you negotiate with). The decision will then have an impact on the outcome in Phase 2.

Decision on whether to attempt to commit to a demand or to remain flexible
Participants will separately have the possibility to attempt to commit to a certain demand. If you choose to attempt to commit, you will specify the size of your demand as 1 to 100% of $30. You can also choose to stay uncommitted and remain flexible.

Any attempt to commit is successful and binding with probability, *P*
With probability *P*, a commitment attempt succeeds and the demand is binding in the next stage. With probability 1-*P*, a commitment attempt fails and the person remains flexible.

These probabilities are the same for both participants, but each participant's probability is *independent* from the other person's.
For example, if both participants attempt to commit, the computer will "roll one dice" for you and separately "roll one dice" for the other participant. There will be four possible outcomes: either only you, only the other participant, both of you, or neither of you will end up having a successful commitment.

Attempting to commit costs $*C* whether or not the commitment succeeds
Attempting to commit to a demand costs $*C*, which you pay whether or not the attempt is successful. In the case you attempt to commit, but do not receive any share of the $30 in the negotiation, the cost will be subtracted from your show-up fee.

If you do not attempt to commit, you will not pay any cost, $*C*.

<u>Phase 2 – Decisions by the Chance Mechanism and Realization of the Negotiation</u>

If any commitments are made – the chance mechanism will decide the outcomes of those commitments in this phase. Then, also using the decisions in Phase 1, an outcome is reached.

Information flow
Each participant who has attempted to commit will find out if their commitment attempt has failed or succeeded.

Each participant will also learn of any successful commitment attempt made by the other participant. (However, a person cannot distinguish whether a person who is flexible is so because he or she made an unsuccessful commitment attempt or chose to remain flexible).

The three scenarios that can occur in Phase 2
Given decisions by each player (and also the chance mechanism if one or more commitment attempts are made), the negotiation can be characterized by *two, one or zero commitment attempts that succeed*. The payoffs in each of these scenarios will be calculated as follows:

i) Two commitment attempts succeed. If both you and the other participant make commitment attempts that succeed, there are three possible outcomes:

- if the sum of your committed demands is more than 100% of $30, none of you will receive a share of $30
- if the sum of your committed demands is equal to 100% of $30, you will each be given your committed demand
- if the sum of you committed demands is less than 100% of $30, you will each be given your demand and will share the remaining part of $30 in a equal split

In all the three sub-cases, both persons will pay the cost of attempting to commit, $*C*.

ii) One commitment attempt succeeds. If one participant has made a commitment attempt that succeeded and the other has not (either because he/she made a commitment attempt that failed or because he/she did not attempt commitment at all), the participant who successfully committed will receive his/her demanded share and the other person will receive $30 minus the first person's demanded share.

Any person who attempted to commit will bear the cost, \$*C*. A person who did not attempt to commit will not bare any cost.

iii) No commitment attempts are made or succeed. In the case that both players are flexible (because they either did not attempt to commit or made commitment attempts that failed), the computer will approximate the symmetric nature of such a negotiation by giving you half of the $30 each.

A person who attempted to commit will bear the cost, \$*C*. A person who did not attempt to commit will not bare any cost.

Concluding remark
As can be seen above: while the chance mechanism "acts" and the outcome of the negotiation is decided in Phase 2, you and the person you negotiate with make your decisions in Phase 1.

References

Asheim, Geir B. & Andrés Perea. 2009a. "Algorithms for cautious reasoning in games." Working Paper.

Asheim, Geir B. & Andrés Perea. 2009b. "Behavior in commitment bargaining." Unpublished Comment on Ellingsen & Miettinen (2008) paper, submitted to the *American Economic Review*.

Camerer, Colin F., Teck-Hua Ho & Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics*, 119(3): 861-898.

Crawford, Vincent. 1982. "A Theory of Disagreement in Bargaining." *Econometrica*, 50(3): 607-637.

Ellingsen, Tore. 1997. "The Evolution of Bargaining Behavior." *Quarterly Journal of Economics*, 112(2): 581–602.

Ellingsen, Tore & Topi Miettinen. 2008. "Commitment and Conflict in Bilateral Bargaining." *American Economic Review*, 98(4): 1629–35.

Falk, Armin, Ernst Fehr & Fischbacher, Urs. 2003. "Reasons for Conflict: Lessons from Bargaining Experiments."*Journal of Institutional and Theoretical Economics*, 159: 171–187.

Fehr, Ernst & Klaus M. Schmidt. 1999. "A Theory Of Fairness, Competition, And Cooperation." *Quarterly Journal of Economics*, 114(3): 817-868.

Johnson, Eric, Colin Camerer & Sankar Senc & Talia Rymond. 2002. "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining."*Journal of Economic Theory*, 104: 16–47.

Harsanyi, John & Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games.* Cambridge, MA: MIT Press.

Nash, John. 1953. "Two-Person Cooperative Games." *Econometrica*, 21(1): 128–40.

Schelling, Thomas C. 1956. "An Essay on Bargaining." *American Economic Review*, 46(3): 281-306.

Schelling, Thomas C. 1960. *The Strategy of Conflict.* Cambridge, MA: Harvard University Press.

Schelling, Thomas C. 1966. *Arms and Influence.* New Haven, CT: Yale University Press.

EFI, The Economic Research Institute

Published in the language indicated by the title.
A complete publication list can be found at *www.hhs.se/efi*
Books and dissertations can be ordered from EFI via e-mail:
efi.publications@hhs.se

Reports since 2006

2009

Books

Engwall, Lars. *Mercury meets Minerva: business studies and higher education: the Swedish case.*
Hagberg, Axel. *Bankkrishantering.* Forskning i Fickformat.
Henriksson, Lars. *Marknad eller reglering?: vägval för europeisk telekommunikation.* Forskning i Fickformat.
Holmberg, Carina and Filip Wijkström (ed). *Kunskapsbyggaren: meningsfulla möten och kunskap utan gränser: en vänbok till Åke Danielsson.*
Schriber, Svante. *Att realisera synergier: ledning av värdeskapande vid företagsköp.* Forskning i Fickformat.
Östman, Lars. *Towards a general theory of financial control for organisations.*

Dissertations

Almenberg, Johan. *Difficult choices: essays on economic behavior.*
Amado, Cristina. *Four essays on the econometric modelling of volatility and durations.*
Arbin, Katarina. *Individual information system acceptance behavior: an electronic ordering system case.*
Brettell Grip, Anna-Karin. *Funding and accountability: studies of a Swedish and a British chamber orchestra.*
Broback, Anna. *Den värdefulla nöjdheten?: en studie om kundnöjdhet och upplevt värde med kläder över tid.*
Dreber Almenberg, Anna. *Determinants of economic preferences.*
Hasseltoft, Henrik. *Essays on the term structure of interest rates and long-run risks.*
Hellström, Katerina. *Financial accounting quality in a European transition economy: the case of the Czech republic.*
Hernant, Mikael. *Profitability performance of supermarkets: the effects of scale of operation, local market conditions, and conduct on the economic performance of supermarkets.*
Jamal, Mayeda. *Creation of social exclusion in policy and practice.*
Jansson, Thomas. *Essays on Household Portfolio Choice.*
Lakomaa, Erik. *The economic psychology of the welfare state.*
Lazareva, Olga. *Labor market outcomes during the Russian transition.*
Lee, Samuel. *Information and control in financial markets.*
Lid Andersson, Lena. *Ledarskapande retorik: Dag Hammarskjöld och FN:s övriga generalsekreterare som scen för karisma, dygder och ledarideal.*

Lindqvist, Göran. *Disentangling clusters: agglomeration and proximity effects.*
Melander, Ola. *Empirical essays on macro-financial linkages.*
Melén, Sara. *New insights on the internationalisation process of SMEs: a study of foreign market knowledge development.*
Murgoci, Agatha. *Essays in mathematical finance.*
Salomonsson, Marcus. *Essays in applied game theory.*
Sjöström, Emma. *Shareholder influence on corporate social responsibility.*
Törn, Fredrik. *Challenging consistency: effects of brand-incongruent communications.*
Wennberg, Karl. *Entrepreneurial exit.*
Wetter, Erik. *Patterns of performance in new firms: estimating the effects of absorptive capacity.*
Zubrickas, Robertas. *Essays on contracts and social preferences.*
Åge, Lars-Johan. *Business manoeuvring: a grounded theory of complex selling processes.*

2008

Books

Breman, Anna. *Forskning om filantropi. Varför skänker vi bort pengar?* Forskning i Fickformat.
Einarsson, Torbjörn. *Medlemskapet i den svenska idrottsrörelsen: En studie av medlemmar i fyra idrottsföreningar.* EFI Civil Society Reports.
Helgesson, Claes-Fredrik and Hans Winberg (eds). *Detta borde vårddebatten handla om.*
Jennergren, Peter, Johnny Lind, Walter Schuster and Kenth Skogsvik (eds). *Redovisning i fokus.* EFI:s Årsbok 2008. EFI/Studentlitteratur.
Kraus, Kalle. *Sven eller pengarna? Styrningsdilemman i äldrevården.* Forskning i Fickformat.
Petrelius Karlberg, Pernilla. *Vd under press: om medialiseringen av näringslivets ledare.* Forskning i Fickformat.
Portnoff, Linda. *Musikbranschens styrningsproblematik.* Forskning i Fickformat.
Sjöstrand, Sven-Erik. *Management: från kontorsteknik till lednings- och organisationsteori: utvecklingen på Handelshögskolan under 100 år: 1909–2009.*
Östman, Lars. *Den finansiella styrningens realiteter och fiktioner: de finansiella styrformernas svenska historia, berättelser om Petersson och "Ericsson", finansiell styrning – en ansats till generell teori.*
Östman, Lars. *Mycket hände på vägen från Buchhaltung till Accounting: delar av Handelshögskolan under 100 år.*

Dissertations

Axelson, Mattias. *Enabling knowledge communication between companies: the role of integration mechanisms in product development collaborations.*
Benson, Ilinca. *Organisering av övergångar på arbetsmarknaden: en studie av omställningsprogram.*
Elhouar, Mikael. *Essays on interest rate theory.*
Farooqi Lind, Raana. *On capital structure and debt placement in Swedish companies.*
Granström, Ola. *Aid, drugs, and informality: essays in empirical economics.*
Hvenmark, Johan. *Reconsidering membership: a study of individual members' formal affiliation with democratically governed federations.*

Höglin, Erik. *Inequality in the labor market: insurance, unions, and discrimination.*
Johansson, Marjana. *Engaging resources for cultural events: a performative view.*
Kallenberg, Kristian. *Business at risk. Four studies on operational risk management.*
Kviselius, Niklas Z. *Trust-building and communication in SME internationalization: a study of Swedish-Japanese business relations.*
Landberg, Anders. *New venture creation: resistance, coping and energy.*
Pemer, Frida. *Framgång eller fiasko? En studie av hur konsultprojekt värderas i klientorganisationer.*
Rosengren, Sara. *Facing clutter: on message competition in marketing communication.*
Schilling, Annika. *Kan konsulter fusionera?: en studie av betydelsen av identitet vid en fusion mellan konsultföretag.*
Schriber, Svante. *Ledning av synergirealisering i fusioner och förvärv.*
Sjödin, Henrik. *Tensions of extensions: adverse effects of brand extension within consumer relationship.*
Strandqvist, Kristoffer. *Kritiska år: formativa moment för den svenska flygplansindustrin 1944–1951.*
Strömqvist, Maria. *Hedge funds and international capital flow.*
Söderström, Johan. *Empirical studies in market efficiency.*
Sölvell, Ingela. *Formalization in high-technology ventures.*
Thorsell, Håkan. *The pricing of corporate bonds and determinants of financial structure.*
Ulbrich, Frank. *The adoption of IT-enabled management ideas: insights from shared services in government agencies.*
Östling, Robert. *Bounded rationality and endogenous preferences.*

2007

Books

Andersson, Per, Ulf Essler and Bertil Thorngren (eds). *Beyond mobility.* EFI Yearbook 2007. EFI/Studentlitteratur.
Einarsson, Torbjörn and Filip Wijkström. *Analysmodell för sektorsöverskridande statistik: fallet vård och omsorg.* EFI Civil Society Reports.
Ericsson, Daniel. *Musikmysteriet: organiserade stämningar och motstämningar.*
Samuelson, Lennart (ed). *Bönder och bolsjeviker: den ryska landsbygdens historia 1902–1939.*

Dissertations

Ahlersten, Krister. *Empirical asset pricing and investment strategies.*
Alexius, Susanna. R*egelmotståndarna: om konsten att undkomma regler.*
Andersson, Magnus. *Essays in empirical finance.*
Berg, Bengt Åke. *Volatility, integration and grain bank: studies in harvests, rye prices and institutional development of the parish magasins in Sweden in the 18th and 19th centuries.*
Bianchi, Milo. *Of speculators, migrants and entrepreneurs: essays on the economics of trying your fortune.*
Brodin, Karolina. *Consuming the commercial break: an ethnographic study of the potential audiences for television advertising.*
Elger, Max. *Three essays on investment-specific technical change.*

Hagberg, Axel. *Bankkrishantering: aktörer, marknad och stat.*
Hinnerich, Mia. *Derivatives pricing and term structure modeling.*
Hjalmarson, Hanna. *En växande marknad: studie av nöjdheten med konsumtionsrelaterade livsområden bland unga konsumenter.*
Hjelström, Tomas. *The closed-end investment company premium puzzle: model development and empirical tests on Swedish and British data.*
Kraus, Kalle. *Sven, inter-organisational relationships and control: a case study of domestic care of the elderly.*
Lindqvist, Erik. *Essays on privatization, identity, and political polarization.*
Macquet, Monica. *Partnerskap för hållbar utveckling: systrar av Oikos och guvernanten som blev diplomat.*
Melian, Catharina. *Progressive open source.*
Nilsson, Daniel. *Transactions in cyberspace: the continued use of Internet banking.*
Petrelius Karlberg, Pernilla. *Den medialiserade direktören.*
Portnoff, Linda. *Control, cultural production and consumption: theoretical perspectives, empirical dilemmas, and Swedish music industry practices.*
Sköld, Martin. *Synergirealisering: realisering av produktsynergier efter företagssammanslagningar.*
Sonnerby, Per. *Contract-theoretic analyses of consultants and trade unions.*
Tyrefors, Björn. *Institutions, policy and quasi-experimental evidence.*
Valiente, Pablo. *Re-innovating the existing: a study of wireless IS capabilities to support mobile workforces.*

2006

Books

Lundeberg, Mats, Pär Mårtensson and Magnus Mähring (eds) *IT & business performance: a dynamic relationship.* EFI Yearbook 2006. EFI / Studentlitteratur.
Thodenius, Björn. *Organisering av kunskap: en studie av Wallenberg Consortium North.* EFI Civil Society Reports.
Wijkström, Filip and Torbjörn Einarsson. *Från nationalstat till näringsliv?: det civila samhällets organisationsliv i förändring.*
Wijkström, Filip, Stefan Einarsson and Ola Larsson. *Staten och det civila samhället: idétraditioner och tankemodeller i den statliga bidragsgivningen till ideella organisationer.*
Östman, Lars. *Lysande ögonblick och finansiella kriser: Dramaten under ett sekel.*

Dissertations

Argenton, Cedric. *Quality provision in duopoly.*
Beckerman, Carina. *The clinical eye: constructiong and computerizing an anesthesia patient record.*
Borglund, Tommy. *Aktievärden i fokus: internationell påverkan på intressentrelationer vid förvärv och fusion.*
Breman, Anna. *The Economics of altruism, paternalism and self-control.*
Edquist, Harald. *Technological breakthroughs and productivity growth.*
Eklund, Jana. *Essays on forecasting and Bayesian model averaging.*

Frostenson, Magnus. *Legitimitetskontrollen: en studie av etiska värderingars roll i gränsöverskridande förvärv och fusioner.*

Gaspar, Raquel M. *Credit risk and forward price models.*

Gustafsson, Peter. *Essays on trade and technological change.*

Hopkins, Elisabeth. *Is a higher degree of local currency pricing associated with lower exchange rate pass-through?: a study of import pricing in 51 Swedish industries*.

Kling, Ragnar. *Developing product development in times of brutal change.*

Langenskiöld, Sophie. *Peer influence on smoking: causation or correlation?*

Lychnell, Lars-Olof. *"Och fungerar det inte, gör vi på något annat sätt": en klinisk fallstudie av IT-relaterat förändringsarbete i småföretag*

Meitz, Mika. *Five contributions to econometric theory and the econometrics of ultra-high-frequency data.*

Mendicino, Caterina. *Financial market imperfections, business cycle fluctuations and economic growth.*

Ovanfors, Anna. *Essays on nonlinear time series analysis and health economics.*

Paltseva, Elena. *Essays on commitment and inefficiency in political economy.*

Rogberg, Martin. *Den modeföljande organisationen: om acceptansen av TQM och andra populära managementmodeller.*

Silvennoinen, Annastiina. *Essays on autoregressive conditional heteroskedasticity.*

Sjögren, Ebba. *Reasonable drugs: making decisions with ambiguous knowledge.*

Slinko, Irina. *Essays in option pricing and interest rate models.*

Wilander, Fredrik. *Essays on exchange rates and prices.*