Essays on bargaining and strategic communication

de Groot Ruiz, A.W.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Strategic Communication
This thesis develops and tests game theoretic models of communication and bargaining. These models can be used to analyze strategic aspects of negotiations, such as those that arise between managers and workers, divorcing spouses or political parties. While the foundations may be somewhat technical, the essence can often be captured by a simple example. If you want to experience strategic communication, try your hand at the setting below during the public exposé of this thesis. (You can contact the author at adrian@degrootruiz.nl for questions.)

Do it yourself
You play Thomas, a mathematician from New York, who visits the village of Tamamdrés, along the beautiful coast of Oaxaca (Mexico). Tamamdrés is surrounded by five beaches, all a kilometer apart from each other. Fallen in love with the place, you have decided to open a surf resort at one of the beaches. You want to maximize your resort’s earnings, but face the problem that you do not know where the surf is right. You have heard that there is one ‘surf beach,’ which provides a good surf, and that the other four beaches offer less reliable surf conditions. You do not know which one is the surf beach, but you believe it must be either the most western beach (a. Revolcadero) or the most eastern one (e. Positano). Both are equally likely to be the surf beach in your opinion. The closer your resort is to the surf beach, the better it is for your business. Your resort cannot be more than 2 km away from the surf beach, because otherwise not enough customers will show up.

So you decide to talk things over with the Mayor of the village, Don Miguel. He also happens to be the local surf expert and could tell you without a shadow of a doubt where the surf beach is. What is more, you need his permission to open your surf resort. However, you wonder how helpful Don Miguel will really be. Don Miguel wants to maximize the earnings of his village and faces the following dilemma. A surf resort at the surf beach will cost the village dearly, as tourists will stay away from the village center due to the distance, which is bad for local business. On the other hand, a successful resort closer to the village center would bring in loads of tourists, meaning booming business for his village. If Don Miguel does not give permission, nobody earns or loses money.

You can play this situation as follows with the person sitting next to you, who will play the role of Don Miguel. Both you and Don Miguel base the choices on the (same) table below.

1. You ask Don Miguel which beach is the surf beach.
2. Don Miguel gives you an answer (but beware, Don Miguel might not be completely honest...).
3. You choose at which beach you want to open your resort and you tell Don Miguel your choice.
4. Don Miguel decides whether or not to give you permission.
Location of surf beach

- a. Revolcadero
- b. Ragusa
- c. Montaña Blanca
- d. Venecia
- e. Positano

Expected annual earnings for resort and village (in millions of pesos) depending on Thomas’ choice and Don Miguel’s information

<table>
<thead>
<tr>
<th>Location</th>
<th>Resort at a. Revolcadero</th>
<th>Resort at b. Ragusa</th>
<th>Resort at c. Montaña Blanca</th>
<th>Resort at d. Venecia</th>
<th>Resort at e. Positano</th>
<th>Permission</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Revolcadero</td>
<td>$10</td>
<td>$7</td>
<td>$10</td>
<td>$2</td>
<td>$0</td>
<td>$10</td>
</tr>
<tr>
<td>b. Ragusa</td>
<td>$0</td>
<td>$10</td>
<td>$40</td>
<td>$4</td>
<td>$0</td>
<td>$10</td>
</tr>
<tr>
<td>c. Montaña Blanca</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
<tr>
<td>d. Venecia</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
<tr>
<td>e. Positano</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
</tbody>
</table>

Location of surf resort to be chosen by Thomas

- a. Revolcadero
- b. Ragusa
- c. Montaña Blanca
- d. Venecia
- e. Positano

Expected annual earnings for resort and village (in millions of pesos) depending on Thomas’ choice and Don Miguel’s information

<table>
<thead>
<tr>
<th>Location</th>
<th>Resort at a. Revolcadero</th>
<th>Resort at b. Ragusa</th>
<th>Resort at c. Montaña Blanca</th>
<th>Resort at d. Venecia</th>
<th>Resort at e. Positano</th>
<th>Permission</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Revolcadero</td>
<td>$10</td>
<td>$7</td>
<td>$10</td>
<td>$2</td>
<td>$0</td>
<td>$10</td>
</tr>
<tr>
<td>b. Ragusa</td>
<td>$0</td>
<td>$10</td>
<td>$40</td>
<td>$4</td>
<td>$0</td>
<td>$10</td>
</tr>
<tr>
<td>c. Montaña Blanca</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
<tr>
<td>d. Venecia</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
<tr>
<td>e. Positano</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
</tr>
</tbody>
</table>

This thesis develops and tests game theoretic models of communication and bargaining. These models can be used to analyze strategic aspects of negotiations, such as those that arise between managers and workers, divorcing spouses or political parties. While the foundations may be somewhat technical, the essence can often be captured by a simple example. If you want to experience strategic communication, try your hand at the setting below during the public exposition of this thesis. (You can contact the author at adrian@degrootruiz.nl for questions.)
Strategic Communication
This thesis develops and tests game theoretic models of communication and bargaining. These models can be used to analyze strategic aspects of negotiations, such as those that arise between managers and workers, divorcing spouses or political parties. While the foundations may be somewhat technical, the essence can often be captured by a simple example. If you want to experience strategic communication, try your hand at the setting below during the public exposure of this thesis. (You can contact the author at adrian@degrootruiz.nl for questions.)

Do it yourself
You play Thomas, a mathematician from New York, who visits the village of Tamamdrés, along the beautiful coast of Oaxaca (México). Tamamdrés is surrounded by five beaches, all a kilometer apart from each other. Fallen in love with the place, you have decided to open a surf resort at one of the beaches. You already have a construction permit from the state authorities.

You want to maximize your resort’s earnings, but face the problem that you do not know where the surf is right. You have heard that there is one ‘surf beach,’ which provides a good surf, and that the other four beaches offer less reliable surf conditions. You do not know which one is the surf beach, but you believe it must be either the most western beach (a. Revolcadero) or the most eastern one (e. Positano). Both are equally likely to be the surf beach in your opinion. The closer your resort is to the surf beach, the better it is for your business. Your resort cannot be more than 2 km away from the surf beach, because otherwise not enough customers will show up.

So you decide to talk things over with the Mayor of the village, Don Miguel. He also happens to be the local surf expert and could tell you without a shadow of a doubt where the surf beach is. However, you wonder how helpful Don Miguel will really be. Don Miguel wants to maximize the earnings of his village and faces the following dilemma. A surf resort at the surf beach will cost the village dearly, as tourists will stay away from the village center due to the distance, which is bad for local business. On the other hand, a successful resort closer to the village center would bring in loads of tourists, meaning booming business for his village.

You can play this situation as follows with the person sitting next to you, who will play the role of Don Miguel. Both you and Don Miguel base the choices on the (same) table below.

1. You ask Don Miguel which beach is the surf beach.
2. Don Miguel gives you an answer (but beware, Don Miguel might not be completely honest...).
3. You choose at which beach you want to open your resort.

<table>
<thead>
<tr>
<th>Expected annual earnings for resort and village (in millions of pesos $) depending on Thomas’s choice and Don Miguel’s information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location of surf beach known by Don Miguel</td>
</tr>
<tr>
<td>a. Revolcadero</td>
</tr>
<tr>
<td>e. Positano</td>
</tr>
</tbody>
</table>
Strategic Communication
This thesis develops and tests game theoretic models of communication and bargaining. These models can be used to analyze strategic aspects of negotiations, such as those that arise between managers and workers, divorcing spouses or political parties. While the foundations may be somewhat technical, the essence can often be captured by a simple example. If you want to experience strategic communication, try your hand at the setting below during the public exposé of this thesis. (You can contact the author at adrian@degrootruiz.nl for questions.)

Do it yourself
You play Don Miguel, the Mayor of the village of Tamamdrés and local surf expert. Tamamdrés lies at the beautiful coast of Oaxaca (México) and is surrounded by five beaches, all a kilometer apart from each other. You have heard that a guy called Thomas, a visiting mathematician from New York, has fallen in love with the place and wants to open a surf resort at one of the beaches. He already has a construction permit from the state authorities.

This Thomas wants to maximize his resort’s earnings and faces the following problem. He has heard that there is one ‘surf beach,’ which provides a good surf, and that the other four beaches offer less reliable surf conditions. He does not know which one is the surf beach, but he believes it must be either the most western beach (a. Revolcadero) or the most eastern one (e. Positano). Both are equally likely to be the surf beach in his opinion. The closer the resort is to the surf beach, the better it is for business. The resort cannot be more than 2 km away from the surf beach, because otherwise not enough customers will show up. Hence, Thomas has made an appointment with you to find out the best location for his resort.

You, however, want to maximize the earnings of your village and face the following dilemma. You know the surf beach is . However, a surf resort at that place will cost the village dearly, as tourists will stay away from the village center due to the distance and force local businesses to close. On the other hand, a successful resort closer to the village center would bring in loads of tourists, meaning booming business for your village.

You can play this situation as follows with the person sitting next to you, who will play the role of Thomas. Both you and Thomas base the choices on the (same) table below.

1. Thomas asks you which beach is the surf beach.
2. You give one of the following answers: (i) “it’s a. Revolcadero”, (ii) “It’s e. Positano” or (iii) “I’m not going to tell you” (you do not need to be honest).
3. Thomas chooses at which beach he wants to open the resort.

<table>
<thead>
<tr>
<th>Location of surf beach known by Don Miguel</th>
<th>Location of surf resort to be chosen by Thomas</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>-10</td>
</tr>
<tr>
<td>e. Positano</td>
<td>0</td>
</tr>
<tr>
<td>village earnings</td>
<td>resort earnings</td>
</tr>
<tr>
<td>$0</td>
<td>$7</td>
</tr>
<tr>
<td>$-10</td>
<td>$-10</td>
</tr>
<tr>
<td>$7</td>
<td>$7</td>
</tr>
</tbody>
</table>
ESSAYS ON BARGAINING AND
STRATEGIC COMMUNICATION
ISBN: 978 90 361 0309 1

Copyright © Adrian de Groot Ruiz 2012

Cover design by Chris Weel

All rights reserved. Save exceptions stated by the law, no part of this publication may be reproduced, stored in a retrieval system of any nature, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, included a complete or partial transcription, without the prior written permission of the author, application for which should be addressed to author.

Alle rechten voorbehouden. Niets uit deze opgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar worden gemaakt, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.
Essays on Bargaining and Strategic Communication

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde commissie, in
het openbaar te verdedigen in de Agnietenkapel
op vrijdag 14 september 2012, te 12:00 uur

door

Adrián Wadse de Groot Ruiz

egeboren te Mexico-Stad, Mexico
Promotiecommissie:

Promotor: Prof. dr. T.J.S. Offerman
Co-promotor: Dr. A.M. Onderstal
Overige leden: Prof. dr. E.E.C. van Damme
Dr. N. Kartik
Prof. dr. R. Sloof
Prof. dr. A.J.H.C. Schram
Prof. dr. O.H. Swank
To my parents, Bita and Xochitl.
# Table of Contents

Acknowledgements ix

Chapter 1 Introduction 1

1.1 Overview 1

1.2 Chapter 2: Power and the Privilege of Clarity 3

1.3 Chapter 3: ACDC Rocks When Other Criteria Remain Silent 4

1.4 Chapter 4: An Experimental Study of ACDC 6

1.5 Chapter 5: Formal versus Informal Legislative Bargaining 7

Chapter 2 Power and the Privilege of Clarity 9

2.1 Introduction 9

2.2 Example 14

2.3 Theory 20

2.4 Conclusion 26

2.5 Appendix: Proofs 28

Chapter 3 ACDC Rocks When Other Criteria Remain Silent 35

3.1 Introduction 35

3.2 ACDC 38

3.3 Discrete games 47

3.4 ACDC versus Other Criteria in a Veto Threats Game 49

3.5 Crawford-Sobel Game 54

3.6 Experimental Evidence 59

3.7 Conclusion 62

3.8 Appendix: Proofs 62

Chapter 4 An Experimental Study of Credible Deviations and ACDC 71

4.1 Introduction 71

4.2 Theory 73
| Chapter 4 | Experimental Design and Procedures | 81 |
| Chapter 4 | Experimental Results | 83 |
| Chapter 4 | Dynamics | 88 |
| Chapter 4 | Robustness | 95 |
| Chapter 4 | Power and Information Transmission | 99 |
| Chapter 4 | Conclusion | 102 |
| Chapter 4 | Appendix | 103 |

| Chapter 5 | Formal versus Informal Legislative Bargaining | 133 |
| Chapter 5 | Introduction | 133 |
| Chapter 5 | The Bargaining Problem and Cooperative Solutions | 138 |
| Chapter 5 | Formal and Informal Procedure | 143 |
| Chapter 5 | Experimental Procedures and Design | 152 |
| Chapter 5 | Experimental Results | 156 |
| Chapter 5 | Conclusion | 166 |
| Chapter 5 | Appendix | 168 |

| References | 203 |
| Samenvatting (Dutch Summary) | 213 |
Acknowledgements

I would like to take this opportunity to thank those who have made this thesis possible and those who made my life more enjoyable in the past five years (or both).

I start with the people without whom this thesis would not have been in your hands. I consider myself very lucky to have had Theo and Sander as supervisors. They were true mentors, providing lots of freedom, gently indicating limitations, putting in lots of hours (as co-authors of Chapters 2-4) and being awesome individuals. I would also like to thank Roald and Arthur, as co-authors of the final chapter of this thesis and for ‘recruiting’ me for research in game theory and experimental economics in the first place. Finally, I am grateful to Eric van Damme, Navin Kartik, Randolph Sloof and Otto Swank for accepting to be on my doctoral committee and providing insightful comments.

Then there are two men whom I met at the start of my PhD and who have been close friends and a support throughout: my paronymphs Michel and Thomas. One man brought this thesis to the next artistic level: Chris, thanks - and not only for a great cover.

My academic habitat during my PhD studies has been the Center for Research in Experimental Economics and political Decision making (CREED) at the University of Amsterdam. CREEDers made the doctoral journey very enjoyable and shaped my ideas about human behavior and economics. So, thanks Adam, Ailko, Aljaz, Audrey, Ben, Eva, Frans, Gonul, Jeroen, Joep, Jos, Julian, Karin, Marcelo, Michal, Matthijs, Nadège, Pedro, Roel and Yang. And of course, Jona and Thomas, my partners-in-crime, whose presence I sometimes still can miss in my office.

In addition, I thank Andrew Schotter and CESS for hosting me for two months at New York University. I want to say thanks to Cars, who welcomed me into the economics faculty of the UvA when I graduated from University College Utrecht and was looking for a destination. I am still very much indebted to Piter van Tuinen and others from the Christelijk Gymnasium Utrecht for their unwavering support early on in my Bildung. Finally, I am grateful to Utz
for his advice and patience, and to the economics faculty of Radboud University for having provided me a new and warm academic home.

Then I come to my social support line. I am very fortunate to still see friends from elementary school (Lancelot and Sander), be surrounded by my high school friends (Chris, Wilco, and Jan) and be in touch with those from College (Madalina, Borja, Nynke, Aida, Marianne, Feri, Gabor, Annemarie, Maarten, Joost and Ties). I don’t forget Arthur, Annelies and Daphne. I am also glad to play risk with Alex, Noach, Sjoerd and Sander. I still feel connected to United Netherlands, at least at heart, by people as Willemien, Tjeerd and Stijn. I took up debate at the same time I started my PhD. I had a lot of fun debating with members of Bonaparte such as Reinier and Jeannette amongst many others as well as with people from other societies whom I shall not name out of loyalty to my own tribe. I have enjoyed coming together at (and after) meetings with the young Worldconnectors and very recently the Shapers. For inspiration I owe a lot to Herman, Herman and Hans, Worldconnectors young at heart. In Mexico, I say gracias to Lili, Maria Elena, Maricela, Nelly and Ofe for their care. A few individuals defy categories. Eric: grazie per tutto caro amico. Michel: hope we will be working side-by-side as social entrepreneurs for a long time to come.

A special ‘thank you’ goes out to John and Inge for being great landlords and neighbors for the past few years. At this time of writing, they prevent the Amstel waters that creep into my apartment from tickling my knees.

Those who know me a bit will confirm I am very much a family person and for good reasons. I thank my close relatives on both sides of the Atlantic for their caring interest in me and my activities; in particular Tía Queta, Adriana, Alejandro, Gaby, Queta, Geuby, Tineke, Gerrit and their families.

I end with my existential support system. Xochitl, thank you for being! Papa and mama: I owe you the world. Mama: gracias por todo el amor que me has dado y por compartir conmigo tu cariño por el estudio y la humanidad. Papa: thank you for the countless hours you have spent with me, from teaching me Dutch in Mexico-City all the way to improving many of the texts in this book.

Adrian
Amsterdam, 30 July 2012
Chapter 1  Introduction

1.1  Overview

Communication and bargaining are social activities that permeate many aspects of life. We negotiate with strangers, friends and foes alike: We strike political and business deals, try to cajole our friends into the restaurant we like and haggle in the family about the small (who changes the diapers) and big (who gets the children after divorce). The two activities are intimately linked: Bargaining is done through communication and many instances of communication involve a certain degree of strategic behavior and bargaining.

This thesis is a collection of four essays on bargaining and strategic communication from the perspective of behavioral economics. These two topics have a rich tradition in standard economics, in particular in game theory. In recent years, behavioral economists have been able to improve on the predictions of standard economics by relaxing its assumptions and supplementing theory with experimental data. A common thread running through this thesis is its focus on strategic situations (games) where strict adherence to traditional assumptions of rationality and equilibrium fail to yield helpful predictions. As such, it falls within the domain of behavioral game theory. In particular, it belongs to the branch that focuses on predicting aggregate behavior of boundedly rational individuals in dynamic settings. The first and last study we present are applications of behavioral methods to economic and political questions (in Chapters 2 and 5). Chapter 3 and 4 provide a methodological contribution by introducing and testing a behavioral stability measure for cheap talk equilibria. In the remainder of this introduction, we first provide an overview showing the common threads running through the chapters. Subsequently, we discuss each study and its contribution to the literature in more detail.

In Chapter 2 we study how bargaining power influences the clarity of communication. In other fields the relation between power and communication is widely recognized, but in economics this relation is a largely untouched research
area. By modeling communication as a cheap talk game, we get a remarkable theoretical result: The informed party cannot transmit much information if she has little power, as the uninformed party can use that information to exploit her. Increasing the informed party’s power reduces the extent to which she can be exploited. This, in turn, increases the maximum amount of information transmission possible in equilibrium. The analysis leaves one issue open: increasing the informed party’s power does not eliminate equilibria with little information transmission. Hence, to make a prediction we need an equilibrium refinement. However, cheap talk equilibria are hard to refine and existing refinements have little predictive force in this model.

This leads us to the open problem that no equilibrium refinement exists that is generally predictive in cheap talk games. We take up this challenge in Chapter 3, which introduces the Average Credible Deviation Criterion (ACDC), a stability measure for cheap talk equilibria. ACDC builds on previous rational theories of credible deviations. It manages to be generally predictive by relaxing the common rationalistic assumptions that stability is binary and that an equilibrium is completely upset by any (credible) deviation. We show that the predictions of ACDC are sensible across a wide range of games and that it can organize data well from earlier experiments. One application is that ACDC selects a most informative equilibrium in the class of games studied in Chapter 2. This allows us to substantiate the prediction that an increase in the Sender’s relative power leads to an increase in actual information transmission.

In Chapter 4, we put ACDC to the test in a new set of experiments. This experiment is the first to systematically test whether and to which extent credible deviations matter. We use the class of cheap talk games introduced in Chapter 2, as this allows for large equilibrium sets and hence provides a good testing ground for refinements. We find clear support for the predictions of ACDC. In addition, we designed the experiment in such a manner that we can also directly test the hypothesis of Chapter 2. The data shows that increasing power indeed increases information transmission.

Finally, in Chapter 5 we study how the formality of (legislative) bargaining affects its outcome. Whereas Chapters 2, 3 and 4 focus on information transmission during bargaining, Chapter 5 assumes complete information and zooms in
on the bargaining process itself. The bargaining problem studied here shares important characteristics with that of Chapter 2 (and 4): both have a non-convex outcome set that consists of the real line on which preferences are single peaked and a disagreement point outside of the line. Chapter 5 starts with the observation that political agreement is typically forged in a mix of formal bargaining in parliamentary session and informal bargaining outside of it. A major challenge in studying informal bargaining is the strategic richness of such situations. We show that standard game theory is silent about the influence of formality in the setting we study. Using an experiment, we are able to show that formality matters: informal bargaining is better for the party in the superior bargaining position.

Below we describe our contributions in more detail.

1.2 Chapter 2: Power and the Privilege of Clarity

Chapter 2 addresses an issue that has largely gone unnoticed in economics: the relation between power and communication. In other fields, in contrast, it has been widely recognized that power has an important impact on communication. One particular idea is that clarity is a privilege of the powerful: People belonging to a more powerful social, economic or political group seem to be able to communicate their preferences more clearly than those with less power. Social psychologists have found, for instance, that workers are more assertive in communicating their desires towards lower ranked co-workers than towards higher ranked co-workers (Kipnis, Schmidt & Wilkinson (1980), Yukl & Falbe (1990)). Similarly, gender studies have found that women in patriarchal societies are more hesitant in stating their wishes and interests than men (Baer (1976), Butler (1976), Maltz & Borker (1982), Henley & Kramarae (2001)).

It could be that the relation between power and clarity is entirely historical and cultural. In Chapter 2, however, we explore the possibility that there is a strategic foundation to this relation as well. We model the interaction in an asymmetric information continuous cheap talk (veto threats) bargaining game. It consists of one round of communication between an informed Sender and an
uninformed Receiver, followed by an ultimatum game where the Receiver makes a proposal to the Sender. Using game theory has the virtue that we can operationalize our concepts precisely: We define power as the attractiveness of the outside option and clarity as the degree of information transmission in equilibrium. Our main result is that the maximum amount of information transmission in equilibrium is increasing in the power of the Sender and decreasing in that of the Receiver. The intuition is that having power allows the Sender to be more open: Information transmission is limited because the Receiver can exploit the Sender with information the latter provides and increasing the Sender’s power reduces the extent to which she can be exploited. In Chapter 4, we provide experimental evidence that the Sender’s relative power increases information transmission.

In addition to casting game theoretic light on the relation between power and communication, this chapter also contributes to the existing cheap talk literature. Our model differs from previous ones in that the private information of the Sender does not determine her bargaining power. This allows us to capture the power individuals have due to the social, political or economic position of the group they belong to. This is what sets it apart from Matthew’s model of veto threats (1989) and buyer-seller models (e.g. Matthews & Postlewaite (1989) and Farrell & Gibbons (1989)). The modeling choice also has profound implications for information transmission. Whereas in Matthews’ model, information transmission is limited, a full range of partition equilibria exists in our set-up. In particular, the role of power in our model mirrors the role of interest-alignment in the model of Crawford and Sobel (1982).

Finally, we provide testable implications of our model for labor contracts and for remedy negotiations between firms and competition authorities.

1.3 Chapter 3: ACDC Rocks When Other Criteria Remain Silent

Chapter 3 proposes a solution to the important equilibrium selection problem in cheap talk games. Crawford & Sobel (1982) introduced asymmetric infor-
mation games with cheap talk communication. The main question in such games is how much information can be transmitted in pre-play costless communication between an informed Sender and an uninformed Receiver when interests are partially aligned. This class of games proved to have many real world applications in economics and politics, ranging from stock recommendations (Morgan & Stocken, 2003) to the presidential veto (Matthews, 1989). A major problem in applying such models is that they have multiple equilibria, which differ drastically in their predictions about how much information will be transmitted. Furthermore, the set of equilibria has proven to be very hard to refine, since – in contrast to signaling games – messages are costless. Currently, no refinement exists that successfully selects equilibria across a wider range of cheap talk games. Our model in Chapter 2 is one example where existing refinements are not predictive. We believe that existing refinements lose predictive power because they impose a binary distinction between stable and unstable equilibria, whereas the success of equilibria to organize behavior is typically a matter of degree.

This study proposes a generalization of refinements based on credible deviations, such as neologism proofness (Farrell, 1993) and announcement proofness (Matthews, Okuno-Fujiwara & Postlewaite, 1991). These refinements are grounded on the observation that messages can have a literal meaning. In particular, such messages can urge the Receiver to play an out-of-equilibrium action in a manner that is credible to rational players. Neologism proofness and announcement proofness select equilibria that do not admit credible deviations. Unfortunately, these refinements tend to eliminate all equilibria. According to our Average Credible Deviation Criterion (ACDC), the stability of a cheap talk equilibrium is a continuous quantity determined by the frequency and the size of credible deviations. An equilibrium is an ‘ACDC equilibrium’ if it minimizes the amount of credible deviations. This provides a way to rank equilibria that are unstable in a strict sense and to guarantee the existence of a most plausible equilibrium.

ACDC provides a contribution relative to existing concepts as neologism proofness, announcement proofness, Partial Common Interest (Blume, Kim & Sobel, 1993) and No Incentive to Separate (Chen, Kartik & Sobel, 2008). We
show that ACDC organizes behavioral data equally well as the other criteria in the settings for which these criteria were designed, improves upon them in other settings and makes sensible predictions where all previous criteria are silent.

1.4 Chapter 4: An Experimental Study of ACDC

Chapter 4 puts ACDC to the test in a new experiment. The experiment tests the predictions of ACDC that credible deviations matter and matter gradually. In addition, it tests directly whether the ACDC equilibrium predicts best. In the experiment, we study five games that belong to the model of Chapter 2. These games are a suitable testing ground for selection criteria, as they allow for a clean manipulation of the size and frequency of credible deviations and can have a large equilibrium set. Our key results are that credible deviations matter gradually and that the ACDC equilibrium predicts best in each game. More generally, the data provides evidence that ACDC can predict the stability of equilibria within and across games, even if all equilibria admit credible deviations and if all existing criteria are silent. We introduce a neologism dynamic that supports the conclusions of ACDC and explains important dynamic aspects of our data. In addition, our design is such that the experiment allows us to test the main hypothesis from Chapter 2 about power and clarity. We find that increasing the relative power of the Sender indeed increases information transmission.

Our experiment adds to the recent experimental literature on equilibrium selection in cheap talk games. Blume, DeJong & Sprinkle (2001) study equilibrium selection in discrete games, whereas Dickhaut, McCabe & Mukherji (1995) look at the Crawford-Sobel uniform quadratic game. Both studies support the conclusions of ACDC, although the results of the former are also consistent with the PCI criterion (Blume, Kim & Sobel, 1993) and those of the latter with the NITS criterion (Chen, Kartik & Sobel, 2008). In the current experiment we employ a setting where both other concepts are silent.

---

1.5 Chapter 5: Formal versus Informal Legislative Bargaining

Whereas the previous chapters focused on information transmission, Chapter 5 examines the bargaining process itself. It deals with the question of how the formality of the legislative bargaining process can affect the outcome. When parliament is in session, parliamentary procedures strictly govern what members can do at what time; hence, bargaining is highly formalized. After official sessions have been adjourned, however, bargaining often continues informally in offices, corridors and backrooms, where formal rules barely exist. As a consequence, the outcome of the legislative process is usually a result of both formal and informal bargaining. How the degree of formality affects the bargaining outcome is an open question.

That the bargaining procedure can drastically affect the outcome and favor specific negotiators through the order of voting, agenda-setting power, or proposal and voting rights has been recognized since the research boom on spatial voting in the late 1970s (e.g., McKelvey (1976; 1979), Schofield (1978), McCarty (2000)). The difference between a formal and an informal procedure, however, cannot be captured by a difference in voting or proposal rights. In addition, if both procedures provide equal rights to each player, then no procedure prima facie favors a specific player. Rather, the difference is that informal bargaining provides much more flexibility to all the bargaining parties. Players can use this flexibility both to make more offers and to communicate more, since any proposal also conveys credible information about what a player is willing to accept.

To answer our research question, this study compares a formal alternating offers game to an informal continuous-time bargaining game. Both non-cooperative games correspond to the same cooperative game: a three-player median voter setting with an external disagreement point. The divergence of interests (polarization) determines whether the core is empty (if so, we consider the uncovered set). In the formal game, the (refined) subgame perfect equilibrium converges to the core if this exists. The informal game is strategically so rich that a large range of outcomes can be supported in equilibrium, including the
equilibrium outcomes of the formal game. As a consequence, theory is silent about the influence of formality on the outcome and an experiment can shed light on this issue. Our main experimental finding is that formality matters. In particular, the median player is significantly better off under an informal bargaining procedure. Our interpretation is that the informal game provides the median player more room to exploit her superior bargaining position by allowing her to play off the other players against each other. Another interesting result we get is that the median player is harmed by polarization, even if her ideal is the unique core, probably due to inter-coalitional fairness concerns.

The study in Chapter 5 is the first to provide evidence that the formality of bargaining matters to the outcome. In particular, it suggests that parties in a superior bargaining position are better off under an informal structure. Our finding is relevant for the study of institutional choice, because it suggests that political agents (should) have strategic preferences over the weight they wish to put on formal versus informal bargaining. To put this far-reaching conclusion on a stronger footing, more research is needed as we compare two representative but still specific procedures. Recently, this conclusion has received support from Drouvelis, Montero & Sefton (2010).
Chapter 2  

Power and the Privilege of Clarity

2.1  

Introduction

Clarity seems to be a privilege of the powerful. The less fortunate among us are typically vaguer about their desires and need to think harder about what they say or do not say. Social psychologists have found that workers are more assertive in communicating their desires towards lower ranked co-workers than towards higher ranked co-workers (Kipnis, Schmidt & Wilkinson (1980), Yukl & Falbe (1990)). Gender studies point to a similar pattern in patriarchal societies, where women are found to be more hesitant in stating their wishes and interests than men. The relation between power and clarity could be shaped primarily by history and culture. In the communication literature, the link between power and communication is widely recognized (Keating, 2009) and believed to be strongly mediated by culture (Gudykunst & Lee, 2003). High status individuals are typically approached with more respect and too clear a message by a lower ranked individual about her preferences might simply be seen as ‘disrespectful.’ Similarly, direct communication of preferences may result in the loss of face of the powerful person if it openly contradicts her wishes or of the less powerful person if her wishes are ignored. In contrast to the above disciplines, in economics the relation between power and communication is a largely untouched research area.

In this chapter, we explore the possibility that there is a fundamental strategic foundation to the relation between power and clarity. In particular, we are interested in the communication between members belonging to different groups of a society (or organization or community) with different levels of power. We

---

2 This chapter is based on De Groot Ruiz, Offerman & Onderstal (2011a).

3 When discussing the source of miscommunication between men and women, some authors emphasize the role of power relations while others stress the role of culture (Baer (1976), Butler (1976), Maltz & Borker (1982), Heuley & Kramarae (2001)).

4 As far as we know, in economics the only research touching on this subject concerns how the level of connectedness in a network affects the bargaining power of individuals in bilateral negotiations (Calvó-Armengol, 2001).
focus on bargaining under asymmetric information, as this is a common type of social interaction where clarity matters. One can think of divorce negotiations between men and women, managers and workers discussing the worker’s tasks, members of different castes in India bargaining over the provision of a service or competition authorities discussing merger remedies with multinationals.

How does power come into play in such situations? Importantly, the consequences of disagreement differ among individuals coming from groups with different levels of power. Simply put, people in a more powerful position have better outside options. This can firstly be due to the fact that people who belong to a powerful group benefit from institutional or cultural rules. For instance, in countries with Islamic law, men have more rights than women at divorce. Secondly, people with more power tend to have more social, political or economic resources. Even in communities where women have equal legal rights but do not perform (much) paid work, men tend to have a superior economic position when filing for divorce. In sum, power affects the costs of disagreeing for agents in bargaining settings.

We think about clarity as informational clarity: how much does someone learn about the state of the world from a message? The informational clarity of a message can firstly depend on its literal clarity: the indirectness, inexplicitness, vagueness or ambiguity of the words used (Cheng & Warren (2003), Agranov & Schotter (2010)). In a single interaction with a stranger from a culture one does not know, the literal meaning is all one can go by. Secondly, if people share a cultural and social history, the information messages convey also depends on how messages are used. For instance, the precise statement “I’ll be there at seven o’clock” is in some cultures not at all informative, because people use it under a wide range of intentions as when to come. By contrast, in some countries the ambiguous phrase “I may prefer if you stopped making noise” can be very informative if such a formulation is only used when people are really upset. The more history people share, the more the clarity of messages will depend on their use. In equilibrium, informational meaning is completely determined by use: what message is used in what state of the world?

---

5 One may of course be mainly interested in literal clarity, for instance for linguistic purposes. We are chiefly interested in informational clarity as this determines the actions people take.
Language and its interpretation evolve and as such are subject to strategic forces. Hence, in the long run, there will be a tendency towards some strategic equilibrium. This tendency is strongest in dynamic settings where members of one power-group frequently interact with different members of the other group and reputational concerns play a small role.\(^6\) One implication is that in a stable culture, the informational clarity of messages is largely determined by their use. Hence, one should be careful when providing purely cultural explanations for a lack of informational clarity on the side of individuals with little power. If politeness requires vague messages, then a group may start using several polite (and vague) messages. Over time, these messages may evolve to encode harder information if there is a strategic pressure—such as efficiency gains—to do so. For example, consider a wife and husband who can go to the theater or a concert. The wife knows they both prefer the theater. In a Western society, the wife may reply to her husband’s question “Shall we go to the concert” with “Nah, let’s go to the theater.” In a more patriarchal society where it is impolite for the wife to contradict her husband openly, she may instead say “Well, Sir, are you sure you do not feel like going to the theater?” with the same outcome. Another example is the Iranian practice of Taarof civility, which requires among other things that a shopkeeper says his products are ‘worthless’ when asked for the price. Still, despite such politeness, the price is revealed in the end. Customers have learned they should repeat the question a couple of times to get a real answer and cannot just walk away without paying.\(^7\)

---

\(^6\) Hence, we focus on one-shot situations where people have social information. Other fascinating possibilities are to study the relation between power and clarity in one-shot interactions without social information or in repeated interactions. Although an equilibrium analysis may be informative here, we have to be careful when generalizing our findings to such settings. In interactions without social information, out-of-equilibrium behavior in cheap talk games cannot be ruled out. In such cases, approaches based on rationalizability and some focal meaning of messages, such as Rabin (1990), may be more appropriate. This means that one needs to look at the literal meaning to derive its information content in such cases. One problem is that in these settings we are as of yet only able to theoretically predict only rather conservative lower bounds on information transmission. Another problem is that it is not always clear what the relation between the literal and focal meaning of messages is. Analyzing repeated interactions between the same players is even more challenging. Strictly speaking, it is just a very complex one-shot game, possibly involving an endogenous form of reputation building (Sobel, 1985). In real life, myopic strategic reasoning may give it a dynamic flavor, so that messages may acquire a consistent meaning justifying some equilibrium concept.

\(^7\) For more on Taarof and the power-language relationship in Iran, see Beeman (1986).
It is now possible to translate our original question of how power influences clarity into a precise game-theoretic one: how does bargaining power influence information transmission in equilibrium? We study this question in an elementary bargaining setting. A Sender with private preferences and a Receiver with commonly known preferences bargain over a one-dimensional issue.\textsuperscript{8} The Sender sends a costless message to the Receiver, after which they play an ultimatum game in which the Sender can reject or accept the proposal of the Receiver.

We find that bargaining power is a key determinant of how much information can be transmitted: information transmission is increasing in the Sender’s power and decreasing in that of the Receiver. In other words, the higher the relative power of an informed agent, the clearer she will be. There is one exception in which full revelation is possible. Senders who are closely aligned with the Receivers or have no bargaining power can fully reveal their type since they will be offered the Receiver’s preferred outcome anyway in equilibrium. The range of Senders who can reveal their type without costs decreases with the relative power of the Sender.

We see our results primarily as a proof-of-principle, as many power relations and strategic settings are more complex in practice. At the same time, we believe that the intuition behind our results holds more generally. If you hold little power, it is not in your best interest to reveal too much information, because that can be exploited. Hence, you better be kind of vague and strategic about what you communicate. If you are powerful, the potential for exploitation is limited and you can afford to be clear.

In addition to shedding light on power relations, this chapter contributes to the theoretical literature on bargaining and information transmission. Our model differs from previous models in that the private information of the Sender does not determine her bargaining power. This allows us to capture the power individuals have due to the social, political or economic position of the group they belong to. Our model is close to that of Matthews (1989), who was the first to study veto threats. In Matthews’ model, however, the Sender’s type determines her disagreement payoff, whereas in our model the disagreement payoff is an exogenous variable which is the same for all Sender types. Hence, we can

\textsuperscript{8} We will refer to the Sender as a ‘she’ and the Receiver as ‘he.’
model members of the same group having commonly known and similar power levels but differing in their private preferences. For example, consider men filing for divorce in Saudi Arabia and entering into custody negotiations. They all have the same legal position but differ in their preferences to see their children.

Our modeling choice has profound implications for information transmission. In Matthews’ model, information transmission is limited: the maximum equilibrium size is two. In our setup, a full range of Crawford-Sobel-like partition equilibria exists, potentially allowing for more refined communication. In particular, the role of power in our model mirrors the role of interest-alignment in the Crawford-Sobel game.

The literature on economic bargaining and information transmission has mostly focused on buyer-seller situations, where the outcome-set is zero sum conditional on trade (e.g. Matthews & Postlewaite (1989) and Farrell & Gibbons (1989)). In these models, the bargaining power of the other party is typically unknown, so that power and private information again coincide.9

Finally, our model applies to various interesting situations. Consider, for example, custody negotiations between lawyers of divorcing parents. A common situation is that the mother would like to see the children as much as possible, whereas the father’s preferences are not precisely known. If they do not manage to agree, they will have to go through a costly court procedure. In these situations, one can ask how the power of the father relative to the mother affects the ability of his lawyer to communicate the preferences of his client. In the conclusion, we discuss testable economic implications for labor contracts and for remedies merging firms propose to competition authorities.

The remainder of this chapter has the following structure. Section 2.2 presents a simple example of our model that serves to illustrate our set-up and results. Section 2.3 presents the model and the results. Section 2.4 concludes.

---

9 For a literature-review on bargaining with incomplete information see Ausubel, Cramton & Deneckere (2002).
2.2 Example

2.2.1 Game

Consider a cheap talk game with veto threats between an informed Sender and an uninformed Receiver. The outcome of the game $x$ is a point on the interval $[0,1]$ or the disagreement point $\delta \not\in [0,1]$. The Sender’s payoff on the interval depends on the state of the world $t$ (her type): $U^S(x,t) = -|x - t|$. The larger the distance between the outcome $x$ and her type $t$, the lower the Sender’s payoff. Her type $t$ is private information of the Sender and it is common knowledge that $t$ is drawn from the uniform distribution on $[0,1]$. The Receiver’s payoff on the interval, $U^R(x) = -x$, is independent of $t$: he always prefers smaller outcomes to larger ones. We vary the payoff of the disagreement point to the players: $U^R(\delta) = -d^R$ and $U^S(\delta) = -d^S$ with $d^R, d^S > 0$. (Note that $d^R$ and $d^S$ are the size of the “harm” if bargaining breaks down.) In particular, we have:

\[
U^R(x) - U^R(\delta) = d^R - x \text{ for all } x \in [0,1] \\
U^S(x,t) - U^S(\delta) = d^S - |x - t| \text{ for all } x \in [0,1]
\]

Observe that the Receiver prefers $\delta$ to all outcomes more than $d^R$ away from the origin and that the Sender prefers $\delta$ to all outcomes on the line more than $d^S$ away from her type $t$.

The game proceeds as follows. First, the Sender is informed of her type $t$. Subsequently, she sends a costless message $m \in M$ to the Receiver, where $M$ is some sufficiently rich message set. Then, the Receiver proposes an action $a \in [0,1]$ to the Sender. Finally, the Sender accepts or rejects $a$. If she accepts, $a$ is the outcome and if she rejects, $\delta$ is the outcome.

The game is an elementary bargaining setting under asymmetric information and models some important aspects of real interactions. The one-dimensional bargaining set allows us to capture partially aligned and partially conflicting
interests of the players. The one-sided information asymmetry and single round of ‘pre-play communication’ captures the essence of biased information transmission. The game is similar to Matthews’ (1989), except that the disagreement point lies on the real line in Matthews’ model. The disagreement point being now outside of the line allows us to model differences in bargaining power independent of player’s preferences on the line. Hence, the disagreement payoff reflects power individuals have due to their (commonly known) social, political or economic position and which they share with other members of their group.\footnote{In theory, the disagreement point could be determined both by social position and private preferences. We assume for simplicity that it only depends on a player’s social position, as this is sufficient to address the relationship between power and information transmission.}

The bargaining power of the players in our game is determined by how attractive the disagreement point $d$ is to them. As $d^S$ [$d^R$] becomes smaller, the Sender’s [Receiver’s] payoff of the disagreement point increases and the interval of points that the player prefers to $d$ narrows. Hence, the larger $d^S$ [$d^R$], the smaller the bargaining power of the Sender [Receiver].

2.2.2 Equilibria

We look at a refinement of perfect Bayesian equilibria that restricts the Receiver to pure strategies and lets the Sender consider that she may tremble at the veto stage. Hence, for the Receiver we need to specify which message elicits which action, and for each Sender a probability distribution over messages she sends. In equilibrium, the Receiver best responds to his correctly updated posterior beliefs and each Sender type induces the action(s) that give her the highest payoff, even if she plans to veto anyway (since she might tremble). From now on we refer to a perfect Bayesian equilibrium satisfying these two requirements simply as an equilibrium.

It turns out that all equilibria are partition equilibria. In a partition equilibrium, types separate into disjunct intervals. A partition equilibrium can be characterized by the finite set of actions $a_1 < a_2 < \ldots < a_n$ the Receiver proposes in equilibrium.\footnote{To be precise, the action set characterizes an (infinite) class of essentially equivalent equilibria that induce the same equilibrium outcome and differ only in the messages that are used.} The number of equilibrium actions $n$ is called the size of the
equilibrium. Each type then simply induces that action that is closest to her type, so that the type space can be partitioned into intervals of types who induce the same action. The main intuition for this result is that full separation is impossible in any type-interval contained in $(d^5, 1]$: a revealed type would get zero, such that she has an incentive to mimic a higher type. In fact, there is a minimum distance between any two equilibrium actions, so that the equilibrium action set is finite. Together with the fact that Senders elicit the action closest to their type, this results in a well-ordered partition.

In games with only partition equilibria, the size of the equilibrium provides a natural measure of information transmission, which is invariant to scaling (of the payoff, action or type space). Equilibrium size will correlate well with other measures, such as conditional variance, prediction error, ex-ante efficiency or ex-post efficiency – with the suitability of each depending on the context.\footnote{A problem of these measures is that they lack a natural dimension, such that they are typically not invariant to immaterial transformations of the game. For instance, the average prediction error or the conditional variance is not invariant to scaling of the type space and ex-ante utility is not invariant to immaterial transformation of payoffs of subsets of the type set. The fraction of outcomes that is ex-post efficient does not suffer from this invariance problem, but is a rather crude measure.} In all cheap talk games, a size-1 (pooling) equilibrium exists in which no information is transmitted. We say that more information is transmitted as the size of the equilibrium increases.

In this chapter we focus on the relation between power and the maximum information transmission possible in equilibrium. In particular, we look at how power affects the maximum equilibrium size. Like in other continuous cheap talk games, such as Crawford & Sobel (1982) and Matthews (1989), the maximum size equilibrium seems the most plausible one although the equilibrium set is actually hard to refine.\footnote{Traditional signaling refinements have no bite in cheap talk games, as messages are costless. In Chapter 3, we show that in the current game also the cheap talk refinements neologism proofness (Farrell, 1993), announcement proofness (Matthews, Okuno-Fujiwara & Postlewaite, 1991), communication proofness (Blume & Sobel, 1995), the recurrent mop (Rabin & Sobel, 1996) and NITS (Chen, Kartik & Sobel, 2008) are not selective and the non-equilibrium concepts of Credible Message Rationalizability (Rabin, 1990) and Partial Common Interest (Blume, Kim & Sobel, 1993) are not predictive.} Chapters 3 and 4 provide a theoretical and experimental justification to focus on the maximum size equilibrium. In Chapter 3, we introduce the Average Credible Deviation Criterion (ACDC). This criterion
generalizes credible deviations approaches as neologism proofness (Farrell, 1993) and announcement proofness (Matthews, Okuno-Fujiwara & Postlewaite, 1991). ACDC does not suffer from non-existence and organizes data from previous experiments successfully. We show in Chapter 3 that, under some additional assumptions, ACDC selects in the model we present in this chapter a unique maximum-size equilibrium. In Chapter 4, we conduct an experiment with games belonging to the current model and find that the maximum size equilibrium indeed predicts behavior best.

As an illustration of what the equilibria look like, consider the case where \( d^S = \frac{1}{2} \) and \( d^R = \frac{3}{4} \). This game has two equilibria. In the pooling equilibrium, the Receiver always proposes action \( \frac{3}{4} \). In this equilibrium, the Receiver ignores all messages and best responds to his prior beliefs. The optimal action of the Receiver always involves a trade-off between maximizing the probability that the proposal is accepted and maximizing the payoff of the proposal conditional on acceptance. Senders all have the same message strategy, which is optimal for each type as it does not matter what they send. The game also has a size-2 equilibrium with actions \( a_1 = 0 \) and \( a_2 = \frac{1}{2} \). Senders in \( [0, \frac{1}{4}) \) could, for instance, send the message “My type is low,” inducing the Receiver to propose 0 and those in \( [\frac{1}{4}, 1] \) the message “My type is high,” inducing the Receiver to propose \( \frac{1}{2} \). When the Receiver receives the message “My type is low” he correctly infers that types are in the interval \( [0, \frac{1}{4}) \) and proposes 0, as all types in \( [0, \frac{1}{4}) \) accept 0. If he receives the message “My type is high,” he infers \( t \in [\frac{1}{4}, 1] \). Her expected payoff from proposing action \( a \) now is \( EU^R(a) = \Pr\{a \text{ is accepted}\} \times \left(U^R(a) - U^R(\delta)\right) = \left(\min\{1, a + \frac{1}{2}\} - \max\{\frac{1}{4}, a - \frac{1}{4}\}\right) \times \left(\frac{3}{4} - a\right) \), which is maximized at \( a = \frac{1}{2} \). We can characterize the equilibrium set as follows for general \( d^S, d^R > 0 \).
Equilibrium Characterization Example

Let $\overline{x} = \max\{0, \min\{d^R - 2d^S, 1 - d^S\}\}$. Any equilibrium of the game described in section 2.2.1 is a partition equilibrium that can be described by a natural number $n \in \{1, \ldots, \overline{n}\}$ and a set of equilibrium actions $\{a_1, \ldots, a_n\}$, such that

1. $a_1 = \max\{0, \min\{1 - d^S, d^S, \frac{1}{2}(d^R - d^S)\}\}$ if $n = 1$
2. $a_1 = \min\{d^S, \max\{0, a_2 - 2d^S\}\}$ if $n \geq 2$
3. $a_2 = \min\{\frac{1}{2}(d^R - d^S), 2d^S, 1 - d^S\}$ if $n = 2$ and $a_1 = 0$
4. $a_k = a_{k-1} + 2d^S$ if $a_{k-1}$ exists and $a_{k-1} > 0$
5. $a_n \leq \overline{x}$ if $d^R \geq 4d^S$

The maximum size $\overline{n}$ is equal to 1 if $d^S \geq 1$. If $d^S < 1$, $\overline{n} = \max\left\{2, \left\lfloor \frac{d^R}{2d^S} \right\rfloor \right\}$ if $d^R \leq d^S + 1$ and $\overline{n} = \max\left\{2, \left\lfloor \frac{3}{2} + \frac{1}{2d^S} \right\rfloor \right\}$ otherwise, where $\lfloor . \rfloor$ is the ceiling function.\(^{14}\)

2.2.3 Power and Clarity

We can now turn to the central question of this chapter: how does power affect information transmission in equilibrium? To determine this, we vary $d^S$ and $d^R$. For instance, suppose that relative to $d^S = \frac{1}{2}$ and $d^R = \frac{4}{7}$, we increase the bargaining power of the Sender to $d^S = \frac{1}{2}$. This results in more information transmission that can be supported in equilibrium: the maximum equilibrium size goes from 2 to 3. This illustrates the key result that the maximum information transmission possible in equilibrium increases with the power of the Sender and decreases with that of the Receiver:

**Maximum Equilibrium size.** $\overline{n}$ is decreasing in $d^S$ and increasing in $d^R$.\(^{14}\)

\(^{14}\) We provide the proof at the end of the appendix.
This follows directly from the characterization of the equilibria. If the Sender has no power \( d^S \geq 1 \) and all types would accept the Receiver’s optimal action, only a pooling equilibrium exists. If the Sender has some power \( d^S < 1 \), at least a size 2 equilibrium exists. If the Receiver has very little power \( d^R > d^S + 1 \), the maximum size only depends on (and is increasing in) the power of the Sender. If the Receiver has some power as well \( d^R \leq d^S + 1 \), then the maximum size \( \pi \) increases in the relative power of the Sender:

\[ \pi = \max \left\{ 2, \left\lfloor \frac{d^R}{2d^S} \right\rfloor \right\}. \]

Therefore, \( \pi \) jumps to a higher level if \( \left\lfloor \frac{d^R}{2d^S} \right\rfloor \) increases sufficiently. As the power of the Sender relative to that of the Receiver becomes large, the maximum equilibrium size also becomes large.

The intuition for this result is the following. The highest action the Receiver prefers to \( \delta \) (in our game \( d^R \)) and the type density close to this point impose an upper limit on the highest action. The fact that the Receiver always wants to offer a lower action than the Sender imposes a lower bound on how close equilibrium actions can be together. A Sender prefers \( t \), whereas the Receiver would offer \( \max\{0, t - d^S\} \), if he would know \( t \). In our example, the smallest distance between two positive equilibrium actions is \( 2d^S \).

If the Sender is sufficiently powerful (and the Receiver has some power as well), we can construct the maximum size equilibrium as follows. We set \( a_n \) equal to the highest action possible in an equilibrium, which is \( d^R - 2d^S \). From there, we create the tightest partition by iteratively setting \( a_{i-1} \) as close as possible to \( a_i \) as long as \( a_{i-1} > 0 \). In this example, we need to set \( a_{i-1} = a_i - 2d^S \). Finally, we set \( a_i = 0 \). We can show that this tightest partition is in fact an equilibrium. As a consequence, the maximum equilibrium size is \( \left\lfloor \frac{d^R}{2d^S} \right\rfloor \); namely 1 for \( a_1 = 0 \), plus \( \left\lfloor \frac{d^R - 2d^S}{2d^S} \right\rfloor \) positive actions minus 1 if \( d^R - 2d^S \) is divisible by \( 2d^S \).

Hence, an increase of the Receiver’s power (decreasing \( d^R \)) leads to a decrease of information transmission by lowering the highest possible equilibrium
action \((d^S - 2d^S)\). An increase of the Sender’s power (decreasing \(d^S\)) increases information transmission firstly by increasing the highest possible equilibrium action. Secondly and more importantly, it lowers the minimum distance between equilibrium actions \((2d^S)\). In the next section, we show that this result holds for a broader class of payoff functions and type distributions. The underlying intuition is twofold. First, the power of a Receiver limits the range of potentially mutually profitable actions. Second, the conflict of interest between Sender and Receiver puts a fundamental upper bound on information transmission, as the Receiver has an incentive to exploit the Sender’s information. However, the bargaining power of the Sender raises this upper bound by limiting how much a Receiver can exploit the Sender.

2.3 Theory

2.3.1 Model

A Sender and Receiver play the following game with an outcome in \(X = \mathbb{R} \cup \{\delta\}\). First, the Sender privately observes her one dimensional type \(t\). It is common knowledge that \(t\) is drawn from the uniform distribution on the interval \([0,1]\).\(^{15}\) Second, the Sender sends a message \(m \in M\), where \(\mathbb{R} \subset M\). Third, the Receiver receives \(m\) and proposes an action \(a \in \mathbb{R}\). Finally, the Sender accepts or rejects \(a\). If the Sender accepts [rejects], the proposed action \(a\) [\(\delta\)] is the outcome of the game.

Let \(U^R : X \rightarrow \mathbb{R}\) be the utility function of the Receiver and \(U^S : X \times T \rightarrow \mathbb{R}\) that of the Sender. We model the players’ bargaining power as the payoff of the disagreement point \(U^R(\delta)\) and \(U^S(\delta)\), where we assume \(U^S(\delta, t) = U^S(\delta)\) does not depend on \(t\). \(U^R\) and \(U^S\) satisfy the following assumptions:

\(^{15}\) Given that types \(t\) are drawn from a smooth distribution function \(F\), we can make the assumption of uniformly distributed types without further loss of generality: \(t\) can be replaced by \(\tilde{t} \equiv F(t)\), which is uniformly distributed. Of course, all other variables should be redefined accordingly.
(A1) \( U^R \) is twice continuously differentiable, unimodal with a peak at 0 and concave on \( \mathbb{R} \).

(A2) \( U^S(\cdot, \cdot) \) is continuous, unimodal at each \( t \) on \( \mathbb{R} \); \( U^S(x, t) \) is strictly increasing [decreasing] in \( x \) for \( x < t \) \([x > t]\).

(A3) If a Sender type \( t \) is indifferent between outcomes \( x_1 \) and \( x_2 > x_1 \), then higher types than \( t \) prefer \( x_2 \) and lower types prefer \( x_1 \).

Let the outcomes \( \lambda(t, U^S(\delta)) < t \) and \( \rho(t, U^S(\delta)) > t \) be the indifference points to the left respectively right of \( t \) with respect to the disagreement point, i.e.

\[
U^S(\lambda(t, U^S(\delta)), t) = U^S(\rho(t, U^S(\delta)), t) = U^S(\delta, t).
\]

Let \( \lambda^{-1}(x, U^S(\delta)) \) and \( \rho^{-1}(x, U^S(\delta)) \) be the inverse functions of \( \lambda \) and \( \rho \) with respect to \( t \). Finally, we assume that

(A4) \( \lambda \) and \( \rho \) exist and are twice continuously differentiable and strictly increasing in \( t \).

(A5) \( \frac{\partial}{\partial x} \lambda^{-1} \geq \frac{\partial}{\partial t} \rho^{-1} \), \( \frac{\partial^2}{\partial x^2} \lambda^{-1} \leq 0 \) and \( \frac{\partial^2}{\partial x \partial U^S(\delta)} \lambda^{-1} > 0 \).

If it is clear that we talk about a particular game with fixed \( U^S(\delta) \), we will suppress the dependency on \( U^S(\delta) \) and write \( \lambda(t), \rho(t), \lambda^{-1}(x) \) and \( \rho^{-1}(x) \). A simple condition on the Sender’s preferences such that they satisfy (A2)-(A5) is the following:
\( U^S(x,t) \) can be written as a function \( f(t-x) \), for all \( x \in \mathbb{R}, \ t \in [0,1] \), where \( f \) is continuously differentiable, strictly increasing in \( \mathbb{R}_- \), strictly decreasing in \( \mathbb{R}_+ \) and for all \( y \in \mathbb{R} \) there is a \( z > 0 \) such that \( f(-z) < y \) and \( f(z) < y \). Finally, we require \( U^S(\delta) < f(0) \).\(^{16}\)

Our results will often hold even when the Receiver’s utility is not concave and (A5) does not hold, but these assumptions will greatly facilitate the construction of equilibria in Proposition 2.2. In particular, let \( \rho^R \) be the point \( x \) in \( \mathbb{R} \) where the Receiver is indifferent between \( x \) and \( d \) and define the function

\[
(2.1) \quad h(x) \equiv \lambda^{-1}(x)(U^R(x) - U^R(\delta)) + \left(\lambda^{-1}(x) - \rho^{-1}(x)\right)U^R'(x)
\]

where a prime (‘) denotes a derivative with respect to \( x \).

(A4) and (A5) imply that \( h \) is strictly decreasing on \([0, \rho^R]\) for all values of \( U^R(\delta) \) and \( U^S(\delta) \).\(^{17}\) Since \( h(\rho^R) < 0 \), this implies that for all \( U^R(\delta) \) and \( U^S(\delta) \) there is an \( \overline{\rho} \in [0, \lambda(1)] \) such that

\[
(2.2) \quad \begin{align*}
    h(x) &> 0 \text{ for all } x \in [0, \overline{\rho}] \text{ and } \\
    h(x) &< 0 \text{ for all } x \in (\overline{\rho}, \lambda(1)]
\end{align*}
\]

(2.2) imposes regularity on the Receiver’s best response. Suppose the Receiver infers from a message that a type lies in the interval \([\underline{t}, \overline{t}]\). As long as \( a < \min\{\lambda(\underline{t}), \rho(\underline{t})\} \), increasing \( a \) involves the tradeoff between decreasing \( U^R(a) \) if \( a \) is accepted and increasing the acceptance probability by increasing the

\(^{16}\) Due to the invariance of games to affine payoff transformations, it is actually only required that \( U^S(x,t) \) can be written as a function \( a + b \cdot f(t-x) \), where \( b > 0 \) and \( f \) should adhere to the conditions specified (with \( U^S(\delta) < a + b \cdot f(0) \)).

\(^{17}\) Under (A4) and (A5), all terms in \( h'(x) = \lambda^{-1}\nu(x)(U^R(x) - U^R(\delta)) + \lambda^{-1}(x)U^R'(x) + \left(\lambda^{-1}(x) - \rho^{-1}(x)\right)U^R''(x) \) are negative on \([0, \rho^R]\).
highest type that accepts \( a \). (2.2) will ensure that there is a point \( \bar{x} \) independent of \( t \) and \( \bar{t} \) such that (in a few important cases) for \( a < \min\{\bar{x}, \lambda(\bar{t}), \rho(t)\} \) it pays to increase \( a \) and for \( a > \min\{\bar{x}, \lambda(\bar{t}), \rho(t)\} \) it does not.

We follow Matthews (1989) in the refinement of the perfect Bayesian equilibrium that we employ. First, we restrict the Receiver to play pure strategies. Second, we require that also Sender types who plan to veto any equilibrium action send a message inducing an action \( a \) that maximizes \( U^S(a,t) \). This refinement is motivated on the basis of Selten’s (1975) trembling hand perfection argument: the Sender considers that she might tremble with a small probability and accept the Receiver’s proposed action.

Let for any set \( S \), \( \Delta S \) denote the set of probability distributions on \( S \). A strategy for the Sender consists of a message function \( \mu : T \rightarrow \Delta M \) and an acceptance probability function \( \nu : \mathbb{R} \times T \rightarrow [0,1] \). A strategy of the Receiver is a function \( \alpha : M \rightarrow \mathbb{R} \). Let \( \sigma \equiv \{\mu, \nu, \alpha\} \) be a strategy profile and \( \Sigma \) the set of all strategy profiles. The Receiver has correct prior beliefs \( \beta^0 \). Finally, the Receiver has posterior beliefs \( \beta : M \rightarrow \Delta T \). An equilibrium \( \sigma^* = \{\mu^*, \nu^*, a^*, \beta^*\} \) is characterized by the following four conditions:

\[
\begin{align*}
\text{(i) } & \quad \alpha^*(m) \in \arg \max_{\alpha \in \mathbb{R}} E_t[(U^R(a) - U^R(\delta))\nu(a,t) | \beta^*] \text{ for all } m \\
\text{(ii) } & \quad m \in \arg \max_{m \in M} U^S(\alpha^*(m), t) \text{ for all } m \text{ in the support of } \mu^*(t) \text{ for all } t \\
\text{(iii) } & \quad \nu^*(a,t) = 1 \text{ if } U^S(a,t) > 0 \text{ and } \nu^*(a,t) = 0 \text{ if } U^S(a,t) < 0 \\
\text{(iv) } & \quad \beta^* \text{ is derived from } \mu^* \text{ and } \beta^0 \text{ using Bayes’ rule whenever possible}
\end{align*}
\]

We say that a type \( t \) induces action \( \hat{a} \), and write \( a(t) = \hat{a} \), if \( \alpha(t) = \hat{a} \) for all messages \( m \) that type sends.

### 2.3.2 Results

As in Matthews’ (1989) and Crawford and Sobel’s (1982) cheap talk game, all equilibria of our game are partition equilibria.
Definition 2.1 An equilibrium $\sigma^*$ is called a partition equilibrium if there is a partition $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$ of the type space such that each type in $(t_{i-1}, t_i)$ induces action $a_i$ with $a_1 < a_2 < \ldots < a_n$.

Note that any partition equilibrium outcome can be characterized by the actions $a_1 < a_2 < \ldots < a_n$ the Receiver proposes in equilibrium. Due to (A3), the partition is then fixed by $U^S(a_i, t_i) = U^S(a_{i+1}, t_i)$. The number of actions $n$ is called the size of the equilibrium.

Proposition 2.1 Any equilibrium of the cheap talk game is a partition equilibrium.

The intuition behind the proof is that the conflict of interest between the Sender and the Receiver puts a limit on information transmission. After all, the Receiver would prefer to propose $\max\{0, \lambda(t)\}$ if he knew the Sender’s type, whereas the Sender prefers to receive proposal $t$. This means that the highest type (supremum) inducing an action $a_i > 0$ must get zero payoff in equilibrium, because otherwise the Receiver can do better by proposing a lower action instead of $a_i$. This means that $a_{i+1}$ cannot be smaller than $\rho(\lambda^{-1}(a_i))$, because otherwise the highest type inducing $a_i$ would prefer $a_{i+1}$ over $a_i$. Hence, there is a minimum distance between any two strictly positive equilibrium actions. Since the Receiver never makes a proposal higher than $\rho^R$, this means that the equilibrium action set is finite. As Senders elicit the action closest to their type, this results in a well-ordered partition.

We model an increase in a player’s bargaining power as an increase in her disagreement point payoff. Power influences maximum information transmission:

Proposition 2.2 (i) If the Receiver’s bargaining power increases, the maximum size of the equilibrium decreases. (ii) If the Sender’s bargaining power increases, the maximum size of the equilibrium increases.
Hence, the higher the bargaining power of the Sender relative to that of the Receiver, the more information that can be transmitted in equilibrium. In the proof, we construct a maximum size equilibrium. Using a similar reasoning as in the proof of Proposition 2.1, we show that in equilibrium \( \lambda(t_{k-1}) < a_k \leq \lambda(t_k) \) for all \( k = 2, \ldots, n \) (Lemma 2.1) and \( \rho(t_{k-1}) \leq a_k \) for \( k = 3, \ldots, n \) (Lemma 2.2). In the most interesting case, when \( \lambda^{-1}(0) < \rho^{-1}(\overline{x}) \), at least a size-3 equilibrium exists. Furthermore, the highest possible value an equilibrium action can take is then \( \overline{x} \), the point where \( h(x) \) is zero. We can in this case construct a maximum size equilibrium by

- Setting the highest equilibrium action \( a_n \) equal to \( \overline{x} \);
- Iteratively adding equilibrium actions \( a_{k-1} = \lambda(\rho^{-1}(a_k)) \) until it ‘does not fit anymore’ \( (\lambda(\rho^{-1}(a_k)) < 0) \);
- Setting \( a_1 = 0 \).

Increasing the Receiver’s bargaining power decreases \( \overline{x} \), so that the number of equilibrium actions that ‘fit’ in this equilibrium becomes smaller (or remains the same). Increasing the Sender’s bargaining power firstly increases \( \overline{x} \). In addition, increasing the Sender’s power decreases the distance between equilibrium actions because the function \( a - \lambda(\rho^{-1}(a)) \) becomes smaller.

There are two features of the relation between power and information transmission that deserve further attention. First, Senders who induce 0 could reveal their type. We can use the number of equilibrium actions as a proxy for information transmission in so far as all Sender types who induce an equilibrium action \( a \) send the same message. For each equilibrium outcome, always an equilibrium exists where this holds for all types. However, also pay-off equivalent equilibria exist where types inducing \( a_1 = 0 \) reveal their type. The reason is that they are protected by the fact that the Sender does not want to propose an action below 0. This means that Senders who are closely aligned to the Receiver can fully separate. In addition, this implies that if the Sender has no power (if \( \lambda(1) \leq 0 \)), then an equilibrium exists with full separation. In this case the Receiver always proposes 0 and Senders can send any message they want in
equilibrium. Observe that they do not have an actual incentive to reveal their type.

Second, complete power in the Sender’s hands may lead to less information transmission. If the Sender becomes very powerful ($U^S(\delta)$ approaches $U^S(t,t)$), her incentive to communicate becomes small. In our model this is not relevant as long as $U^S(\delta) \leq U^S(t,t)$. However, one can imagine a model where sending a message has a small positive cost. In this case, increasing the power of the Sender will increase information transmission until the costs of communicating outweigh the benefits.

### 2.4 Conclusion

In this chapter, we examined how power shapes communication under information asymmetry. Our interest was to explore the levels of clarity that are likely to arise between members of groups with different levels of power. We expect cultural patterns to evolve to a strategic equilibrium over time and, hence, we investigated whether there is a strategic relationship between power and clarity. In a game-theoretic bargaining model, we showed that clarity is indeed a privilege of the powerful. When negotiating an outcome, an informed bargainer with (relatively) little bargaining power cannot afford to reveal too much information, as that can be used against her. How much information can be transmitted depends crucially on the relative power of the informed party, the Sender: less information can be transmitted if either the Sender’s power decreases or the Receiver’s power increases.

We see our one-dimensional model with one-sided asymmetric information as a proof of principle providing a more general intuition. The conflict of interest between the two parties imposes an upper bound on information transmission in equilibrium; the informed party cannot reveal too much information, as information allows the uninformed party to exploit her. Crucially, the informed party’s bargaining power limits how much she can be exploited and hence enhances information transmission in equilibrium. In contrast, bargaining power of the uninformed party limits the range of mutually attractive actions where
information transmission is meaningful. Crawford & Sobel (1982) found that information transmission is determined by the alignment of interests when the Sender has no influence on the outcome. We have identified another key determinant of information transmission if the Sender does have some influence on the outcome: bargaining power.

In addition to providing a proof-of-principle, our analysis has testable implications for communication and outcomes. In Chapter 3, we test our predictions in a controlled laboratory experiment and find that the relative power of the Sender increases information transmission. Furthermore, the type of bargaining situations we study often occur in the field.

One application concerns contract negotiations between employers and employees. Here, asymmetric information and bargaining power play a significant role. One can think of an employee’s preference for the work-life balance (salary versus flexibility) or the type of activities she is required to do. For instance, when a department negotiates with a potential new professor about her administrative and teaching duties, the preference of the professor for administration versus teaching are typically unknown. One implication of our analysis is that when the employee has more bargaining power, she will be able to convey her preferences more precisely. As a measure of bargaining power one could use the level of skill of employees or the unemployment rate in a given sector and/or year. Our model predicts that as information transmission increases, the variety of outcomes also increases. As a consequence, our model predicts that the variety of labor contracts in a specific job market should be increasing with the employment rate (in the sector) and the skill-level required for the job.

Another application where our model has testable implications is negotiations between a competition authority and two firms planning a merger. If the merger creates or strengthens market power in the relevant market, the competition authority can demand remedies, such as requiring the firms to sell some production-lines to a third party. Firms always want as few remedies as possible. Preferences of competition authorities are less clear, as they have to weigh economies of scale against market power.\(^\text{18}\) Competition authorities provide

\(^{18}\) Another possible trade-off for competition authorities concerns collusion (Compte, Jenny & Rey, 2002). A merger reduces the number of competing firms, which can make collusion easier.
information about their preferences to firms before they submit their final proposal, often already in the pre-notification phase. Our model predicts that the variability of the proposals a competition authority receives is increasing in her power. A competition authority with little power always gets the same kind of proposal (across comparable cases), for instance a proposal without remedies. A competition authority with more power can expect to receive proposals that sometimes include remedies and sometimes not. Indicators exist about the strength of competition authorities, such as the OECD’s Competition Law and Policy (CLP) indicators (Høj, 2007) or those developed by Voigt (2006) for a broader set of countries. Such indicators include the formal and factual independence of competition authorities. These proxies for bargaining power could be related to how often final proposals include remedies (or even to the variety of remedies included). A relevant comparison would be between competition authorities in (old) EU member states and the US with those in Latin America or Eastern Europe.19

2.5 Appendix: Proofs

Proof of Proposition 2.1 Let \( \sigma^* \) be an equilibrium. First, suppose the Receiver plays at least three actions in equilibrium and let \( a < a' \) be two strictly positive equilibrium actions. Let \( \bar{t}(a) \) be the supremum of types that induce action \( a \). Then \( 0 < a \leq \lambda(\bar{t}(a)) \), because otherwise the Receiver would be better off by playing \( \lambda(\bar{t}(a)) \) instead of \( a \). Furthermore, \( a' \geq \rho(\bar{t}(a)) \), because otherwise \( U^S(a',\bar{t}(a)) > U^S(a,\bar{t}(a)) \). This means that \( a' - a \geq \rho(\bar{t}(a)) - \lambda(\bar{t}(a)) \). Consequently, an upper bound on the number of equilibrium actions is \( 1 + 1 / \eta \), where \( \eta = \min_{t \in [0,1]} \{ \rho(t) - \lambda(t) \} \). This means that the set of equilibrium actions, \( A^* \), is finite. Hence, we can write \( A^* = \{ a_1, a_2, ..., a_n \} \) with \( a_1 < a_2 < ... < a_n \).

However, if a merger involves the largest firm, it can also increase asymmetries in capacity constraints, making collusion more difficult.

19 For comparative work on competition authorities in Latin America see Schatan & Rivera (2008) and Qaqaya & Lipurse (2008); For competition authorities in Europe see Cseres (2010).
Second, (A2)-(A3) imply that for each consecutive action pair \( a_k, a_{k+1} \) a triple of types \( t_{k-1} < t_k < t_{k+1} \) exists such that

\[
U^S(a_k, t_k) = U^S(a_{k+1}, t_k) \quad \text{and} \quad \quad \quad \quad \quad \quad \quad a_t \in \arg \max_{a \in A} U^S(a, t) \quad \text{iff} \quad t \in [t_{l-1}, t_l] \quad \text{for} \quad l = k, k+1
\]

Consequently, a partition of the type space \( 0 = t_0 < t_1 < ... < t_{n-1} < t_n = 1 \) exists such that for each \( k \) all types in \( (t_{k-1}, t_k) \) induce the same action \( a_k \) and that \( k \neq l \) implies that \( a_k \neq a_l \). \( Q.E.D. \)

The following two lemmas are useful for the proof of Proposition 2.2.

**Lemma 2.1** In equilibrium, \( \lambda(t_{k-1}) < a_k \leq \lambda(t_k) \) for all \( k = 2, ..., n \).

**Proof.** Note that by construction, \( a_1 < a_2 < ... < a_n \). It must be the case that \( a_k > a_l \geq 0 \) for all \( k = 2, ..., n \). Moreover, \( a_k > \lambda(t_k) \) cannot occur in equilibrium, because if the Sender type is in the interval \( [t_{k-1}, t_k] \), the Receiver is better off by offering \( a = \lambda(t_k) \) instead of \( a = a_k \). Finally, \( a_k > \lambda(t_{k-1}) \) because otherwise none of the types in \( [t_{k-1}, t_k] \) will accept \( a_k \). \( Q.E.D. \)

**Lemma 2.2** \( \rho(t_{k-1}) \leq a_k \) for \( k = 3, ..., n \).

**Proof.** The proof is by contradiction. Lemma 2.1 shows that \( a_{k-1} \leq \lambda(t_{k-1}) \), so that type \( t_{k-1} \)’s utility is equal to the utility of the disagreement point if she induces \( a_{k-1} \). Suppose that \( \rho(t_{k-1}) > a_k \). Then type \( t_{k-1} \) is strictly better off inducing \( a_k \) instead of \( a_{k-1} \). This constitutes a contradiction, because types just below \( t_{k-1} \) would strictly prefer sending \( a_k \) instead of \( a_{k-1} \), while they induce \( a_{k-1} \) in equilibrium. \( Q.E.D. \)

**Proof of Proposition 2.2** We first prove \((i)\). If \( \lambda(1) \leq 0 \), the game only has a pooling equilibrium in which \( t_0 = 0, \ t_1 = 1 \), and \( a_1 = 0 \). Otherwise, if
\[ \lambda^{-1}(0) \geq \rho^{-1}(x) \], the maximum size equilibrium has size 2 with \( t_0 = 0, \ t_2 = 1 \), \( a_1 = 0 \), and \( t_1 \) and \( a_2 \) simultaneously solve

\[
a_2 \in \arg\max_{a \in [0,1]} \left( \left( U_s^R(a) - U_s^R(\delta) \right) \left( \lambda^{-1}(a) - \max\{t, \rho^{-1}(a)\} \right) \right) \text{ and } U_s^S(t_1, a_2) = U_s^S(t_1, 0). \]

Finally, if \( \lambda^{-1}(0) < \rho^{-1}(x) \), the maximum size equilibrium has at least size 3 and can be constructed using the following algorithm:

1. Let \( \pi \) be some natural number. Define \( t_\pi = 1, \ t_{\pi-1} = \rho^{-1}(x), \ a_\pi = \pi, \) and assign value \( \pi - 2 \) to counter \( k \).
2. Define \( t_k \in (0,1] \) such that \( \rho(t_k) = \lambda(t_{k+1}) \). If such a \( t_k \) does not exist, go to step 3. Otherwise, \( a_{k+1} = \rho(t_k), \ k \leftarrow k - 1 \) and return to step 2.
3. Relabel \( t_{k+1}, \ldots, t_\pi \) and \( a_{k+2}, \ldots, a_\pi \) such that \( k \leftarrow 1, \ k + 1 \leftarrow 2, \ldots, \pi \leftarrow n \).

Let \( t_0 = 0, \ a_1 = 0, \ a_2 = \lambda(t_2) \) and define \( t_1 \) such that \( U_s^S(t_1, 0) = U_s^S(t_1, a_2) \).

In step 2, \( a_{k+1} \) follows from the requirement that actions maximize expected utility \( (U_s^R(x) - U_s^R(\delta))P\left\{ U_s^S(x, t) > U_s^S(\delta, t) \mid t \in [t_{k-1}, t_k] \right\} \), which in our case implies

\[
a_{k+1} \in \arg\max_{a \in [0,1]} \left( \left( U_s^R(a) - U_s^R(\delta) \right) \left( \min\{t_{k+1}, \lambda^{-1}(a)\} - \max\{t_k, \rho^{-1}(a)\} \right) \right)
\]

Due to Lemma 2.1, for all \( k = 2, \ldots, n-1 \) it must hold that \( a_{k+1} \leq \lambda(t_{k+1}) = \rho(t_k) \), and hence the condition reduces to

\[
a_{k+1} \in \arg\max_{a \leq \lambda(t_{k+1})} \left( U_s^R(a) - U_s^R(\delta) \right) \left( \lambda^{-1}(a) - t_k \right).
\]

\[ \text{These two equations have a solution. Let } \tilde{t}_k(a) \text{ be the point where } U_s^S(a, t_k) = U_s^S(0, t_k). \]\n
Then there exists a continuous function \( a_k(a) \) such that \( a_k(a) \in \arg\max_{a \in [0,1]} \left( U_s^R(a) - U_s^R(\delta) \right) \left( \lambda^{-1}(a) - \max\left\{ \tilde{t}_k(a), \rho^{-1}(a) \right\} \right) \). Observe that \( a_2(0) \geq 0 \) and \( a_2(1) \leq \lambda(1) \). Hence \( a_2(a) \) has a fixed point on \([0,1]\).
If \( a_{k+1} = \lambda(t_{k+1}) \) for \( k \geq 2 \), \( a_{k+1} \) maximizes expected utility if

\[
\lambda^{-1}(a)\big(U^R(a) - U^R(\delta)\big) + (\lambda^{-1}(a) - t_k)U^{RF}(a) \geq 0 \text{ for } a \leq \lambda(t_{k+1})
\]

From (2.2) it follows that

\[
\lambda^{-1}(a)\big(U^R(a) - U^R(\delta)\big) + (\lambda^{-1}(a) - t_k)U^{RF}(a) = h(a) + (\rho^{-1}(a) - t_k)U^{RF}(a) \geq h(a) > 0 \text{ for } a < \rho(t_k) = \lambda(t_{k+1}) < \bar{x}.
\]

Similarly, we can justify setting \( a_\pi = \bar{x} \) in step 1 by observing that \( \bar{x} \leq \lambda(1) \) and setting \( a_2 = \lambda(t_2) \) in step 3 by observing that \( a_{k+1} \leq \lambda(t_{k+1}) < \rho(t_k) \).

In addition, no equilibrium exists with an \( a_n > \bar{x} \). Suppose, to the contrary that \( a_n > \bar{x} \). Observe that \( a_n \leq \lambda(1) = \lambda(t_n) \). Furthermore, \( \rho(t_{n-1}) \leq a_n \) by Lemma 2.2, since the equilibrium size is at least 3. We can show that the first order condition for \( a_n \) to be optimal cannot be met, since \( h(a) < 0 \) for \( a \in (\bar{x}, a] \):

First, let \( \rho(t_{n-1}) < a_n \). Now, \( (\lambda^{-1}(a) - \rho^{-1}(a)) \)

\[
\big(U^R(a) - U^R(\delta)\big) + (\lambda^{-1}(a) - \rho^{-1}(a))U^{RF}(a) = h(a) - \rho^{-1}(a)\big(U^R(a) - U^R(\delta)\big) < 0 \text{ for } a \in (\max\{\bar{x}, \rho(t_{n-1})\}, a_n].
\]

(Observe that \( -\rho^{-1}(a)\big(U^R(a) - U^R(\delta)\big) < 0 \) for \( a \in (\max\{\bar{x}, \rho(t_{n-1})\}, a_n]\).) Second, let \( \rho(t_{n-1}) = a_n \). In this case, the derivative for the expected utility for \( a \in [\lambda(t_{n-1}), a_n] \) is

\[
\lambda^{-1}(a)\big(U^R(a) - U^R(\delta)\big) + (\lambda^{-1}(a) - t_{n-1})U^{RF}(a) = h(a) + (\rho^{-1}(a) - t_{n-1})U^{RF}(a).
\]

Since the first term is strictly negative and the second term is 0 for \( a = a_n = \rho(t_{n-1}) \), there is an interval \( (a_n - \varepsilon, a_n] \) where the derivative is negative. Hence, \( a_n \) cannot be optimal if it is larger than \( \bar{x} \).

Because \( \lambda(t) \) and \( \rho(t) \) are monotonic by assumption, this algorithm results in the tightest partition which satisfies the equilibrium properties from Lemma 2.1 and Lemma 2.2 so that it implements an equilibrium of the highest possible size.

Suppose the Receiver’s bargaining power \( U^R(\delta) \) increases from \( \Pi \) to \( \bar{\Pi} \). Now
\[ \lambda^{-1'}(\bar{x}) \left( U^R(\bar{x}) - \Pi \right) + (\lambda^{-1}(\bar{x}) - \rho^{-1}(\bar{x})) U^R(\bar{x}) \]

\[ = \lambda^{-1'}(\bar{x}) \left( U^R(\bar{x}) - \Pi \right) + (\lambda^{-1}(\bar{x}) - \rho^{-1}(\bar{x})) U^R(\bar{x}) + \lambda^{-1'}(\bar{x})(\Pi - \bar{\Pi}) < 0. \]

Hence, for the new threshold \( \bar{x} \) following from (2.2), it holds that \( \bar{x} < \bar{x} \). If \( \lambda^{-1}(0) \geq \rho^{-1}(\bar{x}) \), observe that \( \rho^{-1}(\bar{x}) \leq \rho^{-1}(\bar{x}) \leq \lambda^{-1}(0) \). If \( \lambda^{-1}(0) < \rho^{-1}(\bar{x}) \), then with the above algorithm we can get a maximum size equilibrium with partition \( 0 = \tilde{t}_0 < \tilde{t}_1 < \cdots < \tilde{t}_n = 1 \). Furthermore, when running the above algorithm, all new threshold types will be lower than for the original game \( \tilde{t}_{n-k} < t_{n-k} \) for all \( 1 \leq k < \tilde{n} \). Therefore, the maximum size of the equilibrium decreases if the Receiver’s bargaining power increases.

Now we proof (ii). Suppose that the Sender’s bargaining power \( U^S(\delta) \) increases from \( \Sigma \) to \( \hat{\Sigma} \). If \( \lambda(1, \Sigma) \leq 0 \), observe that \( \lambda(1, \hat{\Sigma}) > \lambda(1, \Sigma) \). If \( \lambda^{-1}(0, \Sigma) \geq \rho^{-1}(\bar{x}, \Sigma) \), note that \( \lambda^{-1}(0, \hat{\Sigma}) - \rho^{-1}(\bar{x}, \Sigma) < \lambda^{-1}(0, \Sigma) - \rho^{-1}(\bar{x}, \Sigma) \), and \( \rho^{-1}(\bar{x}, \Sigma) > \rho^{-1}(\bar{x}, \Sigma) \), since \( \lambda^{-1}(0, \hat{\Sigma}) < \lambda^{-1}(0, \Sigma) \) and \( \rho^{-1}(\bar{x}, \hat{\Sigma}) > \rho^{-1}(\bar{x}, \Sigma) \).

Finally, let \( \lambda^{-1}(0, \Sigma) < \rho^{-1}(\bar{x}, \Sigma) \). Analogous to the proof of (i), we can show that increasing the Sender’s bargaining power results in a higher \( \hat{x} \):

\[ \lambda^{-1'}(\bar{x}, \hat{\Sigma}) \left( U^R(\bar{x}) - U^R(\delta) \right) + (\lambda^{-1}(\bar{x}, \hat{\Sigma}) - \rho^{-1}(\bar{x}, \hat{\Sigma})) U^R(\bar{x}) \]

\[ = \lambda^{-1'}(\bar{x}, \Sigma)) \left( U^R(\bar{x}) - U^R(\delta) \right) + (\lambda^{-1}(\bar{x}, \Sigma)) - \rho^{-1}(\bar{x}, \Sigma)) \right) U^R(\bar{x}) \]

\[ + (\lambda^{-1'}(\bar{x}, \hat{\Sigma}) - \lambda^{-1'}(\bar{x}, \Sigma)) \left( U^R(\bar{x}) - U^R(\delta) \right) \]

\[ + \left( \lambda^{-1}(\bar{x}, \hat{\Sigma}) - \lambda^{-1}(\bar{x}, \Sigma) - \rho^{-1}(\bar{x}, \hat{\Sigma}) + \rho^{-1}(\bar{x}, \Sigma) \right) U^R(\bar{x}) \]

\[ > 0 \]

Observe for the third term in the middle expression that \( \lambda^{-1'}(\bar{x}, \hat{\Sigma}) > \lambda^{-1'}(\bar{x}, \Sigma) \) by (A5); and for fourth term that \( \lambda^{-1}(\bar{x}, \hat{\Sigma}) < \lambda^{-1}(\bar{x}, \Sigma) \) and \( \rho^{-1}(\bar{x}, \hat{\Sigma}) > \rho^{-1}(\bar{x}, \Sigma) \).

32
due to the shrinking interval of points the Sender accepts when her power increases.

Hence, with the above algorithm we can get a new maximum size equilibrium partition $0 = \tilde{t}_0 < \tilde{t}_1 < \cdots < \tilde{t}_n = 1$. There are now two reasons why all threshold types will be higher than for the original game ($\tilde{t}_{n-k} > t_{n-k}$ for all $1 \leq k < n$), possibly resulting in extra equilibrium actions. First, the highest equilibrium action can be higher, since $\tilde{\bar{\tau}} > \bar{\tau}$. Second, the equilibrium actions will be closer together, since for each type it holds that $\lambda(t, \Sigma) < \lambda(t, \hat{\Sigma})$ and $\rho(t, \Sigma) > \rho(t, \hat{\Sigma})$. Hence, the number of equilibrium actions in the maximum size equilibrium is (weakly) higher when $U^S(\delta) = \hat{\Sigma}$ than when $U^S(\delta) = \Sigma$. Q.E.D.

Proof Equilibrium Characterization Example section 2.2. Observe that in the example assumptions (A2)-(A5) are satisfied. By Proposition 2.1, all equilibria are partition equilibria. Condition (iv) can be shown to follow Lemma 2.1 and Lemma 2.2, and the other conditions follow from the proof of Proposition 2.2. The characterization of the maximum equilibrium-size is a direct result from conditions (i)-(v). Observe that if $d^S < 1$, $\bar{n} = \max \left\{ 2, \left\lfloor \frac{\bar{\tau}}{2d^S} + 1 \right\rfloor \right\}$. 
3.1 Introduction

Crawford & Sobel (1982) showed how meaningful costless communication between an informed Sender and an uninformed Receiver can be supported in equilibrium. Their seminal paper inspired many applications ranging from the presidential veto (Matthews, 1989), legislative committees (Gilligan & Krehbiel, 1990) and political correctness (Morris, 2001) to double auctions (Matthews & Postlewaite (1989); Farrell & Gibbons (1989)), stock recommendations (Morgan & Stocken, 2003) and matching markets (Coles, Kushnir & Niederle, 2010). These cheap talk games are characterized by multiple equilibria which differ crucially in their prediction about how much information will be transmitted. Standard signaling refinements such as Kohlberg & Merten’s strategic stability (1986) have no bite in this setting, as messages are costless. Hence, this raises the important issue of equilibrium selection. As of yet, however, no satisfying refinement exists that works well across a wide range of cheap talk games.22

To overcome the selection problem, Farrell (1993) pioneered the approach of what we call credible deviations. Farrell observed that, in contrast to what is assumed in standard game theory, communication in real life is based on a pre-existing natural language.23 Hence, it is possible to send out-of-equilibrium messages that will be understood, although they will not necessarily be believed. Farrell proposes a criterion under which such out-of-equilibrium messages, or neologisms, are credible and, hence, should be believed by a Receiver. Arguably, an equilibrium is not stable if some Senders can send a credible neologism that would induce a rational Receiver to deviate from equilibrium. Equilibria that do

---

21 This chapter is based on De Groot Ruiz, Offerman & Onderstal (2011b).
22 For a comprehensive review of Sender-Receiver games, see Sobel (2010).
not admit credible neologisms are ‘neologism proof.’ Alas, neologism proofness tends to be too effective and regularly eliminates all equilibria. Matthews, Okuno-Fujiwara & Postlewaite (1991) address important conceptual issues with neologism proofness and propose three alternative accounts of what constitutes a credible deviation. Unfortunately, these ‘announcement proofness’ criteria also fail to be predictive in many games (such as the Crawford-Sobel (1982) model). Partly for this reason, several other types of concepts have been proposed that distinguish between stable and unstable equilibria (or profiles), such as Partial Common Interest (PCI) (Blume, Kim & Sobel, 1993), the recurrent mop (Rabin & Sobel, 1996) and No Incentive To Separate (NITS) (Chen, Kartik & Sobel, 2008). These criteria often work well in specific settings, but fail to discriminate successfully across a wider range of cheap talk games.

We take a different tack: we propose a criterion that is based on credible deviations but allows for a continuous instead of a binary stability concept. Our first conjecture is that credible deviations matter, in the sense that an equilibrium which does not admit credible deviations is more stable than one that does. Our second conjecture is that credible deviations matter gradually: of two equilibria that allow for credible deviations, the one with smaller deviations will predict better. A binary stability criterion is appropriate for rational agents, but may unnecessarily lose predictive power when applied to human behavior, which is seldom completely in (or out of) equilibrium.

We formalize this idea in the Average Credible Deviation Criterion (ACDC). According to ACDC, the behavioral stability of an equilibrium is a decreasing function of its Average Credible Deviation (ACD), a measure of the frequency and intensity of credible deviations. The ACD measures the mass of types that can credibly deviate and the size of those induced deviations (as measured by the difference in Sender payoff between the equilibrium and deviating action). Comparable equilibria will perform better if they have a lower ACD on this account. In particular, we call an equilibrium that minimizes the ACD in a game an ‘ACDC equilibrium.’ This allows us to select equilibria, even in games where no equilibrium is completely stable.

We show that an ACDC equilibrium exists under general conditions. In addition, ACDC makes meaningful predictions in a variety of settings that have
previously been analyzed theoretically and experimentally. We look at how ACDC performs in a range of discrete games analyzed by Blume, DeJong, Kim & Sprinkle (2001) in the laboratory and find that ACDC organizes the main features of the data. We illustrate the predictive capability of ACDC in a veto threats model with a large equilibrium set where existing criteria fail. This game belongs to the class of veto threats games introduced in Chapter 2 and we show here that ACDC selects a unique most informative equilibrium in this class. In Chapter 4, we study this class of games experimentally and find that the results corroborate the predictions of ACDC. Furthermore, we show that ACDC selects the unique maximum size equilibrium in the leading uniform quadratic case of the Crawford-Sobel game for a large range of bias parameters. Until now, only NITS (Chen, Kartik & Sobel, 2008) was able to select (this equilibrium) in the Crawford-Sobel setting. In addition, the maximum size equilibrium becomes more stable as the bias parameter becomes smaller according to ACDC, which is not predicted by existing criteria. Both results are qualitatively supported by experimental work on (discrete) Crawford-Sobel games (Dickhaut, McCabe & Mukherji (1995), Cai & Wang (2006) and Wang, Spezio & Camerer (2010)).

ACDC is meant to solve a practical problem: how to select the most plausible equilibrium in cheap talk games? It constitutes an intuitive, generally applicable and experimentally validated selection criterion. In addition to selecting the most plausible equilibrium, ACDC can be informative about how well this equilibrium will perform if it is not completely stable. A major advantage is that it can easily be applied in almost all cheap talk games and provides meaningful results. Wherever experimental evidence exists, the predictions of ACDC are in line with the data: it performs at least as well as other criteria, when they are predictive, and also makes predictions when other criteria are silent. We believe these characteristics make it a valuable contribution to the current literature.

This chapter has the following structure. In section 3.2, we motivate and introduce ACDC and compare it to neologism proofness and announcement proofness. In section 3.3, we illustrate how ACDC works in some simple discrete games. In section 3.4, we compare ACDC to other concepts in a veto threats game with a large equilibrium set. In section 3.5, we analyze the ACDC-properties of the uniform quadratic Crawford-Sobel model. In section 3.6, we
look at the performance of ACDC in experiments. Finally, section 3.7 concludes. All proofs are relegated to Appendix 3.8.

3.2 ACDC

ACDC is based on theories of credible deviations. To illustrate the idea behind credible deviations, we start with Game A, a simple game taken from Farrell (1993). A Sender sends a costless message $m$ to the Receiver, who then takes one of three actions: $a_1, a_2$ or $a_3$. Both players’ payoff functions depend on the Receiver’s action and the Sender’s type. The Sender’s type, drawn from $\{t_1, t_2\}$ with equal probability, is private information of the Sender. Table 3.1 shows the payoffs.

<table>
<thead>
<tr>
<th>$t_1 (\frac{1}{2})$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3, 3</td>
<td>0, 0</td>
<td>2, 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$t_2 (\frac{1}{2})$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0, 0</td>
<td>3, 3</td>
<td>2, 2</td>
</tr>
</tbody>
</table>

Notes: The left column shows the Sender’s type and between brackets the probability that it is drawn. The top row shows the Receiver’s actions. The remaining cells provide the Sender’s payoff in the first entry and the Receiver’s payoffs in the second entry.

The game has two (perfect Bayesian) equilibria: a pooling and a separating equilibrium.\(^{24}\) In the pooling equilibrium, all Senders use the same message strategy and the Receiver always proposes $a_3$, regardless of the message. In the separating equilibrium, $t_1$ sends a message inducing $a_1$ whereas $t_2$ sends a different message that induces $a_2$.\(^{25}\) The separating equilibrium seems more plausible, as it Pareto-dominates the pooling equilibrium. Still, standard signaling refinements in the vein of Kohlberg and Mertens’ (1986) strategic stability

\(^{24}\) An equilibrium outcome specifies which action(s) are played by the Receiver given each type. Each equilibrium outcome is induced by a whole class of essentially equivalent equilibria, which only differ in the messages used. For simplicity, we will refer to an equilibrium outcome simply by ‘equilibrium.’

\(^{25}\) We say a type $t$ induces an action $a$, if she always sends a message to which the Receiver always responds with action $a$. 

38
are silent in cheap talk games, since all messages are equally costless. For a similar reason, the (Agent) Quantal Response Equilibrium (McKelvey & Palfrey, 1998), which can often select equilibria in signaling games, is not predictive in cheap talk games.\(^{26}\)

To overcome the selection problem, Farrell (1993) pioneered the idea of credible deviations and introduced the neologism proofness criterion. Neologisms are out-of-equilibrium messages which are assumed to have a literal meaning in a pre-existing natural language. Farrell considers neologisms which literally say: “play action \( \tilde{a} \), because my type is in set \( N \).” Farrell deems a neologism credible if and only if \( (i) \) all types \( t \) in \( N \) prefer \( \tilde{a} \) to their equilibrium action \( a^*(t) \), \( (ii) \) all types \( t \) not in \( N \) prefer their equilibrium action \( a^*(t) \) to \( \tilde{a} \) and \( (iii) \) the best reply of the Receiver after restricting the support of his prior to \( N \) is to play \( \tilde{a} \). We will denote neologisms by \( \langle \tilde{a}, N \rangle \). According to Farrell, credible deviations lead rational players to deviate from equilibrium. An equilibrium is neologism proof, and stable on this account, if and only if it does not admit any credible neologism.

Matthews, Okuno-Fujiwara & Postlewaite (1991) identify two potential problems with this credibility criterion. First, it can sometimes be too strict, as it does not allow different deviating types to separate. Second, it is sometimes not strict enough. The credibility of a neologism is not affected by the existence of other credible neologisms, although the choice of the Sender about which neologism she sends potentially reveals information. To solve these problems they consider more elaborate credible deviations, called announcements. Like a neologism, an announcement specifies a type-subset \( N \) to which the deviating Sender claims to belong. An announcement also provides a strategy specifying for each types if and how they deviate. Let \( \mathcal{N}(t) \) be the set \( N \) type \( t \) claims to be in and \( \mathcal{N} \) the set of all \( \mathcal{N}(t) \). Let \( \beta_0 \) be the Receiver’s prior beliefs and \( \beta_0^N \)

---

\(^{26}\) The Agent Quantal Response Equilibrium (A-QRE) is the extensive form game variant of the Quantal Response Equilibrium. The problem in cheap talk games is that the A-QRE does not eliminate the pooling equilibrium, which is often implausible. The pooling equilibrium is always a limiting, principal branch A-QRE: For any rationality parameter \( \lambda \), there is an A-QRE where all Senders mix uniformly over the message space and the Receiver ignores all messages. As \( \lambda \) increases, the Senders strategy remains unchanged, and the Receiver’s best response smoothly approaches its actual best response to her prior.
her prior beliefs $\beta_0$ restricted to have support $N$. Finally, let $BR(\beta)$ be the Receiver’s best response set if she has beliefs $\beta$. An announcement is credible if (i) each deviating type prefers any action in $BR(\beta_0^{N(t)})$ to her equilibrium action, (ii) each deviating type prefers any action in $BR(\beta_0^{N(t)})$ to any action in $N_{tBR}^\beta$, (iii) each non-deviating type prefers her equilibrium action to any action in $\bigcup_{N\in N\setminus\{N(t)\}} BR(\beta_0^N)$, (iv) each non-deviating type prefers her equilibrium action to any action in $\bigcup_{N\in N} BR(\beta_0^N)$ and (v) for each deviating type no announcements exist satisfying (i)-(iii) where she earns more than in the current announcement.27 They also define a weaker and a stronger credibility criterion. A weakly credible announcement only satisfies (i)-(iii), whereas a strongly credible announcement also must propose an equilibrium. Matthews and coauthors express a preference for (ordinary) credible announcements.28

In game A, the pooling equilibrium admits two credible neologisms, $\{a_1,t_1\}$ and $\{a_2,t_2\}$. These deviations also form a credible announcement. Hence, the pooling equilibrium is neither neologism proof nor announcement proof. The separating equilibrium, on the other hand, must be neologism and announcement proof since all types receive their maximum payoff. Hence, neologism and announcement proofness provide a compelling strategic reason why rational players would play the separating equilibrium.

Unfortunately, neologism and announcement proofness are often too effective and eliminate all equilibria, as Game B in Table 3.2 illustrates. Now three types can be drawn, $t_1$ and $t_2$ each with probability $(1-\delta)/2$ and $t_3$ with probability $\delta$. The Receiver’s best response is to play $a_i$ if he knows the Sender is $t_i$. His best response is to play $a_i$ if he holds his prior beliefs and to play $a_4$ if he restricts his prior to have support $\{t_2,t_3\}$ (for $\delta < 1/2$). $t_1$ prefers $a_5$ over all other actions and hence would prefer to pool with the other types. $t_3$ prefers to be identified as herself, as she prefers $a_3$. $t_2$ prefers to pool with $t_3$ (as she prefers

27 We use a somewhat simpler definition than Matthews et al. for ease of exposition.

28 Another approach similar to credible neologisms is Myerson’s (1989) credible negotiation statements. Myerson is able to obtain a solution concept that guarantees existence but at the cost of assuming the presence of a mediator.
a_1) and the worst case for her is to be identified. The game has two equilibria. In the pooling equilibrium, all Senders induce a_5. In the semi-separating equilibrium, t_1 induces a_1, whereas t_2 and t_3 induce a_4. Neither equilibrium is neologism or announcement proof (if 0 < \varepsilon < 1). The pooling equilibrium admits the credible neologism \{a_1, \{t_2, t_3\}\}. The semi-separating equilibrium admits the credible neologism \{a_3, \{t_3\}\}. In this game, (weakly) credible announcements coincide with credible neologisms.  

<table>
<thead>
<tr>
<th>TABLE 3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GAME B</strong></td>
</tr>
<tr>
<td>t_1 (t_2)</td>
</tr>
<tr>
<td>t_2 (t_3)</td>
</tr>
<tr>
<td>t_3 (\delta)</td>
</tr>
</tbody>
</table>

*Notes: We use the same notation as in Table 3.1. 0 < \delta < \frac{1}{2} and 0 < \varepsilon < 1*

For entirely rational agents, the fact that neither equilibrium is stable might be all there is to be said. When explaining or predicting human behavior, however, we feel we can go further. Human behavior is hardly ever completely in or out of equilibrium, and by imposing a binary distinction between stable and unstable equilibria one may lose predictive power.

In game B, even though the separating equilibrium is not entirely stable, it seems more plausible than the pooling equilibrium if either t_3 is infrequent (\delta small) or t_3 has a very small incentive to deviate (\varepsilon small). If \delta is small, then the separating equilibrium will be upset with a small probability, whereas the pooling equilibrium will be upset almost half of the time. Similarly, if \varepsilon is small,  

---

29 Neither equilibrium admits strongly credible announcements, since the deviations profile does not comprise an equilibrium in itself. This is typical for many cheap talk games. Due to the strong credibility requirement of strongly credible announcements, in most games no equilibrium admits them.

30 We consider equilibrium, when applied to human behavior, to be most meaningful in a dynamic context, where members of a group interact frequently with different other members. In this context language evolves and behavior is shaped by strategic forces in the direction of equilibrium. For a one-shot game between rational individuals without social information, an approach based on rationalizability and some focal meaning of messages, such as that in Rabin (1990), may be appropriate.
then \( t_3 \) has a small incentive to deviate in the separating equilibrium and may choose to stick to it, lest she be misunderstood and get a payoff lower than she gets by sticking to equilibrium. Hence, we would expect to observe behavior close to the separating equilibrium more frequently than behavior close to the pooling equilibrium.

Our intuition is that the behavioral stability of an equilibrium is a decreasing function of the average intensity of the credible deviations it admits. This depends, firstly, on the mass of types that can credibly induce a deviation and, secondly, on the intensity of the deviation, measured by the incentive the Sender has to deviate. As a consequence, if the deviating mass and the induced deviations from equilibrium are small, the equilibrium is likely to be a good predictor of behavior. We formalize this intuition in our ACDC criterion.

We will now define ACDC for a general setting that allows for a discrete and continuous type set, and covers both pure advising and veto threats bargaining games. We consider the following two-player cheap talk game. The game is played by a Sender and a Receiver. Nature draws the Sender type \( t \) from distribution \( f \) on \( T \), where \( T \) is a compact metric space. The Sender then privately observes her type \( t \) and chooses a message \( m \in M \). After having observed the Sender’s message, the Receiver chooses an action \( a \in A \), where \( A \) is a compact metric space. \(^{31}\) After seeing the action, the Sender chooses between accepting \(( v = 1 )\) or rejecting \(( v = 0 )\) the action. If she rejects, the outcome is the disagreement point \( \delta \). If \( \delta \in A \), the game has an interior veto threat and otherwise it has an exterior veto threat. In cheap talk games without veto threats, the Sender is forced to accept the action so that \( v \) is always equal to 1. The outcome set is \( X = A \cup \{ \delta \} \). Let \( U^R : X \times T \rightarrow \mathbb{R} \) be the utility function of the Receiver \( U^S : X \times T \rightarrow \mathbb{R} \) that of the Sender. We assume both are bounded from above and below. For a set \( S \), let \( \Delta S \) denote the set of probability distributions on \( S \). Then, a strategy for the Sender consists of a function \( \mu : T \rightarrow \Delta M \) and a function \( \nu : A \times T \rightarrow [0,1] \) and a strategy of the Receiver is a function

\(^{31}\) This representation allows for \( T \) and \( A \) to be de facto discrete, by having \( U^R(t,x) \) and \( U^R(t,x) \) be constant on regions of the type and outcome space.
\(\alpha : M \rightarrow \Delta A\). Let \(\Sigma^S\) be the set of Sender strategies and \(\Sigma^R\) the set of Receiver strategies. Let \(\sigma \equiv \{\mu, \nu, \alpha\}\) be a strategy profile and \(\Sigma\) the set of all strategy profiles. \(m(t)\) will denote the random variable of the Sender’s message after learning \(t\) (determined by \(\mu\)) and \(a(m)\) will denote the random variable of the Receiver’s action after receiving \(m\) (determined by \(\alpha\)). Finally, let the Receiver have prior and posterior beliefs (a function \(\beta : M \rightarrow \Delta T\)) over the Sender’s type, \(\beta^0 = f\) and \(\beta(m)\) respectively.

The Sender’s expected payoff is \(EU^S(t, \mu, \nu \mid \alpha) = \int_{\Sigma^S} (\nu(a(m(t)), t)U^S(a(m(t)), t) + (1 - \nu(a(m(t)), t))U^S(\delta, t)) \, d\sigma.\) The Receiver’s expected payoff given her beliefs and message \(m\) is \(EU^R(\alpha, m \mid \nu, \beta) = \int_{\Sigma^R} \int_{[0,1]} \left(\nu(a(m(t)), t)U^R(a(m), t) + (1 - \nu(a(m(t)), t))U^R(\delta, t)\right)(d\beta(m))d\nu d\alpha.\) A perfect Bayesian equilibrium \(\sigma^* = \{\mu^*, \nu^*, \alpha^*, \beta^*\}\) specifies a strategy profile together with Receiver beliefs and is characterized by the following three conditions:

\[
\{\mu^*, \nu^*\} \in \arg\max_{(\mu, \nu) \in \Sigma^S} EU^S(t, \mu, \nu \mid \alpha^*) \text{ for all } t \in T
\]
\[
\alpha^* \in \arg\max_{\alpha \in \Sigma^S} EU^R(\alpha, m \mid \nu^*, \beta^*) \text{ for all } m \in M
\]
\[
\beta^*(m) \text{ is derived from } \mu \text{ and } \beta^0 \text{ using Bayes Rule whenever possible}
\]

Let \(\Sigma^*\) be the set of perfect Bayesian equilibria (from now on just ‘equilibria’). ACDC provides a stability measure and a selection criterion for equilibria. The starting point of ACDC is a theory of credible deviations \(\gamma\). Such a theory associates a deviating profile \(\gamma(\sigma^*) \in \Sigma\) with an equilibrium \(\sigma^*\). (For convenience, we write \(\gamma(\sigma^*) = \{\mu^*, \nu^*, \alpha^*\}\) A deviating profile specifies firstly which Sender types would deviate and in which way, and secondly, how the Receiver would react. If no type can send a credible deviation according to \(\gamma\), then \(\gamma(\sigma^*) = \sigma^*\).
The following step is to measure the intensity of each type’s credible deviation, which we denote by \( CD_\gamma(t, \sigma^*) \). We would like this measure to have some properties. It should be

- invariant to affine transformations of payoffs;
- increasing in the difference between the deviating and equilibrium payoff;
- 0 if the difference between deviating and equilibrium payoff is zero;
- 1 if the difference between deviating and equilibrium payoff is maximal.

We then define the Average Credible Deviation (ACD) of an equilibrium \( \sigma^* \) relative to \( \gamma \) as:

\[
(3.2) \quad ACD_\gamma(\sigma^*) = E_t[CD_\gamma(t, \sigma^*)]
\]

Specifications of \( CD_\gamma(t, \sigma^*) \) adhering to the properties above will lead to similar conclusions. We propose the following function for \( CD_\gamma(t, \sigma^*) \). Let \( \Sigma^t \) be the set of rationalizable strategy profiles. Then, \( \underline{U}^S(t) \equiv \inf_{\alpha \in \Sigma^t} EU^S(t, \mu, v \mid \alpha) \) and \( \overline{U}^S(t) \equiv \sup_{\alpha \in \Sigma^t} EU^S(t, \mu, v \mid \alpha) \) are the lowest and highest rationalizable payoff for Sender type \( t \). We define the credible deviation of type \( t \) in equilibrium \( \sigma^* \) relative to \( \gamma \) as

\[
(3.3) \quad CD_\gamma(t, \sigma^*) = \frac{EU^S(t, \mu^*, v^* \mid \alpha^*) - EU^S(t, \mu^*, v^* \mid \alpha^*)}{\overline{U}^S(t) - \underline{U}^S(t)}
\]

if \( EU^S(t, \mu^*, v^* \mid \alpha^*) > \underline{U}^S(t) \). If \( EU^S(t, \mu^*, v^* \mid \alpha^*) = \overline{U}^S(t) \), then \( CD_\gamma(t, \sigma^*) = 1 \), as in this case the Sender has no incentive to adhere to her equilibrium strategy.

A deviation theory can be based on credible neologisms or credible announcements (or, in principle, on any theory of credible deviations). In the case that types can send multiple credible deviations, a deviating theory should also specify which one(s) Senders would use.\(^{32}\) In many games, credible neologisms

\(^{32}\) In the case of credible neologisms, a reasonable specification is that types coordinate on the credible neologism with the highest total intensity (mass-weighted sum of credible deviations). Results will in most games be qualitatively the same for different coordination rules.
and credible announcements coincide. Which theory predicts behavior best when they do not coincide, is an empirical question about which actually very little is known. We prefer credible announcements for ‘simple’ games and credible neologisms for ‘complex’ games. Credible neologisms have the virtue of simplicity, whereas credible announcements meet the soundest criterion of credibility. In simple games, human agents will be able to reason according to the logic of credible announcements and credible announcements seem most appropriate. In addition, a deviation theory based on credible announcements has the elegant property that it need not specify which credible announcement a type will play if she can send more than one, as they all must yield an identical payoff for each deviating Sender. In complex games, however, the complexity of announcements and the strict credibility criteria seem too demanding for boundedly rational agents. This can lead to counterintuitive predictions. For example, in the Crawford-Sobel game, announcement proofness may only select the pooling equilibrium, which seems an unlikely outcome for people to play. Hence, here the simplicity of credible neologisms seems more appropriate. A rule of the thumb is to use credible announcements in discrete games and credible neologisms in continuous games. We adhere to this rule of thumb in this chapter and, hence, when we apply ACDC we will suppress $\gamma$.

Based on the ACD, we formulate the ACDC-Criterion (ACDC), which says that an equilibrium $\sigma^*$ will on average predict better than equilibrium $\sigma$ if $ACD_{\sigma^*} < ACD_{\gamma}(\sigma)$. In particular, based on ACDC we can formulate the following selection criterion:

For instance, in the uniform quadratic game for $b \in \left(\frac{1}{31}, \frac{1}{12}\right)$, the pooling equilibrium is announcement proof, while the size-2 and size-3 equilibria are not. For these values of $b$, the pooling equilibrium admits the weakly credible announcement composed of the neologisms at the beginning and end, characterized by the set of intervals of deviating types sending the same message $\{[0, \frac{1}{3} + \frac{4}{3} b], [\frac{1}{2} - \frac{1}{2} b, 1]\}$. In addition, however, it admits the weakly credible announcement $\{[0, \frac{1}{3} - \frac{4}{3} b], [\frac{1}{2} - \frac{1}{2} b, \frac{1}{2} + \frac{1}{2} b], [\frac{1}{2} - \frac{1}{2} b, \frac{1}{2} + \frac{1}{2} b], \{\frac{1}{2} - \frac{1}{2} b, 1]\}$. Since for all weakly credible announcements deviating types exist that prefer another weakly credible announcement, none is a credible announcement. The size-2 and size-3 equilibria only admit weakly credible announcements composed of the non-overlapping credible neologisms, which are thus credible announcements. Observe that the computational demands on agents to determine whether credible announcements exist and how they look like are quite high in this game.

Admittedly, this rule is somewhat coarse, but does prevent ex-post determination as to what constitutes a ‘simple’ game.
Definition 3.1 An equilibrium $\sigma^*$ is an ACDC equilibrium relative to deviation theory $\gamma$ if $ACD_\gamma(\sigma^*) \leq ACD_\gamma(\sigma)$ for all $\sigma \in \Sigma$.

Note that this selection criterion selects the equilibrium that will predict best on average rather than the equilibrium that will always be played.

A simple implication is that $\sigma^*$ is an ACDC equilibrium if $\gamma(\sigma^*) = \sigma^*$. Hence, the prediction of ACDC coincides with that of the underlying deviation theory if the latter identifies a stable equilibrium. The following result is immediate.

Proposition 3.1 If the number of equilibrium outcomes is finite, the cheap talk game has an ACDC equilibrium relative to $\gamma$.

Hence, existence of an ACDC equilibrium is guaranteed by a finite set of equilibrium-outcomes. This is a relevant result, as Park (1997) has shown that finite Sender-Receiver games have a finite set of equilibrium outcomes under generic conditions. Before, Crawford and Sobel (1982) showed a similar result for their setting with a continuous type-space.

Even when games do not have a finite outcome set (as our continuous veto threats model in section 3.4), mild conditions can be formulated in order to guarantee existence of an ACDC equilibrium:

Proposition 3.2 Let $s$ be an equilibrium outcome and $ACD_\gamma(s)$ the ACD of equilibria inducing $s$. Suppose the equilibrium outcome set $S$ can be represented by a finite union of compact metric spaces $S = \bigcup_{i \in N} S_i$, such that $ACD_\gamma(s)$ is continuous in $s$ on all subsets $S_i$. Then, an ACDC equilibrium exists with respect to $\gamma$.

For instance, this means that continuous games with an equilibrium set consisting of partition equilibria that are well-behaved with respect to their ACD will
have an ACDC equilibrium. In almost all applications we have come across, a unique ACDC equilibrium outcome exists. Still, one can construct games with multiple ACDC equilibrium outcomes (for instance when multiple outcomes are completely stable). In such cases, we are content with the conclusion that several equilibrium outcomes are equally plausible.

### 3.3 Discrete games

To illustrate ACDC, let us first put it to work in Game B (Table 3.2). In the pooling equilibrium, all types induce $a_5$. It admits the credible announcement where types $t_2$ and $t_3$ deviate to $a_4$. (Credible announcements and credible neologisms coincide in all games presented in this section.) In a discrete game, the ACD of a pure equilibrium $\sigma$ (with a pure $\gamma(\sigma)$) reduces to

$$\text{(3.4)} \quad \sum_{t \in T} f(t) \frac{U^S(a^{\gamma(\sigma)}(t), t) - U^S(a^\sigma(t), t)}{U^S(t) - U^S(t)},$$

where $f(t)$ is type $t$’s prior probability, and $a^\sigma(t)$, $a^{\gamma(\sigma)}(t)$ the equilibrium respectively deviating action type $t$ induces. Hence, the ACD of the pooling equilibrium is $$\frac{(1 - \delta)(2 - 1)}{2} - \frac{2 - 1}{2 + \varepsilon - 0} = \frac{1}{4} + \frac{\delta - 2 - \varepsilon}{8 + 4\varepsilon}.$$ The ACD of the separating equilibrium is $$\frac{\delta(2 + \varepsilon - 2)}{2 + \varepsilon - 0} = \frac{\delta\varepsilon}{2 + \varepsilon}.$$ The separating equilibrium is always ACDC. In addition, the ACD of the separating equilibrium goes to zero if $\delta$ or $\varepsilon$ goes to zero.

One may be worried that ACDC always selects the most informative equilibrium. This is not the case, as the following two examples show. First consider

---

35 An equilibrium of a game with a one-dimensional type and action set is a partition equilibrium if there exists a partition $t_0 < t_1 < \cdots < t_{n-1} < t_n$ of $T$ such that each type in $(t_{i-1}, t_i)$ induces action $a_i$ with $a_1 < a_2 < \cdots < a_n$. Hence, a partition equilibrium is characterized by a vector $a = (a_1, \ldots, a_n)$ and a partition equilibrium outcome set can be represented by a finite union of compact subsets of $\mathbb{R}^1 \ldots \mathbb{R}^n$. 47
Game C in Table 3.3, which is a reproduction of game 2 in Farrell (1993) and example 2 of Matthews, Okuno-Fujiwara & Postlewaite (1991). In this game, the two types occur with equal probability. The game has a separating equilibrium where type $t_1$ induces $a_1$ and type $t_2$ induces $a_2$. In addition, the game has a pooling equilibrium where both types elicit $a_3$. Here, the separating equilibrium is not announcement proof because it admits the credible announcement “I am not going to tell you my type.” In contrast, no type would want to send an announcement in the pooling equilibrium, which is thus announcement proof.

We share the intuition of Farrell and Matthews et al. that the announcement (and neologism) proofness criteria (and hence ACDC) appropriately reject the communication outcome.

<table>
<thead>
<tr>
<th>TABLE 3.3</th>
<th>GAME C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1(\frac{1}{2})$</td>
<td>1, 3</td>
</tr>
<tr>
<td>$t_2(\frac{1}{2})$</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

Game D (Table 3.4) extends the previous game to three Sender types with a twist. For $\varepsilon = 0$, we get an equivalent result as in the previous game. Game D then has a separating equilibrium, where $t_i$ induces $a_i$ and which admits the credible announcement where $t_1, t_2$ and $t_3$ deviate to $a_4$. It also has an announcement proof pooling equilibrium, where all types induce $a_4$. For small but positive $\varepsilon$, the two equilibria remain intact, but the pooling equilibrium admits a credible announcement where $t_1$ and $t_2$ deviate to $a_5$. Hence, announcement proofness is silent here. ACDC selects the pooling equilibrium: The ACD of the separating equilibrium is $\frac{1}{2} - \frac{1}{6 + 3\varepsilon}$ and that of the pooling equilibrium $\frac{2}{6 + 3\varepsilon}$. 

48
### TABLE 3.4
**GAME D**

<table>
<thead>
<tr>
<th></th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>$a_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1(\frac{1}{3})$</td>
<td>1, 3</td>
<td>0, 0</td>
<td>0, 0</td>
<td>2, 2</td>
<td>$2 + \varepsilon, 2 + \varepsilon$</td>
</tr>
<tr>
<td>$t_2(\frac{1}{3})$</td>
<td>0, 0</td>
<td>1, 3</td>
<td>0, 0</td>
<td>2, 2</td>
<td>$2 + \varepsilon, 2 + \varepsilon$</td>
</tr>
<tr>
<td>$t_3(\frac{1}{3})$</td>
<td>0, 0</td>
<td>0, 0</td>
<td>1, 3</td>
<td>2, 2</td>
<td>$2 - \varepsilon, 0$</td>
</tr>
</tbody>
</table>

*Notes: $0 \leq \varepsilon < 1.$

### 3.4  ACDC versus Other Criteria in a Veto Threats Game

We illustrate the comparative advantage of ACDC over other criteria in a simple continuous external veto threats game (Game E in Table 3.5). At the end of this section, we show that ACDC selects a unique equilibrium in the class of external veto threats games to which this game belongs. The model is close to Matthews’ (1989) model, who introduced veto threat games. The difference is that in our model, the disagreement point is external, in the sense that the Sender’s payoff does not depend on her type. Whereas in Matthews’ model the maximum equilibrium size is two, these games can have a large equilibrium set, with fine partitions and continua of equilibria. This makes them a good testing ground for refinements.

In Game E (Table 3.5), the outcome of the game $x$ is a point on the interval $[0,1]$ or the disagreement point $\delta \notin [0,1]$. The Sender’s payoff on the interval depends on $t$: $U^S(x,t) = -|x - t|$. The larger the distance between the outcome $x$ and her type $t$, the lower the Sender’s payoff. The Sender’s type $t$ is drawn from the uniform distribution on $[0,1]$. The Receiver’s payoff on the interval, $U^R(x) = -x$, is independent of $t$: he always prefers smaller outcomes to larger ones. It is a veto threats game, because the Sender can veto the Receiver’s action, in which case the outcome is the disagreement point $\delta$. The disagreement point payoffs are $U^R(\delta) = -\frac{1}{4}$ and $U^S(\delta,t) = U^S(\delta) = -\frac{1}{4}$. Observe that
the Receiver prefers \( \delta \) to all outcomes on the line larger than \( \frac{5}{4} \) and that the Sender prefers \( \delta \) to all outcomes on the line more than \( \frac{1}{5} \) away from her type \( t \). 

### Table 3.5

**Game E**

| \( t \sim U[0,1] \) | \( U^R(x) = -x \) for \( x \in \mathbb{R} \) | \( U^S(x,t) = -|x-t| \) for \( x \in \mathbb{R} \) | \( U^R(\delta) = -\frac{5}{4} \) and \( U^S(\delta) = -\frac{1}{5} \) |

Like Matthews (1989), we look at a refinement of perfect Bayesian equilibrium that restricts the Receiver to pure strategies and lets the Sender take into account that she may tremble at the veto stage (with epsilon probability). As we showed in Chapter 2, in this game all equilibria are partition equilibria which can be characterized by a finite set of Receiver actions \( a_1 < a_2 < \ldots < a_n \), where \( n \) is called the size of the equilibrium. Senders simply induce that action closest to their type and accept it if it gives them positive payoff.

The game has a unique pooling equilibrium, in which the Receiver always proposes action \( \frac{1}{5} \). This optimal action involves a trade-off between maximizing the probability that the action is accepted and maximizing the payoff of the action once it is accepted. In addition, the game has continua of size-2 and size-3 equilibria. The set of size-2 equilibria is characterized by \( \{a_1, a_2 = a_1 + \frac{1}{5}\} \), with \( a_1 \in [0,\frac{4}{5}] \), whereas the set of size-3 equilibria is characterized by \( \{a_1 = 0, a_2, a_3 = a_2 + \frac{1}{5}\} \) with \( a_2 \in (0,\frac{4}{5}] \). In this game, credible neologisms coincide with (weakly and ordinary) credible announcements. Both equilibria admit credible neologisms. All equilibria have a credible neologism ‘at the end’: \( \langle \frac{a}{5}, [\frac{1}{5}, 1] \rangle \) in the pooling equilibrium and \( \langle \frac{1}{5}, (\frac{1}{5} + \frac{1}{2} a_n, 1) \rangle \) in the other equilibria. If \( a_1 > 0 \), also a credible neologism ‘at the beginning’ of the form \( \langle 0, [0, \frac{1}{5} a_1] \rangle \) exists. The game has a unique ACDC equilibrium, which is \( \{0, \frac{1}{5}, \frac{4}{5}\} \). The intuition is that this equilibrium does not have a neologism at the beginning and minimizes both the frequency and intensity of deviations at the end by maximizing \( a_n \).
We now turn to other refinements and non-equilibrium solution concepts. From our analysis above, it follows that no equilibrium in Game E is neologism or announcement proof. Rabin & Sobel (1996) propose the recurrent mop criterion, which can select equilibria that, although not impervious to credible deviations, are likely to recur in the long run, because they are frequently deviated to. The authors restrict their definition of the recurrent mop to games with a finite number of actions as it may run into problems in continuous games, amongst others because the deviation correspondence may not converge in these settings. In Game E, it is hard to evaluate the solution concept. In a similar game with just two equilibria (e.g. when $U^R(\delta) = \frac{3}{2}$ and $U^S(\delta) = \frac{1}{4}$) one can show that even if the deviation correspondence would converge, neither equilibrium is stable and both are recurrent.

We now turn to approaches that are not based on credible deviations. The simplest are ex-ante efficiency and influentiality, both of which are not very attractive. Ex-ante efficiency selects the equilibrium with the highest ex-ante payoff for the Sender. One problem with this is that it is not clear how the Sender type can commit ex-post to an ex-ante optimal strategy. Furthermore, which equilibrium is ex-ante optimal is not invariant to affine transformations of payoffs of a subset of types (which should not affect the game). Hence, as long as some types receive less in equilibrium A than in Equilibrium B and the other way around, one can make either equilibrium ex-ante more efficient. In Game E, we can make any size-3 equilibrium efficient by multiplying the payoffs of types around $\alpha_2$ and $\alpha_3$ with a suitably large number. (In other games, e.g. with $U^R(\delta) = \frac{3}{2}$ and $U^S(\delta, t) = \frac{1}{4}$, one can also make the pooling equilibrium ex-ante efficient.) Influentiality selects the equilibrium with the largest equilibrium action set. As Games C and D showed, the most influential equilibria need not be the most plausible. Furthermore, Game E has a whole continuum of size-3 equilibria.

The Communication Proofness (CP) criterion of Blume & Sobel (1995) singles out equilibria that would not be destabilized if new opportunities to com-

---

36 As in most cheap talk games, neither equilibrium admits strongly credible announcements.
37 In particular, one can show that for both equilibria it must hold that neither equilibrium lies in the deviation correspondence of that equilibrium.
municate arose. CP looks for partitions that distinguish good and bad equilibria, where in each partition good equilibria cannot be destabilized by other good equilibria. An equilibrium survives CP if it is a good equilibrium in some such partition. In Game E all equilibria are stable according to CP, as in each equilibrium some Sender-type in each partition-element receives her maximum payoff, so that no equilibrium can ever be destabilized.

The No Incentive to Separate (NITS) criterion (Chen, Kartik, & Sobel, 2008) is up till now the only refinement based on some notion of stability that can successfully select an equilibrium in the Crawford-Sobel (1982) setting. NITS starts by specifying a ‘lowest type,’ a type with the property that all other types prefer to be revealed as themselves rather than as that lowest type. An equilibrium survives NITS if the lowest type has no incentive to separate, i.e. if the lowest type prefers her equilibrium outcome to the outcome she would get if she could reveal her type. In the Crawford-Sobel model, only the maximum size equilibrium outcome satisfies NITS (under some general monotonicity assumption). In Game E, such a ‘lowest type’ cannot easily be formulated. All types in $[0, \frac{4}{5}]$ are lowest types according to Chen et al.’s definition. Still, one can argue that $\delta = 0$ is a natural lowest type in our game. Under this assumption, the size-2 equilibrium $\{0, \frac{4}{5}\}$ and all size-3 equilibria survive NITS. (By making $U^S(\delta) / U^R(\delta)$ arbitrarily close to zero, one can make the maximum size arbitrarily high and there will be a NITS equilibrium of each size $2, ..., n$.)

Also non-equilibrium concepts exist. Rabin (1990) introduced the concept of Credible Message Rationalizability (CMR). This non-equilibrium concept proposes conditions under which communication can be guaranteed to happen. It assumes that rational players take truth-telling as a focal point, but use the strategic incentives of the game to check whether truth-telling is rational. In Game E, CMR can only guarantee that the 0 type can send a credible message (and is silent about what other types do). CMR requires that all Sender-types who send a credible message receive an action in which they achieve their maximum payoff. This would imply that the Receiver does not best respond to
credible messages (of all types except 0), which cannot be the case under CMR.\textsuperscript{38}

Blume, Kim & Sobel (1993) put forward the Partial Common Interest (PCI) concept. A partition of the typeset satisfies PCI “if types in each partition element unambiguously prefer to be identified as members of that element, and there is no finer partition with that property.” PCI does not make a definite prediction in Game E, as no partition of the type space (except $0 = t_0 < t_1 = 1$) satisfies PCI. The main reason is that the highest Sender-type of a partition-element always prefers the Receiver to believe that the upper boundary is higher than the true boundary (except for types $t = 0$ or $t = 1$). Finally, the ‘partition’ 0 and $(0, B]$ is not PCI, as 0 (which is the best response if the Sender is 0) is also a best response to some Receiver-beliefs with support on the interval $(0,1]$.

We finish this section by showing that the uniqueness of the ACDC equilibrium is not an artifact of the specific characteristics of Game E. Game E belongs to a wider class of veto threat games introduced in Chapter 2. Here we show that ACDC, under somewhat stricter conditions, selects a unique equilibrium. This result is interesting in its own right, as these games model relevant settings of information transmission under power differences. We assume the Sender’s type $t$ is uniformly distributed on the interval $[0,1]$. We model the player’s bargaining power as the payoff of the disagreement point $U^R(\delta)$ and $U^S(\delta)$, where we assume $U^S(\delta, t) = U^S(\delta)$ does not depend on $t$. $U^R$ and $U^S$ satisfy the following assumptions:

(3.5) $U^R$ on $\mathbb{R}$ is twice continuously differentiable, unimodal with a peak at 0 and concave.

\textsuperscript{38} Rabin also introduces an equilibrium version of CMR, Credible Message Equilibria (CME), but as a consequence of the previous analysis, neither equilibrium in Game E can be a CME.
(3.6) \( U^S(x,t) \) can be written as a function \( f(t - x) \), for all \( x \) in \( \mathbb{R} \), \( t \) in \([0,1]\), where \( f \) is continuously differentiable, symmetric, concave, strictly increasing in \( \mathbb{R}_+ \) and for all \( y \in \mathbb{R} \) there is a \( z > 0 \) such that \( f(z) < y \) and \( f(-z) < y \); \( U^S(\delta) < f(0) \).\(^{39, 40}\)

In Chapter 2, we show that only partition equilibria exist. Here we show that there is a unique ACDC equilibrium:

**Proposition 3.3** Under assumptions (3.5) and (3.6), the unique ACDC equilibrium is the maximum size equilibrium with the highest equilibrium action.

In sum, Game E illustrates that current criteria can fail to be predictive for various reasons. Our intuition is that this is due to the fact that they are based on a dichotomous notion of stability. In contrast, ACDC selects a unique equilibrium in the class of games Game E belongs to.

### 3.5 Crawford-Sobel Game

In this section, we apply ACDC to select equilibria in the leading uniform-quadratic case of Crawford & Sobel’s (1982) cheap talk game (henceforth ‘CS game’). In this CS game, types are uniformly distributed on \([0,1]\), the action space is \([0,1]\), \( U^R(a,t) = -(a - t)^2 \) and \( U^S(a,t) = -(a - (t + b))^2 \), with \( b > 0 \) capturing the Sender bias. (The Sender has no veto in this game.)

Crawford and Sobel (1982) show that this game only has (perfect Bayesian) partition equilibria and that the maximum equilibrium size \( n(b) \) is the largest integer \( n \) for which

---

\(^{39}\) Due to the invariance of games (and the ACDC) to affine payoff transformations, it is actually only required that \( U^S(x,t) \) can be written as a function \( a(t) + b(t) \cdot f(t - x) \), where \( b > 0 \) and \( f \) should adhere to the conditions in (3.6) (with \( U^S(\delta) < a + b \cdot f(0) \)).

\(^{40}\) Observe that (3.6) implies assumptions (A2)-(A5) in Chapter 2. Our assumptions here are stricter. In particular, they require a uniform type distribution and a symmetric and concave payoff function for the Sender.
(3.7) \[ 2n(n-1)b < 1. \]

In addition, the game has a unique size-\( n \) equilibrium for each \( n \in \{1,\ldots,n(b)\} \).

Let

(3.8) \[ t_i^n \equiv \frac{i}{n} - 2bi(n - i). \]

for \( i = 0,\ldots,n \) and \( n = 1,\ldots,n(b) \). In the size-\( n \) equilibrium, types in \([t_{i-1}^n, t_i^n] \) send the same equilibrium message, which induces the Receiver to choose action

(3.9) \[ a_i^n = \frac{1}{2} (t_{i-1}^n + t_i^n), \quad i = 1,\ldots,n. \]

We start by deriving all credible neologisms the equilibria admit. For each credible neologism \( \langle \tilde{a}, N \rangle \), the set of deviating types \( N \) turns out to be an interval between some \( \tau \) and \( \bar{\tau} \). Hence, we can characterize neologisms by \([\tau, \bar{\tau}]\) alone, since the Receiver’s best response is \( \tilde{a} = \frac{\tau + \bar{\tau}}{2} \). An equilibrium can admit three types of credible neologisms. First of all, there may be a credible neologism which includes \( t = 0 \). If this credible neologism exists, then it has the shape \([0, \tau_0^n] \) where

\[ \tau_0^n = \frac{2}{3} a_i^n - \frac{4}{3} b = \frac{1}{3n} - \frac{2}{3} b(n + 1). \]

Chen, Kartik & Sobel (2008) show that an equilibrium that fails NITS has a credible neologism of this kind. They also prove that only the size-\( n(b) \) equilibrium satisfies NITS, so that the credible neologism \([0, \tau_0^n] \) exists if and only if \( n < n(b) \).
Second, Farrell (1993) shows that if $b < \frac{1}{2}$, the game has a credible neologism on the right-end of the type space of the form $(\tau^+_n, 1)$ where

$$\tau^+_n = 1 - \frac{1}{3n} - \frac{2}{3} b(n + 1).$$

Finally, if $n \in \{2, ..., n(b) - 1\}$, there are $n - 1$ credible neologisms “in the middle.” These take the form $(\tau^+_n, \tau^+_i)$ for $i = 1, ..., n - 1$, where $\tau^+_n [\tau^+_i]$ is indifferent between the equilibrium action $a^+_i [a^+_{i+1}]$ and the neologism action $\tilde{a}^+_i = (\tau^+_n + \tau^+_i) / 2$. We obtain:

$$\tau^+_n = \frac{3}{4} a^+_i + \frac{1}{4} a^+_{i+1} - 2b \quad \text{and}$$

$$\tau^+_i = \frac{1}{4} a^+_i + \frac{3}{4} a^+_{i+1} - 2b,$$

$i = 1, ..., n - 1$. If $n = n(b)$, the game has the same types of credible neologisms “in the middle,” with the exception that the neologism $(\tau^{n(b)}_n, \tau^{n(b)}_1)$ need not exist.\(^{41}\) Observe that $\tau^{n-1}_i < \tau^n_n$ for $i = 1, ..., n$, so that none of the credible neologisms overlap. Figure 3.1 illustrates the results for $b = \frac{1}{18}$.

It seems intuitive that the highest size equilibrium is the ACDC equilibrium, since the deviations seem to get smaller and smaller as the size increases. This indeed turns out to be the case. Although one can obtain analytical results for the ACD for specific parameter values, finding the ACDC equilibrium for general $b$ defies an analytical approach. Hence, we calculated the ACD for a very fine grid of $b$ and obtain the following result.

\(^{41}\) If (and only if) $2bm(b)^2 \geq 1$, there is no credible neologism of the form $(\tau^{n(b)}_i, \tau^{n(b)}_1)$ because $\tau^{n(b)}_i = \frac{3}{4} a^{n(b)}_i + \frac{1}{4} a^{n(b)}_{i+1} - 2b = -\frac{3}{4m(b)} (2bm(b)^2 - 1) \leq 0$, which is inconsistent with all types being in the interval $[0, 1]$ or the interval $(\tau^{n(b)}_i, \tau^{n(b)}_1) = (0, t^{n(b)}_i)$ being a neologism.
**Proposition 3.4** For all $b \in \left\{ \frac{1}{10000}, \frac{2}{10000}, \ldots, \frac{1}{4} \right\}$ it holds that the ACD of the size-$n$ equilibrium in the CS game is decreasing in $n$.

**Corollary 3.1** For all $b \in \left\{ \frac{1}{10000}, \frac{2}{10000}, \ldots, \frac{1}{4} \right\}$, the size-$n(b)$ equilibrium is the unique ACDC equilibrium.

We also derive the following property of the maximum size equilibrium (for which we do not need to calculate the ACD’s for each $b$):

**Proposition 3.5** The ACD of the size-$n(b)$ equilibrium tends to zero if $b$ tends to zero in the CS game.

Hence, the ACD of this equilibrium converges to zero if $b$ approaches zero, i.e. if the interests of the players are almost perfectly aligned. This finding is intuitive because the Sender obtains almost her ideal outcome when $b$ is close to zero, so she will not gain much in the case of deviation, and even if she deviates, the deviation will hardly change the equilibrium.

Proposition 3.4 is in line with NITS (Chen, Kartik & Sobel, 2008), which also selects the size-$n(b)$ equilibrium. If the bias parameter is large, however, the maximum size equilibrium can still have a large ACD, so it may not be all that stable. The prediction of ACDC that the maximum size equilibrium becomes more stable as $b$ becomes smaller (Proposition 3.5) is not made by previous concepts. NITS does not predict this, as it assumes that no type separates if the lowest type does not separate.
The size-1, size-2 and (maximum) size-3 equilibria with the credible neologisms they admit for $b = \frac{1}{18}$. The area of the neologisms give an impression of their contribution to the ACD, although there height contributes quadratically to the ACD.

Figure 3.1


3.6 Experimental Evidence

In this section, we look at the experimental support for ACDC.

3.6.1 Discrete games

Blume, DeJong, Kim & Sprinkle (2001) provide an experimental analysis of 4 discrete cheap talk games, in which they compare the predictive power of refinements as neologism proofness, influentiality and ex-ante efficiency with PCI. (Credible announcements coincide with credible neologisms in Blume et al.’s games.) They find that PCI is a reliable predictor of when communication takes place and that the equilibrium refinements sometimes but not always improve on PCI. In their Games 1 and 3, the predictions of PCI and announcement proofness (and ACDC) are very much aligned, and borne out by the data. In their Game 2 (see Table 3.6) announcement proofness predicts complete separation while the finest partition consistent with PCI entails partial separation. The data are in line with separation, as a clear majority of 88% of the outcomes is consistent with the separating equilibrium. One could argue that this result does not contradict PCI, because PCI allows multiple patterns including separation (see their footnote 10). As the authors note (in footnote 19), one needs to add neologism proofness to PCI to actually predict that separation happens.

| Table 3.6 |
| Reproduction of Games 2 and 4 of Blume et al. (2001) |

<table>
<thead>
<tr>
<th></th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>800, 800</td>
<td>100, 100</td>
<td>0, 0</td>
</tr>
<tr>
<td>$a_2$</td>
<td>100, 100</td>
<td>$x$, 100</td>
<td>0, 0</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0, 0</td>
<td>0, 0</td>
<td>500, 800</td>
</tr>
<tr>
<td>$a_4$</td>
<td>500, 500</td>
<td>0, 400</td>
<td>0, 0</td>
</tr>
<tr>
<td>$a_5$</td>
<td>0, 400</td>
<td>0, 400</td>
<td>0, 400</td>
</tr>
</tbody>
</table>

Notes: All the three types $\{t_1, t_2, t_3\}$ of the Sender are equally likely and the Receiver can implement one of the actions $\{a_1, ..., a_5\}$. Entry $i,j$ represents $U^S(t_i, a_j), U^R(t_i, a_j)$. Games 2 and 4 are identical, except that $x = 100, y = 300$ in game 2, whereas $x = 300, y = 100$ in game 4.

Their Game 4 (Table 3.6) is interesting because no equilibrium is announcement proof while PCI makes a prediction. This game has two equilibrium
outcomes. Besides the pooling equilibrium where action $a_5$ is induced there is a partially separating equilibrium where types $t_1$ and $t_2$ send a common message that differs from the message of $t_3$. Types $t_1$ and $t_2$ induce $a_4$ while type $t_3$ induces $a_3$. Full separation is not an equilibrium because $t_2$ prefers to mimic $t_1$. None of the equilibria satisfies announcement proofness. PCI predicts meaningful communication because the finest partition consistent with PCI is given by $\{\{t_1,t_2\},\{t_3\}\}$. The partially separating equilibrium only has a credible announcement where $t_1$ deviates to $a_1$. Thus, its ACD equals $\frac{1}{3} \frac{(800 - 500)}{800} = \frac{1}{8}$.

The pooling equilibrium admits the credible announcement where $t_1$ and $t_2$ deviate to $a_4$ and $t_3$ deviates to $a_3$. Consequently, its ACD is $\frac{1}{3} \left( \frac{(500 - 0)}{800} + \frac{(500 - 0)}{500} + \frac{(500 - 0)}{500} \right) = \frac{7}{8}$. So ACDC predicts that the partially separating equilibrium will be the most observed equilibrium outcome but that it will not be completely stable.

In line with this prediction, Blume et al. find that 37% of the outcomes are consistent with the partially separating equilibrium but no outcome is consistent with the pooling equilibrium. Thus, of the two equilibria, the one with the lowest ACD performs best. Consistent with the ACD measures, much fewer outcomes are in line with the equilibrium selected by ACDC in game 4 than in game 2. In line with the fact that types $t_1$ have a credible announcement, they turn out to be the ones that are able to credibly identify themselves.

Our conclusion is that our ACDC concept improves the predictions of announcement proofness and that it does at least as well as PCI in explaining the data of Blume, DeJong, Kim & Sprinkle (2001). The extra mileage for ACDC comes from continuous games like the Crawford-Sobel game and the veto threat game. PCI fails to predict any communication at all in these settings, while in accordance with ACDC subjects are able to communicate meaningfully to a large extent, as we will see next.
3.6.2 Crawford-Sobel game

Several experimental studies have been done in discrete versions of the CS game. Cai & Wang (2006) test a discrete version of this game where the Sender’s type is uniformly drawn from \{1, 3, 5, 7, 9\}. They report the results of four treatments that differ in the disalignment between the Sender’s and Receiver’s preferences. The treatment with the smallest disalignment has the full range of equilibria from pooling to fully separating, while the treatment with the largest disalignment parameter only allows for the pooling equilibrium. In the two treatments in between, the most informative equilibrium is a size-2 equilibrium. Both NITS and ACDC select a most informative equilibrium in each treatment.\(^\text{42}\)

The experimental results are relatively closest to the most informative equilibrium. Except in the case where complete separation is supported in equilibrium, subjects over-communicate compared to the most informative equilibrium, and overcommunication increases as the bias parameter increases. Hence, in agreement with ACDC, behavior departs more from the most informative equilibrium, as its ACD increases.\(^\text{43}\)

3.6.3 Veto Threats Game

In Chapter 4, we present new experimental data. We test the predictions of ACDC in five games belonging to the class of veto threat games discussed in section 3.4. We find that ACDC organizes the data well. The ACDC equilibrium performs best, even if it admits credible deviations. Furthermore, in comparable games, the ACDC equilibrium performs better as its ACD decreases. Finally, in

\(^{42}\) In their treatment with the disalignment parameter equal to 1.2 there is an additional most informative equilibrium \{1\}, \{3579\}, besides the reported most informative equilibrium \{13\}, \{579\}. Both equilibria are NITS, while ACDC selects the latter equilibrium. The data are not sufficiently informative to discriminate between these equilibria.

\(^{43}\) The results of Dickhaut, McCabe & Mukherji (1995) on a Crawford-Sobel game are similar to those reported by Cai and Wang, although they do not interpret their results in terms of overcommunication. More recently, Wang, Spezio & Camerer (2010) replicate the results of Cai & Wang (2006) and find that look-up patterns of Senders (as measured by eye-tracking) reveals a significant amount of information about their type.
a setting with continua of equilibria similar to Game E, we find that the closer
equilibria are to the ACDC equilibrium, the better they perform.

3.7 Conclusion

This chapter generalizes refinements based on credible deviations, in particu-
lar neologism proofness and announcement proofness. We started with an
intuition for why the frequency and size of credible deviations could affect
equilibrium stability in a continuous rather than a binary manner. Consequently,
we formalized this intuition in ACDC, which measures the (in)stability of cheap
talk equilibria and determines which are most plausible. We show an ACDC
equilibrium exists under general conditions unlike existing concepts. Further-
more, the predictions of ACDC are meaningful in previously analyzed settings
and organize the data of previous experiments well. Finally, ACDC makes
predictions in settings where other concepts cannot.

In Chapter 4, we find support for ACDC in a new experimental setting where
other criteria remain silent.

3.8 Appendix: Proofs

3.8.1 Proposition 3.2

Proof of Proposition 3.2 $\text{ACD}_i(s)$ achieves a minimum on each compact
subset $S_i$ and hence achieves a minimum on $S$. As a consequence, also
$\min_{\sigma} \text{ACD}_{i}(\sigma)$ is nonempty, so that at an ACDC equilibrium exists. Q.E.D.

3.8.2 Proposition 3.3

For the proof of Proposition 3.3, we introduce some definitions and results
from Chapter 2, and derive two helpful lemmas.
Observe that in this game, a neologism $\langle \tilde{a}, N \rangle$ is credible relative to equilibrium $\sigma^*$ if and only if

$$\tilde{a} \in \arg \max_{a \in \mathbb{R}} P \left\{ U^S(a,t) \geq 0 \big| t \in N \right\} \left( U^R(a) - U^R(\delta) \right),$$

and for all $k = 1,...,n$ it holds that $t \in [t_{k-1},t_k] \cap N \Rightarrow U^S(\tilde{a},t) > U^S(a_k,t)$ and $t \in [t_{k-1},t_k] \setminus N \Rightarrow U^S(\tilde{a},t) \leq U^S(a_k,t)$.

**Lemma 3.1** If $\langle \tilde{a}, N \rangle$ is a credible neologism relative to equilibrium $\sigma^*$, then $N$ is an interval.

**Proof.** The proof is by contradiction. Suppose $0 \leq t^1 < t^2 < t^3 \leq 1$, $t^1, t^3 \in N$ and $t^2 \not\in N$. Suppose further that in equilibrium, type $t^i$ obtains action $a^i$, $i = 1,2,3$. The fact that the a type’s utility is strictly decreasing in the distance between $t - a$ implies $a^1 \leq a^2 \leq a^3$. If $\tilde{a} \leq t^2$ then it must be the case that $\tilde{a} \leq a^2$ (otherwise type $t^2$ would prefer $\tilde{a}$ over $a^2$). As a consequence, $\tilde{a} \leq t^3 \leq a^3 = a^2$ because type $t^3$ must prefer $\tilde{a}$ over $a^3$ and $a^3$ over $a^2$. A contradiction is established, because the fact that the indifference points $t - d$ and $t + d$ are strictly increasing in $t$ implies that type $t^2$ strictly prefers $\tilde{a}$ over $a^2$. This is in conflict with the definition of a credible neologism. Analogously, $\tilde{a} > t^2$ can be ruled out, so that $N$ is an interval. Q.E.D.

From (3.6), it follows that there is a $d > 0$ such that for all $t$ and $a \in \mathbb{R}$, $U^S(a,t) \geq U^S(\delta)$ if and only if $a \in [t-d,t+d]$. Hence, $t - d$ and $t + d$ are the Sender’s indifference points as to whether she accepts action $a$. From Lemma’s 1 and 2 in Lemma 2.1 and Lemma 2.2 it follows that in equilibrium

$$a_i \geq 0, \quad t_{k-1} - d < a_k \leq t_k - d \quad \text{for all} \quad k = 2,...,n \quad \text{and} \quad t_{k-1} + d \leq a_k \quad \text{for} \quad k = 3,...,n.$$ (3.11)

We can now show that under (3.6), it holds that
**Lemma 3.2** In equilibrium, \( a_k + d = t_k = a_{k+1} - d \) for \( k = 2, \ldots, n - 1 \).

**Proof.** Due to the \( t \) being uniformly distributed and (3.6), the indifference points \( t - d \) and \( t + d \) are uniformly distributed as well. This means that if the Receiver receives a message that identifies Sender types to be in the interval \([t_k, t_{k+1}]\) \((k = 0, \ldots, n - 1)\), the probability the Sender accepts an action is not higher for an action \( a > t_k + d \) than for action \( a' = t_k + d \), while \( U^R(a) < U^R(a') \). Hence, for the equilibrium action \( a_k \) it holds true that \( a_k \leq t_{k-1} + d \) and by (3.11), this means \( a_k = t_{k-1} + d \leq t_k - d \) for \( k = 3, \ldots, n \).

Now, suppose that \( t_{k-1} + d < t_k - d \) for some \( k = 3, \ldots, n \). This means that \( a_k < t_k - d \) and hence \( U^S(a_k, t_k) < 0 \). Since \( U^S(a_k, t_k) = U^S(a_{k+1}, t_k) \), this implies, however, that \( a_{k+1} > t_k + d \), which for \( k = 3, \ldots, n - 1 \) is a contradiction with \( a_k \leq t_{k-1} + d \) for \( k = 3, \ldots, n \). Hence, \( a_k = t_{k-1} + d = t_k - d \) for \( k = 3, \ldots, n - 1 \).

Consequently, \( a_k + d = t_k = a_{k+1} - d \) for \( k = 3, \ldots, n - 1 \).

Furthermore, from the discussion above we have that \( t_2 = a_2 - d \) and that \( a_2 \leq t_1 + d \). In addition, from (3.11) it follows that \( a_2 \leq t_2 - d \). Hence, a necessary condition on \( a_2 \) is that \( a_2 \in \arg \max_{a_2} U^R(a) \left( \min\{a + d, 1\} - \frac{1}{2} a_2 \right) \). Analogously to our discussion in the proof of Proposition 2.2, one can show that this implies that \( a_2 \) must be equal to \( t_2 - d \). As a result, \( a_2 + d = t_2 = a_3 - d \).

Q.E.D.

**Proof of Proposition 3.3** Suppose that the game has more than one equilibrium outcome. If \( \bar{x} \leq 2d \), then consider the equilibrium outcome \( \sigma^* \) with \( a_1 = 0 \) and \( a_2 \) such that \( a_2 \in \arg \max_{a_2} U^R(a) \left( \min\{a + d, 1\} - \frac{1}{2} a_2 \right) \). If \( \bar{x} > 2d \), let \( n \) be the natural number for which \( \bar{x} - 2dn \leq 0 \) and \( \bar{x} - 2d(n-1) > 0 \), and consider the following \( \sigma^* : a_1 = 0; a_k = \bar{x} - 2d(n - k - 2), k = 2, \ldots, n \). We now show that \( \sigma^* \) has the maximum equilibrium size and is the unique ACDC equilibrium outcome.
From (2.2) and (3.6), it follows that there exists an \( x \in \mathbb{R} \) such that

\[
U^R(x) - U^R(\delta) + 2dU^R(x) \geq 0 \text{ for all } x \in [0, \bar{x}) \text{ and }
\]

\[
U^R(x) - U^R(\delta) + 2dU^R(x) < 0 \text{ for all } x \in (\bar{x}, 1 - d],
\]

where a prime ('') denotes a derivative with respect to \( x \). Let \( a^* \) denote the highest equilibrium action \( a_n \) in \( \sigma^* \). Using (3.12), it can be verified that \( \sigma^* \) constitutes the highest size equilibrium, analogously to the proof of Proposition 2.2. Similarly, it can be verified that the highest action \( a^{**} \) in any other equilibrium \( \sigma^{**} \) must be smaller than \( a^* \):

\[
a^{**} \leq a^* \leq 1 - d.
\]

If \( a^{**} < 1 - d \), \( \sigma^{**} \) has at least one credible neologism: Types in the interval \((\tau^{**}, 1]\) are willing to send a credible neologism \( \{a^{**}, (\tau^{**}, 1]\} \), where

\[
\tau^{**} = \frac{1}{2}(a^{**} + \bar{a}^{**}), \text{ and }
\]

\[
\bar{a}^{**} \in \arg \max_{a \in (a^{**}, 1]} \left(U^R(a) - U^R(\delta)\right) \frac{a + d - \tau^{**}}{1 - \tau^{**}}.
\]

To proof that \( \sigma^* \) is an ACDC equilibrium, we first show it has at most one credible neologism (claim 1) and this credible neologism, if it exists, maximizes \( \tau^{**} \) and minimizes \( \bar{a}^{**} - a^{**} \) (claim 2).

In order to prove claim 1, suppose that \( \sigma^* \) has another credible neologism. By Lemma 3.1, the set of types that send the credible neologism relative to equilibrium \( \sigma^* \) is an interval. We can exclude neologisms that induce the Receiver to propose \( a = 0 \), because \( a_i = 0 \) is already an equilibrium action. Hence, the neologism \( \bar{a} \) (with supremum neologism type \( \bar{\tau} \)) is in between two equilibrium actions \( a_{k-1} \) and \( a_k \). Due to Lemma 3.1, \( a_{k-1} < \bar{a} < \bar{\tau} < a_k \). This implies that \( U^S(\bar{a}, \bar{\tau}) \leq 0 \), because if \( U^S(\bar{a}, \bar{\tau}) > 0 \), action \( \bar{\tau} - d \) would be better for the
Receiver than \( \tilde{a} \) after receiving the neologism. Consequently, 
\( U^S(a_{k-1}, \hat{\tau}) < U^S(\tilde{a}, \hat{\tau}) \leq 0 \) and \( U^S(a_k, \hat{\tau}) < U^S(\tilde{a}, \hat{\tau}) \leq 0 \). This means that an \( \varepsilon > 0 \) exists such that a types in \( (\hat{\tau} - \varepsilon, \hat{\tau} + \varepsilon) \) receive 0 payoff in equilibrium. Since this is not the case in \( \sigma^* \), \( \sigma^* \) has no other neologisms.

The proof of claim 2 proceeds as follows. Note that \( \tilde{a}^{**} = \min\{\alpha^{**}, 1 - d\} \), where \( \alpha^{**} = \arg \max_{a \in \mathbb{R}} \left( U^R(a) - U^R(\delta) \right) \frac{a + d - \zeta^{**}}{1 - \zeta^{**}} \). We know \( \alpha^{**} > a^{**} \), because the solution to \( \arg \max_{a \in \mathbb{R}} \left( U^R(a) - U^R(\delta) \right) \frac{a + d - t}{1 - t} \) is increasing in \( t \) and \( a^{**} \) is the solution for \( t = t_{n-1} \), and \( \alpha^{**} \) is the solution to the problem with \( t \geq a^{**} > t_{n-1} \). Moreover,

\[
U^R(\alpha^{**}) - U^R(\delta) + U^R(\alpha^{**}) \left( \alpha^{**} + d - \zeta^{**} \right) = U^R(\alpha^{**}) - U^R(\delta) + U^R(\alpha^{**}) \left( \frac{\alpha^{**} - a^{**}}{2} + d \right)
\]

= 0 implies that

\[
\alpha^{**} - a^{**} = -2 \frac{U^R(\alpha^{**})}{U^R(\alpha^{**})} - 2d.
\]

From the concavity of \( U^R \) it follows that \( \frac{U^R(a)}{U^R(a)} \) is increasing in \( a \). Hence, \( \alpha^{**} - a^{**} \) is decreasing in \( a^{**} \). In particular, this implies that \( \tilde{a}^{**} - a^{**} \) is decreasing in \( a^{**} \). Moreover, \( \zeta^{**} \) is increasing in \( a^{**} \).

Finally, to show that \( \sigma^* \) is an ACDC equilibrium, we show that it has the lowest ACD. By Lemma 3.2, for equilibrium \( \sigma^{**} \) it must then hold that \( a_i^{**} > 0 \) or \( a^{**} < a^* \). If \( a_i^{**} > 0 \), then a neologism \( \{\tilde{a}_0, [0, \bar{\tau}]\} \) exists with \( \tilde{a}_0 < a_i^{**} \). If \( a^{**} < a^* \), then a neologism \( \{\tilde{a}_0, (0, \bar{\tau}]\} \) exists with \( \tilde{a}_0 > a^{**} \).
evident that \( ACD(\sigma^*) = 0 < ACD(\sigma^{**}) \). Hence, suppose that \( \sigma^* \) admits the credible neologism \( \{\bar{a}^*, [\tau^*, 1]\} \).

We can now compare the ACD of \( \sigma^* \) and \( \sigma^{**} \). First, \( CD^{\sigma^*}(t) = 0 \) for \( t \in [0, \tau^*] \). Second, we show that \( U^S(\bar{a}^{**}, t) - U^S(a^{**}, t) > U^S(\bar{a}^*, t) - U^S(a^*, t) \) for \( t \in [\tau^*, \bar{a}^{**} + d] \). Due to claim 2 \( \bar{a}^{**} - a^{**} > \bar{a}^* - a^* \) and \( \tau^{**} < \tau^* \). If \( t \leq \bar{a}^{**} < \bar{a}^* \), then \( U^S(a^{**}, t) < U^S(a^*, t) \) and \( U^S(\bar{a}^{**}, t) > U^S(\bar{a}^*, t) \), so that the result is immediate. Assume now that \( \bar{a}^{**} > t \). By (3.6), \( U^S(a, t) \) is concave in \( a \), such that for \( x < y \leq t \) and \( b, c > 0 \) it holds that:

\[
U^S(y, t) - U^S(x, t) \leq U^S(y - b, t) - U^S(x - b, t) < U^S(y - b, t) - U^S(x - b - c, t).
\]

Hence, for \( t \in [\tau^*, \bar{a}^*] \) we have that \( U^S(\bar{a}^*, t) - U^S(a^*, t) \leq U^S(\bar{a}^{**}, t) - U^S(a^{**}, t) < U^S(\bar{a}^{**} + t + \bar{a}^{**}, t) < U^S(\bar{a}^{**}, t) - U^S(a^{**}, t) \). (Observe that \( t - a^* < \bar{a}^* - a^* < a^{**} - a^{**} \).) Similarly, for \( t \in (\bar{a}^*, a^{**} + d] \), \( U^S(\bar{a}^{**}, t) - U^S(a^{**}, t) \leq U^S(\bar{a}^{**}, t) - U^S(a^{**}, t) < U^S(\bar{a}^{**}, t) - U^S(a^{**}, t) \). As a consequence, \( CD^{\sigma^{**}}(t) > CD^{\sigma^*}(t) \) for \( t \in [\tau^{**}, a^{**} + d) \). Finally, \( CD^{\sigma^{**}}(t) = 1 \geq CD^{\sigma^*}(t) \) for \( t \in [a^{**} + d, 1] \). Together, this implies that \( ACD(\sigma^{**}) = E_t[CD^{\sigma^{**}}(t)] > E_t[CD^{\sigma^*}(t)] = ACD(\sigma^*) \).

In sum, if \( \sigma^{**} \) is different from \( \sigma^* \), then either \( a_0^{**} > 0 \) or \( a^{**} < a^* \) and in both cases \( ACD(\sigma^{**}) > ACD(\sigma^*) \). Therefore, \( \sigma^* \) is the unique ACDC equilibrium. Q.E.D.

### 3.8.3 Proposition 3.4 and Proposition 3.5

We did the analysis in the following way. First, we obtain closed-form solutions for the ACD for each value of \( b \) and second we calculate the ACD for a fine grid of \( b \).

The ACD of equilibrium \( \sigma \) in the CS game is equal to

\[
ACD(\sigma) = E_t\left[ \frac{U^S(\bar{a}^*(t), t) - U^S(a^*(t), t)}{U^S(t) - U^S(t)} \right]
\]
\[
\int_0^1 \frac{U^s(\bar{a}^s(t), t) - U^s(a^s(t), t)}{U^s(\min\{t + b, 1\}, t) - \min\{U^s(0, t), U^s(1, t)\}} \, dt
\]
\[
= \int_0^1 \frac{(a^s - (t + b))^2 - (\bar{a}^s - (t + b))^2}{-\max\{0, t + b - 1\}^2 + \max\{(t + b)^2, (t + b - 1)^2\}} \, dt.
\]

Note that \((t + b - 1)^2 > (t + b)^2\) if and only if \(t < \frac{1}{2} - b\). Suppose \(a^s(t)\) and \(\bar{a}^s(t)\) are constant and \(\bar{U}^s(t) = 0\) on the interval \([\underline{t}, \bar{t}]\). Let \(\hat{t} \equiv \max\{\underline{t}, \min\{\bar{t}, \frac{1}{2} - b\}\}\). Then, \(\int_\underline{t}^{\bar{t}} CD(t, \sigma) dt\) is equal to

\[
h(b, a^s, \alpha^s, \underline{t} \bar{t}) \equiv \int_\underline{t}^\bar{t} \frac{(a^s - (t + b))^2 - (\bar{a}^s - (t + b))^2}{\max\{(t + b)^2, (t + b - 1)^2\}} \, dt \\
= \int_\underline{t}^\bar{t} \frac{(a^s - (t + b))^2 - (\bar{a}^s - (t + b))^2}{(t + b - 1)^2} \, dt + \int_\hat{t}^\bar{t} \frac{(a^s - (t + b))^2 - (\bar{a}^s - (t + b))^2}{(t + b)^2} \, dt \\
= (a^s - \bar{a}^s) \left( \frac{(a^s + \bar{a}^s - 2)(\hat{t} - t)}{(b - 1 + \hat{t})(b - 1 + t)} + 2\log\left[ \frac{b - 1 + t}{b - 1 + \hat{t}} \right] \right) \\
+ (a^s - \bar{a}^s) \left( \frac{(a^s + \bar{a}^s)(\bar{t} - \hat{t})}{(b + \hat{t})(b + \bar{t})} + 2\log\left[ \frac{b + \hat{t}}{b + \bar{t}} \right] \right).
\]

As noted before, an equilibrium of size \(n\) can have a neologism in the beginning \(\bar{a}^n_0\), a neologism at the end \(\bar{a}^n_n\) and at most \(n - 1\) neologisms in the middle, \(\bar{a}^n_i, i = 1, \ldots, n - 1\). The size-1 equilibrium has a neologism at the beginning and at the end. The maximum size \(n(b)\) equilibrium has a neologism at the end and neologisms in the middle \(\bar{a}^n_i, i = \hat{i}(b), \ldots, n - 1\), where \(\hat{i}(b) = 1\) if \(2bn(b)^2 < 1\) and \(\hat{i}(b) = 2\) if \(2bn(b)^2 \geq 1\). Size- \(n\) equilibria with \(1 < n < n(b)\) admit all neologisms specified above. Observe that \(\tau^s_{n-1} < 1 - b\), such that \(\bar{U}^s(t) = 0\) except for the highest types of the highest neologism, such that \(h(b, a^s, \bar{a}^s, \underline{t} \bar{t})\) can be used to calculate the contribution to the ACD for neologisms \(\hat{i}(b) = 1, \ldots, n - 1\). For the highest neologism, the contribution to the ACD is equal to
\( \overline{h}(b, n) = h(b, a_n^b, \bar{a}_n^b, \bar{\pi}_n^b, b - 1) + \int_{b-1}^1 \left( a_n^b - (t + b) \right)^2 - \left( \bar{a}_n^b - (t + b) \right)^2 \frac{dt}{(t + b - 1)^2 + (t + b)^2} \)

\( = h(b, a_n^b, \bar{a}_n^b, \bar{\pi}_n^b, b - 1) + \frac{1}{2} (a_n^b - \bar{a}_n^b)(a_n^b + \bar{a}_n^b - 1) \log[2b + 1]. \)

Let \( \sigma_n^b \) be the size-\( n \) equilibrium of the game with bias parameter \( b \). Then, the ACD of the pooling equilibrium is

\[ ACD(\sigma_n^1) = h(b, a_1, \bar{a}_1, 0, \bar{\pi}_1^1) + \overline{h}(b, 1). \]

The ACD of the maximum-size equilibrium is

\[ ACD(\sigma_n^{u(b)}) = \sum_{i=\lceil n(b) \rceil}^{i=n(b)-1} [h(b, a_i^{u(b)}, \bar{a}_i^{u(b)}, \bar{\pi}_i^{u(b)}, t_i^{u(b)}) + h(b, a_{i+1}^{u(b)}, \bar{a}_{i+1}^{u(b)}, t_{i+1}^{u(b)}, \bar{\pi}_{i+1}^{u(b)})] + \overline{h}(b, n(b)). \]

The ACD of a size-\( n \) equilibrium with \( 1 < n < n(b) \) is equal to

\[ ACD(\sigma_n^n) = h(b, a_n^n, \bar{a}_n^n, 0, \bar{\pi}_0^n) + \sum_{i=1}^{i=n-1} [h(b, a_i^n, \bar{a}_i^n, \bar{\pi}_i^n, t_i^n) + h(b, a_{i+1}^n, \bar{a}_{i+1}^n, t_{i+1}^n, \bar{\pi}_{i+1}^n)] + \overline{h}(b, n) \]

**Proof of Proposition 3.4** For each \( b \in \left\{ \frac{1}{10000}, \frac{2}{10000}, ..., \frac{1}{4} \right\} \), one can calculate the (closed-form) value of \( ACD(\sigma_n^n) \) for all \( 1 \leq n \leq n(b) \), and verify that the ACD of the size-\( n \) equilibrium in the CS game is decreasing in \( n \).

**Proof of Proposition 3.5** Let \( \sigma(b) \equiv \sigma_n^{u(b)} \) be the maximum size equilibrium for \( b \). Then,

\[ \lim_{b \uparrow b_0} ACD(\sigma(b)) \leq \lim_{b \downarrow b_0} E_t \left[ \frac{U^S(\bar{a}^{\sigma(b)}(t), t) - U^S(a^{\sigma(b)}(t), t)}{\min_{t \in T} \{U^S(t) - U^S(t)\}} \right] \leq \lim_{b \downarrow b_0} E_t \left[ \frac{0 - U^S(a^{\sigma(b)}(t), t)}{\frac{1}{4}} \right]. \]
Equality 1 follows from the specification of \( \text{SEU} \) in Crawford & Sobel (1982). Inequality 2 follows from \( n(b) = \left[ \frac{1}{2} + \frac{1}{2} \sqrt{2 / b + 1} \right] - 1 \) due to (3.7). The other manipulations are straightforward. \( Q.E.D. \)
Chapter 4  An Experimental Study of Credible Deviations and ACDC

4.1  Introduction

In this chapter, we test the Average Credible Deviation Criterion (ACDC), which we introduced in Chapter 3. As we saw there, many equilibrium refinements and solution concepts have been proposed for cheap talk games. None of these, however, can select equilibria across a wider range of cheap talk games and few have been tested experimentally. In contrast, ACDC is predictive across a wide range of cheap talk games and has shown to organize behavior well in experiments meant to study other concepts. Hence, it is valuable to test ACDC rigorously in a new experiment.

The idea behind ACDC is that the credible deviation approach is sound, but that the insistence on a binary distinction between stable and unstable equilibria is problematic. In particular, ACDC generalizes the binary stability criteria of neologism proofness (Farrell, 1993) and announcement proofness (Matthews, Okuno-Fujiwara & Postlewaite, 1991). ACDC assumes that credible deviations matter for the stability of equilibria but that they matter in a gradual manner. In this study, we test this assumption. In addition, to determine its value added more sharply, we test the predictions of ACDC in games where existing criteria are not predictive. We use the class of continuous external veto threat-games introduced in Chapter 2. These games allow for a clean manipulation of the size and frequency of credible deviations. Furthermore, they can have a large equilibrium set, which previous concepts cannot refine.

Our experimental design consists of five veto threat games and allows us to test four hypotheses. First, we test whether credible deviations matter at all: do neologism proofness and announcement proofness have any bite indeed? Second, we test whether the ACDC equilibrium performs best if all equilibria are unsta-

45 This chapter is based on De Groot Ruiz, Offerman & Onderstal (2011c).
ble according to neologism proofness and announcement proofness. Third, we look at whether the ACDC equilibrium in similar games performs worse when its stability according to ACDC decreases. Finally, we test whether ACDC can explain behavior in case there is a large set of equilibria. The experimental results are supportive of all four hypotheses. This provides evidence that ACDC is able to predict well.

In addition, our design allows us to test a comparative static result concerning the relationship between power and communication. In Chapter 2, we showed that equilibrium size is increasing in the Sender’s relative bargaining power. As ACDC selects the maximum size equilibrium in these games, we hypothesize that the Sender’s relative power increases actual information transmission. One challenge is that while the equilibrium size is an appropriate theoretical measure for information transmission, it is less appropriate for real data. Hence, we develop an empirical measure of information transmission. We find that increasing the Sender’s relative power indeed increases information transmission.

In order to better explain the dynamics of our data, we introduce a ‘neologism dynamic.’ The neologism dynamic is a simple best response dynamic with the additional feature that Senders send credible neologisms, which are also believed by Receivers. In contrast to a best response dynamic and a level-\(k\) analysis, the neologism dynamic is predictive. In particular, the neologism dynamic supports the conclusions of ACDC and can explain the main dynamic characteristics of the data.

Turning to the literature, we see that relatively little experimental work exists on equilibrium selection in cheap talk games. Blume, DeJong & Sprinkle (2001) test the predictions of PCI in a series of discrete games and compare it with neologism proofness. ACDC can explain Blume et al.’s experimental data at least as well as PCI (see Chapter 3.) Experimental work on the Crawford-Sobel uniform quadratic game provides evidence that the most informative equilibrium performs best (Dickhaut, McCabe & Mukherji (1995), Cai & Wang (2006) and Wang, Spezio & Camerer (2010)).

\(46\) This is predicted by NITS (Chen, Kartik &

---

\(46\) The focus of this chapter is on what makes communication of private information credible. A different strand of the experimental literature deals with the question how players can
Sobel, 2008) as well as ACDC. Furthermore, it shows that as the bias parameter becomes smaller, the most informative equilibrium performs better, which is also predicted by ACDC. The present study is the first systematic experimental test of whether and to what extent credible deviations matter for the stability of cheap talk equilibria. In addition, it presents a rigorous test of ACDC in new experiments. Our contribution on power and information transmission adds to a rich experimental bargaining literature. Even so, to the best of our knowledge we are the first to experimentally address the link between bargaining power and information transmission.

This chapter has the following structure. In section 4.2, we discuss the theory we require. We present the experimental games we study, introduce ACDC and discuss the issue of equilibrium selection. In section 4.3, we provide the experimental design. In section 4.4 we present the experimental results in relation to equilibrium selection. In section 4.5, we look at the dynamic aspects of our data and discuss the neologism dynamic. In section 4.6, we introduce and discuss two additional treatments to test the robustness of our results. In section 4.7 we present the experimental results in relation to power and information transmission. Section 4.8 concludes.

4.2 Theory

In this section, we develop the theory we use to construct the hypotheses for our experiment. In subsection 4.2.1 we introduce the game \( G(B) \), on which our main treatments are based, and apply existing refinements to this game. In 4.2.2, we introduce ACDC and show how it works out in \( G(B) \). Since this credibly communicate their intentions (Davis & Holt (1988), Cason (1995), Charness (2000), Ellingsen & Johannesson (2004), Gneezy (2005), Charness & Dufwenberg (2006), Lundquist, Ellingsen, Gribbe & Johannesson (2009), Serra-Garcia, Van Damme & Potters (2010)). Crawford (1998) provides a survey of papers on cheap talk experiments and DellaVigna & Gentzkow (2010) review field evidence on persuasive communication.

section contains all theory needed in this chapter, the material here overlaps with the material in the previous two chapters.

4.2.1 $G(B)$

$G(B)$ is a two-player veto threats cheap talk game between an informed Sender and an uninformed Receiver. The outcome of the game $x$ is a point in the interval $[0,B]$ or the disagreement point $\delta \notin \mathbb{R}$, where $B$ stands for boundary. The payoffs of the Receiver and the Sender are given by $U^R(x)$ and $U^S(x,t)$:

$$U^R(x) = 60 - \frac{2}{5}x \text{ for all } x \in [0,B]$$

$$U^S(x,t) = 60 - |x - t| \text{ for all } x \in [0,B]$$

(4.1) $$U^R(\delta) = U^S(\delta) = 0$$

On the interval, the Receiver prefers smaller outcomes to larger outcomes. The payoffs of the Sender on the interval depend on her type $t$, which is drawn from a uniform distribution on $[0,B]$. The larger the distance between $t$ and $x \in [0,B]$, the lower the Sender’s payoff. Both players receive a payoff of 0 if $\delta$ is the outcome, regardless of $t$. The Receiver prefers all outcomes on the line smaller than 150 to $\delta$; the Sender prefers $\delta$ to all outcomes on the line more than 60 away from her type $t$.

At the start of the game, nature draws a type $t$. Everything is common knowledge, except $t$. The game then proceeds as follows. Nature informs the Sender of $t$. Subsequently, the Sender sends a message $m$ to the Receiver. Next, the Receiver proposes action $a \in [0,B]$. Finally, the Sender accepts or rejects. If the Sender accepts, $a$ is the outcome of the game, and if she rejects, $\delta$ is the outcome. Note that all messages are costless for the Sender. We assume $B \geq 120$ because under this condition the boundary does not affect the set of equilibrium actions. The model is close to the cheap talk game with veto threats.
of Matthews (1989). The main difference is that in our model the disagreement point does not lie on the interval.

We consider the following type of equilibria. Following Matthews, we require the Receiver to play pure strategies and require Senders to always induce actions that are payoff-maximizing (in the spirit of Selten’s (1975) trembling hand perfection). From now on, we will use ‘equilibrium’ to refer to a perfect Bayesian equilibrium satisfying these two requirements. As we showed in Chapter 2, all equilibria in such veto threat games are partition equilibria. A partition equilibrium can be characterized by the finite set of actions \( A^* = \{a_1, a_2, \ldots, a_n\} \) the Receiver proposes in equilibrium, where \( a_1 < a_2 < \ldots < a_n \). The number of equilibrium actions \( n \) is called the size of the equilibrium. Each type induces an action \( a_i \in A^* \) which maximizes her payoff and accepts it if and only if \( U^S(a_i, t) \geq 0 \). (In \( G(B) \) the payoff maximizing action is simply the action closest to her type.) We say a type \( t \) induces an action \( a' \), and write \( a(t) = a' \), if the Receiver proposes \( a' \) after any message \( m \) the Sender sends. This means that a partition of the type space \( 0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1 \) exists such that each type in \( (t_{i-1}, t_i) \) induces \( a_i \).

It is straightforward to check that a set of actions \( a_1 < a_2 < \ldots < a_n \) characterizes a partition equilibrium if and only if

\[
a_i \in \arg\max_{a \in A^*} \int_{t_{i-1}}^{t_i} U^R(a) f_{\{t \mid U^S(a, t) \geq 0\}}(t) dt \quad \text{for all } i = 1, \ldots, n, \text{ where}
\]

\[
t_0 = 0, t_n = 1 \text{ and } U^S(a_i, t_i) = U^S(a_{i+1}, t_i) \text{ for all } i = 1, \ldots, n - 1.
\]

The game has two equilibria: a pooling (size-1) equilibrium and a (semi)separating size-2 equilibrium. For both equilibria, the set of equilibrium actions the Receiver takes does not depend on \( B \):

---

48 There is an infinite number of equilibria that induce the same equilibrium outcome. These essentially equivalent equilibria just differ in the messages that are used. For simplicity, we refer to a class of equilibria inducing the same equilibrium outcome simply as ‘an equilibrium.’

49 There will be a set of measure zero of types for which \( U^S(a_i, t) = U^S(a_{i+1}, t) \) for some \( i \). It does not matter which action they induce.
Proposition 4.1 \( G(B) \) has two equilibria: a pooling equilibrium \( \sigma^p \) \( \{a_1 = 45\} \) and a separating equilibrium \( \sigma^s \) \( \{a_1 = 0, a_2 = 60\} \).

In the pooling equilibrium, all types induce 45, whereas in the separating equilibrium all Sender types in the interval \([0, 30)\) induce action \(a = 0\) and all Sender types in the interval \((30, B]\) induce action \(a = 60\). In the separating equilibrium, the Sender always accepts, and in the pooling equilibrium all Senders in \([0, 105]\) accept.

The intuition behind the proposition is the following. Since all equilibria are partition equilibria, the Receiver’s posterior beliefs consist of intervals. If the Receiver believes the Sender’s type is uniformly distributed on an interval \([t, \overline{t}]\), he faces the following trade-off when looking for a best response. As the proposed action increases (up to \(\min\{t + 60, \overline{t} - 60\}\) ), the probability of acceptance increases but the utility conditional on acceptance decreases. Senders best respond by inducing the action closest to their type. As in any cheap talk game, there is a pooling equilibrium in which all Senders employ the same message strategy and the Receiver ignores all messages. In \( G(B) \), also a size-2 equilibrium exists. Higher size equilibria do not exist, roughly because there is a minimum distance between two positive equilibrium actions and a maximum to the value an equilibrium action can take.

At this point two questions arise. Is one equilibrium more plausible than the other? And does \( B \) influence the stability of the equilibria? We first turn to theories of credible deviations. The neologism proofness criterion of Farrell (1993) is based on the concept of neologisms. Neologisms are out-of-equilibrium messages which are assumed to have a literal meaning in a pre-existing natural language. Farrell considers neologisms which literally say: “propose action \( \tilde{a} \), because my type is in set \( N \).” Farrell deems a neologism credible if and only if \( i \) all types \( t \) in \( N \) prefer \( \tilde{a} \) to their equilibrium action \( a(t) \), \( ii \) all types \( t \) not in \( N \) prefer their equilibrium action \( a(t) \) to \( \tilde{a} \) and \( iii \) the best response of the
Receiver after restricting the support of his prior to \( N \) is to play \( \tilde{a} \).\(^{50}\) We will denote credible neologisms by \( \langle \tilde{a}, N \rangle \). An equilibrium is neologism proof if it does not admit any credible neologism. Farrell argues that only neologism proof equilibria are stable, since rational players would move away from equilibria which admit credible neologisms. (The predictions of announcement proofness (Matthews, Okuno-Fujiwara & Postlewaite, 1991) are equivalent to those of neologism proofness in \( G(B) \).\(^{51}\)

In our game two types of credible neologisms can exist. A ‘low’ neologism which roughly says “I am a low type and prefer 0 to the lowest equilibrium action and so do you, so play 0” and ‘high’ neologism which roughly says “I am a high type, and it is probable that I will not accept the highest equilibrium action, so it is better for both of us if you propose something higher.” As the following proposition shows, the pooling equilibrium is never neologism proof and the separating equilibrium is only neologism proof if \( B = 120 \).

**Proposition 4.2** The pooling equilibrium admits the credible neologisms \( \langle 0, [0, 22.5] \rangle \) and \( \langle \min \{ B - 60, 75 \}, \min \{ \frac{B - 15}{2}, 60 \}, B \rangle \). The separating equilibrium is neologism proof if \( B = 120 \). For \( B > 120 \), the separating equilibrium admits the credible neologism \( \langle \min \{ B - 60, 80 \}, \min \{ \frac{B}{2}, 70 \}, B \rangle \).

So, for \( B > 120 \) neologism proofness (and announcement proofness) are silent about the stability of the separating equilibrium relative to the pooling equilib-

\(^{50}\) Farrell does not consider cheap talk games with veto threats. In line with the trembling hand refinement, we assume that types induce a neologism if \( U^S(\tilde{a}, t) > U^S(a(t), t) \).

\(^{51}\) Matthews, Okuno-Fujiwara & Postlewaite (1991) consider more elaborate messages, called announcements, and propose three types of credible deviations. Weakly credible announcements are similar to neologisms, but allow deviating types to distinguish amongst themselves. A weakly credible announcement that should be believed if the Receiver’s realizes that types can send multiple announcements is a credible announcement. A credible announcement that survives a rigorous Stiglitz-critique is strongly credible. Equilibria that admit no (weakly/strongly) credible announcements are called (strongly/weakly) announcement proof. In our game, strong announcement proofness, announcement proofness and neologism proofness coincide. The reason is that all weakly credible announcements are equivalent to a credible neologism (for a similar reason that there is at most a size-2 equilibrium) and that all types can send at most one credible neologism. All equilibria in \( G(B) \) are weakly announcement proof, as in almost all cheap talk games.
rium. In addition, they are silent about whether the separating equilibrium is more stable if \( B = 121 \) than, say, if \( B = 210 \).

The same holds for other cheap talk refinements including communication proofness (Blume & Sobel, 1995) and the recurrent mop (Rabin & Sobel, 1996), as well as the non-equilibrium concepts of Credible Message Rationalizability (Rabin, 1990) and Partial Common Interest (Blume, Kim & Sobel, 1993). (See Chapter 3 for details.)

The NITS (Chen, Kartik & Sobel, 2008) criterion, which successfully predicts in the Crawford-Sobel (1982) game, is partially predictive in \( G(B) \). NITS starts by specifying a ‘lowest type,’ a type with the property that all other types prefer to be revealed as themselves rather than as that lowest type. An equilibrium survives NITS if the lowest type has no incentive to separate, i.e. if the lowest type prefers her equilibrium outcome to the outcome she would get if she could reveal her type. In our game, such a ‘lowest type’ cannot easily be formulated. All types in \([0,60]\) are in fact lowest types according to Chen et al.’s definition. The pooling equilibrium survives NITS relative to types in \([22.5,105]\), whereas the separating equilibrium survives NITS relative to types in \([0,30]\). Still, one can argue that \( t = 0 \) is a natural lowest type in our game. Under this assumption, for each \( B \) only the separating equilibrium is NITS in our game. Hence, NITS would predict that the separating equilibrium is always stable regardless of \( B \).

Finally, some may argue that the most influential equilibrium (i.e. the equilibrium which induces the largest number of actions) is the most plausible equilibrium, aside of any stability considerations. In our game, this criterion also selects the separating equilibrium regardless of \( B \).

In sum, existing criteria provide no or a partial answer to the question how stable equilibria in \( G(B) \) are for \( B > 120 \).

4.2.2 ACDC in \( G(B) \)

Our conjecture is that two aspects will affect the behavioral stability of an equilibrium. The first concerns the mass of types that can credibly induce a deviation. The smaller this mass becomes, the less unstable an equilibrium will
be, as it will be disturbed less frequently. The second aspect concerns how much the deviation profile differs from the equilibrium profile in terms of Sender payoffs. The smaller this difference becomes, the smaller both the Sender’s incentive to deviate and the perturbation to the equilibrium if she deviates will be.

For instance, the separating equilibrium is not neologism proof if \( B = 121 \). However, we do not expect behavior in the game \( G(121) \) to be very different to behavior in \( G(120) \). After all, the induced deviations from equilibrium are very small: types in \([60.5,121] \) induce 61 instead of 60. Hence, Senders can at most earn 1 by deviating and, if they deviate, the resulting profile is very similar to the equilibrium profile. In contrast, in the pooling equilibrium the neologism deviations are substantial: types from 0 to 22.5 deviate from 45 to 0, and types from 53 to 121 deviate from 45 to 61. As a consequence, Senders have a large incentive to deviate to a profile which is very different from the equilibrium profile. Furthermore, the separating equilibrium seems more stable if \( B = 121 \) than if, say, \( B = 210 \), when types in \([70,210] \) can credibly induce 80 rather than 60.

In Chapter 3, we formalized these ideas in the concept of the Average Credible Deviation (ACD).\(^{52}\) Let \( a^\ast(t) \) be the equilibrium action induced by type \( t \) in equilibrium \( \sigma \); and let \( \bar{a}^\ast(t) \) be the deviating action type \( t \) induces if she plays a credible neologism.\(^{53}\) Let \( \bar{a}^\ast(t) = a^\ast(t) \) if Sender type \( t \) cannot play a credible neologism. Finally, we define \( \underline{U}^\ast(t) \) and \( \overline{U}^\ast(t) \) as the lowest and highest payoff a Sender can get if both players play a rationalizable strategy. Now, for each Sender type \( t \), we specify the size of the credible deviation from equilibrium, \( CD(t,\sigma) \). The ACD is the expected value of the credible deviations. We measure the size of a credible deviation by the Sender’s incentive to deviate relative to the largest (rationalizable) incentive possible, so that it lies on a scale between 0 and 1. The higher this incentive is, the higher the probability that a Sender will

\(^{52}\) There we also provided a more general and rigorous treatment of the concept. Here we restrict ourselves to the definitions needed in the current setting.

\(^{53}\) In \( G(B) \), ACDC gives equivalent results under weakly or ordinary credible announcements as under credible neologisms, as is the case in many games. For cases where the theories differ, we prefer (ordinary) credible announcements for discrete games and credible neologisms for continuous games.
deviate and the larger the upheaval such a deviation can cause to an equilib-

r

rium. In particular, we define the credible deviation for type $t$ as

\begin{equation}
CD(t, \sigma) \equiv \frac{U^S(t, \alpha^*(t)) - U^S(t, a^*(t))}{U^S(t) - \underline{U}^S(t)}
\end{equation}

whenever $U^S(t, a^*(t)) > \underline{U}^S(t)$. If $U^S(t, a^*(t)) = \underline{U}^S(t)$, the Sender has no incentive to adhere to her equilibrium strategy, as she can do no worse by deviating, and we set $CD(t, \sigma) \equiv 1$. The ACD of equilibrium $\sigma$ is now defined as

\begin{equation}
ACD(\sigma) = E_t[CD(t, \sigma)]
\end{equation}

Observe that $ACD(\sigma) \in [0,1]$. We formulate the ACD-Criterion (ACDC) that an equilibrium $\sigma$ is more stable than an equilibrium $\sigma'$ if $ACD(\sigma) < ACD(\sigma')$. Using ACDC, we can select equilibria. In particular, we call an equilibrium an ACDC equilibrium if there is no other equilibrium in the game with a lower ACD. We consider an ACDC equilibrium as the most plausible equilibrium, i.e. that which will predict best on average, rather than the equilibrium that will always be played all of the time. ACDC can select equilibria when neologism proofness is silent and reduces to the latter if neologism proof equilibria exist.

The following proposition gives the results of ACDC for $G(B)$.

**Proposition 4.3** The separating equilibrium $\sigma^S$ is the unique ACDC equilib-

r

rium. Furthermore, the ACD of the separating equilibrium is 0 for $B = 120$ (in which case the equilibrium is neologism proof) and strictly increasing in $B$.

We can now see why $G(B)$ provides a good testing ground for our ideas. It contains the features that make (continuous) cheap talk games difficult to refine. In contrast to Crawford & Sobel’s (1982) and Matthews’ (1989) cheap talk models, however, in our game a parameter value exists such that there are multiple equilibria, of which only one is neologism proof. Hence, our model allows us to test the relevance of credible deviations in a continuous setting.
Furthermore, it allows us to test the idea that stability is not all-or-nothing. First, we can compare within a game two equilibria that are not neologism proof. Second, across games we can gradually increase the number of types that can credibly deviate in an equilibrium by increasing $B$.

### 4.3 Experimental Design and Procedures

We ran five treatments. In three treatments, we ran $G(B)$ with an increasing boundary: $G(120)$, $G(130)$ and $G(210)$. In addition we ran two additional robustness treatments T4 and T5 to which we return in section 4.6. Table 4.1 summarizes the theoretical properties of the experimental treatments. For each treatment, we have six matching groups, each consisting of 10 subjects (5 Senders and 5 Receivers).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$U^R(\delta)$</th>
<th>$U^S(\delta)$</th>
<th>$B$</th>
<th>Equilibria $^1$</th>
<th>ACD $^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>G(120)</td>
<td>0</td>
<td>0</td>
<td>120</td>
<td>${45}$, ${0, 60}$**</td>
<td>0</td>
</tr>
<tr>
<td>G(130)</td>
<td>0</td>
<td>0</td>
<td>130</td>
<td>${45}$, ${0, 60}$*</td>
<td>0.22</td>
</tr>
<tr>
<td>G(210)</td>
<td>0</td>
<td>0</td>
<td>210</td>
<td>${45}$, ${0, 60}$*</td>
<td>0.50</td>
</tr>
<tr>
<td>T4</td>
<td>30</td>
<td>30</td>
<td>120</td>
<td>${45/2}$, ${0, 30}$*</td>
<td>0.56</td>
</tr>
<tr>
<td>T5</td>
<td>30</td>
<td>30</td>
<td>120</td>
<td>${30}$, ${a_t, a_i + 60}$, ${0, a_2 + 60}$**</td>
<td>3,4</td>
</tr>
</tbody>
</table>

Notes: In each game, $t$ was uniformly distributed on the integers in $[0, B]$. $U^R(x) = 60 - \frac{2}{3}x$ and $U^S(x, t) = 60 - |t - x|$. $^1$An equilibrium has a * if it is ACDC and ** if it is neologism proof as well. $^2$ The ACD of the ACDC equilibrium. $^3$ $a_t \in [0, 30]$ and $a_i \in (0, 30]$. $^4$ Only $\{0, 30, 60\}$ is ACDC.

We used a standard procedure to recruit subjects from the student population of the University of Amsterdam. The computerized experiment was run at the CREED lab. The software was written with z-Tree (Fischbacher, 2007).

At the start of the experiment, subjects were randomly assigned to the role of Sender (‘chooser’ in the terminology of the experiment) or Receiver (‘proposer’). Subjects kept the same role throughout the whole experiment. Subjects read the role-specific instructions on paper at their own pace. (See appendix 4.9.4 for the
instructions.) After reading the instructions, subjects had to answer several questions testing their understanding of the instructions. Only when all subjects had answered all questions correctly, the experiment started.

Subjects received a starting capital of 100 points. In addition, subjects earned points with their decisions in each of the 50 periods. (Subjects were informed that the experiment would last for approximately 50 periods.) At the end of the experiment, total point earnings were exchanged to euros at a rate of 1.5 euros for 100 points. In a session, we ran 2 matching groups simultaneously, each consisting of 5 Senders and 5 Receivers. In every period, each Sender was randomly rematched with a Receiver in the own matching group. In total, 300 subjects participated, who on average earned 28.30 euros in approximately 2 hours, with a minimum of 10.10 euros and a maximum of 40.47 euros. Each subject only participated once.

The procedure within a period was as follows. In each period, the Sender was informed of her own type. All subjects knew that each individual Sender’s type in each period was an independent draw from the uniform distribution on $[0, B]$. After having been informed of the own type, each Sender sent a message (‘suggestion’ in the terminology of the experiment) to the Receiver. The Receiver was informed of the message but not of the Sender’s type. Then the Receiver chose an action (‘made a proposal’) that was either accepted or rejected by the Sender. Types, messages and actions were confined to integers in $[0, B]$.\footnote{To maximize the comparability of the treatments, we drew three sets of types for one treatment and then rescaled these sets for each of the other treatments.}

Payoffs were then calculated according to the payoffs in Table 4.1. At the end of the period, Senders and Receivers were informed of the state of the world (the Sender type) and all the decisions made by the pair they were part of. In addition, each subject was shown her own payoff and how it was calculated. At any moment, subjects were provided with information about the social history in order to facilitate learning.\footnote{We chose for this restricted message space instead of a free chat in order to be able to provide a history screen, facilitate learning and have data that can be interpreted clearly. Notice that the message space is rich enough for the communication of all credible neologisms in both equilibria, as in our game a neologism action uniquely identifies a credible neologism.}

At the bottom of the screen they saw how play...
had unfolded in the 15 most recent periods in their own matching group. For Senders the information was organized as follows. The left-hand side showed a table summarizing the choices of the pairs in the own matching group. Each row contained a pair’s suggestion (message), proposal, acceptance and preferred outcome (type of the Sender). The table was first sorted on suggestion, then on proposal, acceptance and finally on preferred outcome. The right-hand side showed the corresponding graph that listed the proposals as function of the suggestions. Figure 4.1 shows an example of the information that Senders received.

<table>
<thead>
<tr>
<th>Suggestion</th>
<th>Proposal</th>
<th>Acceptance</th>
<th>Preferred outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>35</td>
<td>Yes</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>15</td>
<td>No</td>
<td>70</td>
</tr>
<tr>
<td>50</td>
<td>40</td>
<td>No</td>
<td>20</td>
</tr>
<tr>
<td>55</td>
<td>65</td>
<td>Yes</td>
<td>50</td>
</tr>
</tbody>
</table>

Figure 4.1
Example of social history screen History Screen (Senders)

For Receivers the information was communicated in a slightly different way. In their table, each row listed a pair’s suggestion, preferred outcome, proposal and acceptance. The table was first sorted on message, then on preferred outcome, proposal and finally on acceptance. In their graph, preferred outcomes were shown as function of the suggestions.

### 4.4 Experimental Results

In this section, we deal with the issue of equilibrium performance. Therefore we focus exclusively on the final 15 periods of the experiment. In section 4.5, we
will take a closer look at the dynamics in the data and also use the data from early periods. All statistical tests have been done treating each matching group as one independent data point. For comparisons within a treatment we use Wilcoxon signed rank tests, and for comparisons between treatments we use Mann-Whitney ranksum tests. *, **, *** indicate significance at the 10%, 5% and 1% level respectively (for two-tailed tests).

We start with some descriptive statistics of the final 15 periods of the experiment.

Figure 4.2 plots the Receivers’ actions as a function of the Senders’ types for each of the three treatments together with the equilibrium predictions. Consistent with the separating equilibrium we find that Senders with low types tend to elicit an action of 0, while the Senders with high types tend to trigger a high action. Types close to the equilibrium indifferent type of 30 show a more continuous separation than the equilibrium step-function. In the G(120) treatment, where the separating equilibrium is neologism proof, the distribution of actions high type Senders trigger is similar and centered around 60. This is in agreement with the equilibrium prediction. We see a similar pattern in G(130), although the mean high action is now somewhat higher than the equilibrium action, as expected if high types send a neologism. In G(210), the elicited actions by high types are far off from equilibrium and much more dispersed.

The equilibrium analysis assumes that Senders do not leave money on the table, i.e., they are supposed to accept any action that gives them positive
payoff. Overall, Senders accepted proposed actions that would give them a positive payoff in 96% of the cases. Table 4.2 presents the actual acceptance frequencies as function of the Sender’s payoff (in the rows) and the Sender’s share in the total payoff (in the columns). It is remarkable that Senders almost always accept ‘decent’ proposals that give them at least 10, independent of their share in the total payoff. The share in total payoff only matters when Senders receive ‘peanuts’ proposals with payoffs below 10. As a result, the equilibrium assumption about Senders’ acceptance behavior is by and large supported in the data. Notice that the picture about acceptance rates differs from results in ultimatum games, where subjects tend to reject proposed actions more often (Oosterbeek, Sloof & Van der Kuilen, 2004). A crucial difference between our bargaining game and the ultimatum game is that in our game the Receiver is not informed of the type of the Sender. Therefore, unlike in the ultimatum game, it is unclear whether an unfavorable proposal is made intentionally.

### TABLE 4.2

<table>
<thead>
<tr>
<th>Payoff Sender</th>
<th>Total</th>
<th>0-10%</th>
<th>10-20%</th>
<th>20-30%</th>
<th>30-40%</th>
<th>40-50%</th>
<th>&gt;50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0</td>
<td>0</td>
<td>211</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>=0</td>
<td>50</td>
<td>6</td>
<td>50</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0-10</td>
<td>70</td>
<td>112</td>
<td>40</td>
<td>30</td>
<td>73</td>
<td>49</td>
<td>90</td>
</tr>
<tr>
<td>10-20</td>
<td>92</td>
<td>122</td>
<td>-</td>
<td>100</td>
<td>2</td>
<td>87</td>
<td>69</td>
</tr>
<tr>
<td>20-30</td>
<td>99</td>
<td>135</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>13</td>
<td>100</td>
</tr>
<tr>
<td>30-40</td>
<td>100</td>
<td>182</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>28</td>
</tr>
<tr>
<td>40-50</td>
<td>100</td>
<td>250</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>50-60</td>
<td>100</td>
<td>332</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>Total when payoff&gt;0</td>
<td>40</td>
<td>36</td>
<td>75</td>
<td>51</td>
<td>89</td>
<td>111</td>
<td>99</td>
</tr>
</tbody>
</table>

Notes: the cells show the acceptance rates (in percentages) as a function of the surplus payoff the Sender would receive if she would accept the proposed action (in the rows) and her share of the surplus (in the columns, cases where the surplus was not positive were dropped). The number of observations pertaining to the cell is listed in italics. The Sender’s share is defined as $100 \times \frac{U^S(a,t)}{U^S(a,t) + U^A(a)}$.  

Table 4.3 reports how often actual play was close to equilibrium. We say that an outcome is consistent with equilibrium (‘correctly predicted’) if the actual action lies within a bandwidth of 10 and if the acceptance decision was correctly
predicted. The absolute numbers in the table are obviously dependent on the chosen bandwidth. Here, we are interested in the relative magnitudes of the numbers, for which the exact level of the bandwidth turns out not to matter. In addition, we also look at the average (absolute) prediction error of the equilibria, reported in Figure 4.3. Let $a^s(t)$ be the equilibrium action of the Receiver given type $t$ and $\hat{a}_i(t)$ the observed action for observation $i$. The average prediction error (for a set of $n$ observations $I$) is then $\frac{1}{n} \sum_{i \in I} |\hat{a}_i(t) - a^s(t)|$. The percentage of outcomes that are correctly predicted is an intuitive measure of predictive success, whereas the average prediction error a parameter-free and precise measure. The results are qualitatively identical (and equally significant) for both measures of predictive success.

### TABLE 4.3

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Equilibria</th>
<th>Diff Equilibria</th>
<th>Observations with $t &lt; 120$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooling</td>
<td>Separating</td>
<td>Pooling</td>
</tr>
<tr>
<td>G(120)</td>
<td>.39</td>
<td>.69</td>
<td>.30**</td>
</tr>
<tr>
<td>G(130)</td>
<td>.13</td>
<td>.57</td>
<td>.43**</td>
</tr>
<tr>
<td>G(210)</td>
<td>.09</td>
<td>.18</td>
<td>.09**</td>
</tr>
<tr>
<td>Dif 120-130</td>
<td>.25**</td>
<td>.12</td>
<td>.25**</td>
</tr>
<tr>
<td>Dif 120-210</td>
<td>.30***</td>
<td>.51***</td>
<td>.29**</td>
</tr>
<tr>
<td>Dif 130-210</td>
<td>.05**</td>
<td>.39***</td>
<td>.05</td>
</tr>
</tbody>
</table>

*Notes: The table shows per treatment the median (over matching groups) of the fraction of correctly predicted outcomes by the equilibrium. We classified a prediction as correct if both (i) the distance between the predicted and observed action was not larger than 10 and (ii) the acceptance decision was correctly predicted. We used the data of the last 15 periods. ‘Dif’ denotes ‘difference.’*

We first deal with the question whether credible deviations (and hence neologism proofness and announcement proofness) have a bite. For this question, two comparisons are relevant. First, within treatment G(120) the separating equilibrium is neologism proof while the pooling equilibrium is not. In accordance with neologism proofness, more outcomes are consistent with the separating equilibrium than with the pooling equilibrium. The difference is both substantial and statistically significant. Second, when we move from treatment G(120) to G(210), the separating equilibrium ceases to be neologism proof. While the
separating equilibrium does a good job in G(120), it organizes only a bleak minority of the data in G(210). This conclusion is valid when we take the data for all types as well as when we condition on the outcomes with types less than 120. So also when we compare the behavioral stability of the same equilibrium across treatments, we find support for neologism proofness.

To investigate whether credible deviations matter gradually and ACDC is relevant across games, we compare G(120), G(130) and G(210). In G(130), the separating equilibrium is no longer neologism proof but the ACD measure remains rather small relative to G(210). So if ACDC makes sense, the results of G(130) should be closer to G(120) than to G(210). Table 4.3 confirms that this is indeed the case. Like in G(120), the separating equilibrium is much more successful than the pooling equilibrium. The separating equilibrium predicts behavior a bit less in G(130) than in G(120), but the difference is not significant.\textsuperscript{57} In contrast, the difference between G(130) and G(210) is much larger. In G(210), significantly fewer cases are consistent with the separating equilibrium than in G(130). So even though the separating equilibrium is not neologism proof in either treatment, it traces the data much better in G(130) than in G(210), as predicted by ACDC.

Regarding equilibrium selection, observe that in all treatments the ACDC (separating) equilibrium predicts significantly better than the non-ACDC (pooling) equilibrium. This is also the case for G(130) and G(210), where neither equilibrium is neologism proof. Still, in G(210), the difference between the separating and the pooling equilibrium is much smaller than in the other treatments, in line with ACDC.

The power of ACDC is further illustrated in Figure 4.3. This figure lists the average prediction error of a particular equilibrium and treatment as a function of its ACD. In agreement with ACDC, the higher the ACD measure, the larger the average prediction error tends to be. Notice in particular that the average prediction error of the separating equilibrium only rises slightly when it just ceases to be neologism proof (relative to the differences with G(210)).

\textsuperscript{57} Since this result also holds when the analysis is restricted to observations with \( t \leq 120 \), this is not a measurement artifact due to a change in the interval of measurement.
In this section, we look at the dynamics. In subsection 4.5.1, we describe the most important dynamic features of the data. In 4.5.2, we observe that an elementary best response model is not predictive. In 4.5.3, we introduce the neologism dynamic, which is able to explain important parts of the data.

4.5 Dynamics

In this section, we look at the dynamics. In subsection 4.5.1, we describe the most important dynamic features of the data. In 4.5.2, we observe that an elementary best response model is not predictive. In 4.5.3, we introduce the neologism dynamic, which is able to explain important parts of the data.

4.5.1 Dynamics in Experiment

In this section we deal with the question how subjects adapted their behavior during the experiment. Figure 4.4 plots messages conditional on Sender type (left-hand side), actions conditional on message received (middle) and actions conditional on Sender type (right-hand side). We present plots for the first 15 and last 15 periods in each treatment. The type-message plots show that Senders’ messages are higher than their types and that Senders learn gradually to

Figure 4.3
The figure plots for each equilibrium in each treatment its theoretical ACD against its empirical prediction error. We report the median (over matching groups) of the average prediction error. We used the data of the last 15 periods.
exaggerate more. In the last 15 periods of each treatment, Senders overstate the true state more than in the first 15 periods. Thus, there is ‘language inflation.’

Receivers’ action-message plots provide the mirror image of Senders’ type-message plots. That is, in the first part of the experiment Receivers tend to propose actions slightly below the messages received. In the final part of the experiment, Receivers have learned to subtract larger amounts from the messages received. The type-action plots on the right hand side illustrate how close the actually triggered actions are to the equilibrium predictions. For treatments G(130) and G(120), the data are closer to the separating equilibrium in the final part of the experiment than in the first part of the experiment. A similar trend is not observed in G(210). To the contrary, in this treatment the data remain far from equilibrium throughout the whole experiment.

We now turn to the questions how easily subjects reached the separating equilibrium in the different treatments and how likely it was that they stayed there. Table 4.4 presents the relevant statistics separately for the first part (first 15 periods) and the final part (last 15 periods) of the experiment. In the first part of the experiment, subjects more easily reached the separating equilibrium from a state of disequilibrium in treatments G(120) and G(130) than in G(210). When subjects were approximately playing according to the separating equilibrium in the previous period, they were much more likely to stay there in treatments G(120) and G(130) than in treatment G(210). The lower part of the table shows that the differences between treatment G(210) and the other treatments became even more pronounced in the final part of the experiment. In particular, in G(120) and G(130) the separating equilibrium attracts more outcomes in the final part than in the first part (and in G(120) significantly so). In contrast, in G(210) the separating equilibrium attracts less outcomes in the final part; in fact, it hardly attracts any outcomes in the final part.
Figure 4.4

This figure compares the chosen strategies (type-message, message-action) and the resulting profile for the first 15 and last 15 rounds. The bubble plots are clustered on a grid of 10. In the last column, the solid line represents the separating equilibrium and the dotted line the pooling equilibrium.
TABLE 4.4
FA RATION OF MATCHING GROUP OBSERVATIONS IN SEPARATING EQUILIBRIUM CONDITIONAL ON PREVIOUS STATE

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Previous state</th>
<th>Previous state</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no equilibrium</td>
<td>equilibrium</td>
</tr>
<tr>
<td><strong>First Part</strong> (Periods 1-15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G(120)</td>
<td>.46</td>
<td>.60</td>
</tr>
<tr>
<td>G(130)</td>
<td>.41</td>
<td>.51</td>
</tr>
<tr>
<td>G(210)</td>
<td>.10</td>
<td>.09</td>
</tr>
<tr>
<td>Dif 120-130</td>
<td>.05</td>
<td>.09</td>
</tr>
<tr>
<td>Dif 120-210</td>
<td>.36***</td>
<td>.51***</td>
</tr>
<tr>
<td>Dif 130-210</td>
<td>.31***</td>
<td>.42***</td>
</tr>
<tr>
<td><strong>Final Part</strong> (Periods 36-50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G(120)</td>
<td>.64</td>
<td>.74</td>
</tr>
<tr>
<td>G(130)</td>
<td>.50</td>
<td>.72</td>
</tr>
<tr>
<td>G(210)</td>
<td>.03</td>
<td>.00</td>
</tr>
<tr>
<td>Dif 120-130</td>
<td>.14</td>
<td>.02</td>
</tr>
<tr>
<td>Dif 120-210</td>
<td>.60***</td>
<td>.74***</td>
</tr>
<tr>
<td>Dif 130-210</td>
<td>.46***</td>
<td>.72***</td>
</tr>
<tr>
<td><strong>Difference between First and Final Part</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G(120)</td>
<td>.18**</td>
<td>.14*</td>
</tr>
<tr>
<td>G(130)</td>
<td>.09</td>
<td>.21</td>
</tr>
<tr>
<td>G(210)</td>
<td>.07</td>
<td>-.09</td>
</tr>
</tbody>
</table>

**Notes:** A matching group observation in a given period is classified as consistent with the separating equilibrium prediction if and only if the acceptance decision was predicted correctly and \( |\text{equilibrium action} - \text{observed action}| \leq 10 \) for at least 3 of the 5 pairs in the matching group; the middle column displays the fraction of equilibrium observations given that the previous observation was not in equilibrium; the right column displays the fraction of equilibrium observations given that the previous observation was in equilibrium.

Summarizing, the two main features of the dynamics in the data are (i) there is language-inflation and (ii) the separating equilibrium attracts behavior over time in G(120) and to a slightly lesser extent in G(130), but not at all in G(210).

**4.5.2 Best Response Dynamic**

A first avenue to look to explain the data is a simple best response model. This, however, cannot distinguish between the two equilibria or the effects of \( B \) using natural initial conditions. (We get equivalent results for a level-\( k \) analysis.) Consider the simplest best response dynamic, in which Sender and Receiver best respond to the other’s strategy in the previous period. We again assume that Senders induce the action they prefer most. If Senders are indifferent, they
randomize between their optimal actions. The outcome of the best response model depends very much on the initial conditions and we will look at the two natural starting points: a babbling strategy and a naive strategy. In the babbling strategy no information is transmitted: Senders randomize in the interval $[0, B]$ and Receivers take the optimal prior action 45 regardless of the message (this corresponds to a random level-0 in a level-\(k\) analysis). In the naive strategy all information is transmitted: Senders report their type and this is believed by Receivers (this corresponds to a truthful level-0 in a level-\(k\) analysis). It is readily verified that if players babble in the first period, then the dynamic forever stays in this pooling equilibrium, regardless of the boundary. Similarly, if players use a naive strategy in the first periods, it can be shown that the dynamic converges to the separating equilibrium, regardless of the boundary.

### 4.5.3 Neologism Dynamic

We introduce a small twist to create a ‘neologism dynamic’: Sender types who can send a credible neologism with respect to the Receiver’s strategy in the previous round will do so and such a credible neologism will be believed. In all other respects, the dynamic is the same as above. If we analyze this dynamic for our experimental treatments ($G(120)$, $G(130)$, $G(210)$), we get entirely different results. First, the outcome becomes less dependent on the initial conditions. Second, the dynamic converges to behavior that resembles the separating equilibrium when the ACD is small and only when it is small.\(^{58}\) Finally, in the attractor of $G(130)$ and $G(210)$, types close 30 do not separate neatly as they would in the separating equilibrium. (A level-\(k\) with neologisms analysis yields qualitatively the same result as the best response dynamic.)

In $G(120)$, the dynamic converges to a steady state that corresponds to the separating equilibrium for both random and naive first-period strategies. If players have a naive strategy in period 1, then Senders realize in period 2 that they should send their type plus 60, leading to an inflation of language. Types higher than 60 pool at the highest message of 120. In period 3, Receivers recog-

\(^{58}\) Although (in line with ACDC) the dynamic converges in all cases to behavior that is closer to the separating than to the pooling equilibrium.
nize the language inflation and propose 0 to any message smaller than 120. In addition, they propose 60 if they receive 120. In period 4, the players are already in the separating equilibrium. Note that as long as the Receiver proposes 0 and an action higher or equal to 60, no neologism can be played. Suppose players start with a babbling strategy. Then in period 2, Senders in \([0,22.5)\) send a low neologism of 0 and Senders in \((57.5,120]\) send a high neologism of 60. In period 3, the Receiver realizes that types who do not send a neologism accept 0, and propose 0 to them and 60 to others. As a result, in period 4 equilibrium is reached.

In G(130), the dynamic starts out (for both initial conditions) similar to G(120) but does not converge to the separating equilibrium. Instead, the dynamic converges to a four-cycle that, nonetheless, stays pretty close to the separating equilibrium. In G(210), the dynamic converges to a (non-steady) state, where the Receiver proposes actions 0, 30 and 90, and the Senders in \((90,210]\) send a neologism of 90. Hence, the dynamic does not come close to the separating equilibrium. (Although it comes closer to the separating than to the pooling equilibrium.) We summarize the findings in Figure 4.5. We have held the discussion of the neologism dynamic informal here. For details and proofs, we refer to appendix 4.9.3. The calculations are straightforward, but tedious.

Observe that the results of the neologism dynamic are in line with the two main dynamic features of the experiment: (i) there is language-inflation and (ii) the separating equilibrium attracts behavior in G(120) and to a slightly lesser extent in G(130), but not at all in G(210). (See Figure 4.4.) In line with the neologism dynamic is, furthermore, the fact that the prediction error of the separating equilibrium decreases over time in G(120) and G(130), but not in G(210). Finally, note that in G(130) and G(210) types close to 30 (the indifference type in the separating equilibrium) do not separate neatly into low and high types. Hence, the dynamic predicts a messy separation close to the equilibrium indifference type for these treatments, which we observe in the data.

Our neologism dynamic is a parsimonious behavioral model that organizes the main patterns in the data even though it ignores some features that may also play a role. Firstly, it abstracts away from noise, which is present in the data (as is the case in most experiments). Hence, adding noise to our analysis,
in the spirit of Quantal Response (McKelvey & Palfrey, 1995; 1998) would
certainly make it more realistic. Nonetheless, in this case formalizing noise
would not teach us much more about the data except that agents best respond
in a noisy manner. In particular, an Agent Quantal Response Equilibrium (A-
QRE) analysis per se cannot (meaningfully) select equilibria in cheap talk
games: The pooling equilibrium is always a limiting principal branch A-QRE,
while it is often implausible.\(^{59}\) Secondly, there is somewhat more separation close
to the indifference type than predicted by the separating equilibrium. Although
this is predicted by the neologism dynamic when that equilibrium is not neolo-
gism proof, other forces may also be in play here. In particular, one could think
in the direction of lying aversion or some naivety on the part of the Receivers,
such as in the model of Kartik, Ottaviani & Squintani (2007).\(^{60}\)

---

\(^{59}\) The pooling equilibrium of a cheap talk game is always a limiting, principal branch A-
QRE: for any rationality parameter \(\lambda\), there is an A-QRE where all Senders mix uniformly over
the message space and the Receiver ignores all messages. As \(\lambda\) increases, the Senders strategy
remains unchanged, and the Receiver’s best response smoothly approaches its actual best
response to her prior.

\(^{60}\) Lying aversion is a topical and fascinating research field (e.g. Gneezy (2005), Kartik
(Strategic Communication with Lying Costs, 2009), Hurkens & Kartik (2009), Lundquist,
Ellingsen, Gribbe & Johannesson (2009), Serra-Garcia, Van Damme & Potters (2011)). Credible
deviations and ACDC currently do not take any potential aversion to lying people may have into
account. We believe abstracting away from lying aversion is justified here, as a first approxima-
tion of behavior, for two reasons. First, Senders can avoid lying by sending messages containing
suggestions to what the Receiver should do rather than information about the state of the
world. In the experiment, we have restricted subjects to send messages containing suggestions.
Second, ACDC is an equilibrium theory for dynamic settings where language evolves and the
meaning of a message, except for credible deviations, is determined by how it is used. Although
lying may be highly relevant for one-shot interactions, it is arguably less so in such a dynamic
setting. With language evolving or having evolved, it is not at all clear what lying could entail.
In particular, Senders either want to use messages that have acquired a pragmatic meaning by
its use or messages with a literal but credible meaning. In both cases, lying does not seem to be
well-defined. Furthermore, in equilibrium, players hold correct beliefs about each other, so that
lying in equilibrium is a particularly tricky concept. Still, if one succeeds in constructing a
consistent framework of dynamic communication with lying, then incorporating lying aversion in
ACDC may well increase its predictive power.
4.6 Robustness

We ran two additional treatments, T4 and T5, to perform two types of robustness checks. As in section 4.4, we focus on the last 15 periods and we also use the same statistical methods and notation as in that section.

4.6.1 T4

One possible experimental risk of manipulating $B$ in treatments G(120), G(130) and G(210) is that our results could be influenced (merely) by increasing the action and type space. In particular, one may be worried that confused subjects simply choose something close to the midpoint of the message or action space. Therefore, we ran T4, which is strategically equivalent to the game G(240), but where we do not shift the boundary (we keep it at 120), but change the disagreement point payoff instead. T4 is identical to G(120), except that disagreement payoff for both players is 30: $U^R(\delta) = U^S(\delta) = 30$. We have:

**Proposition 4.4** T4 has a size-1 equilibrium $\{a_1 = \frac{45}{2}\}$ and a size-2 equilibrium $\{a_1 = 0, a_2 = 30\}$. In addition, T4 is strategically equivalent to G(240).
Figure 4.6 shows a type-action bubble plot of T4 (left panel) and Table 4.5 shows the ACDC properties of T4 together with those of G(120), G(130) and G(210). The results of T4 are in line with those of the three G(B) treatments. In particular, the data of T4 are close to G(210), except that outcomes in T4 are even more dispersed, which is in line with the higher ACD of the separating equilibrium.

![Type-action bubble plot for G(210) and T4](image)

**FIGURE 4.6**

This figure shows the type action bubble plot for G(210) and T4. The bubble plots are clustered on a grid of 10. The solid line represents the ACDC equilibrium outcome.

<table>
<thead>
<tr>
<th></th>
<th>ACD</th>
<th>Average Prediction Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooling</td>
<td>Separating</td>
</tr>
<tr>
<td>G(120)</td>
<td>0.39</td>
<td>0</td>
</tr>
<tr>
<td>G(130)</td>
<td>0.45</td>
<td>0.22</td>
</tr>
<tr>
<td>G(210)</td>
<td>0.63</td>
<td>0.50</td>
</tr>
<tr>
<td>T4 ~ G(240)</td>
<td>0.65</td>
<td>0.56</td>
</tr>
</tbody>
</table>

**TABLE 4.5**

**Performance Pooling versus (ACDC) Separating Equilibrium**

Notes: This table shows the theoretical ACD and the observed prediction error of each equilibrium. The prediction error of T4 is scaled (doubled) to make it comparable with the other treatments.

4.6.2 T5

There are a few issues we cannot test in G(B). First, we cannot test whether ACD also organizes data in a more complicated setting, where the equilibrium set is large. Secondly, our conjecture that credible deviations matter gradually
implies that, even if a neologism proof equilibrium exists, other equilibria admitting few credible deviations may not perform that much worse. Finally, in $G(B)$ we cannot discriminate between ACDC on the one hand and NITS and influentiality on the other in terms of selection properties. For these two reasons, we study $T5$.

$T5$ is the same as $G(120)$, except that the Sender’s disagreement point payoff is raised from 0 to 30: $U^R(\delta) = 0, U^S(\delta) = 30$. In $T5$, the maximum equilibrium-size is 3 and continua of equilibria exist:

**Proposition 4.5** $T5$ has a size-1 equilibrium $\{a_1 = 30\}$. In addition, it has a set of size-2 respectively size-3 equilibria characterized by $\{a_1, a_2 = a_1 + 60\}$, with $a_1 \in [0,30]$ and $\{a_1 = 0, a_2, a_3 = a_2 + 60\}$ with $a_2 \in (0,30]$. The ACDC equilibrium, which is also neologism proof, is $\{0,30,90\}$.

Observe that all size-3 equilibria survive the influentiality criterion. The size-2 equilibrium $\{0,60\}$ and all size-3 equilibria survive NITS (relative to lowest type $t = 0$): the lowest type has no incentive to separate, because it obtains its highest possible utility in equilibrium.

As Figure 4.7 shows, in $T5$, behavior roughly follows the predictions of the ACDC equilibrium, although there is excess separation. Primarily, types close to the boundaries between the intervals of the ACDC equilibrium tend to elicit different actions than in equilibrium.
The ACDC equilibrium predicts significantly better than the pooling equilibrium. (The prediction error of the separating equilibrium (11.4) is significantly smaller at the 5% level than that of the pooling equilibrium (34.3).) Furthermore, Figure 4.8 shows that in T5 the ACDC equilibrium outperforms the other equilibria as well. The left hand side displays the theoretical ACD for the size-2 and size-3 equilibria. The equilibrium that is characterized by the actions (0,30,90) minimizes the ACD and is thus the ACDC equilibrium. The right hand side of the figure shows that for this equilibrium the average prediction error of the action is minimized. In addition, the rank of an equilibrium’s prediction error (right panel in Figure 4.8) roughly follows the rank an equilibrium’s ACD (left panel in Figure 4.8). (The two plots have a different curvature though.) This is also interesting since the ACDC equilibrium is neologism proof. Equilibria that are not neologism proof, but have a small ACD perform quantitatively but not qualitatively worse. In sum, ACDC organizes the data quite well.
4.7 Power and Information Transmission

Our additional treatments T4 and T5 also allow us to test a comparative statics hypothesis regarding the relation between bargaining power and information transmission. In Chapter 2, we studied a general external veto threats model to which all games in this chapter belong, and found a relation between power and information transmission. There we model a player’s bargaining power by her disagreement point payoff. The idea is that the more power a player gets, the lower the harms of disagreement are for her. We take the size of an equilibrium as a measure of information transmission. Our main result in Chapter 2 is that the maximum equilibrium size is increasing in the Sender’s bargaining power and decreasing in that of the Receiver. In Chapter 3 we showed that ACDC selects the maximum size equilibrium. Hence, we hypothesized that the relative power of the Sender affects actual information transmission positively. With our data we can test this.

Observe that treatments T4, G(120) and T5 only differ in the bargaining power of the players (See Table 4.6). In T5, the maximum equilibrium size is 3. If we decrease the Sender’s bargaining power relative to T5, we get G(120) and the maximum equilibrium size is 2. Similarly, if we increase the Receiver’s
bargaining power relative to T5, we get T4 and the maximum equilibrium size is again 2.

**TABLE 4.6**

**COMPARING BARGAINING POWER**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Disagreement payoffs</th>
<th>Equilibrium Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sender</td>
<td>Receiver</td>
</tr>
<tr>
<td>T5</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>G(120)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T4</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

The empirical challenge we face is that while equilibrium size is an appropriate theoretical measure of information transmission for partition equilibria, it is less suitable to judge the amount of information transmitted in actual data. To overcome that, we operationalize information transmission in a manner that can be applied to empirical data as well. We identify information with the informational value messages have to the Receiver. If no information is transmitted at all, then the Receiver’s expected payoff $EU^R$ is equal to that if he used his prior $EU^R_{prior}$ (which is equal to his expected payoff in the pooling equilibrium). If all information is transmitted, he earns the payoff $EU^R_{perfect}$ he would earn if he always perfectly knew the Sender’s type. Information transmission in some (observed) profile $\sigma$, $EU^R_{\sigma}$, can be written as the convex combination of these extremes:

$$EU^R_{\sigma} = (1-\tau(\sigma))EU^R_{prior} + \tau(\sigma)EU^R_{perfect},$$

Hence

$$\tau(\sigma) = \frac{EU^R_{\sigma} - EU^R_{prior}}{EU^R_{perfect} - EU^R_{prior}}$$

is a measure of information transmission, where $\tau(\sigma) = 0$ indicates no information and $\tau(\sigma) = 1$ perfect information transmission.$^{61}$

Table 4.7 shows the information transmission analysis for our data. We look at the total information transmission by calculating $\tau$ using the payoffs actually earned by the Receivers. In addition, we look at the informativeness of Sender’s

$^{61}$ Observe that $\tau$ is invariant to positive affine transformations of the payoffs. In addition, note that $\tau$ could be negative in case the Receiver draws wrong inferences from messages, for instance if he is successfully misled systematically.
messages. Here, we do not use what Receivers actually earned, but estimate what Receivers would have earned if they optimally used the information in their history screen. (The table provides details about how we calculated the optimal action.) The difference between the two can be used as a measure of how successful the Receiver was in extracting the information out of the messages, i.e. his ‘decoding success.’ We also provide the benchmark values for information transmission in the ACDC equilibrium.

Our main finding here is that there was significantly more information transmission in the T5 than in the other two treatments. The same holds true if we look at the informativeness of messages. The informativeness of the messages in the experiment is very close to its ACDC equilibrium benchmark. However, in all treatments, Receivers do not manage to perfectly decode the information in messages. Therefore, in all treatments, actual information transmission is lower than the equilibrium benchmark.
TABLE 4.7

INFORMATION TRANSMISSION

<table>
<thead>
<tr>
<th></th>
<th>Information transmission</th>
<th>Informativeness Sender</th>
<th>Decoding success Receiver</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NE</td>
<td>Data</td>
<td>NE</td>
</tr>
<tr>
<td>T5</td>
<td>.60</td>
<td>.47**</td>
<td>.60</td>
</tr>
<tr>
<td>G(120)</td>
<td>.30</td>
<td>.17**</td>
<td>.30</td>
</tr>
<tr>
<td>T5</td>
<td>.17</td>
<td>.01**</td>
<td>.17</td>
</tr>
<tr>
<td>ΔT5-G(120)</td>
<td>.30***</td>
<td>.29***</td>
<td>-.05</td>
</tr>
<tr>
<td>ΔT5-T4</td>
<td>.48***</td>
<td>.50***</td>
<td>.01</td>
</tr>
<tr>
<td>ΔG(120)-T4</td>
<td>.18</td>
<td>.21</td>
<td>.06</td>
</tr>
</tbody>
</table>

Notes: The table shows the informativeness of the Senders’ messages, the total information transmission and the decoding success of the Receiver. Per treatment, we provide the benchmark of the ACDC equilibrium, the statistics for the data and whether the statistics differ significantly from the benchmark. In addition, we calculate the treatment effects and whether they are significant.

1 The information transmission (calculated per data point using the last 15 periods) is given by \( \tau(data) \), where we used the average of the Receiver’s payoffs across observations as \( U_{data} \).

2 The informativeness of the Senders’ messages (calculated per data point using the last 15 periods) is given by \( \tau(optimal) \), where we used the average of the Receiver’s optimal payoffs across observations as \( U_{optimal} \). We determine the optimal action of the Receiver given their history screen (all observations in the previous 15 periods \( \{t_i,m_i\}_{i \in I} \) in the matching group) as follows. If a Receiver gets message \( m \), we estimate the probability distribution of the Sender type \( t \) given this message \( m \). Since the history screen is sparse (and not all estimates are identified without assumptions), we smooth the history screen as follows. The probability of a Sender being type \( t \) conditional on sending message \( m \) given the set of previous observations \( \{t_i,m_i\}_{i \in I} \) is equal to \( p(m,t) / \sum_{s \in T} p(m,s) \), where \( p(m,s) = \sum_{i \in I} (1 + |s - s_i|)^{-2} (1 + |m - m_i|)^{-2} \). Then, using this probability distribution, we calculate the optimal action and determine the Receiver’s payoff (assuming that Senders accept proposals if they give positive payoff).

3 The decoding success is given by \( 1 - (\tau(optimal) - \tau(data)) \) and measures how much less the Receivers earned compared to what they could have maximally earned.

4.8 Conclusion

In this chapter we presented an experimental test of ACDC. Our main conclusion is that the results provide systematic support for ACDC: credible deviations matter and they matter gradually. In addition, we find that a neologism dynamic can organize the main dynamic characteristics of the data. In Chapter 3, we saw ACDC performs as well as existing criteria in previously conducted
CHAPTER 4. AN EXPERIMENTAL STUDY OF CREDIBLE DEVIATIONS AND ACDC

experiments. This study supports the predictions of ACDC in a new setting where existing concepts are silent.

Finally, we used our design to study the relation between bargaining power and information transmission we analyzed theoretically in Chapter 2. We found that, as predicted, less information is transmitted as the Sender’s relative power decreases.

4.9 Appendix

4.9.1 Proofs of Proposition 4.1, Proposition 4.4 and Proposition 4.5

G(B), T4 and T5 belong to the following game \( \Gamma(d^R, d^S) \), which is the uniform linear case of the veto threats model we study in Chapter 2. \( \Gamma(d^R, d^S) \) proceeds as game \( G(B) \). However, the Sender’s type \( t \) is drawn from the interval \([0,1]\). The Receiver’s and Sender’s payoff on the real line are given by \( U^R(x) = -x \) and \( U^S(x,t) = -|x-t| \). The disagreement point payoff is \( U^R(\delta) = -d^R \) and \( U^S(\delta) = -d^S \) with \( d^R, d^S > 0 \). In Chapter 2, we provide a characterization of the equilibria of \( \Gamma(d^R, d^S) \).

From Proposition 3.3 it follows that

Lemma 4.1 The unique ACDC equilibrium of \( \Gamma(d^R, d^S) \) with respect to credible neologisms is the maximum size equilibrium with the highest equilibrium action.

G(B) corresponds to \( \Gamma(d^R, d^S) \) with \( d^R = \frac{5}{4} \frac{160}{R} \) and \( d^S = \frac{5}{4} \frac{160}{R} \). T4 corresponds to \( \Gamma(\frac{5}{4}, \frac{1}{4}) \) and T5 to \( \Gamma(\frac{5}{4}, \frac{1}{4}) \). Proposition 4.1, Proposition 4.4 and Proposition 4.5 are direct corollaries of the characterization of the equilibria in Chapter 2 and Lemma 4.1.
4.9.2 Proofs of Proposition 4.2 and Proposition 4.3

Proof of Proposition 4.2 Let \( \tilde{a}(t) \) characterize the equilibrium outcome. In our game, \( \langle \tilde{a}, [\tau, \tilde{\tau}] \rangle \) is a credible neologism iff \( U^S(\tilde{a}, t) < U^S(a(t), t) \ \forall t \not\in [\tau, \tilde{\tau}] \), \( U^S(\tilde{a}, t) > U^S(a(t), t) \forall t \in (\tau, \tilde{\tau}) \) and \( \tilde{a} = a^*[\tau, \tilde{\tau}] \). Hence \( \tilde{a} < a_i \) implies \( \tau = 0 \) and \( \tilde{a} > a_n \) implies \( \tau = B \).

First, let us look at pooling equilibrium \( \sigma^P \). Consider a low credible neologism \( \tilde{a}^L < a = 45 \). Now, \( \tau^L = 0 \). Furthermore, \( \tilde{\tau}^L = \frac{1}{2}(\tilde{\tau}^L + 45) < 60 \). Hence, \( \tilde{a}^L = a^*[0, \tilde{\tau}] = 0 \) and \( \tilde{\tau}^L \) must be 22.5. Next, consider a high credible neologism \( \tilde{a}^H > 45 \), \( \tilde{\tau}^H = B \) and \( \tau^H = \frac{1}{2}(\tilde{\tau}^H + 45) \). Solving \( a^*[\frac{1}{2}(\tilde{\tau}^H + 45), B] = \tilde{a}^H \) yields \( \tilde{a}^H = \min\{B - 60, 75\} > 45 \). Consequently, \( \tilde{\tau}^H = \min\{\frac{B - 15}{2}, 60\} \).

Second, let us look at the separating equilibrium \( \sigma^S \). There can be no credible equilibrium \( \tilde{a} < a_i \) as \( a_i = 0 \). Consider a credible neologism \( \tilde{a} > 60 \). Now, \( \tilde{\tau} = B \) and \( \tau = \frac{1}{2}(d + 60) \). Solving \( a^*[\frac{1}{2}(d + 60), B] = \tilde{a} \) yields \( \tilde{a} = \min\{B - 60, 80\} \). Hence, \( \tau = \min\{\frac{B}{2}, 70\} \). If \( B = 120 \), then \( \tilde{a} = 60 \) and it is no neologism. If \( B > 120 \), it is a neologism. Finally, consider some neologism with \( a_i = 0 < \tilde{a} < a_2 = 60 \). Since \( \tilde{a} < 60 \), it must be that \( \tilde{\tau} < 60 \). However, if \( \tilde{\tau} < 60 \), then \( a^*[\tilde{\tau}, \tilde{\tau}] = 0 \). Hence \( \tilde{a} \) cannot be a neologism. Q.E.D.

Proof of Proposition 4.3 First, we show that \( ACD(\sigma^P) > ACD(\sigma^S) \). Let \( \tilde{\tau}^H \) be the lowest deviating type of the high neologism in the pooling equilibrium \( \sigma^P \) and \( \tilde{\tau} \) the lowest deviating type of the neologism in the separating equilibrium \( \sigma^S \). Due to the low credible neologism, \( CD(t, \sigma^P) > CD(t, \sigma^S) = 0 \) for
$t \in \left[0, \frac{45}{2}\right)$. For $t \in \left[\frac{45}{2}, \tau^u\right)$, $CD(t,\sigma^p) = CD(t,\sigma^s) = 0$. Since the distance between the neologism action and the equilibrium action is larger in the pooling equilibrium than in the separating equilibrium and $\tau^u < \tau$, it must hold that $CD(t,\sigma^p) > CD(t,\sigma^s)$ for $t \in (\tau^u,120]$. Furthermore, $CD(t,\sigma^p) = CD(t,\sigma^s) = 1$ for $t \in [120, B]$. Hence, $E_t[CD(t,\sigma^p)] > E_t[CD(t,\sigma^s)]$.

For the second result, observe that the set of rationalizable actions for the Receiver is $[0,B-60]$ and that the Sender can always guarantee herself a payoff of 0 by rejecting the proposed action. This means that $\bar{U}^S(t) = U^S(\min\{t,B-60\},t)$ and $U^S(t) = \max\{0,\min\{U^S(0,t), U^S(B-60,t)\}\}$. Using Proposition 4.1 and Proposition 4.2, we get for the ACD of the separating equilibrium

$$ACD(\sigma^s) = \frac{1}{B} \int_{\min\{70,B/2\}}^{120} \left\{ \frac{(60 - t - (\min\{80,B-60\}) - (60 - |t - 60|)}{U^S(t) - \bar{U}^S(t)} \right\} dt + \frac{B-120}{B}.$$  

It is readily verified that this function is zero for $B = 120$ and strictly increasing in $B$ for $B \geq 120$. Q.E.D.

### 4.9.3 Neologism Dynamic

We first describe the standard, simple, best response dynamic. In each period all Sender types and the Receiver choose a strategy. We assume that the Sender in the acceptance stage accepts all actions that yield her nonnegative payoff: $\nu(a,t) = 1$ if $U^S(a,t) \geq 0$ and $\nu(a,t) = 0$ otherwise. The strategy of the Sender in period $r$ is then given by $\mu_r : T \rightarrow \Delta M$ and that of the Receiver by $\alpha_r : M \rightarrow A$. Let $m_r(t)$ denote the message Sender type $t$ sends (which may be a random variable) and $a_r(m)$ denote the Receiver’s action after receiving message $m$. Both players best respond to the strategy of the other player in the previous round. First, the support of $\mu_r(t)$ is equal to $\arg\max_{m \in M} U^S(a_{r-1}(m),t)$. In particular, we assume the Sender randomizes uniformly over the set of best responses. Second, $a_r(m) = \arg\max_{a \in A} U^K(a,t)E_t[\nu(a,t) | \beta^r(m)]$, where $\beta^r(m)$ is
derived from $\mu_{r-1}$ by Bayes rule whenever possible.\textsuperscript{62} If $\beta'$ cannot be derived from $\mu_{r-1}$, then $\beta' = \beta'(m)$ for some randomly chosen $m \in \bigcup_{t \in T} \text{supp} \mu_{r-1}(t)$.

The neologism dynamic differs from the best response dynamic on one crucial aspect: Senders can send credible neologisms, which will be believed. We define $\langle \tilde{a}, N \rangle$ as a credible neologism with respect to Receiver strategy $\alpha_r$ if (i) $U^S(\tilde{a}, t) > \max_{a \in A} U^S(a_r(m), t) \cdot v(a_r(m), t)$ for all $t \in N$, (ii) $U^S(\tilde{a}, t) \leq \max_{a \in A} U^S(a_r(m), t) \cdot v(a_r(m), t)$ for all $t \notin N$ and (iii) $\tilde{a} = \max_{a \in A} U^R(a, t) E_t[v(a, t) | t \in N]$.\textsuperscript{63}

Now, in the neologism dynamic all Senders that can send a credible neologism in round $r$ with respect to $\alpha_{r-1}$, will do so and such credible neologisms will be believed by the Receivers in round $r$. In all other cases, the dynamic is identical to a best response dynamic. We call the neologism dynamic $f(\mu, \alpha_r)$.

This best response dynamic bears similarities to a level-$k$ analysis. The difference is that in level-$k$, in each iteration just one player (Sender or Receiver) changes her strategy. In the best response dynamic, both players change their strategy each period. Still, the best response dynamic converges in all cases below to very similar outcomes as the outcomes a level-$k$ analysis would converge to.

Before analyzing the dynamic, we characterize the best responses and neologisms. The Sender’s best response is simply to induce the action closest to her type. We call the Receiver’s best response $a^*[t, \bar{t}[$ if Sender types are uniformly distributed in the interval $[t, \bar{t}[$. $a^*[t, \bar{t}[$ is single-valued and equal to $\min\{\bar{t} - 60, 45 + \frac{1}{2} t\}$. Let $\bar{a} = \max_{a \in A} \{a_r(m)\}$ be the highest action of a Receiver’s strategy. Then, for $B = 120$ and $B = 130$, there exists a high credible neologism with respect to $a_r$ if and only if $\bar{a} < B - 60$. In particular, it is equal

\textsuperscript{62} We assume (for ease of exposition) that there is one unique best response for the Receiver, which is generically the case in our game. In case there are more optimal actions one could let the Receiver randomize.

\textsuperscript{63} We need to point out the following subtlety. If a credible neologism was used in the previous period, it becomes just a message (which may have acquired a new ‘meaning’). If the same credible neologism has to be made in the following period, it cannot be the same literal message, as then it would not be a neologism. Hence, the Sender can add for instance Really! or Really, Really! etc. to make it a neologism and distinguish it from the old message.
to \( \left( B - 60, \frac{B - 60 + \bar{a}}{2}, 130 \right) \) if \( 3(B - 120) \leq \bar{a} < B - 60 \) and
\( \left( 60 + \bar{a} / 3, \left( \frac{2}{3}(45 + \bar{a}), 130 \right) \right) \) if \( \bar{a} < 3(B - 120) \). For \( B = 210 \), there exists a high credible neologism if and only if \( \bar{a} < 90 \), and in this case it is equal to
\( \left( 60 + \frac{1}{3} \bar{a}, \frac{2}{3}(45 + \bar{a}), 130 \right) \).

We restrict our analysis to two natural initial strategy profiles: babbling (where no information is transmitted) and naive (where all possible information is transmitted).

For \( G(120), G(130) \) and \( G(210) \), we (i) give the attractor,\(^{64}\) (ii) show that both the babbling and naive initial profiles lie in its basis of attraction and (iii) calculate the average prediction error of the pooling and separating equilibria for the attractor.

\( G(120) \)

For \( G(120) \) it is easy to check that the equilibrium profile is a steady state of the neologism dynamic: \( m_s(t) = m^1 \) for \( t \in [0, 30] \) and \( m_s(t) = m^2 \neq m^1 \) for \( t \in [30, 120] \), and \( a_s(m^1) = 0 \) and \( a_s(m^2) = 60 \). It is a steady state of the best response dynamic and no neologism relative to \( a_s \) exists.

If we start with a babbling profile in period 1, the neologism dynamic proceeds as follows:

<table>
<thead>
<tr>
<th>Strategy Sender period 1 (Babbling)</th>
<th>Strategy Receiver period 1 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1(t) \sim U[0,120] ) if ( t \in [0,120] )</td>
<td>( a_1(m) = 45 ) for all ( m \in [0,120] )</td>
</tr>
</tbody>
</table>

where all Senders randomize uniformly over the interval \([1,120]\).

\(^{64}\) An attractor is roughly speaking a set in the phase-space the neighborhood of which the dynamic evolves to after sufficient time. This can be, for instance, a steady state or a higher n-cycle.
Hence, from period 4, the dynamic is and stays in the separating equilibrium. If we start with a naive profile in period 1, the neologism dynamic proceeds as follows:

<table>
<thead>
<tr>
<th>Strategy Sender period 1 (Naive)</th>
<th>Strategy Receiver period 1 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_i(t) = t$ if $t \in [0, 120]$</td>
<td>$a_i(m) = 0$ if $m \in [0, 60]$</td>
</tr>
<tr>
<td></td>
<td>$a_i(m) = m - 60$ if $m \in [60, 120]$</td>
</tr>
</tbody>
</table>

where all Senders randomize uniformly over the interval [1, 120].
CHAPTER 4. AN EXPERIMENTAL STUDY OF CREDIBLE DEVIATIONS AND ACDC

<table>
<thead>
<tr>
<th>Strategy Sender period 4 (Naive)</th>
<th>Strategy Receiver period 4 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_s(t) \sim U[0,120]$ if $t \in [0,30)$</td>
<td>$a_r(m) = 0$ if $m \in [0,120)$</td>
</tr>
<tr>
<td>$m_s(t) = 120$ if $t \in [30,120]$</td>
<td>$a_r(m) = 60$ if $m = 120$</td>
</tr>
</tbody>
</table>

Hence, from period 4, the dynamic is and stays in the separating equilibrium.

Now we turn to the prediction error. Let the equilibrium profile be $\sigma^e$ and the attracting profile $\sigma^a$. Then, the average (or expected) prediction error of an equilibrium for the attracting profile is $E[|a^e(m^e(t)) - a^a(m^a(t))|]$. The average prediction error of the separating equilibrium is obviously 0. The prediction error of the pooling equilibrium is $\frac{1}{120} \left( \int_0^{30} 45 - 0 \, dt + \int_{30}^{120} 45 - 60 \, dt \right) = \frac{45}{2}$.

**$G(130)$**

For $G(130)$, consider the following state $r'$:

<table>
<thead>
<tr>
<th>Strategy Sender period $r'$</th>
<th>Strategy Receiver period $r'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_s(t) = m^1$ if $t \in [0,t^1)$</td>
<td>$a_r(m) = 0$ if $m \in {m^1,m^2}$</td>
</tr>
<tr>
<td>$m_s(t) = m^2$ if $t \in [t^1,t^2)$</td>
<td>$a_r(m) = a^1$ if $m = m^3$</td>
</tr>
<tr>
<td>$m_s(t) = m^3$ if $t \in [t^2,130]$</td>
<td></td>
</tr>
</tbody>
</table>

with the restriction that $0 \leq t^1 < t^2 < 50$ and $50 < a1 < 70$. $m^1,m^2,m^3$ can be any three messages.

Then, by straightforwardly applying the neologism dynamic, we get the following for rounds $r' + 1$, $r' + 2$, $r' + 3$ and $r' + 4$.
where \( n^1 \) is the credible neologism \( \{70,(35 + a^1 / 2,130)\} \). Furthermore, a Sender in \( [0,a^1 / 2) \), will randomize uniformly over \( m^1 \) and \( m^2 \).

Hence, if player type is in \( [0,a^1 / 2) \), then she will randomize uniformly over \( m^1 \) and \( m^2 \).
Hence, starting at period \( r' \), we can characterize \( f^t \) by \( a^t_{p+1} = a^t_p / 8 + 225 / 4 \), \( t^t_p \) , \( t^t_{p+1} = t^t_p / 8 - 5 / 4 \) and \( t^t_{p+1} = 135 / 4 + t^t_{p+1} / 8 \) (as long as \( 0 \leq t^t_p < t^t_{p+1} < 50 \) and \( 50 < a^t_p < 70 \)).

\( a^t_p = 450 / 7 , t^t_p = 270 / 7 \) and \( t^t_p = 25 / 7 \) is a steady state and attractor to which the dynamic converges monotonically. Hence, if in some period the strategy profile meets the conditions in \( r' \), then \( f \) converges to the 4-cycle characterized by above values.

We proceed to give the first periods of the neologism dynamic for the babbling and naive initial conditions. We end as soon as the dynamic meets the sufficient conditions for their respective attractors specified above.

If we start with a babbling profile in period 1, the neologism dynamic proceeds as follows:

<table>
<thead>
<tr>
<th>Strategy Sender period 1 (Babbling)</th>
<th>Strategy Receiver period 1 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1(t) \sim U[0,130] ) if ( t \in [0,130] )</td>
<td>( a_1(m) = 45 ) for all ( m \in [0,130] )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 2 (Babbling)</th>
<th>Strategy Receiver period 2 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_2(t) = n^1 ) if ( t \in [0,45/2) )</td>
<td>( a_2(m) = 0 ) if ( m = n^1 )</td>
</tr>
<tr>
<td>( m_2(t) \sim U[0,130] ) if ( t \in [45/2,115/2] )</td>
<td>( a_2(m) = 45 ) if ( m \in [0,130] )</td>
</tr>
<tr>
<td>( m_2(t) = n^2 ) if ( t \in (115/2,2130] )</td>
<td>( a_2(m) = 70 ) if ( m = n^2 )</td>
</tr>
</tbody>
</table>

where \( n^1 = \langle 0, [0, 45 / 2) \rangle \) and \( n^2 = \langle 70, (115 / 2, 2130] \rangle \).

<table>
<thead>
<tr>
<th>Strategy Sender period 3 (Babbling)</th>
<th>Strategy Receiver period 3 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_3(t) = n^1 ) if ( t \in [0,45/2) )</td>
<td>( a_3(m) = 0 ) if ( m \in [0,130] \cup {n^1 } )</td>
</tr>
<tr>
<td>( m_3(t) \sim U[0,130] ) if ( t \in [45/2,115/2] )</td>
<td>( a_3(m) = 70 ) if ( m = n^2 )</td>
</tr>
<tr>
<td>( m_3(t) = n^2 ) if ( t \in [115/2,2130] )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 4 (Babbling)</th>
<th>Strategy Receiver period 4 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_4(t) \sim U[0,130] \cup n^1 ) if ( t \in [0,35] )</td>
<td>( a_4(m) = 0 ) if ( m \in [0,130] \cup {n^1 } )</td>
</tr>
<tr>
<td>( m_4(t) = n^2 ) if ( t \in [35,130] )</td>
<td>( a_4(m) = 70 ) if ( m = n^2 )</td>
</tr>
<tr>
<td>Strategy Sender period 5 (Babbling)</td>
<td>Strategy Receiver period 5 (Babbling)</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>( m_5(t) \sim U(0,130] \cup n^1 ) if ( t \in [0,35) )</td>
<td>( a_r(m) = 0 ) if ( m \in [0,130]\cup n^1 )</td>
</tr>
<tr>
<td>( m_5(t) = n^2 ) if ( t \in [35,130] )</td>
<td>( a_r(m) = 125/2 ) if ( m = n^2 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 6 (Babbling)</th>
<th>Strategy Receiver period 6 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_6(t) \sim U(0,130] \cup {n^1} ) if ( t \in [0,125/4) )</td>
<td>( a_r(m) = 0 ) if ( m \in [0,130] \cup {n^1} )</td>
</tr>
<tr>
<td>( m_6(t) = n^2 ) if ( t \in [125/4,265/4] )</td>
<td>( a_r(m) = 125/2 ) if ( m = n^2 )</td>
</tr>
<tr>
<td>( m_6(t) = n^3 ) if ( t \in [265/4,130] )</td>
<td>( a_r(m) = 70 ) if ( m = n^3 )</td>
</tr>
</tbody>
</table>

where \( n^3 = \langle 70,(265/4,130) \rangle \).

<table>
<thead>
<tr>
<th>Strategy Sender period 7 (Babbling)</th>
<th>Strategy Receiver period 7 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_7(t) \sim U(0,130] \cup {n^1} ) if ( t \in [0,25/8) )</td>
<td>( a_r(m) = 0 ) if ( m \in [0,130] \cup {n^1} )</td>
</tr>
<tr>
<td>( m_7(t) = n^2 ) if ( t \in [25/8,305/8) )</td>
<td>( a_r(m) = 25/4 ) if ( m = n^2 )</td>
</tr>
<tr>
<td>( m_7(t) = n^3 ) if ( t \in [305/8,130] )</td>
<td>( a_r(m) = 70 ) if ( m = n^3 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 8 (Babbling)</th>
<th>Strategy Receiver period 8 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_8(t) \sim U(0,130] \cup {n^1,n^2} ) if ( t \in [0,25/8) )</td>
<td>( a_r(m) = 0 ) if ( m \in [0,130] \cup {n^1,n^2} )</td>
</tr>
<tr>
<td>( m_8(t) = n^2 ) if ( t \in [25/8,585/16) )</td>
<td>( a_r(m) = 25/4 ) if ( m = n^2 )</td>
</tr>
<tr>
<td>( m_8(t) = n^3 ) if ( t \in [585/16,130] )</td>
<td>( a_r(m) = 1025/16 ) if ( m = n^3 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 9 (Babbling)</th>
<th>Strategy Receiver period 9 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_9(t) \sim U(0,130] \cup {n^1,n^2} ) if ( t \in [0,25/8) )</td>
<td>( a_r(m) = 0 ) if ( m \in [0,130] \cup {n^1,n^2} )</td>
</tr>
<tr>
<td>( m_9(t) = n^2 ) if ( t \in [25/8,585/16] )</td>
<td>( a_r(m) = 1025/16 ) if ( m = n^3 )</td>
</tr>
<tr>
<td>( m_9(t) = n^3 ) if ( t \in [585/16,130] )</td>
<td>( a_r(m) = 70 ) if ( m = n^4 )</td>
</tr>
</tbody>
</table>

where \( n^4 = \langle 70,(2145/32,130) \rangle \).
### Chapter 4. An Experimental Study of Credible Deviations and ACDC

<table>
<thead>
<tr>
<th>Strategy Sender period 11 (Babbling)</th>
<th>Strategy Receiver period 11 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{11}(t) \sim U[0, 130] \cup {n^1, n^2} ) if ( t \in [0, 125/4) )</td>
<td>( a_{11}(m) = 0 ) if ( m \in [0, 130] \cup {n^1, n^2} )</td>
</tr>
<tr>
<td>( m_{11}(t) = n^3 ) if ( t \in [125/4, 265/4) )</td>
<td>( a_{11}(m) = 225/32 ) if ( m = n^3 )</td>
</tr>
<tr>
<td>( m_{11}(t) = n^4 ) if ( t \in [265/4, 4130] )</td>
<td>( a_{11}(m) = 70 ) if ( m = n^4 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 12 (Babbling)</th>
<th>Strategy Receiver period 12 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{12}(t) \sim U[0, 130] \cup {n^1, n^2} ) if ( t \in [0, 225/64) )</td>
<td>( a_{12}(m) = 0 ) if ( m \in [0, 130] \cup {n^1, n^2} )</td>
</tr>
<tr>
<td>( m_{12}(t) = n^3 ) if ( t \in [225/64, 2465/64) )</td>
<td>( a_{12}(m) = 25/4 ) if ( m = n^3 )</td>
</tr>
<tr>
<td>( m_{12}(t) = n^4 ) if ( t \in [2465/64, 4130] )</td>
<td>( a_{12}(m) = 70 ) if ( m = n^4 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 13 (Babbling)</th>
<th>Strategy Receiver period 13 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{13}(t) \sim U[0, 130] \cup {n^1, n^2} ) if ( t \in [0, 25/8) )</td>
<td>( a_{13}(m) = 0 ) if ( m \in [0, 130] \cup {n^1, n^2, n^3} )</td>
</tr>
<tr>
<td>( m_{13}(t) = n^3 ) if ( t \in [25/8, 305/8) )</td>
<td>( a_{13}(m) = 8225/128 ) if ( m = n^3 )</td>
</tr>
<tr>
<td>( m_{13}(t) = n^4 ) if ( t \in [305/8, 4130] )</td>
<td>( a_{13}(m) = 8225/128 ) if ( m = n^4 )</td>
</tr>
</tbody>
</table>

Now, \( t_{13}^1 = 25/8 < t_{13}^2 = 305/8 < 50 \) and \( 50 < a_{13}^1 = 8225/128 < 70 \). Hence, period 13 meets the requirements of round \( r' \) and the dynamic converges to the attracting four-cycle.

If we start with a naive profile in period 1, the neologism dynamic proceeds as follows:

<table>
<thead>
<tr>
<th>Strategy Sender period 1 (Naive)</th>
<th>Strategy Receiver period 1 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1(t) = t ) if ( t \in [0, 130] )</td>
<td>( a_1(m) = 0 ) if ( m \in [0, 60] )</td>
</tr>
<tr>
<td>( a_1(m) = m - 60 ) if ( m \in [60, 130] )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 2 (Naive)</th>
<th>Strategy Receiver period 2 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_2(t) \sim U[0, 60] ) if ( t = 0 )</td>
<td>( a_2(m) = 0 ) if ( m \in [0, 60] )</td>
</tr>
<tr>
<td>( m_2(t) = t + 60 ) if ( t \in (0, 70) )</td>
<td>( a_2(m) = m - 60 ) if ( m \in [60, 130] )</td>
</tr>
<tr>
<td>( m_2(t) = 130 ) if ( t \in [70, 130] )</td>
<td></td>
</tr>
<tr>
<td>Strategy Sender period 3 (Naive)</td>
<td>Strategy Receiver period 3 (Naive)</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>$m_3(t) \sim U[0,60]$ if $t = 0$</td>
<td>$a_3(m) = 0$ if $m \in [0,120)$</td>
</tr>
<tr>
<td>$m_3(t) = t + 60$ if $t \in [0,70)$</td>
<td>$a_3(m) = m - 120$ if $m \in [120,130)$</td>
</tr>
<tr>
<td>$m_3(t) = 120$ if $t \in [70,130]$</td>
<td>$a_3(m) = 70$ if $m = 130$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 4 (Naive)</th>
<th>Strategy Receiver period 4 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_4(t) \sim U[0,120]$ if $t = 0$</td>
<td>$a_4(m) = 0$ if $m \in [0,120)$</td>
</tr>
<tr>
<td>$m_4(t) = t + 120$ if $t \in [0,10)$</td>
<td>$a_4(m) = m - 120$ if $m \in [120,130)$</td>
</tr>
<tr>
<td>$m_4(t) = 130 - \epsilon$ if $t \in [10,40)$</td>
<td>$a_4(m) = 70$ if $m = 130$</td>
</tr>
<tr>
<td>$m_4(t) = 130$ if $t \in [40,130]$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 5 (Naive)</th>
<th>Strategy Receiver period 5 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_5(t) \sim U[0,130]$ if $t \in [0,65/2)$</td>
<td>$a_5(m) = 0$ if $m \in [0,130)$</td>
</tr>
<tr>
<td>$m_5(t) = 130$ if $t \in [65/2,135/2]$</td>
<td>$a_5(m) = 65$ if $m = 130$</td>
</tr>
<tr>
<td>$m_5(t) = n_i$ if $t \in (135/2,2130]$</td>
<td>$a_5(m) = 70$ if $m = n_i$</td>
</tr>
</tbody>
</table>

where $n_i = (70,(135/2,130])$.

<table>
<thead>
<tr>
<th>Strategy Sender period 6 (Naive)</th>
<th>Strategy Receiver period 6 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_6(t) \sim U[0,130]$ if $t \in [0,65/2)$</td>
<td>$a_6(m) = 0$ if $m \in [0,130)$</td>
</tr>
<tr>
<td>$m_6(t) = 130$ if $t \in [65/2,135/2]$</td>
<td>$a_6(m) = 65$ if $m = 130$</td>
</tr>
<tr>
<td>$m_6(t) = n_i$ if $t \in (135/2,2130]$</td>
<td>$a_6(m) = 70$ if $m = n_i$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 7 (Naive)</th>
<th>Strategy Receiver period 7 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_7(t) \sim U[0,130]$ if $t \in [0,65/2)$</td>
<td>$a_7(m) = 0$ if $m \in [0,130)$</td>
</tr>
<tr>
<td>$m_7(t) = 130$ if $t \in [65/2,135/2]$</td>
<td>$a_7(m) = 15/2$ if $m = 130$</td>
</tr>
<tr>
<td>$m_7(t) = n_i$ if $t \in (135/2,2130]$</td>
<td>$a_7(m) = 70$ if $m = n_i$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 8 (Naive)</th>
<th>Strategy Receiver period 8 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_8(t) \sim U[0,130]$ if $t \in [0,15/4)$</td>
<td>$a_8(m) = 0$ if $m \in [0,130)$</td>
</tr>
<tr>
<td>$m_8(t) = 130$ if $t \in [15/4,155/4]$</td>
<td>$a_8(m) = 15/2$ if $m = 130$</td>
</tr>
<tr>
<td>$m_8(t) = n_i$ if $t \in (155/4,4130]$</td>
<td>$a_8(m) = 70$ if $m = n_i$</td>
</tr>
</tbody>
</table>
Now, $t'_9 = 15 / 4 < t^*_9 = 155 / 4 < 50$ and $50 < a'_9 = 515 / 8 < 70$. Hence, period 9 meets the requirements of round $r'$ and the dynamic converges to the attracting four-cycle.

Finally, we turn to the prediction errors for the attracting four-cycle. First the pooling equilibrium. In the same way as above, it can be straightforwardly calculated that prediction error of the pooling equilibrium in periods $r'$, $r' + 1$, $r' + 2$ and $r' + 3$ is respectively equal to $\frac{17145}{637}$, $\frac{2585}{91}$, $\frac{304}{91}$ and $\frac{2640}{91}$. Hence, the average prediction error of the pooling equilibrium over the four cycle is $\frac{18750}{637} \approx 29.4$. The prediction error of the separating equilibrium in periods $r'$, $r' + 1$, $r' + 2$ and $r' + 3$ is respectively equal to $\frac{4440}{637}$, $\frac{635}{91}$, $\frac{1825}{91}$ and $\frac{7625}{91}$. Hence, the average prediction error of the separating equilibrium over the four cycle is $\frac{29285}{2548} \approx 11.5$.

$G(210)$

We continue with $G(210)$. Consider the following state $r'$:
ESSAYS ON BARGAINING AND STRATEGIC COMMUNICATION

<table>
<thead>
<tr>
<th>Strategy Sender period $r'$</th>
<th>Strategy Receiver period $r'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_r(t) = m^1$ if $t \in [0, t^1)$</td>
<td>$a_r(m) = 0$ if $m \in {m^1, m^2}$</td>
</tr>
<tr>
<td>$m_r(t) = m^2$ if $t \in [t^1, t^2)$</td>
<td>$a_r(m) = a^1$ if $m = m^3$</td>
</tr>
<tr>
<td>$m_r(t) = m^3$ if $t \in [t^2, t^3)$</td>
<td>$a_r(m) = a^2$ if $m = m^4$</td>
</tr>
<tr>
<td>$m_r(t) = m^4$ if $t \in [t^3, t^4)$</td>
<td>$a_r(m) = a^3$ if $m = m^5$</td>
</tr>
<tr>
<td>$m_r(t) = m^5$ if $t \in [t^4, t^5]$</td>
<td>$a_r(m) = a^4$ if $m = n^1$</td>
</tr>
<tr>
<td>$m_r(t) = n^1$ if $t \in (t^5, 210]$</td>
<td>$a_r(m) = n^1$ if $t \in (t^5, 210]$</td>
</tr>
</tbody>
</table>

where $t^1 < t^2 < t^3 < t^4 < t^5$ with $0 < t^1 < 15$, $t^2 < 60$ and $t^5 < 90$; $0 < a^1 < a^2 < a^3 < a^4$ with $a^2 < 30$ and $a^4 < 90$ and $n^1 = \{a^4, [t^5, 210]\}$.

Then, by straightforwardly applying the neologism dynamic, we get for round $r' + 1$:

<table>
<thead>
<tr>
<th>Strategy Sender period $r' + 1$</th>
<th>Strategy Receiver period $r' + 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{r+1}(t) = m^1$ if $t \in [0, a^1 / 2)$</td>
<td>$a_{r+1}(m) = 0$ if $m \in {m^1, m^2, m^3}$</td>
</tr>
<tr>
<td>$m_{r+1}(t) = m^3$ if $t \in [a^1 / 2, (a^1 + a^2) / 2)$</td>
<td>$a_{r+1}(m) = t^4 - 60$ if $m = m^4$</td>
</tr>
<tr>
<td>$m_{r+1}(t) = m^4$ if $t \in [(a^1 + a^2) / 2, (a^2 + a^3) / 2)$</td>
<td>$a_{r+1}(m) = t^5 - 60$ if $m = m^5$</td>
</tr>
<tr>
<td>$m_{r+1}(t) = m^5$ if $t \in [(a^2 + a^3) / 2, (a^3 + a^4) / 2)$</td>
<td>$a_{r+1}(m) = 45 + t^5 / 2$ if $m = n^1$</td>
</tr>
<tr>
<td>$m_{r+1}(t) = n^1$ if $t \in [(a^3 + a^4) / 2, 2 / 3(45 + a^4)]$</td>
<td>$a_{r+1}(m) = 60 + a^4 / 3$ if $m = n^2$</td>
</tr>
<tr>
<td>$m_{r+1}(t) = n^2$ if $t \in (2 / 3(45 + a^4), 210]$</td>
<td>$a_{r+1}(m) = 60 + a^4 / 3, (2 / 3(45 + a^4), 210]$</td>
</tr>
</tbody>
</table>

where $n^2 = \left\{60 + a^4 / 3, (2 / 3(45 + a^4), 210]\right\}$.

Hence, for period $r \geq r'$ we can describe $f$ by $a_{r+1} = 60 + a^4 / 3$,

$t_{r+1}^5 = \frac{2}{3}(45 + a^4 / 3), \quad a_{r+1}^3 = 45 + t_{r+1}^5 / 2, \quad t_{r+1}^4 = \frac{1}{2}(a_{r+1}^3 + a_{r+1}^4), \quad a_{r+1}^2 = t_{r+1}^5 - 60,$

$t_{r+1}^3 = \frac{1}{2}(a_{r+1}^2 + a_{r+1}^3), \quad a_{r+1}^1 = t_{r+1}^4 - 60, \quad t_{r+1}^2 = \frac{1}{2}(a_{r+1}^1 + a_{r+1}^2)$ and $t_{r+1}^1 = \frac{1}{2}a_{r+1}^1$ (as long as $a_{r+1}^1, \ldots, a_{r+1}^5$ and $t_{r+1}^1, \ldots, t_{r+1}^5$ meet the above conditions).
Since \( a^4_{t+1} = 60 + a^4_t / 3 \), \( a^4_t \) converges monotonically to 90. Consequently, it follows that
\[
a^4_t = 90, \quad t^4_t = 90, \quad a^3_t = 90, \quad t^3_t = 90, \quad a^2_t = 30, \quad t^2_t = 60, \quad a^1_t = 30, \quad t^1_t = 30 \quad \text{and} \quad t^1 = 15
\]
is an attractor for this dynamic to which converges. (It is not a steady state, as if \( a^4_t = 90 \), then no neologism could be made. Nonetheless, the profile is never reached and all points in its neighborhood converge to it.)

We now proceed to give the first periods of the neologism dynamic for the babbling and naive initial conditions.

If we start with a babbling profile in period 1, the neologism dynamic proceeds as follows:

<table>
<thead>
<tr>
<th>Strategy Sender period 1 (Babbling)</th>
<th>Strategy Receiver period 1 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1(t) \sim U[0, 210] ) if ( t \in [0, 210] )</td>
<td>( a_1(m) = 45 ) for all ( m \in [0, 210] )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 2 (Babbling)</th>
<th>Strategy Receiver period 2 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_2(t) = n^1 ) if ( t \in [0, 45 / 2) )</td>
<td>( a_2(m) = 0 ) if ( m = n^1 )</td>
</tr>
<tr>
<td>( m_2(t) \sim U[0, 210] ) if ( t \in [45 / 2, 60] )</td>
<td>( a_2(m) = 45 ) if ( m \in [0, 210] )</td>
</tr>
<tr>
<td>( m_2(t) = n^2 ) if ( t \in (60, 210] )</td>
<td>( a_2(m) = 75 ) if ( m = n^2 )</td>
</tr>
</tbody>
</table>

where \( n^1 = \langle 0, [0, 45 / 2) \rangle \) and \( n^2 = \langle 75, (60, 210] \rangle \).

<table>
<thead>
<tr>
<th>Strategy Sender period 3 (Babbling)</th>
<th>Strategy Receiver period 3 (Babbling)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_3(t) = n^1 ) if ( t \in [0, 45 / 2) )</td>
<td>( a_3(m) = 0 ) if ( m \in [0, 210] \cup {n^1} )</td>
</tr>
<tr>
<td>( m_3(t) \sim U[0, 210] ) if ( t \in [45 / 2, 60] )</td>
<td>( a_3(m) = 75 ) if ( m = n^2 )</td>
</tr>
<tr>
<td>( m_3(t) = n^2 ) if ( t \in (60, 80] )</td>
<td>( a_3(m) = 85 ) if ( m = n^3 )</td>
</tr>
<tr>
<td>( m_3(t) = n^3 ) if ( t \in (80, 210] )</td>
<td></td>
</tr>
</tbody>
</table>

where \( n^3 = \langle 85, (80, 210] \rangle \).
Now, 0 < \( t_6^1 = 10 < 15, t_6^5 = 115 / 2 < 60, t_6^9 = 2420 / 27 < 90, a_6^2 = 260 / 9 < 30 \)
and \( a_6^4 = 2425 / 27 < 90 \) Hence, period 6 meets the requirements of round \( r' \) and the dynamic converges to the attractor.

If we start with a naive profile in period 1, the neologism dynamic proceeds as follows:
<table>
<thead>
<tr>
<th>Strategy Sender period 1 (Naive)</th>
<th>Strategy Receiver period 1 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1(t) = t ) if ( t \in [0,210] )</td>
<td>( a_s(m) = 0 ) if ( m \in [0,60] )</td>
</tr>
<tr>
<td></td>
<td>( a_s(m) = m - 60 ) if ( m \in [60,210] )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 2 (Naive)</th>
<th>Strategy Receiver period 2 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_2(t) \sim U[0,60] ) if ( t = 0 )</td>
<td>( a_s(m) = 0 ) for all ( m \in [0,60] )</td>
</tr>
<tr>
<td>( m_2(t) = t + 60 ) if ( t \in (0,150) )</td>
<td>( a_s(m) = m - 60 ) for all ( m \in [60,210] )</td>
</tr>
<tr>
<td>( m_2(t) = 210 ) if ( t \in [150,210] )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 3 (Naive)</th>
<th>Strategy Receiver period 3 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_3(t) \sim U[0,60] ) if ( t = 0 )</td>
<td>( a_s(m) = 0 ) if ( m \in [0,120] )</td>
</tr>
<tr>
<td>( m_3(t) = t + 60 ) if ( t \in (0,150) )</td>
<td>( a_s(m) = m - 120 ) if ( m \in [120,210] )</td>
</tr>
<tr>
<td>( m_3(t) = 210 ) if ( t \in [150,210] )</td>
<td>( a_s(m) = 120 ) if ( m = 210 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 4 (Naive)</th>
<th>Strategy Receiver period 4 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_4(t) \sim U[0,120] ) if ( t = 0 )</td>
<td>( a_s(m) = 0 ) if ( m \in [0,120] )</td>
</tr>
<tr>
<td>( m_4(t) = t + 120 ) if ( t \in (0,90) )</td>
<td>( a_s(m) = m - 120 ) if ( m \in [120,210] )</td>
</tr>
<tr>
<td>( m_4(t) = 210 - \epsilon ) if ( t \in [90,105) )</td>
<td>( a_s(m) = 120 ) if ( m = 210 )</td>
</tr>
<tr>
<td>( m_4(t) = 210 ) if ( t \in [105,210] )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 5 (Naive)</th>
<th>Strategy Receiver period 5 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_5(t) \sim U[0,120] ) if ( t = 0 )</td>
<td>( a_s(m) = 0 ) if ( m \in [0,180] )</td>
</tr>
<tr>
<td>( m_5(t) = t + 120 ) if ( t \in (0,90) )</td>
<td>( a_s(m) = m - 180 ) if ( m \in [180,210 - \epsilon) )</td>
</tr>
<tr>
<td>( m_5(t) = 210 - \epsilon ) if ( t \in [90,105) )</td>
<td>( a_s(m) = 45 ) if ( m = 210 - \epsilon )</td>
</tr>
<tr>
<td>( m_5(t) = 210 ) if ( t \in [105,210] )</td>
<td>( a_s(m) = 195 / 2 ) if ( m = 210 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy Sender period 6 (Naive)</th>
<th>Strategy Receiver period 6 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_6(t) \sim U[0,180] ) if ( t = 0 )</td>
<td>( a_s(m) = 0 ) if ( m \in [0,180] )</td>
</tr>
<tr>
<td>( m_6(t) = t + 180 ) if ( t \in (0,30) )</td>
<td>( a_s(m) = m - 180 ) if ( m \in [180,210 - \epsilon) )</td>
</tr>
<tr>
<td>( m_6(t) = 210 - 2\epsilon ) if ( t \in [30,75/2) )</td>
<td>( a_s(m) = 45 ) if ( m = 210 - \epsilon )</td>
</tr>
<tr>
<td>( m_6(t) = 210 - \epsilon ) if ( t \in [75/2,285/4) )</td>
<td>( a_s(m) = 195 / 2 ) if ( m = 210 )</td>
</tr>
<tr>
<td>( m_6(t) = 210 ) if ( t \in [285/4,4210] )</td>
<td></td>
</tr>
<tr>
<td>Strategy Sender period 7 (Naive)</td>
<td>Strategy Receiver period 7 (Naive)</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>( m_s(t) \sim U[0,180] ) if ( t = 0 )</td>
<td>( a_r(m) = 0 ) if ( m \in [0,210-2\epsilon) )</td>
</tr>
<tr>
<td>( m_s(t) = t + 180 ) if ( t \in (0,30) )</td>
<td>( a_r(m) = 45 / 4 ) if ( m = 210 - \epsilon )</td>
</tr>
<tr>
<td>( m_s(t) = 210 - 2\epsilon ) if ( t \in [30,75 / 2) )</td>
<td>( a_r(m) = 645 / 8 ) if ( m = 210 )</td>
</tr>
<tr>
<td>( m_s(t) = 210 - \epsilon ) if ( t \in [75 / 2,2,285 / 4) )</td>
<td>( a_r(m) = 645 / 8 ) if ( m = 210 )</td>
</tr>
<tr>
<td>( m_s(t) = 210 ) if ( t \in [285 / 4,2,120] )</td>
<td>( a_r(m) = 645 / 8 ) if ( m = n^1 )</td>
</tr>
</tbody>
</table>

where \( n^1 = \{695 / 8, 335 / 4, 4, 210\} \).

<table>
<thead>
<tr>
<th>Strategy Sender period 8 (Naive)</th>
<th>Strategy Receiver period 8 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_s(t) \sim U[0,210-2\epsilon) ) if ( t \in [0,45 / 8) )</td>
<td>( a_s(m) = 0 ) if ( m \in [180,210-2\epsilon) )</td>
</tr>
<tr>
<td>( m_s(t) = 210 - \epsilon ) if ( t \in [45 / 8,8,735 / 16) )</td>
<td>( a_s(m) = 45 / 4 ) if ( m = 210 - \epsilon )</td>
</tr>
<tr>
<td>( m_s(t) = 210 ) if ( t \in [735 / 16,335 / 4) )</td>
<td>( a_s(m) = 645 / 8 ) if ( m = 210 )</td>
</tr>
<tr>
<td>( m_s(t) = n^1 ) if ( t \in [335 / 4,2,1055 / 24] )</td>
<td>( a_s(m) = 695 / 8 ) if ( m = n^1 )</td>
</tr>
</tbody>
</table>

where \( n^1 = \{695 / 8, 335 / 4, 4, 210\} \).

<table>
<thead>
<tr>
<th>Strategy Sender period 9 (Naive)</th>
<th>Strategy Receiver period 9 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_s(t) \sim U[0,210-2\epsilon) ) if ( t \in [0,1055 / 12,210] )</td>
<td>( a_s(m) = 0 ) if ( m \in [210-\epsilon) )</td>
</tr>
<tr>
<td>( m_s(t) = 210 - \epsilon ) if ( t \in [885 / 16,1055 / 12) )</td>
<td>( a_s(m) = 95 / 4 ) if ( m = 210 )</td>
</tr>
<tr>
<td>( m_s(t) = 210 ) if ( t \in [1055 / 12,3215 / 36) )</td>
<td>( a_s(m) = 695 / 8 ) if ( m = n^1 )</td>
</tr>
<tr>
<td>( m_s(t) = n^1 ) if ( t \in [3215 / 36,2,120] )</td>
<td>( a_s(m) = 2135 / 24 ) if ( m = n^2 )</td>
</tr>
</tbody>
</table>

where \( n^1 = \{695 / 8, 335 / 4, 4, 210\} \).

<table>
<thead>
<tr>
<th>Strategy Sender period 10 (Naive)</th>
<th>Strategy Receiver period 10 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_s(t) \sim U[0,210-\epsilon) ) if ( t \in [0,1055 / 12,210] )</td>
<td>( a_{10}(m) = 0 ) if ( m \in [210-\epsilon) )</td>
</tr>
<tr>
<td>( m_s(t) = 210 ) if ( t \in [95 / 8,8,1055 / 16) )</td>
<td>( a_{10}(m) = 95 / 4 ) if ( m = 210 )</td>
</tr>
<tr>
<td>( m_s(t) = n^1 ) if ( t \in [885 / 16,1055 / 12) )</td>
<td>( a_{10}(m) = 335 / 12 ) if ( m = n^1 )</td>
</tr>
<tr>
<td>( m_s(t) = n^2 ) if ( t \in [1055 / 12,3215 / 36) )</td>
<td>( a_{10}(m) = 2135 / 24 ) if ( m = n^2 )</td>
</tr>
<tr>
<td>( m_s(t) = n^3 ) if ( t \in [3215 / 36,2,120] )</td>
<td>( a_{10}(m) = 6455 / 72 ) if ( m = n^3 )</td>
</tr>
</tbody>
</table>

where \( n^1 = \{6455 / 72,3215 / 36,210\} \).
### Chapter 4. An Experimental Study of Credible Deviations and ACDC

<table>
<thead>
<tr>
<th>Strategy Sender period 11 (Naive)</th>
<th>Strategy Receiver period 11 (Naive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{11}(t) \sim U[0, 210 - \epsilon]$ if $t \in [0, 95/8)$</td>
<td>$a_{11}(m) = 0$ if $m \in [0, 210]$</td>
</tr>
<tr>
<td>$m_{11}(t) = 210$ if $t \in [95/8, 8155/6)$</td>
<td>$a_{11}(m) = 335/12$ if $m = n^1$</td>
</tr>
<tr>
<td>$m_{11}(t) = n^1$ if $t \in [155/6, 935/16)$</td>
<td>$a_{11}(m) = 1055/36$ if $m = n^2$</td>
</tr>
<tr>
<td>$m_{11}(t) = n^2$ if $t \in [935/16, 3215/36)$</td>
<td>$a_{11}(m) = 6455/72$ if $m = n^3$</td>
</tr>
<tr>
<td>$m_{11}(t) = n^3$ if $t \in [3215/36, 9695/108]$</td>
<td>$a_{11}(m) = 19415/216$ if $m = n^4$</td>
</tr>
<tr>
<td>$m_{11}(t) = n^4$ if $t \in (9695/108, 210]$</td>
<td></td>
</tr>
</tbody>
</table>

where $n^4 = \{19415/216, 9695/108, 210\}$.

Now, $t^1_{11} = 95/8 < 15$, $t^3_{11} = 935/16 < 60$, $t^5_{11} = 9695/108 < 90$, $a^1_{11} = 1055/36 < 30$ and $a^4_{11} = 19415/216 < 90$. Hence, period 11 meets the requirements of round $r'$ and the dynamic converges to the attractor.

Finally, we turn to the prediction errors for of the equilibria with respect to the attractor. The average prediction error of the pooling equilibrium is equal to $\frac{285}{7} \simeq 40.7$. The average prediction error of the separating equilibrium is equal to $\frac{195}{7} \simeq 27.9$.

#### 4.9.4 Experimental Instructions

We include the experimental instructions (including check questions) of the G(120) treatment for both the “Chooser” (Sender) and “Proposer” (Receiver) roles. The instructions of the G(130) and G(210) treatments are very similar.

**Instructions Chooser**

INSTRUCTIONS

Welcome to this decision-making experiment. Please read these instructions carefully. We will first provide you with an outline of the instructions and then we will proceed with a detailed description of the instructions.
OUTLINE

Experiment
- At the start of the experiment you will receive a starting capital of 100 points. In addition, you can earn points with your decisions.
- At the end of the experiment, you receive 1,5 (one-and-a-half) euro for each 100 points earned.
- The experiment consists of around 50 periods.
- Your role in the whole experiment is: CHOOSER.
- In each period, you will be randomly paired with a different participant who performs the role of Proposer.

Sequence of events
- In each period, you and the Proposer will bargain over an outcome, which can be any number between 0 and 120.
- Your preferred outcome is a number between 0 and 120. Any number between 0 and 120 is equally likely. The Proposer’s preferred outcome is always 0.
- Each period you will receive a new (random) preferred outcome. You are the only one who is informed about your preferred outcome.
- After learning your preferred outcome, you will send a SUGGESTION for a proposal (between 0 and 120) to the Proposer.
- The Proposer is informed of your suggestion and makes a PROPOSAL (between 0 and 120) for the outcome.
- After you have been informed of the proposal, you accept or reject it.
- At the end of a period, you are informed of the points you earned (your payoff).

Payoffs
- When you accept a proposal, your payoff is 60 minus the distance between your preferred outcome and the proposal.
- The Proposer’s payoff is 60 minus 0.4 times the proposal in this case.
- When you reject a proposal, you receive 0 points and the Proposer receives 0 points.

History Overview
When making a decision, you may use the History Overview, which provides an overview of the results of the other Chooser/Proposer pairs (including your own pair) in the 15 most recent periods. The left part of the overview is a Table with four columns SUGGESTION, PROPOSAL, ACCEPTANCE and PREFERRED OUTCOME. In a row, you will find a particular pair’s suggestion, the corresponding proposal, whether the Chooser accepted or rejected the proposal and the preferred outcome of that Chooser. On the right, you find a Graph where the most recent results are represented by blue squares. On the horizontal axis you can read the value of the suggestion and on the vertical axis the value of the corresponding proposal.

DETAILED INSTRUCTIONS
Now we will describe the experiment in detail. At the start of the experiment you will receive a starting capital of 100 points. During the experiment you will be asked to make a number of decisions. Your decisions and the decisions of the participants you will be paired with will determine how much money you earn. The experiment consists of around 50 periods. In each period, your earnings will be denoted in points. Your final earnings in the experiment will be equal to the starting capital plus the sum of your earnings in all periods. At the end of the
experiment, your earnings in points will be transferred to money. For each 100 points you earn, you will receive 1.5 (one-and-a-half) euro. Your earnings will be privately paid to you in cash. In each period, all participants are paired in couples. One participant within a pair has the role of CHOOSER, the other participant performs the role of PROPOSER. In all periods you keep the same role.

Your role is: CHOOSER.

MATCHING PROCEDURE
For the duration of the experiment, you will be in a fixed matching group of five Proposers and five Choosers (hence 10 participants in total, including yourself). In each period you are randomly matched to another participant in this matching group with the role of Proposer. You will never learn with whom you are matched.

BARGAINING AND PREFERRED OUTCOMES
In each period, you and the Proposer with whom you are coupled will bargain over an outcome. The Proposer’s preferred outcome is always 0. Your preferred outcome is a number between (and including) 0 and 120. Any number between 0 and 120 is equally likely. Each period you will receive a new preferred outcome that does not depend on your preferred outcome of any previous period. You are the only one who is informed about your preferred outcome. The Proposer only knows that your preferred outcome is a number between 0 and 120 (and that each such number is equally likely).

SEQUENCE OF EVENTS IN A PERIOD
After you have learned your preferred outcome in a period, you will send a SUGGESTION for a proposal to the Proposer. You may send any suggestion between (and including) 0 and 120. It is up to you to decide whether and how you let your suggestion depend on your preferred outcome. Then, the Proposer with whom you are coupled is informed of your suggestion (but not of your preferred outcome). Subsequently, the Proposer makes a PROPOSAL for the outcome. A proposal is any number between (and including) 0 and 120. Finally, you will choose to accept or reject the proposal.
At the end of a period, you are informed of the payoff (points you earned) that you made. This payoff is automatically added to your total earnings (or in case that you make a loss, it is subtracted from your total earnings). The Proposer is informed of the outcome, your preferred outcome and her or his own payoff.

Please note that the experiment will only continue from one phase to another after everybody has pressed OK/PROCEED. For this reason, please press OK/PROCEED as soon as you have made your decision.

PAYOFFS WHEN YOU ACCEPT THE PROPOSAL
When you accept the proposal, you will receive a payoff of 60 minus the distance between your preferred outcome and the proposal:

\[ \text{Your payoff} = 60 - \text{distance(your preferred outcome and proposal)} \]

When you accept the proposal, the Proposer’s payoff is 60 minus 0.4 times the proposal:

\[ \text{Payoff Proposer} = 60 - 0.4 \times \text{proposal} \]

It is possible to reject a proposal.
PAYOFFS WHEN YOU REJECT THE PROPOSAL
When you reject a proposal, then the outcome is the status quo. In this case, you will receive 0 points and the Proposer will receive 0 points.
Notice that accepting an offer gives you a higher payoff than rejecting it if and only if the distance between the proposal and your preferred outcome is smaller than 60. The Proposer’s payoff is higher when you accept than when you reject in all cases.

EXAMPLE 1. Suppose your preferred outcome is 80 and you receive a proposal of 100. Then, the distance between your preferred outcome and proposal is 100 - 80 = 20.
If you accept, your payoff is 60 - 20 = 40. The Proposer’s payoff in this case is 60 – 0.4*100 = 20.
If you reject, your payoff is 0 and the Proposer’s payoff is 0.

EXAMPLE 2. Suppose your preferred outcome is 80 and you receive a proposal of 10. Then, the distance between your preferred outcome and the proposal is 80 - 10 = 70.
If you accept, your payoff is 60 - 70 = -10. The Proposer’s payoff in this case is 60 – 0.4*10 = 56.
If you reject, your payoff is 0 and the Proposer’s payoff is 0.

HISTORY OVERVIEW
When making a decision, you may use the History Overview, which fills the lower part of the screen. The History Overview summarizes the results of the most recent 15 periods. (If less than 15 periods have been completed, this history overview contains results of all completed periods.)
Apart from your own results in the previous periods, the History Overview also contains the results of the other Chooser/Proposer pairs in your matching group. In total you are thus informed about the past results of the same matching group of five Chooser/Proposer pairs. All other Choosers and Proposers in your matching group will have the same information. The presentation of information for Proposers is different than for Choosers.

TABLE
Below you see an example of the history overview. THE NUMBERS IN THE HISTORY OVERVIEW DO NOT INDICATE WHAT YOU SHOULD DO IN THE EXPERIMENT. The left part of the history overview is a Table with four columns. The first column labelled SUGGESTION contains the suggestions made by the Choosers in the recent previous periods. The second column labelled PROPOSAL gives the proposal that was made by the Proposer as a response to the suggestion in the same row. The third column labelled ACCEPTANCE shows whether the Chooser accepted or rejected the proposal. The fourth column labelled PREFERRED OUTCOME shows the preferred outcome of the Chooser.
The results shown in the history overview will be sorted on the basis of suggestion in ascending order. (The lower the suggestion, the higher the place in the table.) When the suggestion is the same for two or more different results, these observations will be sorted on the basis of proposal, again in ascending order. In the example above, this applies to the third and the fourth row, where two Choosers chose the same suggestion but the corresponding Proposers chose different proposals. More generally, observations have been sorted first on suggestion, then on proposal, then on acceptance or rejection and finally on preferred outcome.

**GRAPH**

On the right of the history overview, the most recent results are represented in a graph. The horizontal axis presents the suggestion and the vertical axis presents the proposal. Each previous observation is represented by a blue square. On the horizontal axis you can read the value of the suggestion for a particular result and on the vertical axis you can read the value of the corresponding proposal. (Proposers will see preferred outcomes on the vertical axis, rather than proposals.)

**EXAMPLE.** Consider the square that is displayed in the lower left corner of the Graph shown above. Here, the Chooser made a suggestion of 30. The Proposer responded with a proposal of 10.

You have now reached the end of the instructions. The next page contains some questions concerning the experiment. When all participants have answered all questions correctly, we will proceed with the experiment.
QUESTIONS

Please answer the following questions. THE VALUES USED IN SOME QUESTIONS DO NOT INDICATE WHAT YOU SHOULD DO IN THE EXPERIMENT. RATHER, THEY HAVE BEEN CHOSEN TO FACILITATE CALCULATIONS.

1. Is the following statement correct? ‘In each period I am coupled with the same Proposer.’

2. Is the following statement correct? ‘My preferred position will be observed by the Proposer before (s)he makes her or his proposal.’

3. (A) What is the highest value your preferred outcome can take on?  
   (B) What is the highest value a suggestion of yours can take on?  
   (C) What is the highest value a proposal can take on?

4. Consider a period in which your preferred outcome is 50. You chose to send a suggestion of 40. The Proposer made a proposal of 20, which was accepted by you.  
   (A) What are your own earnings in this period?  
   (B) How much does the Proposer to whom you are paired earn?

5. Consider a period in which your preferred outcome is 90. You chose to send a suggestion of 100. The Proposer made a proposal of 0, which was accepted by you.  
   (A) What are your own earnings in this period?  
   (B) How much does the Proposer to whom you are paired earn?

6. Consider a period in which your preferred outcome is 30. You chose to send a suggestion of 40. The Proposer made a proposal of 10, which was rejected by you.  
   (A) What are your own earnings in this period?  
   (B) How much does the Proposer to whom you are paired earn?

When you are ready answering the questions, please raise your hand.

Instructions Proposer

O U T L I N E

Experiment

INSTRUCTIONS

Welcome to this decision-making experiment. Please read these instructions carefully. We will first provide you with an outline of the instructions and then we will proceed with a detailed description of the instructions.
At the start of the experiment you will receive a starting capital of 100 points. In addition, you can earn points with your decisions. At the end of the experiment, you receive 1,5 (one-and-a-half) euro for each 100 points earned. The experiment consists of around 50 periods. Your role in the whole experiment is: PROPOSER. In each period, you will be randomly paired with a different participant who performs the role of Chooser.

Sequence of events
- In each period, you and the Chooser will bargain over an outcome, which can be any number between 0 and 120.
- Your preferred outcome is always 0.
- The Chooser’s preferred outcome is a number between 0 and 120. Any number between 0 and 120 is equally likely.
- Each period, each Chooser will receive a new (random) preferred outcome. The Chooser is the only one who is informed about her or his preferred outcome.
- After learning her or his preferred outcome, the Chooser with whom you are matched will send a SUGGESTION for a proposal (between 0 and 120) to you.
- You are informed of the Chooser’s suggestion and make a PROPOSAL (between 0 and 120) for the outcome.
- After the Chooser has been informed of the proposal, she or he accepts or rejects it.
- At the end of a period, you are informed of the points you earned (your payoff).

Payoffs
- When the Chooser accepts your proposal, your payoff is 60 minus 0.4 times the proposal.
- The Chooser’s payoff is in this case 60 minus the distance between her or his preferred outcome and the proposal.
- When the Chooser rejects your proposal, you receive 0 points and the Chooser 0 points.

History Overview
When making a decision, you may use the History Overview, which provides an overview of the results of five Chooser/Proposer pairs (including your own pair) in the 15 most recent periods. The left part of the overview is a Table with four columns SUGGESTION, PREFERRED OUTCOME, PROPOSAL and ACCEPTANCE. In a row, you will find a particular pair’s suggestion, the preferred outcome of the Chooser, the proposal made by the Proposer and whether the Chooser accepted or rejected the proposal. On the right, you find a Graph where the most recent results are represented by blue squares. On the horizontal axis you can read the value of the suggestion and on the vertical axis the value of the corresponding preferred outcome of the Chooser.

DETAILED INSTRUCTIONS

Now we will describe the experiment in detail. At the start of the experiment you will receive a starting capital of 100 points. During the experiment you will be asked to make a number of decisions. Your decisions and the decisions of the participants you will be paired with will determine how much money you earn. The experiment consists of around 50 periods. In each period, your earnings will be denoted in points. Your final earnings in the experiment will be equal to the starting capital plus the sum of your earnings in all periods. At the end of the experiment, your earnings in points will be transferred to money. For each 100 points you earn, you will receive 1,5 (one-and-a-half) euro. Your earnings will be privately paid to you in cash.
In each period, all participants are paired in couples. One participant within a pair has the role of
CHOOSER, the other participant performs the role of PROPOSER. In all periods you keep the
same role.

Your role is: PROPOSER.

MATCHING PROCEDURE
For the duration of the experiment, you will be in a fixed matching group of five Proposers and
five Choosers (hence 10 participants in total, including yourself). In each period you are
randomly matched to another participant with the role of Chooser. You will never learn with
whom you are matched.

BARGAINING AND PREFERRED OUTCOMES
In each period, you and the Chooser with whom you are coupled will bargain over an outcome.
Your preferred outcome is always 0. The Chooser’s preferred outcome is a number between
(and including) 0 and 120. Any number between 0 and 120 is equally likely. Each period, each
Chooser will receive a new preferred outcome that does not depend on a preferred outcome of
any previous period. The Chooser is the only one who is informed about her or his preferred
outcome. You only know that the Chooser’s preferred outcome is a number between 0 and 120
(and that each such number is equally likely).

SEQUENCE OF EVENTS IN A PERIOD
After the Chooser with whom you are matched has learned her or his preferred outcome in a
period, she or he will send a SUGGESTION for a proposal to you. The Chooser may send any
suggestion between (and including) 0 and 120. It is up to the Chooser to decide whether and
how she or he lets her or his suggestion depend on her or his preferred outcome. Then, you are
informed of the Chooser’s suggestion (but not of her or his preferred outcome). Subsequently,
you make a PROPOSAL for the outcome. A proposal is any number between (and including) 0
and 120. Finally, the Chooser will choose to accept or reject the proposal.
At the end of a period, you are informed of the outcome of the period and the preferred outcome
of the Chooser you were paired with. Finally, you are informed of the payoff (points you
earned) that you made. This payoff is automatically added to your total earnings (or in case that
you make a loss, it is subtracted from your total earnings).

Please note that the experiment will only continue from one phase to another after everybody
has pressed OK/PROCEED. For this reason, please press OK/PROCEED as soon as you have
made your decision.

PAYOFFS WHEN THE CHOOSER ACCEPTS THE PROPOSAL
When the Chooser accepts your proposal, your payoff is 60 minus 0.4 times the proposal:

Your payoff = 60 – 0.4 * proposal.

When the Chooser accepts your proposal, the Chooser will receive a payoff of 60 minus the
distance between her or his preferred outcome and the proposal:

Payoff Chooser = 60 – distance(her or his preferred outcome and proposal).

It is possible for a Chooser to reject a proposal.

PAYOFFS WHEN THE CHOOSER REJECTS THE PROPOSAL
When the Chooser rejects a proposal, then the outcome is the status quo. In this case, you will
receive 0 points and the Chooser will receive 0 points.
Notice that accepting an offer gives the Chooser a higher payoff than rejecting it if and only if the distance between the proposal and her preferred outcome is smaller than 60. Your payoff is higher when the Chooser accepts than when she or he rejects in all cases.

EXAMPLE 1. Suppose the Chooser’s preferred outcome turns out to be 80 (which you cannot know) and you make a proposal of 100. Then, the distance between her preferred outcome and your proposal is 100 - 80 = 20.
If the Chooser accepts, your payoff is 60 – 0.4*100 = 20. The Chooser’s payoff in this case is 60 - 20 = 40.
If the Chooser rejects, your payoff is 0 and the Chooser’s payoff is 0.

EXAMPLE 2. Suppose the Chooser’s preferred outcome turns out to be 80 and you make a proposal of 10. Then, the distance between her preferred outcome and your proposal is 80 - 10 = 70.
If the Chooser accepts, your payoff is 60 – 0.4*10 = 56. The Chooser’s payoff in this case is 60 - 70 = -10.
If the Chooser rejects, your payoff is 0 and the Chooser’s payoff is 0.

HISTORY OVERVIEW
When making a decision, you may use the History Overview, which fills the lower part of the screen. The History Overview summarizes the results of the most recent 15 periods. (If less than 15 periods have been completed, this history overview contains results of all completed periods.)
Apart from your own results in the previous periods, the history overview also contains the results of the other Chooser/Proposer pairs in your matching group. In total you are thus informed about the past results of the same group of five Chooser/Proposer pairs. All Choosers and Proposers in your matching group will have the same information. The presentation of information is different for Choosers than for Proposers.

TABLE
Below you see an example of the history overview. THE NUMBERS IN THE HISTORY OVERVIEW DO NOT INDICATE WHAT YOU SHOULD DO IN THE EXPERIMENT. The left part of the history overview is a Table with four columns. The first column labelled SUGGESTION contains the suggestions made by the Choosers in the recent previous periods. The second column labelled PREFERRED OUTCOME shows the preferred outcome of the Chooser. The third column labelled PROPOSAL gives the proposal that was made by the Proposer as a response to the suggestion in the same row. The fourth column labelled ACCEPTANCE shows whether the Chooser accepted or rejected the proposal.
The results shown in the history overview will be sorted on the basis of suggestion in ascending order. (The lower the suggestion, the higher the place in the table.) When the suggestion is the same for two or more different results, these observations will be sorted on the basis of preferred outcome, again in ascending order. In the example above, this applies to the third and the fourth row, where two Choosers chose the same suggestion but had different preferred outcomes. More generally, observations have been sorted first on suggestion, then on preferred outcome, then on proposal and finally on acceptance or rejection.

**GRAPH**

On the right of the history overview, the most recent results are represented in a graph. The horizontal axis presents the suggestion and the vertical axis presents the proposal. Each previous observation is represented by a square. On the horizontal axis you can read the value of the suggestion for a particular result and on the vertical axis you can read the value of the corresponding proposal. If the square is green, the particular proposal was accepted and if the square is red with white inside, the particular proposal was rejected. (Choosers will see proposals on the vertical axis.)

**EXAMPLE 1.** Consider the square that is displayed in the lower left corner of the Graph shown above. Here, the Chooser made a suggestion of 20. This Chooser’s preferred outcome was 30.

You have now reached the end of the instructions. The next page contains some questions concerning the experiment. When all participants have answered all questions correctly, we will proceed with the experiment.
QUESTIONS

Please answer the following questions. THE VALUES USED IN SOME QUESTIONS DO NOT INDICATE WHAT YOU SHOULD DO IN THE EXPERIMENT. RATHER, THEY HAVE BEEN CHOSEN TO FACILITATE CALCULATIONS.

1. Is the following statement correct? ‘In each period I am coupled with the same Chooser.’

2. Is the following statement correct? ‘I will observe the Chooser’s preferred position before I make my proposal.’

3. (A) What is the highest value the preferred outcome of a Chooser can take on?
   (B) What is the highest value a suggestion of a Chooser can take on?
   (C) What is the highest value a proposal of yours can take on?

4. Consider a period in which the Chooser’s preferred outcome is 50. The Chooser chose to send a suggestion of 40. You made a proposal of 20, which was accepted by the Chooser.
   (A) What are your own earnings in this period?
   (B) How much does the Chooser to whom you are paired earn?

5. Consider a period in which the Chooser’s preferred outcome is 90. The Chooser chose to send a suggestion of 100. You made a proposal of 0, which was accepted by the Chooser.
   (A) What are your own earnings in this period?
   (B) How much does the Chooser to whom you are paired earn?

6. Consider a period in which the Chooser’s preferred outcome is 30. The Chooser chose to send a suggestion of 40. You made a proposal of 10, which was rejected by the Chooser.
   (A) What are your own earnings in this period?
   (B) How much does the Chooser to whom you are paired earn?

When you are ready answering the questions, please raise your hand.
Chapter 5  Formal versus Informal Legislative Bargaining

5.1  Introduction

The outcome of a legislative bargaining process is usually a result of both formal and informal bargaining. When parliament is in session, parliamentary procedures strictly govern what members can do at what time; hence, bargaining is highly formalized. After official sessions have been adjourned, however, bargaining often continues informally in offices, corridors and backrooms, where formal rules barely exist. That bargaining occurs at different levels of formality likely has historical and functional reasons: informal bargaining is arguably faster, whereas formal bargaining provides transparency and legitimacy to the democratic process. The question we address in this chapter is whether the formality of bargaining also systematically affects the bargaining outcome. This is important for understanding institutional choice and parliamentary procedures.

That the bargaining procedure can drastically affect the outcome has been recognized at least since the research boom on spatial voting in the late 1970s. If the procedure favors specific negotiators (e.g., through the order of voting, agenda-setting power, or proposal and voting rights), the outcome may crucially depend on it (e.g., McKelvey (1976; 1979), Schofield (1978), McCarty (2000)). The effect of formality seems different on at least two accounts, however. First, the difference between formal and informal bargaining cannot be captured in terms of changing the agenda or proposal or voting rights. Second, moving from a formal to informal bargaining or vice versa does not prima facie favor specific negotiators in any obvious way. The difference between the two is that informal bargaining provides much more flexibility to the bargaining parties. It does not give more flexibility to some parties than to others, however.

65 This chapter is based on De Groot Ruiz, Ramer & Schram (2011).
Intuitively, the choice of how much weight to put on formal versus informal procedures may be determined by strategic considerations (Elster (1998), Stasavage (2004)). For instance, parties with a strong bargaining position may prefer backrooms and wish to reserve formal voting for well negotiated deals. On the other hand, parties with more extreme positions might prefer to avoid backrooms and follow the more formal procedures in order to allow their proposals to have a chance of success. This study intends to help us better understand such preferences.

More specifically, we compare two bargaining procedures, which we believe are representative for formal and informal bargaining in the field. To obtain a clean comparison, in both cases the bargaining procedure is ‘fair’ in the sense that it does not prima facie favor any negotiator. In this important way, our study differs from the legislative bargaining literature of the 1970s discussed above. The main question we address is whether the increased flexibility of the informal compared to the formal procedure affects the legislative outcome. In addition, if it does, does it do so for purely strategic reasons or do psychological effects play a role? To provide an answer to these questions we analyze legislative bargaining both theoretically and in a controlled laboratory experiment.

In the informal procedure, players can freely make and accept proposals at any time. We did not choose for a completely unstructured face-to-face setting, but instead opted for a computerized setting where players can make and accept proposals in continuous time. This allows us to analyze the procedure as a non-cooperative game and to collect data on the bargaining process. We believe that the procedure is sufficiently unrestricted to be representative for informal bargaining like that which takes place in parliamentary backrooms. As we will see, the procedure is also not restrictive in the sense that it imposes no strategic constraints on the players. In the formal procedure, proposals and voting are regulated by a finite, closed-rule Baron-Ferejohn (1987) alternating offers

---

In the 1970s, several experiments used informal bargaining procedures to compare the many cooperative solution concepts that had been proposed. Amongst the first were Fiorina & Plott (1978). The procedures used tend to be rather different from ours, however. More importantly, these studies do not compare their informal procedure to a formal procedure, nor do they model it as a non-cooperative game.
scheme. Though there are potentially very many fair formal procedures, the Baron Ferejohn framework is widely taken to be a suitable model for studying formal legislative bargaining. Our procedure is an elementary Baron-Ferejohn scheme.

We study the effects of formality in the context of a three-player legislative bargaining setting. The game is a straightforward extension of the standard one-dimensional median voter setting (Black, 1948; 1958) and has the following motivation. In the standard setting, the median player’s ideal point is the unique (strong) core outcome irrespective of the location of others’ ideal points (as long as they are on the same dimension). However, intuitively one may expect that the outcome of a legislative bargaining process or the coalition supporting this outcome is less stable if preferences are far apart —i.e., if polarization is strong—, even if the policy space seems unidimensional. One explanation is that the disagreement point may well lie outside of the line on which all policy proposals are defined. This is a point, we believe, that has hardly been appreciated in the literature. Such a situation may occur for various reasons. First, a decision often involves a new type of policy or project so that the status quo may not fall in the space under consideration. Second, if the disagreement point consists in the termination of a project or a coalition, then it may involve significant transaction costs (e.g., involving new elections). If so, the disagreement point will be of a qualitatively different nature than the issue under negotiation.

67 Baron & Ferejohn (1987) compare open and closed amendment rules and find distinct equilibria. Note that both settings constitute formal bargaining procedures.
68 See, in addition to the work by Baron and Ferejohn, amongst many others, Banks & Austen Smith (1988), Merlo & Wilson (1995), McCarty (2000), Diermeier, Eraslan & Merlo (2003), Battaglini & Coate (2008) and Banks & Duggan (2000; 2006). These models tend to reach similar conclusions about agenda setting power. The first proposal is typically always accepted in equilibrium, since players know which proposals would subsequently be accepted or rejected. This gives a great advantage to the player chosen to make the first proposal (Palfrey 2006). Experiments, however, only partly corroborate these theoretical findings (McKelvey (1991), Diermeier & Morton (2005), Frechet, Morelli & Kagel (2005)). The first proposer does indeed have an advantage, but this is not as large as theoretically predicted and, in fact, the first proposal is often rejected, leading to ‘delay.’
69 The only exception we are aware of is Romer & Rosenthal (1978), who make a similar observation when they compare competitive majority rule to a controlled agenda setting mechanism. They do not consider polarization.
An example serves to illustrate the environments we are thinking of. Imagine a legislature that consists of three factions (doves, moderates and hawks) and is deciding on the renewal of a budget for an ongoing war. No single faction holds a majority and any coalition of two does. Doves prefer a reduction of the current war budget, moderates want no change and hawks would like an increase. Preferences are single-peaked with respect to budget revisions. The option to end the war (‘retreat’) serves as a disagreement point, which cannot simply be represented as a budget revision. (Retreating is qualitatively distinct from a reduction of the budget to zero and, furthermore, spending 5 billion on retreating is quite different from spending this amount on war efforts.) Preferences are such that all parties prefer some revisions to retreating. Polarization is then defined as the distance between the ideal revisions of the factions (relative to the attractiveness of retreating\textsuperscript{70}) and captures the extent of divergence of interests. Polarization will most likely drive the stability of the outcome. If polarization is weak, then the factions’ preferences lie close together and retreating is a relatively unattractive agreement. Hence, all coalitions will prefer the median ideal to retreating. In this case, we can use Black’s Median Voter Theorem (1948, 1958) to predict that the moderates’ ideal point will prevail. If the ideal revisions are very far apart, then polarization is strong. In this case retreating is relatively attractive and there are no revisions that any coalition prefers to retreating. With moderate polarization, we get a cyclical pattern. Both doves and hawks prefer retreating to the unaltered budget; however, moderates and doves prefer some negative revisions to retreating; and moderates and hawks prefer the unaltered budget to negative revisions. In this case, it is not clear what the legislature may decide. Intuitively, one may expect the moderates to have the highest bargaining power, as doves and hawks can only coordinate on retreating.

Our model and experimental design take the same basic form as this example. The point of departure is a bargaining problem in a median voter setting that is modified to have an exterior disagreement point. Then, we introduce the formal bargaining procedure to obtain the formal bargaining game and the

\textsuperscript{70} In particular, relative to the distance between a player's ideal budget and the budget(s) she finds equally attractive as retreating.
informal bargaining procedure to obtain the *informal* bargaining game. We analyze cooperative solutions for the bargaining problem and for both games we derive non-cooperative equilibrium predictions.

The bargaining outcome will typically depend on specific characteristics of the bargaining problem (*i.e.*, the extent of polarization). As a consequence, the effect of the procedure may also depend on these characteristics. More specifically, the relevance of formality may arguably be dependent on whether or not the bargaining problem has a core.\(^7\) Our setup allows us to obtain distinct outcomes with respect to the core by varying the level of polarization. When polarization is weak, the core consists of the median ideal and with strong polarization the core is the disagreement point. With moderate polarization, the core is empty.

The non-cooperative predictions depend on the procedure.\(^7\) For the formal game, we derive a unique (refined) subgame perfect equilibrium (SPE) that converges (with the number of bargaining rounds) to the core element when this exists. It typically does not converge at all when the core is empty. In the informal game the disagreement point and all points between the players’ ideal points can be supported as an SPE-outcome, irrespective of the extent of polarization. In addition, the equilibrium set cannot be refined in any standard way.

\(^7\) When the core is empty, all outcomes can be supported by some agenda-setting institution (McKelvey (1976; 1979), Schofield (1978)). This is important, because the core will be empty unless extreme symmetry conditions are satisfied (Gillies (1953), Plott (1967), Riker (1980), Le Breton (1987), Saari (1997)). If the core is non-empty, the (non-cooperative) equilibrium outcome for many procedures tends to lie in it (Perry & Reny (1993; 1994), Baron (1996), Banks & Duggan (2000)). There is indeed experimental evidence on the stability of core-outcomes (Fiorina & Plott (1978), McKelvey & Ordeshook (1984), Palfrey (2006)). On the other hand, the outcome is sometimes sensitive to fairness considerations (Isaac & Plott (1978), Eavey & Miller (1984)). Structure may matter even if there is a core, for instance, if some procedures are considered fairer than others (Bolton, Brandts, & Ockenfels, 2005).

\(^7\) Two theoretical breakthroughs have allowed us to overcome challenges in studying the effects of formality. First, for a long time many cooperative solution concepts have been advanced for situations in which the core is empty but none found broad theoretical and empirical support. Miller’s (1980) uncovered set as a generalization of the core drew theoretical support (Shepsle & Weingast (1984), Banks (1985), Cox (1987), Feld, Grofman, Hartley, Kilgour & Miller (1987)). However, systematic empirical tests were problematic, as it was impossible to compute the uncovered set in most cases. By developing an algorithm to find the uncovered set, Bianco, Lynch, Miller & Sened (2006; 2008) managed to find solid empirical support using data from many old and new experiments. Second, continuous time bargaining has made the non-cooperative analysis of low-structure settings possible (Simon & Stinchcombe (1989), Perry & Reny (1993; 1994)).
Our interpretation is that informality offers so much strategic flexibility that strategic considerations alone cannot identify an outcome.

From our experiments, we have two main findings: polarization matters and formality matters. Polarization has a strong impact on the outcome. In accordance with theory, the median player is significantly worse off with moderate than with weak polarization. However, we find that increased polarization hurts the median player and does so even at weak levels when her most preferred outcome remains the unique core element. Our experimental findings suggests this is due to intra-coalitional fairness considerations. Such considerations become less important as negotiators gain more experience, however. After players have repeatedly played the game (in ever-changing groups), competition between coalitions is strengthened and the position of the median player also becomes stronger.

Our second result is that the formality of the bargaining procedure matters. The median player is significantly better off with the informal than with the formal procedure. One plausible cause seems to be that flexibility in making proposals at any time increases her ability to exploit her superior bargaining position, as observed by Drouvelis, Montero & Sefton (2010) in a different setting. This points to the more general idea that parties in a superior bargaining position will prefer institutions that impose less structure on the bargaining procedure.

The remainder of this chapter is organized as follows. Section 5.2 models the bargaining problem as a cooperative game and derives solutions for it. Section 5.3 describes and solves the non-cooperative games for the formal and informal bargaining procedures. Our experimental design is presented in section 5.4 and the experimental results are presented in section 5.5. Section 5.6 concludes.

5.2 The Bargaining Problem and Cooperative Solutions

Formally, the bargaining problem is represented by $\Gamma = \Gamma(N, Z, u, W)$ and consists of a finite set $N$ of players, thought of as factions in a legislature; a
collection $W$ of subsets of $N$, thought of as winning coalitions; a set $Z$ of alternatives; and utility functions $u_i$, one for each player $i \in N$ representing $i$’s preferences over $Z$. Note that although winning coalitions have been specified, nothing as yet has been said about the decision making process itself. Procedures governing this process will be described and formalized in the next section.

In the bargaining problems studied here, three players ($N = \{1,2,3\}$) bargain over the set of alternatives represented by $Z \equiv \mathbb{R} \cup \delta$, with $\mathbb{R}$ denoting the set of real numbers and $\delta$ the disagreement point. Each player $i \in N$ has an ideal point $z_i \in \mathbb{R}$. Without loss of generality we normalize by setting $z_1 = -a < 0$, $z_2 = 0$, and $z_3 = b > 0$, with $b \geq a$. Hence, the ideal point of player 2, the median player, is $z = 0$. For players 1 and 3, the wing players, $z_1$ is normalized to lie closer to 0 than $z_3$. The distance, $a$, between the closer wing player and the median player will be interpreted as a measure of the polarization of players’ preferences. We shall distinguish three cases of respectively, weak ($a \leq 1$), moderate ($1 < a < 2$), and strong polarization ($a \geq 2$).

Preferences of all players are single-peaked on $\mathbb{R}$ and represented by piecewise linear von Neumann-Morgenstern utility functions $u_i(z) = 1 - |z - z_i|$. We further assume that the utility attributed to the disagreement point is normalized at 0, that is, $u_i(\delta) = 0$, for all $i \in N$. Hence, each player has an open interval of outcomes with strictly positive values; to wit, $(-a - 1, -a + 1)$ for player 1, $(-1, 1)$ for player 2, and $(b - 1, b + 1)$ for player 3. Note that the endpoints of these intervals yield utility of 0, while the outcomes outside of these intervals are strictly worse for the respective players than the disagreement point $\delta$. Figure 5.1 depicts this payoff structure.
As for winning coalitions $W$, we assume that any majority of two players can implement any $z \in Z$ as the outcome. This can be achieved in various ways, determined by the structure of the bargaining process (see section 5.3).

$\Gamma$ can be regarded as a cooperative game and more precisely as a coalitional game without transferable payoff.\(^{73}\) We start by defining the dominating and covering relations for any given $\Gamma(N, Z, u, W)$.

**Definition 5.1** Let $z, z' \in Z$. We say that

(i) $z'$ dominates $z$, and write $z' \succ z$, if there is a winning coalition $M \in W$ such that all members in $M$ strictly prefer $z'$ to $z$;

(ii) $z'$ covers $z$, and write $z'Cz$ if $z' \succ z$ and $z'' \succ z' \Rightarrow z'' \succ z$, for all $z'' \in Z$.

The assumption that the disagreement point $\delta$ does not coincide with a point on the line $\mathbb{R}$ is important. If $\delta$ did lie in $\mathbb{R}$, then we would have a standard median voter setting and the median ideal $0$ would dominate all other possible outcomes.

The counter-positive equivalent of Definition 5.1 reads as follows.

\(^{73}\) Note that individual players cannot achieve any outcome by themselves and hence the payoffs available to singleton coalitions are not independent of the actions of the complementary coalition. Hence, under some definitions it would fall outside of the class of coalitional games with non-transferable utility.
**Definition 5.2** For an alternative $z \in Z$ we say that

(i) $z$ is undominated if for every $z'$ the set of players who strictly prefer $z'$ over $z$ is not a winning coalition;

(ii) $z$ is uncovered if for every $z'$ which dominates $z$ there is $z''$ which dominates $z'$ and does not dominate $z$.

To obtain a solution we look at the core and, when the core is empty, at the uncovered set, which is a generalization of the core. These are defined by:

**Definition 5.3**

(i) The core $c(\Gamma)$ of $\Gamma$ is the set of all points in $Z$ that are undominated;

(ii) The uncovered set $u(\Gamma)$ of $\Gamma$ is the set of all points in $Z$ that are uncovered.

Intuitively, the uncovered set is a ‘two step core.’ If an outcome $z$ is uncovered, there might be an outcome $z'$ that dominates it, but this outcome $z'$ will itself be dominated by an outcome $z''$ that does not dominate $z$. This means for instance, that forward-looking negotiators might be hesitant to move away from a point in the uncovered set. The uncovered set has several appealing theoretical properties. It is never empty, is equal to the core if the latter is nonempty and strict (Miller, 1980), contains all Von Neumann Morgenstern sets (McKelvey, 1986) and it subsumes the Banks set (Banks, 1985).

McKelvey argued that the uncovered set could be seen as a “useful generalization of the core when the core does not exist” (1986). Recently, this concept has also attracted significant empirical support (Bianco, Lynch, Miller & Sened 2006; 2008).

Whereas the uncovered set is typically large and difficult to calculate, in our bargaining problem it is small and simple. Table 5.1 gives the core and uncovered set for our bargaining problems.

---

74 These different papers prove these relations under slightly differing conditions.

75 In addition, in our game the uncovered set is refined in a nice way by the von Neumann Morgenstern set and the bargaining set, both of which are unique. More details are given in Appendix 5.7.1.
# TABLE 5.1

## COOPERATIVE SOLUTIONS

<table>
<thead>
<tr>
<th>Polarization</th>
<th>$\mathcal{C}(\Gamma)$</th>
<th>$\mathcal{U}(\Gamma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>weak: $a \leq 1$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>moderate: $1 &lt; a &lt; 2$</td>
<td>$\varnothing$</td>
<td>${-a + 1, 0, \delta}$</td>
</tr>
<tr>
<td>$a = b$</td>
<td>${-a + 1, 0, b - 1, \delta}$</td>
<td></td>
</tr>
<tr>
<td>strong: $a \geq 2$</td>
<td>$\delta$</td>
<td>$\delta$</td>
</tr>
</tbody>
</table>

Notes: For $\Gamma = (N, Z, u_i, W)$, the table gives the elements in the core ($\mathcal{C}(\Gamma)$) and uncovered set ($\mathcal{U}(\Gamma)$) for the levels of polarization distinguished in the first column.

When polarization is weak or strong, the core is nonempty and coincides with the uncovered set. It always holds that players 1 and 2 prefer the median ideal to all points right of it, whereas players 2 and 3 prefer the median ideal to all points left of it. In addition, when polarization is weak, players 1 and 2 also prefer the median ideal to the disagreement point. Hence it dominates all points, and, as a consequence, also covers them all. Thus, if polarization is weak, the median ideal is the singleton core and uncovered set. When polarization is strong, no point on the line exists that two players prefer to the median ideal. Hence, the disagreement point dominates all points on the line and constitutes the core and uncovered set. From a behavioral perspective, this case seems less interesting, as no alternative outcome to the disagreement point seems viable.

When polarization is moderate, we get a circular dominance pattern and the core is empty: The median ideal dominates all other points on the line, but is dominated by the disagreement point. The disagreement point itself is in turn dominated by some points on the line (which are dominated by the median ideal etc.). This also means that the median ideal and the disagreement point do not cover each other; they are uncovered. Furthermore, the point closest to the median player that is not dominated by the disagreement point, $-a + 1$, is uncovered as well. (The same holds for $a - 1$ if $a = b$). Consequently, in this case the uncovered set consists of three or four elements: the median ideal, the disagreement point, $-a + 1$ and, if $a = b$, $a - 1$. 

---

ESSAYS ON BARGAINING AND STRATEGIC COMMUNICATION

142
For technical details on the dominance relations and the cooperative solutions, see Appendix 5.7.1.

5.3 Formal and Informal Procedure

We now impose procedures on the bargaining problem described in section 5.2 and analyze the resulting non-cooperative games. One may think of this as the legislature selecting exactly one element of the set of feasible alternatives \( Z \) by means of a procedure established (or agreed upon) in advance. Formally, such a procedure can be regarded as an extensive game. We shall present and discuss two such games, exemplary for two important frameworks for legislative bargaining, voting and open bargaining. We do so by introducing a ‘formal’ and an ‘informal’ bargaining procedure to \( \Gamma = \Gamma(N, Z, u, W) \).

5.3.1 Formal procedure

We begin with a formal procedure for the selection of an outcome in \( Z \), represented by a sequential voting game \( \Gamma^F_T \). The game — similar to that in Baron & Ferejohn (1989) — consists of multiple rounds, with a predetermined maximum of \( T \) rounds. Each round comprises of three stages.

At stage 1, one player is randomly selected with equal probability across players. At stage 2, the selected player \( i \) submits her proposal. At stage 3, players vote independently on this proposal. It becomes the final choice if it is accepted by at least two players. Because the player who submitted it supports her own proposal (by assumption), support by one other player suffices to pass the proposal and end the game. Whenever the proposal is voted down by the two other players, it is off the table and the game proceeds to round \( t + 1 \), where a player is selected to submit a new proposal, and so on. If the game reaches round \( T \) and the final proposal is also rejected, the game ends and the disagreement point \( \delta \) is implemented.

For any given \( T \) and bargaining problem \( \Gamma \), the game \( \Gamma^F_T \) is an extensive form game of finite length with random moves by nature at the first stage and simul-
taneous moves by all three players at the third stage of each round.\[^{76}\] Actions played at any stage are observed before the next stage or round begins. In general, players’ best responses will not be unique so one can expect multiple Subgame Perfect Nash Equilibria (SPE), possibly with distinct outcomes. In order to select a single best response at each stage, and ultimately to select consistently a single SPE for every given $T$, we assume players vote as if they are pivotal and shall adopt a number of tie-breaking rules known from the literature on voting games (Baron & Ferejohn (1989), Baron (1996), Banks & Duggan (2006)):

(i) A player accepts a proposal submitted in round $t$ if it provides to her a payoff equal to her expected equilibrium payoff in the subgame beginning at stage 1 of round $t + 1$.

(ii) Whenever a player has two best proposals, one that will be accepted and one that will be rejected, she submits the proposal that will be accepted.

(iii) Whenever $-c$ and $c$ are both best proposals for player 2 she submits each of them with an equal probability.

(iv) Whenever $\delta$ and $c \in \mathbb{R}$ are both best proposals for a player she submits $\delta$.

The first assumption guarantees that an SPE exists, the remaining assumptions imply that it is unique.\[^{77}\] From now on, we will refer to this equilibrium simply as ‘the equilibrium’ of $\Gamma^F_T$. Note that equilibrium strategies in a round depend neither on what happened in previous rounds nor on the total number of rounds ($T$), but only on the number of rounds left before the game ends.

The equilibrium outcome of $\Gamma^F_T$ can be characterized by the probability distribution of the equilibrium outcomes $\mu^T : Z \rightarrow [0,1].\[^{78}\]$ If all equilibrium proposals are accepted in the first round, $\mu^T$ simply allots equal probability to each of the players’ equilibrium proposals in the first round. There is, however, the

\[^{76}\] Though the player that made a proposal has an action set consisting of one element (‘accept’) at the 3rd stage.

\[^{77}\] Assumption 4 is particular to our game (as $\delta \notin \mathbb{R}$) and is a convenient tie breaking rule for the case $a = 2$. For other parameter values it is not essential which rule one assumes for these ties.

\[^{78}\] $\mu^T$ is a probability mass function. As we show in Appendix 5.7.2, $\mu^T$ has countable support.
possibility of delay. Though for some values of parameters $a$, $b$ and $T$, all three equilibrium proposals are immediately accepted, for other values an equilibrium proposal, in particular that of player 3, will be rejected. The equilibrium outcome of $\Gamma^F_T$ may depend in complicated ways on the number of bargaining rounds, $T$.

Hence, our approach is to look at whether the equilibrium outcome converges as $T$ increases. We say that for given values of $a$ and $b$ the equilibrium outcome converges if there exists a probability distribution $\mu^*_{a,b}$ on $\mathbb{Z}$ such that

$$\lim_{T \to \infty} \mu^T = \mu^*_{a,b}$$

in the sense of weak convergence of probability measures (Billingsley, 1999). If no such limit exists, then we say that the outcome does not converge. The equilibrium outcome converges to some single $z \in \mathbb{Z}$, if $\mu^*_{a,b}$ is concentrated at $z \in \mathbb{Z}$, i.e.

$$\mu^*_{a,b}(z) = 1.$$ 

This allows us to summarize the SPE of $\Gamma^F_T$ in the following proposition:

**Proposition 5.1**

(i) If $0 \leq a < 1$ or $a = b = 1$, the equilibrium outcome converges to 0. For $T$ sufficiently large, the first round proposals are accepted without delay.

(ii) If $1 \leq a \leq b < 2$ and $b > 1$, the equilibrium outcome does not converge except for some patches of the values of $a$ and $b$, and never to a single outcome in $\mathbb{Z}$.

(iii) If $a \geq 2$, the equilibrium outcome is $\delta$ for sufficiently large $T$.

**Proof:** The proof is given in Appendix 5.7.2. It is a terse and long exercise in backward induction, since the non-convexity of the outcome set precludes the use of standard techniques and results.

---

79 Numerically, the equilibrium outcome can be calculated for each value of $a$, $b$ and $T$. Simulations show that it always appears to converge to a cycle in $T$, the length of which depends in erratic ways on $a$ and $b$. To illustrate, Appendix 5.7.2 provides simulation results that show how the period of the cycles depends on $a$ and $b$. 

145
Comparing Proposition 5.1 to Table 5.1, we conclude that the equilibrium outcome converges to the single element of the core for \( a < 1 \) and \( a \geq 2 \), and that it does not converge to a single outcome if the core is empty.\(^{80}\)

Proposition 5.1 specifies the parameter configurations for which we have robust predictions. Note that legislative bargaining in the field can go through a substantial number of rounds, but that the number is often finite (due to time constraints, for example). If the outcome converges as the number of rounds increases, this allows for a stable prediction for such cases. If there is a sufficient number of rounds, the prediction will not depend significantly on the exact number, on who gets to propose first or on the precise values of \( a \) and \( b \). If the outcome does not converge, however, then the outcome will depend on all these parameters, and typically in a very sensitive and non-linear way.\(^{81}\) For instance, in our experiment we will use \( T = 10 \). For this case, we can derive a unique prediction, whether the core is empty or not (See Table 5.3 below). When the core is nonempty, it does not matter much whether \( T \) is 9, 10, 11, 20 or 100 or whether \( a=.4 \) or \( a=.41 \) or whether player 1 or player 3 starts.\(^{82,83}\) When the core is empty, however, then the outcome depends crucially on all these parameters. Though we have predictions for the specific parameters of our experiment, they will be strongly affected by small changes. This makes the outcome hard to predict in practice. In addition, this arguably makes it more difficult for players to coordinate on the equilibrium. Hence, Proposition 5.1 helps us understand when specific predictions for finite \( T \) are robust (i.e., this is the case for (i) and (iii)).

### 5.3.2 Informal Procedure

Now, we turn to the informal bargaining procedure where all players can make and accept proposals in continuous time, which we denote by \( I_T \). The

---

\(^{80}\) In the non-generic case of \( a = 1 \) and \( b > 1 \) the core consists of 0 but the equilibrium outcomes do not converge as \( T \) goes to infinity.

\(^{81}\) See footnote 79.

\(^{82}\) In addition, in the field (as in our experiment) the outcome set will not be exactly continuous but discrete and very fine-mazed. In this case, the outcome even converges in a finite number of rounds if the core is non-empty.

\(^{83}\) Parameter values exist (especially those close to values where the core is empty), where more than 10 rounds are required to see converging behavior.
basic tenet of $\Gamma^T_i$ is a triple of proposals $\left(p^i_1, p^i_2, p^i_3\right)$ on the table at all times $t \in [0, T]$ until one of the proposals is accepted lest the game ends with the disagreement point $\delta$ at time $T$. It has by now been well established that such a game with continuous time cannot be solved without further assumptions, however (Simon & Stinchcombe (1989), Perry & Reny (1993; 1994)). Drawing on Perry & Reny’s game (1993), we introduce a reaction and waiting time. Our game requires different rules than those used by Perry & Reny (1993; 1994), nonetheless, as it concerns a more complex bargaining problem. In contrast to Perry & Reny (1993), $\Gamma$ is not a two-player game; $\Gamma$ is different from the bargaining problems Perry & Reny (1994) consider in that $\Gamma$ can have an empty core and does not have transferable payoffs.

The rules of the game are as follows. Player $i$ can either be silent ($\varsigma$), $p^i_t = \varsigma$, have a proposal on the table, $p^i_t = z, z \in \mathbb{Z}$, or accept the proposal of another player $j$, $p^j_t = a_j$. For each player $i$, $p^i_t$ as a function of time is assumed to be piecewise constant and to be right-continuous. We say that a player moves at time $t$ when $p^i_t = p^i_t + \equiv \lim_{t \uparrow t} p^i_t$. It is natural to only allow only such discrete changes in proposals, since actual negotiations (face-to-face or computerized) consist of discrete actions (‘I propose $x$', ‘I accept’, ‘I withdraw $y$’) in a continuous time.

Players start with no proposal on the table: $p^0_i \equiv \varsigma$. We introduce a uniform reaction and waiting time, $\rho$. In particular, if some player moves at time $t$, no player can move at $t \in (s, s + \rho)$. This models the fact that players cannot react (or act again) immediately after a player has moved and that the time it takes to process information, make a decision and execute it is roughly the same for all players at all times. Essential is that we allow $\rho$ to be arbitrarily small.\textsuperscript{85}

\textsuperscript{84} We also allow a player to resubmit her old proposal and induce the reaction time $\rho$. Intuitively, this is the strategic move “I still propose $z$.” Technically, if $p^i_t = z$, then we define the move that resubmits the same proposal as $p^i_t = z$. (We set $z'' \equiv z$, such that if $p^i_t = z''$, then resubmitting $z$ is $p^i_t = z$.) If accepted, $z''$ just induces $z$ as outcome.

\textsuperscript{85} In our model the waiting time is exactly equal to the reaction time, unlike in Perry and Reny (1993). What is important is that we exclude the possibility of making a proposal and then withdrawing it before it can be accepted. For this purpose, any reaction time smaller than
Player $i$ accepts $j$'s proposal by setting $p'_i = a_j$. In order to ensure that a player knows which proposal she accepts, if player $i$ plays $p'_i = a_j$ she accepts $p'^{-}_j$. To ensure a unique, well-defined outcome a player $i$ can only accept a proposal at time $t$ if she is silent herself ($p'^{-}_i = \zeta$). In addition, one can (naturally) not accept a proposal from a player who is silent. As soon as a proposal has been accepted, the accepted proposal is the outcome of the game. If no proposal has been accepted before or at $t = T$, then the outcome is the disagreement point $\delta$. After a proposal has been accepted or when $t > T$, no player can move anymore. Formally, we always let the game end at $t = T + \rho$.\footnote{This is because at time $t$, it is not yet known what happens at $t$ itself. Any time after $T$ would do.}

To define strategies and derive equilibria, we need to introduce some further definitions.

**Definition 5.4**

(i) A history, $h$, consists of a specification of $\{p'_{i}(t), p'_{i}(t), p'_{i}(t)\}$ for $t \in [0, \tau(h))$, where $\tau(h) \in [0, T + \rho]$ is the history’s end.

(ii) $\bar{t}(h) \equiv \sup\{t < \tau : p'_i(h) \neq p'^{-}_i(h)\}$ is the last moment before $\tau(h)$ that any player moved (if no player has moved, we conveniently set $\bar{t}(h) \equiv -\rho$).

(iii) A proposal function is right-continuous and piecewise linear if for all $\tau < T$ and each $i$, there is an $\varepsilon > 0$ such that $p'_i(h) = p'_i(h) \forall s \in [t, t + \varepsilon)$. As discussed above, we will only consider such proposals.

(iv) History $h$ is an active history if $\bar{t}(h) + \rho \leq \tau(h) \leq T$ and no proposal has been accepted.

\footnote{If a player were not required to be silent when accepting a proposal, her proposal could be accepted while she were accepting another. Essentially, we are requiring that a player removes her own proposal before accepting another. Given that $\rho$ can be arbitrarily small, this assumption is not behaviorally restrictive.}
The set of admissible proposals at each history for player \( i \) is called \( Z_i(h) \), with \( Z_i(h_r) \subset \tilde{Z} \), where \( \tilde{Z} = Z \cup \{\xi, a_1, a_2, a_3\} \).\(^{88}\) \( Z_i(h) \) is subject to the following restrictions:

- \( a. \) \( Z_i(h_r) = \{p_i^{r-}(h)\} \) if \( h_r \) is not an active history
- \( b. \) \( Z \cup \{\xi, p_i^{r-}(h)\} \subset Z_i(h_r) \) if \( h_r \) is an active history
- \( c. \) \( a_j \in Z_i(h_r) \) if \( h_r \) is an active history \( i \neq j \), \( p_i^{r-}(h_r) \neq \xi \) and \( p_i^{r-}(h_r) = \xi \)

The outcome of a history \( h \) is \( z(h) \):

\[
z(h) = \left\{ \begin{array}{ll}
\text{The unique element of } \{p_j^{r-}(h) : p_i^{r}(h) = a_j \text{ for some } t < \tau(h)\} \text{ if it exists}
& \text{if } p_j^{r-}(h) \ni \text{is not an active history}
\text{or } \{p_j^{r-}(h) : p_i^{r}(h) = a_j \text{ for some } t < \tau(h)\} = \emptyset \text{ and } \tau(h) \leq T
& \text{if } p_j^{r-}(h) \ni \text{is an active history}
\delta \text{ if } \{p_j^{r-}(h) : p_i^{r}(h) = a_j \text{ for some } t < \tau(h)\} = \emptyset \text{ and } \tau(h) > T
\end{array} \right.
\]

Let \( \overline{H} \) be the set of histories in which all proposals are right-continuous and piecewise linear and admissible and have \( \tau(h) = T + \rho \). Call any \( \overline{h} \in \overline{H} \) a resolved history.

Let \( H \) be the set of histories in which all proposals are right-continuous, piecewise linear and admissible, and that have no outcome \((z(h_r) = \emptyset)\). Any \( h_r \in H \) is called an unresolved history.

\( h \) is a subhistory of \( h' \), or \( h \subseteq h' \), if \( \tau(h) \leq \tau(h') \) and \( p_i'(h) = p_i'(h') \) for each \( i \) and for all \( t \in [0, \tau(h)] \). Furthermore, \( h \) is a proper subhistory of \( h' \), or \( h \subset h' \), if \( h \subseteq h' \) and \( \tau(h) < \tau(h') \).

\( h_s(h) \) is the unique history such that \( \tau(h_s(h)) = s, h \subseteq h_s(h) \) and no player moves in \([\tau(h), s)\).

\( H_s \equiv \{h \in H : \tau(h) = \tau'\} \) is the set of histories ending at \( \tau' \).

\( \tilde{H} \equiv \{h \in H : \tau(h) \geq \tilde{t}(h) + \rho\} \) is the set of active histories in \( H \).

\(^{88}\) Because we allow players to repeat their previous proposal, we actually have \( \tilde{Z} = Z \cup Z' \cup \{\xi', a_1, a_2, a_3\} \), where \( Z' \) includes all asterisked outcomes (cf. fn 84).
A strategy of player $i$ is a mapping from the set of unresolved histories to the set of admissible proposals, $\sigma_i : H \rightarrow \hat{Z}$. It meets the following two conditions:

(S1) For all $h_r \in H$, $\sigma_i(h_r) \in Z_i(h_r)$

(S2) For all $h_r \in H$, a time $\varepsilon > 0$ exists such that $\sigma_i(h_r) = \sigma_i(h_r) \forall h_r$ with $h_r \subseteq h_r \subseteq h_{r+\varepsilon}(h_r)$.

(S1) ensures moves are in the set of admissible proposals. (S2) ensures that the strategies result in well-defined histories with piecewise constant, right-continuous paths. By $\Sigma$ we denote the set of all profiles that meet (S1) and (S2).

Proposition 5.2 shows that $\sigma \in \Sigma$ induces from any well-defined unresolved history $h_r \in H$ a unique well-defined resolved history $\overline{h} \in \overline{H}$:

**Proposition 5.2** The game $\Gamma^I_t$ is a well-defined mapping $\Gamma^I_t : H \times \Sigma \rightarrow \overline{H}_{T+\varepsilon}$.

**Proof:** See Appendix 5.7.3.

The intuition underlying the proof is that for a given strategy profile and unresolved history $h_r$, either no player would do anything after $h_r$ until the game ends or one can find a well-defined first action at or after $\tau$. This action yields a new history, which is either a resolved history or an unresolved history. Hence, one can repeat this procedure of searching for the first action until it yields a resolved history.

We can now define subgames and equilibria of $\Gamma^I_T$. A subgame $\Gamma^I_{T+h_r} : \{h \in H : h \supseteq h_r\} \times \Sigma \rightarrow \overline{H}_{T}$, represents the game that starts at $h_r$. Let $U_i(\sigma_i; \sigma_{-i} \mid h_r) \equiv u_i(z(\Gamma^I_{T+h_r}(h_r, \sigma)))$ be the payoff player $i$ receives from the outcome that $\sigma$ induces on $h_r$ and let $U_i(\sigma_i; \sigma_{-i} \mid h_r) \equiv U_i(\sigma_i; \sigma_{-i} \mid h_r)$. A Nash Equilibrium of $\Gamma^I_{T+h_r}$ is a strategy profile $\sigma$ such that $U_i(\sigma_i; \sigma_{-i} \mid h_r) \geq U_i(\sigma'_i; \sigma_{-i} \mid h_r)$ for each $i$ and all $\sigma'_i \in \Sigma_i$. $\sigma$ is a Subgame Perfect Equilibrium (SPE) of $\Gamma^I_T$ if it is a Nash Equilibrium in all of its subgames: $U_i(\sigma_i; \sigma_{-i} \mid h_r) \geq U_i(\sigma'_i; \sigma_{-i} \mid h_r)$ for each $i$ and all $\sigma'_i \in \Sigma_i$ and in all $h_r \in H$. 

150
Finally, we are able to show by an explicit construction that every point of the interval \([-a, b]\) \(\subset \mathbb{R}\) and the disagreement point \(\delta\) itself can be an outcome of a subgame-perfect equilibrium of \(\Gamma^l_r\). This yields Proposition 5.3.

**Proposition 5.3** The set of SPE outcomes contains \([-a, b] \cup \{\delta\}\) for every continuous game \(\Gamma^l_r\) with \(T > \rho\).

**Proof:** See Appendix 5.7.3.

Many of the SPEs in Proposition 5.3 may seem unintuitive. For instance, the at first sight unlikely outcome \(b\) (which is the ideal point of the wing player furthest from the median player) can be supported by an equilibrium in which players 1 and 2 always propose \(b\). Player 3 will accept \(b\) as soon as she can, while players 1 and 2 cannot individually profitably deviate, as the other player will anyhow propose \(b\), which will be readily accepted by player 3.

This large set of SPE cannot be refined in any standard way. A simple set of standard tie-breaking rules, such as those used for \(\Gamma^f_r\), would be much too weak to have any effect. Stationarity only has a very small bite (\(b\) and \(\delta\) can, for instance, be sustained by stationary strategies for any \(a\) and \(b\)). A procedure of iterated elimination of weakly dominated strategies, as proposed by Moulin (1979), and used by Baron & Ferejohn (1989), is of little avail in our case, due to the fact that typically in many subgames of \(\Gamma^l_r\) multiple actions per player will survive, so that hardly any strategy will eventually be eliminated in the complete game. In addition, a refinement based on trembling-hand perfection (Selten, 1975), if it can be adapted to continuous time and space, will not eliminate these unintuitive equilibrium-outcomes either. For instance, the reason that player 1 does not propose 0 instead of \(b\) in the equilibrium discussed above, could be that she is afraid that player 2 might tremble and play \(b + \varepsilon\) instead of \(b\). Hence, there are few strategic restrictions on the equilibrium strategies in the informal game.

Nonetheless, there are some points in the outcome set that strike us as more ‘likely’ than others (for instance points in the uncovered set). Their plausibility might be the result of their *focal* nature due to the constellation of preferences.
and winning coalitions (which is captured by the cooperative solution concepts of the bargaining problem).

### 5.3.3 Overview of Theoretical Results

In Table 5.2 we summarize the main results obtained for the outcomes of the equilibria of the two strategic games analyzed in this section, $\Gamma_T^F$ and $\Gamma_T^I$, together with the solutions of the cooperative game $\Gamma$.

<table>
<thead>
<tr>
<th></th>
<th>Polarization</th>
<th>Weak: $a &lt; 1$</th>
<th>Moderate: $1 &lt; a &lt; 2$</th>
<th>Strong: $a &gt; 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperative Game $\Gamma (N, X, u_i, W)$</td>
<td>Core</td>
<td>0</td>
<td>$\emptyset$</td>
<td>$\delta$</td>
</tr>
<tr>
<td></td>
<td>Uncovered Set</td>
<td>0</td>
<td>${1-a, 0, a-1^*, \delta}$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>Formal (convergence), $\Gamma_T^F$</td>
<td>0</td>
<td>No convergence **</td>
<td>$\delta$</td>
<td></td>
</tr>
<tr>
<td>Informal (for $T \geq \rho$), $\Gamma_T^I$</td>
<td>$[-a, b] \cup {\delta}$***</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Cells give the solution concepts for the two games as derived above for the three generic cases of polarization ($a < 1$, $1 < a < 2$, $a > 2$). Solution concepts used are described in the previous subsections.

* $a-1$ is only included if $a = b$; ** There are some exceptions, in which case the outcome may converge but never to a single outcome in $\mathbb{Z}$. ***Outcomes can also lie in the interval $[\max\{b-1, -b\}, 0]$ if $\max\{b-1, -b\} < -a$.

### 5.4 Experimental Procedures and Design

The experiment was run at the Center for Research in Experimental Economics and political Decision making (CREED) of the University of Amsterdam. It was computerized using z-Tree (Fischbacher, 2007). An English translation of the Dutch experimental instructions is provided in Appendix 5.7.4. Subjects had to correctly answer a quiz before proceeding to the experiment. In total, 102 subjects were recruited from CREED’s subject pool. The subject pool consists of around 2000 individuals who have voluntarily registered. Almost all of these are undergraduate students at the University of Amsterdam, of which approxi-
up fee plus on average €11.65 in 90-120 minutes. In the experiment, payoffs are in ‘francs’. The cumulative earnings in francs are exchanged for euros at the end of the session at a rate of 1 euro per 10 francs.

We ran six sessions. Each session consists of 24 periods. In each period subjects are rematched in groups of three. We use matching groups of 6 or 9 subjects. After groups have been formed, subjects are randomly appointed the role of player ‘A’, ‘B’, or ‘C’. To avoid focality, players do not play the normalized game described above (e.g. B’s position is not set equal to 0 and it is not necessarily the case that A’s ideal value is closer to B’s value than C’s is). For analysis, the bargaining problem subjects face can easily be normalized to correspond to the model of section 5.2.

Each player is appointed an ‘ideal value’, which is an integer between 0 and 100 (inclusive). Player A’s ideal value is always the smallest and player C’s the largest. Players know all ideal values. Each group has to choose an integer between 0 and 100 (inclusive). If the group chooses a player’s ideal value, this player receives 20 francs. For every unit further from her ideal point, one franc is subtracted from her earnings. Hence, earnings are negative for a player if the group chooses a number that is more than 20 larger or smaller than her ideal value. To avoid negative total earnings at the end of the experiment, each subject starts with a positive balance of 100 francs.

The procedure was varied in a between subjects design, which consisted of a formal and an informal treatment. We ran three sessions per treatment, so that we have six matching groups per treatment. In both treatments, proposals are

---

mately 40% major in Economics or Business. When the experiment was announced, all received an invitation to sign up and participation was on a first-come, first-serve basis.

90 Subjects are told that they are in a session with 15 or 18 participants and will be rematched in every round.

91 The restriction to natural numbers is done for practical purposes. It is sufficiently fine-mazed to avoid affecting the equilibria described in the previous section in any relevant way. One difference is that if the outcome converges in $T$ to 0, it converges in finite time and, for our parameter values, in fact it already converges for $T < 10$. In Table 5.3, we report the equilibria of the high structure game with a discretized line and 10 rounds (which we calculated numerically).

92 Still, five subjects ended their session with negative earnings. They were sent off with no pay other than the €5 show-up fee. Data which involved these individuals were deleted from the sample due to possible incentive problems. Including these individuals makes little difference, except that statistical results become somewhat less conclusive due to the extreme behavior of one subject who would have earned –14.70 euros and showed erratic behavior after his earnings became negative.
made consisting of any integer between 0 and 100 (inclusive) or δ (called “end”). If the disagreement point is the outcome, each player receives a payoff of zero.93

In the formal treatment, subjects play the game Σ of section 5.3.1 with \( T = 10 \); i.e., negotiations were held for a maximum of 10 rounds per period. We use the strategy method (for proposals) where in every round every player is asked to make a proposal. One proposal is randomly selected and put to the other two group members to vote on. If at least one of the two accepts this proposal, it becomes the group choice and the game ends. If the proposal is rejected by both players, a new round begins, unless 10 rounds have been finished. In the latter case, the outcome is the disagreement point.

In the informal treatment, subjects are given two-and-a-half minutes to reach an agreement. At any time, any group member can make a proposal, change a previous own proposal or accept one made by another member. They do so by typing a number (proposal) and clicking on an ok-button, respectively selecting another member’s proposal and clicking on an accept-button. As soon as a proposal has been accepted, this becomes the group choice for the period and negotiations are finished. If no proposal is accepted within the time-span, the disagreement point is the outcome. This treatment follows our informal model closely. We do not impose a reaction or waiting times. In the model, these times do not represent procedural restrictions but rather cognitive and physical restrictions, which are allowed to be arbitrarily small. Nor do we require that players need to retract their own proposal before they can accept a proposal.94 Otherwise, the rules are identical as in \( \Sigma \).

To help a subject in determining her choices during the negotiations for a group decision, her screen always shows a history of previous rounds.95 current

---

93 In both treatments, a round of bilateral messages precedes group negotiations: each player may send a private message (consisting of a number between 0 and 100 or δ) to either or both other player(s). This is meant to reflect pre-negotiation lobbying. This cheap-talk does not affect the theoretical analysis presented in section 5.3.

94 In the model, we need this requirement to guarantee a well-defined outcome. In the experiment, we do not require this, as the probability that a player accepts a proposal at the exact same time her own proposal is accepted is zero. Implementing the additional restriction would not make a big behavioral difference (it would take two mouse clicks instead of one to accept a proposal), but would make the interface unnecessarily more cumbersome.

95 The history showed for each previous round what happened in the group the player participated in. In particular, it specified (i) the ideal point for each role, (ii) the role the player herself had, (iii) the outcome and (iv) the earnings for all three roles.
earnings, a scrollable help-box with instructions, a history of offers in the current round and a device to calculate payoffs for any hypothetical proposal.

Polarization is varied in a within-subjects design by using 12 sets of ideal values. Each set was used once in the first half (first 12 periods) and once in the second half (last 12 periods) of a session. The sets were chosen such that for the normalized parameters there were six with $a<1$ and six with $1<a<2$ (cf. Table 5.3, below). We chose not to use parameters with $a\geq2$ in the experiment because it seems obvious that participants will always agree on the disagreement point of no earnings if there is no outcome where at least two players have positive earnings.

Table 5.3 gives the (normalized) parameters used, the periods in which they were used and the theoretical predictions for each set. We can conclude a few things from this table about the predictions of the cooperative solutions and the equilibrium of the formal game. First, as long as $a<1$ (weak polarization) the median’s payoff does not depend on the level of polarization. Second, when $a>1$ the median’s payoff can be expected to decrease with polarization. Third, when the core is empty, there are many instances where the SPE-outcomes of the formal game are not in the uncovered set.
### TABLE 5.3
PARAMETERS AND PREDICTIONS

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Periods</th>
<th>Cooperative</th>
<th>Formal game</th>
<th>Informal game</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$b$</td>
<td></td>
<td>1 starts</td>
<td>2 starts</td>
</tr>
<tr>
<td>0.2</td>
<td>1.4</td>
<td>5, 23</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>1.1</td>
<td>3, 21</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>1.7</td>
<td>11, 13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>1, 19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.8</td>
<td>1.4</td>
<td>9, 15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.8</td>
<td>2</td>
<td>8, 24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.1</td>
<td>1.1</td>
<td>7, 17</td>
<td>${-0.1, 0, 0.1, \delta}$</td>
<td>$0.3$</td>
</tr>
<tr>
<td>1.1</td>
<td>1.7</td>
<td>10, 22</td>
<td>${-0.1, 0, \delta}$</td>
<td>$-0.4$</td>
</tr>
<tr>
<td>1.1</td>
<td>2.3</td>
<td>2, 14</td>
<td>${-0.1, 0, \delta}$</td>
<td>$-0.45$</td>
</tr>
<tr>
<td>1.4</td>
<td>1.4</td>
<td>12, 20</td>
<td>${-0.4, 0, 0.4, \delta}$</td>
<td>$\delta$</td>
</tr>
<tr>
<td>1.4</td>
<td>2</td>
<td>4, 16</td>
<td>${-0.4, 0, \delta}$</td>
<td>$-0.55$</td>
</tr>
<tr>
<td>1.7</td>
<td>1.7</td>
<td>6, 18</td>
<td>${-0.7, 0, 0.7, \delta}$</td>
<td>$\delta$</td>
</tr>
</tbody>
</table>

**Notes:** Cells give the theoretical prediction (cf. Table 5.1) applied to the experimental parameter set. The predictions for the formal procedure are for the game with a discretized outcome set and 10 periods, as played in the experiment.

1 For $a<1$ the prediction is given by the core (=uncovered set); for $a>1$ it is given by the uncovered set.

2 The column gives the (refined) SPE conditional on the player chosen to make the first offer. Player 2 is defined as the median position, 1 is the other player closest to the median.

### 5.5 Experimental Results

We focus on how formality (and its interaction with polarization) affects the ability of the median player to reach agreements close to her ideal point. All tests used below are two-sided and non-parametric, using each matching group (of six or nine participants) as one independent data point. We use the Wilcoxon signed rank tests for within comparisons and the Mann-Whitney test for between comparisons. Whenever we report statistically significant results for pooled Formal/Informal data only, the results are also significant at the 0.05 level for the disaggregated data where Formal and Informal are tested separately. $P$-values that are (unrounded) smaller than 0.05 are marked by an asterisk.
5.5.1 Earnings

We start with players’ earnings from negotiations. Figure 5.2 shows the payoffs for different levels of polarization (captured by $a$) and the two treatments. Most relevant are the payoffs of the median player. First consider the effect of polarization. Theory predicts that for weak polarization ($a < 1$) the median player will be able to secure her maximum payoff (of 1), whereas moderate levels ($1 < a < 2$) of polarization would hurt her (cf. Table 5.3).

![Figure 5.2](image)

This figure shows payoffs. The bars show the average payoffs of players per period. Player 2 is the median player and player 1 is the other player closest to her.

The experimental results show no obvious change at $a = 1$. Increasing polarization clearly affects the median player (player 2) negatively, even when $a < 1$. For example, Player 2 earns approximately 0.9 (close to the maximum of 1) when $a = 0.2$ (for both Formal and Informal) but only just over 0.79 for $a = 0.8$ in the informal setting. Her earnings are significantly lower for $a = 0.8$ than for $a = 0.2$ ($p = 0.01^*$). As predicted by theory, the median’s payoff is significantly lower for moderate ($a > 1$) than for weak ($a < 1$) polarization ($p < 0.01^*$).
Second, formality also has a clear effect. The median player is (significantly) better off in the *Informal* treatment than in the *Formal* treatment ($p=0.03^*$). The difference between treatments seems to increase with the extent of polarization. When polarization is very weak ($a=0.2$) the procedure does not affect player 2's earnings from negotiations. When it is relatively strong ($a=1.7$) the median earns more than twice as much in *Informal* than in *Formal*. Next, we further explore what drives these results.

### 5.5.2 How Polarization & Formality Affect The Median Player

We start by looking at whether participants manage to reach an agreement before the deadline. Figure 5.3 shows the number of proposals needed to reach agreement.

![Figure 5.3](image)

This figure shows the rounds or proposals before agreement. Bars show the fraction of agreements using the number of proposals depicted on the horizontal axis for Formal (top panel) and Informal (bottom panel). In Formal, a proposal in any round could only be made by the player selected to do so and there was a maximum of 10 rounds. In Informal, any player could make a proposal at any time during a period of at most 150 seconds.
In both treatments, agreement was reached within the limit (150 seconds or 10 rounds, respectively) in 99% of all cases. Hence, it almost never occurred that the disagreement point was forced upon the negotiators for missing their limit. Moreover, agreement was generally reached very quickly. In Formal, agreement was reached in at most 3 rounds in 88% of the cases and in Informal agreement was reached in at most 30 seconds in 82% of the cases. Consequently, binding (time) limits do not appear to be of any influence (for treatment effects). Players make significantly more proposals in Informal (4) than in Formal (2) ($p<0.01^*$), however.

The outcome of the game can be characterized by three dimensions. First, whether it is a real number (as opposed to disagreement). If so, second, its value (‘location’) and, third, its distance to the median position, i.e., its absolute value (‘distance’). We will look at each of these in turn. Figure 5.4 shows the percentages of outcomes that were a real number (‘frequency’). As long as polarization is weak ($a<1$), virtually all outcomes are real numbers and polarization is immaterial.

![Figure 5.4](image)

This figure shows the frequency (of outcomes were a real number). The bars show the percentage of outcomes that were a real number.
The frequency (of real number agreements) is, however, clearly and statistically significantly lower for moderate than for weak polarization \((p<0.01^*)\). Hence, a decrease in real number agreements may partly explain why moderate polarization is worse for the median player than weak polarization. However, it cannot explain why she cannot obtain her optimal payoff even when polarization is weak. Furthermore, there is no clear treatment effect. Real number outcomes are somewhat less likely in *Formal*, but the effect is small and insignificant \((p=0.33\) for \(a<1\) and \(p=0.22\) for \(a>1\)). Hence, the percentage of real number outcomes cannot explain why the median player is better off in *Informal*.

In search of such an explanation, we take a closer look at the real number outcomes players agreed upon. For completeness sake, we first depict the location of real number agreements in Figure 5.5 (although the location per se is not relevant for player 2’s payoff). This clearly shows that the average agreement is typically between the ideal points of the median player and player 1 (the other player closest to the median). In fact, there are only two cases with \(a\neq b\) where the average agreement lies between the ideal points of the median player and player 3. In both cases, player 1 still earns more than player 3. We will discuss the coalitions observed in more detail further on.

Given that an outcome is a real number, the median player’s payoff is fully determined by its distance to the median ideal (0). This is shown in Figure 5.6. This figure clearly shows that the distance increases with polarization. Distance is significantly higher for moderate than for weak polarization \((p<0.01^*)\). Distance matters even within weak levels of polarization: it is significantly higher for \(a=0.8\) than for \(a=0.2\) \((p=0.01^*\) (pooled), \(p=0.12\) (*Low*), \(p=0.05^*\) (*High*)).
CHAPTER 5.  FORMAL VERSUS INFORMAL LEGISLATIVE BARGAINING

Figure 5.5
This figure shows the location of real number outcomes. Bars show the average normalized location of agreements, when groups agreed on a real number. Negative (positive) numbers indicate an agreement in between the ideal points of players 1 (3) and 2. The median position is an agreement at 0. Whenever $a=b$, the non-median players are randomly located as players 1 and 2, so any agreement is equally likely to be normalized to a positive or negative number.

Figure 5.6
This figure shows the distance of real number outcomes. Bars show the average absolute distance between agreements and the median point, when groups agree on a real number.

Figure 5.6 also shows a clear treatment-effect. Distance is significantly lower for Informal than for Formal. Hence, player 2 seems to exploit her superior
bargaining position better in *Informal* than in *Formal*. A possible explanation is that players are freer to make proposals in the *Informal* negotiations, so that they can negotiate better. Recall that players make significantly more proposals in *Informal* (4.0) than in *Formal* (2.0) ($p<0.01^*$). In addition, we find that in *Formal* players use slightly less proposals in the last 12 periods (1.9) than in the first 12 periods (2.1). In contrast, in *Informal*, players use significantly more proposals in the last 12 periods (4.5) than in the first 12 periods (3.5) ($p=0.03^*$).

We conclude that the main driving force underlying the higher profits for the median player in the *Informal* treatment is that the more flexible bargaining procedure allows her to secure real number agreements closer to her preference.

### 5.5.3 Intra-coalitional Fairness vs. Inter-coalitional Competition

One intriguing question that remains is why even weak polarization hurts the median player, while her ideal is the unique core element. To address this question, we consider coalitions and the way in which outcomes distribute payoffs within them. Figure 5.7 shows the distribution of real number agreements divided by $a$. Hence, $-1$ represents an agreement at $-a$ (i.e., player 1’s ideal point), 0 represents the median ideal 0 and 1 represents $a$. 
This figure shows the distribution of real number outcomes and learning effects. Bars show the fraction of real number outcomes that are within 0.05 of the outcome depicted on the horizontal axis. The horizontal axis gives the normalized outcome divided by \(a\). The left panels show the distribution for \(a<1\), the right for \(a>1\). The top panels show the distribution for the first half (first 12 periods), the bottom for the last half (last 12 periods).

Strikingly, almost all real number outcomes lie between \(-a/2\) and \(a/2\), with \(-a/2\) being one of the most frequently chosen outcomes. Note that \(-a/2\) equalizes payoffs between players 1 and 2, but is a rather unfair outcome for player 3; in fact worse than the median preference. From a fairness perspective it might seem remarkable that the players in the coalition do not seem to care much about the player outside of the coalition. Nonetheless, this is in line with the findings in the three-person ultimatum games (Güth & Van Damme (1998), Bolton & Ockenfels (1998)), where the third powerless person (who can neither propose nor reject) is given little consideration.

It seems that players 1 and 3 in many cases demand some part of the ‘surplus’ in a coalition with player 2. However, player 2 does not give more than the fair split to player 1. Furthermore, player 3 does not obtain a better outcome than \(a/2\), since player 2 probably feels that she can certainly obtain \(-a/2\) in a coalition with player 1. Such considerations of intra-coalitional fairness yield real
number agreements increasing in $a$, even for weak levels of polarization, as we observe. Note, however, that as $a$ increases it becomes more costly to the median player to give her coalition partner a ‘fair share.’

Figure 5.7 also shows that there is a strong learning effect: the distribution of outcomes in the first half (first 12 periods) is very different from that in the last half (last 12 periods). In the first half, intra-coalitional fairness considerations seem to play an important role, certainly within coalitions of players 1 and 2. Furthermore, coalitions tend to consist of players 1 and 2, in particular in *Formal* (see Figure 5.8).

![Coalitions for $a<1$ and $a>1$](image)

*Figure 5.8*

This figure shows coalitions and learning effects. Stacked bars show the distribution of distinct coalitions. A coalition $ij$ is defined as an outcome proposed by $i$ and accepted by $j$ or vice versa. A coalition $ijk$ is an outcome with 2 yes votes (only possible in formal).

In the course of the experiment, inter-coalitional competition becomes more important. In the second half, more coalitions arise of players 2 and 3 than in the first half ($p=0.03^*$ (pooled), $p=0.46$ (*Informal*), $p=0.05^*$ (*Formal*)), resulting in a more even spread of positive and negative agreements. Furthermore, for $a>1$ more coalitions are formed in the second half than in the first half between players 1 and 3 ($p=0.03^*$ (pooled), $p=0.03^*$ (*Informal*), $p=0.21$ (*Formal*)).

Note that having the viable ‘outside option’ of a coalition with player 3 means that the median player can offer less to player 1. Median players appear to realize this remarkably well in the second half of the experiment. Agreements
tend to be closer to 0 in the last half (see Figure 5.7). In particular, the number of fair 1-2 compromises drops considerably \( p<0.01^* \) (pooled), \( p=0.04^* \) (Informal), \( p=0.05 \) (Formal) with an accompanying increase in the number of outcomes at the median ideal \( p=0.01^* \) (pooled), \( p=0.03^* \) (Informal), \( p=0.14 \) (Formal).

Figure 5.9

This figure shows the learning effects in frequency, location and distance. The left chart shows the percentage of outcomes that were real numbers. The middle chart shows the average location of real number agreements and the right chart the average distance between real number outcomes and the median preference.

Figure 5.9 splits the data depicted in Figure 5.4-Figure 5.6 and shows the frequency of real number agreements, their location and distance separately for the first and last half. From Figure 5.9 we learn firstly that the greater number of coalitions between players 1 and 3 results in a significantly lower frequency of real number agreements in the second half when \( a>1 \) \( p=0.04^* \) (pooled), \( p=0.03^* \) (Informal), \( p=0.25 \) (Formal). Secondly, the more equal spread of 1-2 and 2-3 coalitions also results in an average location closer to zero \( p=0.04^* \), (pooled), \( p=0.12 \) (Low), \( p=0.18 \) (High). Finally, we see that the strongest learning effect is that the distance of real number outcomes to the median ideal becomes significantly lower \( p<0.01^* \). Hence, it is the median player who benefits most from the increase in inter-coalitional competition.
5.6 Conclusion

In this chapter, we have addressed the question whether the outcome of the legislative process is affected by the formality of the bargaining procedure. We compared an informal bargaining procedure where players can freely make and accept proposals to a formal bargaining procedure where agenda-setting and voting is regulated by a Baron-Ferejohn alternating offers scheme. We studied the effect of formality in a legislative bargaining problem that consisted of a three-player median voter setting modified to have an external disagreement point. This allowed us to study formality both when the core exists and when it is empty, and to study whether an external disagreement point can explain why the polarization of a legislature can affect the legislative outcome. We derived cooperative solutions for the bargaining problem, studied the equilibrium properties of the formal and informal bargaining games, and tested the two procedures in the laboratory.

Our first result pertains to polarization. Theoretically, we find that polarization should matter when there is an external disagreement point and in our experiment, we find that this is indeed the case. In particular, polarization hurts the median player. As predicted by theory, in our experiments the median player is significantly worse off at moderate than at weak levels of polarization. In contrast to what theory predicts, however, more polarization hurts the median player even when her ideal is the unique core element. This seems to be the result of intra-coalitional fairness considerations. Over time, inter-coalitional competition appears to attenuate these fairness considerations and polarization hurts the median player less.

Our second and main result is that formality matters. Theoretically, it is difficult to analyze the effects of formality, as a key characteristic of informal bargaining is that it imposes very few strategic restrictions on the negotiators. We find that in the informal game, all plausible outcomes are supported as subgame perfect equilibrium points. This is an important motivation to run experiments. The data show a clear treatment effect of formality. The median

96 Recall that polarization does not matter in the classic median voter setting (Black, 1948; 1958).
player in our experiment is significantly better off under an informal bargaining procedure without rules about the timing of proposal and acceptance decisions. Outcomes in the *Informal* treatment are significantly more often the median ideal and significantly less often the fair compromise between players 1 and 2. It appears that the informal procedure gives the median player more flexibility to exploit its superior bargaining position. This result supports the armchair observation that players in a better bargaining position prefer less regulation of negotiations. To put this result on a stronger footing, more research is needed as we compare two representative but still specific procedures. Recently, this result has received support from Drouvelis, Montero & Sefton (2010).

Our results are relevant for the application of game theoretic models to the legislative process. The fact that formality influences the payoffs of certain players and the performance of specific predictions means that ‘neutral’ simplifying assumptions (i.e., assumptions that do not favor any player *prima facie*) made to obtain tractable results need not be as innocuous as is often assumed. For instance, a highly stylized alternating offers game may not be a suitable model of the legislative process if a significant part of the bargaining is informal.

Finally, understanding the influence of formality is relevant for studying institutional choice and parliamentary procedure. In particular, legislatures have to decide on a bargaining *procedure* –either from scratch or from a set of previously established procedures– before they can decide on the outcome itself. Even if the extent of formality may seem like a neutral parameter, it can significantly influence the bargaining outcome. Consequently, parties may have preferences for a formal or informal bargaining procedure. For instance, parties in the center of a political spectrum may prefer to prolong backroom discussions until agreement has been reached. Our results point to the more general idea that parties in a superior bargaining position will prefer less formal bargaining institutions, as these give them more room to exploit their bargaining position.
5.7 Appendix

5.7.1 Cooperative Solution Concepts

This appendix gives an overview of the dominance relations and the cooperative solutions concepts in our bargaining problem. The dominance relations (as shown in Table 5.4) depend mainly on the polarization parameter \( a \) and to some degree on \( b \).

| \( z' \succ z \) | \( |z'| < |z| \) for \( z, z' \in \mathbb{R} \) | Polarization     |
|-----------------|-----------------|-----------------|
| \( z \succ \delta \) | \( z \in (-1, 1) \) | \( a + b < 2 \) |
|                  | \( z \in (-1, -a + 1) \cup (b - 1, 1) \) | \( b < 2 \) |
|                  | \( z \in (-1, -a + 1) \) | \( a < 2 \) |
| \( \delta \succ z \) | \( z \in (-a + 1, b - 1) \) | \( a, b \leq 2, a + b > 2 \) |
|                  | \( z \in (-a + 1, 1) \) | \( a \leq 2, b > 2 \) |
|                  | \( z \in [-1, 1] \) | \( a > 2 \) |
|                  | \( z \in (-\infty, -a - 1) \cup (b + 1, \infty) \) | all \( a, b \) |

Notes: The table summarize dominance relations between alternatives in the cooperative game.

For example, the second row states that when two alternatives are real numbers, the one closest to the median ideal of 0 dominates the alternative further away. The third and fourth rows compare real numbers to the disagreement point. For example, row 3 shows for \( a<2 \) that any real number between \(-1\) and \(-a+1\) dominates \( \delta \) (because it gives players 1 and 2 strictly positive utility). On the other hand, if \( a>2 \), any real number between \(-1\) and 1 gives both wing players negative utility, so they both prefer the disagreement point, which then dominates the real number (row 4).

With these dominance relations, we can analyze the set of cooperative solution concepts for our bargaining problem. Aside from the core (\( \mathcal{C}(\Gamma) \)) and uncovered set (\( \mathcal{N}(\Gamma) \)), this includes two refinements of the uncovered set, the von
Neumann Morgenstern set \( \mathcal{L}(\Gamma) \)\(^{97}\) and the bargaining set \( \mathfrak{B}(\Gamma) \)\(^{98}\). Both are unique.

All four solutions are finite sets for all values of \( a \) and \( b \). The size of each depends on the polarization parameter \( a \). The four solutions coincide whenever the core is non-empty. This is the case in both extremes, i.e., for weakly and for strongly polarized preferences. In the remaining case of moderately polarized preferences (\( 1 < a < 2 \)), the other three solution sets are non-empty and satisfy the following inclusions.

\[
\emptyset = \mathcal{C}(\Gamma) \subset \mathfrak{B}(\Gamma) \subset \mathcal{L}(\Gamma) \subset \mathcal{U}(\Gamma)
\]

It turns out that all inclusions are strict. Under the general additional assumption that \( a < b \), the bargaining set \( \mathfrak{B}(\Gamma) \) consists of a single point, the von Neumann Morgenstern set \( \mathcal{L}(\Gamma) \) consists of two points, and the uncovered set \( \mathcal{U}(\Gamma) \) consists of three points. In the special case when \( 1 < a = b < 2 \) all three sets contain an additional solution point, due to symmetry considerations. The solution sets are listed in Table 5.5.

### TABLE 5.5

<table>
<thead>
<tr>
<th>Polarization</th>
<th>( \mathcal{C}(\Gamma) )</th>
<th>( \mathfrak{B}(\Gamma) )</th>
<th>( \mathcal{L}(\Gamma) )</th>
<th>( \mathcal{U}(\Gamma) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>weak: ( a \leq 1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>moderate ( 1 &lt; a &lt; 2 )</td>
<td>( \emptyset )</td>
<td>( {-a + 1} )</td>
<td>( {-a + 1, \delta} )</td>
<td>( {-a + 1, 0, \delta} )</td>
</tr>
<tr>
<td>( a = b )</td>
<td>( {-a + 1, b - 1} )</td>
<td>( {-a + 1, 0, \delta} )</td>
<td>( {-a + 1, 0, b - 1, \delta} )</td>
<td></td>
</tr>
<tr>
<td>strong: ( b \geq 2 )</td>
<td>( \delta )</td>
<td>( \delta )</td>
<td>( \delta )</td>
<td>( \delta )</td>
</tr>
</tbody>
</table>

**Notes:** For \( \Gamma = \Gamma(N, Z, u, W) \), the table gives the elements in the core (\( \mathcal{C}(\Gamma) \)), bargaining set \( \mathfrak{B}(\Gamma) \), vNM set \( \mathcal{L}(\Gamma) \) and uncovered set \( \mathcal{U}(\Gamma) \) for the levels of polarization distinguished in the first column.

---

\(^{97}\) Formally, a subset \( L \) of \( Z \) is a von Neumann Morgenstern set if elements of \( L \) do not dominate each other and every element of \( Z \setminus L \) is dominated by at least one element of \( L \).

\(^{98}\) The bargaining set is the set of efficient points \( z \) in \( Z \) such that for any \( z' \) which dominates \( z \) and player \( k \in N \) who prefers \( z \) over \( z' \) there exists \( z'' \) which weakly dominates \( z' \) and is for player \( k \) at least as good as \( z \). (We use Maschler’s (1992) definition of the bargaining set.)
The results for weak and strong polarization are straightforward. In the former case, the median ideal (0) dominates all other alternatives and the same holds for the disagreement point in the latter case. Here, we briefly explain the arguments underlying the results for moderate polarization, specifically for \( a \) and \( b \) such that \( 1 < a < 2 < b \).\(^9\) The results for parameters where equalities hold in this relationship are straightforward.

(c) All \( z \in \mathbb{R} \) unequal 0 are dominated by 0, 0 itself is dominated by \( \delta \), which in turn is dominated by any \( z \) between \(-1\) and \((-a + 1)\). Hence, the core is empty.

(b) The proposal \(-a + 1\) is dominated by any alternative \( z' \) closer to 0 than itself. Player 1 who prefers \(-a + 1\) above \( z' \) has a counter-objection \( z'' = \delta \) which dominates \( z' \) and which gives him the same utility of 0 as the original proposal \( z \). For any other proposal \( z \in Z \) an objection exists for which there is no counter-objection justifying the original proposal. Hence, \((-a+1)\) is the unique element of the bargaining set \( B(\Gamma) \).

(f) Player 1 is indifferent between \((-a + 1)\) and \( \delta \) while players 2 and 3 have opposite preferences for these alternatives, hence, they do not dominate each other. Points on \( \mathbb{R} \) between \((-a + 1)\) and \((b - 1)\) are dominated by \( \delta \), those beyond these limits are dominated by \((-a + 1)\) itself. Hence, \\{\(-a + 1, \delta\)\} is a vNM set \( L(\Gamma) \). One can easily verify that in \( \Gamma \) there is no other vNM set.

(h) (i) \( z = -a + 1 \) is only dominated by \( z' \in (-a +1, 1- a) \), which in turn are dominated by \( z'' = \delta \) which does not dominate \( z \),

(ii) \( z = 0 \) is dominated only by \( z' = \delta \), which in turn is dominated by, for example, \( z'' = -a / 2 \) which does not dominate \( z \),

\(^9\) The properties of the core and the uncovered set known from the literature on cooperative games are invariably obtained under the assumption of convexity of the values of all coalitions, which does not hold here. The bargaining sets and the von Neumann Morgenstern sets have been studied mostly in the context of TU games. Hence, all results in Table 5.5 must be verified case by case. We do not claim validity for any of these relations beyond the scope of bargaining problems as described here, with a one-dimensional set of alternatives augmented with a disagreement point and single-peaked preferences for the trio of players.
(iii) $z = \delta$ is dominated by any $z' \in (-1, -a +1)$, all of which are dominated by $z'' = 0$ which does not dominate $z$,

(iv) all $z$ with $|z| > -a +1$ are covered by $-a + 1$, while all $z \in (-a +1,0) \cup (0,a -1]$ are covered by 0.

Hence, the uncovered set $\mathcal{U}(\Gamma)$ consists of the triple \{-a + 1, 0, \delta\}.

### 5.7.2 Proof of Proposition 5.1 (Formal)

In this appendix, we provide the proof of Proposition 5.1, which characterizes the equilibrium outcome of $\Gamma_T^F$ (when it converges). Since $\Gamma_T^F$ is (highly) non-convex due to the exterior disagreement point, we cannot use standard results and techniques to derive equilibria; rather it involves a **tour de force** in backward induction. We also ran simulations, which illustrate (and corroborate) the results of the proposition. In particular, they shed some light on what happens if the outcome does not converge. At the end, we provide a figure that illustrates the cyclic dependence of the outcome on $a$ and $b$ for $1 < a < 2$ and $a < b < 3$ (as obtained by simulations).

**Proof**

Before we can determine the equilibrium, we need to introduce some notation. Due to backward induction and players having a unique best response at each information set, the equilibrium proposal and voting strategies only depend on how many rounds are ahead. Hence, we will count the rounds by the remaining number of rounds $r \equiv T - t + 1$. Hence, the first round has $r = T$ and the last round $r = 1$. Furthermore, this implies that the equilibrium strategy for round $r$ is the same for each game $\Gamma_T^F$ with $T \geq r$. Hence, it is meaningful to talk in general about the (sub)game $\Gamma_r^F$. The equilibrium (behavioral) strategy for player $i$, $\sigma_i$, specifies for each round $r \leq T$ (i) for the proposal stage, a
probability distribution over possible proposals \( \pi'_i : Z \to [0,1], p'_i \mapsto \pi'_i(p'_i) \),
and \((ii)\) for the voting stage, an acceptance function \( \nu'_i : Z \to \{0,1\}, p' \mapsto \nu'_i(p') \).

The equilibrium outcome of \( \Gamma'_i \) can be characterized by the probability distribution of the equilibrium outcomes \( \mu' : Z \to [0,1] \). The continuation value

\[
EU'_i = E_{\mu'}[u_i(z)] = \sum_{\text{supp} \mu'} \mu'(z)u_i(z)
\]

is the expected utility of player \( i \) of the (sub)game \( \Gamma'_i \). We can conveniently express \( EU'_i \) in terms of \( f' \equiv \mu'(\delta) \), \( E \equiv E_{\mu'}[z \mid z \in R] \) and \( D' \equiv E_{\mu'}[|z| \mid z \in R] \). (Note that \( D' \geq |E| \)). Define the indicator function \( I_R(x) \equiv [1 \text{ if } x \in R, 0 \text{ if } x \not\in R] \), the acceptance probability \( \xi'_i(x) \equiv 1-(1-\nu'_i(p))(1-\nu'_i(p)) \) and the probability of delay \( P'_i[\text{delay}] \equiv 1-\frac{1}{3} \sum \sum \pi'_i(x)\xi'_i(x) \). Then, we get:

\[
f' = \frac{1}{3} \sum \sum \left( \pi'_i(x)\xi'_i(x)I_R(x) \right) + P'_i[\text{delay}]f'^{-1}, \quad f'^0 \equiv 0
\]

If \( f' = 0 \), then \( E' \equiv 0, D' \equiv 0 \). Otherwise:

\[
L' \equiv \frac{1}{f'} \left( \frac{1}{3} \sum \sum \left( \pi'_i(x)\xi'_i(x)x \right) + P'_i[\text{delay}]f'^{-1}L'^{-1} \right)
\]

\[
D' \equiv \frac{1}{f'} \left( \frac{1}{3} \sum \sum \left( \pi'_i(x)\xi'_i(x)|x| \right) + P'_i[\text{delay}]f'^{-1}D'^{-1} \right)
\]

Since \( E_{\mu'}[u_i(z)] = Pr_{\mu'}[z = \delta] E_{\mu'}[u_i(z) \mid z = \delta] + Pr_{\mu'}[z \neq \delta] E_{\mu'}[u_i(z) \mid z \neq \delta] \) and \( u_i(z) \) are linear in \( z \) for \( z \in [-a,b] \cup \delta \), from \( u_i(z) = 1-|z-z_i| \) we get that (for \( \text{supp} \mu' \subseteq [-a,b] \cup \delta \)):

---

100 \( |\text{supp} \pi'_i| \leq 2 \)

101 \( \mu' \) is a probability mass function and has countable support:

\( |\text{supp} \mu'| \leq 3r \max_{x \in T} |\text{supp} \pi'_i| = 6r \).
\[ E_{U_i}^r = f^r(1 - a - L^r) \]
\[ E_{U_2}^r = f^r(1 - D^r) \]
\[ E_{U_3}^r = f^r(1 - b + L^r) \]

A player \( i \) will accept a proposal \( p \) in round \( r + 1 \) if and only if \( u_i(p) \geq E_{U_i}^r \).

This allows us to characterize for round \( r + 1 \):
1. (i) the largest proposal player 1 accepts, \( L_{i1}^{r+1} \),
2. (ii) the smallest proposal player 3 accepts, \( L_{31}^{r+1} \),
3. (iii) the largest absolute value a proposal can have for player 2 to accept it:

\[ L_{i1}^{r+1} = (1 - f^r)(1 - a) + f^r L^r \]
\[ L_{31}^{r+1} = (1 - f^r)(b - 1) + f^r L^r \]
\[ D_2^{r+1} = (1 - f^r) + f^r D^r \]

Players will only delay if they cannot make a proposal that will be accepted and gives them at least their continuation value. Player 1 or 2 will only delay in round \( r + 1 \) if \( L_{i1}^{r+1} \leq -D_2 \), which is equivalent to \((1 - f^r) (2 - a) + f^r (L^r + D^r) < 0\). This can only hold if \( a > 2 \) and \( f^r < 1 \). Hence, players 1 and 2 will never delay if \( a \leq 2 \) and, by the same reasoning, player 3 will never delay if \( b \leq 2 \).

Note that if \( p_i^{r+1} \in R, i \neq 2 \) is accepted in equilibrium, it must be accepted by player 2 in round \( r + 1 \) and, hence, \( u_2(p_i^{r+1}) \geq E_{U_2}^r \). If \( a < 2 \), then player 2 will propose \( 1 - a \) (or \( a - 1 \)) in round \( T \). This means that \( E_{U_2}^r > 0 \) for all \( r \) and that she will never accept nor propose \( \delta \). Furthermore, if \( a < 1 \), player 1 can always propose 0 so that she will never propose \( \delta \). Finally, \( E_{U_2}^r > 0 \) implies \( L^r < 1 \), and hence \( E_{U_3}^r < 0 \) if \( b \geq 2 \). In the following Lemma, we summarize these facts and some conditions that are easily derived.

**Lemma 5.1** For \( a < 2 \), the equilibrium \( \{\sigma_1, \sigma_2, \sigma_3\} \) is determined by:

1. \( \nu_i^r(p) = 1 \) for \( p \in [-a,b] \) iff \( p \leq L_{i1}^r \) and \( \nu_i^r(\delta) = 1 \) iff \( E_{U_i}^{r-1} \leq 0 \).
2. \( \nu_2^r(p) = 1 \) for \( p \in [-a,b] \) iff \( |p| \leq D_2^r \) and \( \nu_2^r(\delta) = 1 \) iff \( E_{U_2}^r \leq 0 \).
3. \( \nu_3^r(p) = 1 \) for \( p \in [-a,b] \) iff \( p \geq L_{31}^r(\sigma) \) and \( \nu_3^r(\delta) = 1 \) iff \( E_{U_3}^{r-1} \leq 0 \).
4. \( \pi_i'(\delta) = 1 \) iff \(-D > 1 - a\) and \(EU^{-1}_i \leq 0\); \( \pi_i'(-a) = 1 \) iff \(-D < -a\);
\[ \pi_i'(-D) = 1 \] iff \(-a \leq -D < 1 - a\) or \(-D \geq -a\) and \(EU^{-1}_i > 0\);
5. \( \pi_i'\left(0 \right) = 1 \) iff \(\overline{L}_1 > 0\) or \(L'_1 < 0\); \( \pi_i'\left(L'_1\right) = 1 \) iff \(\overline{L}_1 \leq 0\), \(L'_1 > 0\) and \(|\overline{L}_1| < L_1\);
\[ \pi_i'\left(L'_1\right) = 1 \] iff \(\overline{L}_1 < 0\), \(L'_1 \geq 0\) and \(\overline{L}_1 > L_1\);
\[ \pi_i'\left(L'_1\right) = \pi_i'\left(L''_1\right) = \frac{1}{2} \] iff \(\overline{L}_1 \leq 0\), \(L'_1 \geq 0\) and \(|\overline{L}_1| = L_1\);
6. \( \pi_i'(\text{delay}) = 1 \) iff \(L'_1 > \overline{D}_2\) and \(EU'_i > 0\) \(\text{(only if } f^{-1} < 1\text{);}
\[ \pi_i'\left(\delta\right) = 1 \] iff \(\overline{D}_2 \leq b - 1\) and \(EU^{-1}_i \leq 0\); \( \pi_i'(\delta) = 1 \) iff \(\overline{D}_2 > b - 1\);
\[ \pi_i'\left(\overline{D}_2\right) = 1 \] iff \(b - 1 < \overline{D}_2 \leq b\) or \(EU'_i > 0\) and \(\overline{D}_2 \leq b\).

Now we are ready to look at whether the equilibrium outcome converges. Let \(x^* \equiv \lim_{r \to \infty} x^r\). The probability distribution \(\mu^*\) is the limit of \(\mu^r\) if it holds that \(\lim_{r \to \infty} \mu^r(z) = \mu^*(z)\) for all \(z\) in the support of \(\mu^*\). As defined in section 5.3.1, we say that the equilibrium outcome \(\mu^r\) converges to \(\mu^*\) if \(\lim_{r \to \infty} \mu^r = \mu^*\); if this limit does not exist, we say that \(\mu^r\) does not converge.

The equilibrium outcome converges to 0 if \(\mu^*(0) = 1\), which is equivalent to \(f^* = 1\) and \(\overline{D}^* = 1\). The equilibrium outcome converges to \(\delta\) if \(\mu^*(\delta) = 1\), which is equivalent to \(f^* = 0\) and \(\limsup_{r \to \infty} D^r \in \mathbb{R}\). Finally, it is straightforward that \(\mu^r\) does not converge if (i) \(f^*\) does not exist or (ii) \(f^* > 0\) and \(\overline{L}\) or \(\overline{D}\) do not exist.

**Proposition 5.1**

(i) If \(0 \leq a < 1\) or \(a = b = 1\), then the equilibrium outcome converges to 0

(ii) If \(a \leq b < 2\) and \(b > 1\), the equilibrium outcome does not converge, unless \(\frac{10}{7} \leq a = b < 2\) or \(b = \frac{3}{2}, \frac{5}{4} \leq a < \frac{3}{2}\) or \(\frac{8}{3} \leq b < \frac{32}{11}, a = \frac{5}{7}b - \frac{4}{7}\) or
\[ \frac{2}{3} \leq b < \frac{2}{5}, \max\{\frac{2}{3} - \frac{1}{5}b, \frac{2}{5}b - \frac{2}{3}\} < a < \frac{2}{5} + \frac{1}{5}b. \] In these latter cases the outcome may converge, but never to a single outcome in \( Z \).

(iii) If \( a \geq 2 \), the equilibrium outcome is \( \delta \).

Proof:

(i.a) We show that if \( 0 \leq a < 1 \), then \( f^* = 1 \) and \( D^* = 0 \).

Throughout the proof, we will use the following sufficient condition for convergence: For \( 0 \leq a < 1 \), \( f^* = 1 \) and \( D^* = 0 \) if

\[ \text{(SC)} \quad \text{there exists a round } r^1 \text{ such that } f^{r^1} = 1 \text{ and } EU_1^{r^1} > 0 \]

Let (SC) hold for \( r^1 \). \( EU_1^{r^1} > 0 \) (\( \iff L^{r^1} < 1 - a \)) implies that player 1 will not accept nor propose \( \delta \) in round \( r^1 + 1 \). \( f^{r^1} = 1 \) implies that player 3 will not delay and that either player 1 or 3 accept 0 in round \( r^1 + 1 \). Consequently, \( p_2^{r^1+1} = 0 \), \( p_1^{r^1+1} = -D^{r^1+1} = -D^* \) and \( p_3^{r^1+1} = \bar{D}^{r^1+1} = D^* \). Hence, \( f^{r^1+1} = 1 \), \( L^{r^1+1} = \frac{2}{3}(-D^* + 0 + D^*) = 0 < 1 - a \) and \( D^{r^1+1} = \frac{2}{3}(D^* + 0 + D^*) = \frac{2}{3}D^* \). Thus, (SC) holds for \( r^1 + 1 \) and by induction for all \( r \geq r^1 \). As a result, \( f^* = 1 \) and \( D^* = \lim_{m \to \infty} (\frac{2}{3})^m D^* = 0 \). A sufficient condition for (SC) to hold is that:

\[ \text{(SC')} \quad \text{there is a round } r' \text{ such that} \]

\[ (i) \bar{L}_i^r \leq D_i^r \leq a \text{ and } L_i^r < 1 - a \text{ and } (ii) L_i^r \geq -\bar{L}_i^r \text{ or } L_i^r < 3(1 - a) \]

\[ L_3^r \leq D_2^r \leq a \text{ and } L_3^r < 1 - a \] imply that \( p_1^r = -D_1^r \) and \( p_3^r = \bar{D}_2^r \). Hence, \( f^r = 1 \) and \( L^r = \frac{1}{3}E[p_2^r] \). If \( L_3^r \geq -\bar{L}_3^r \) or \( L_3^r < 3(1 - a) \), then \( E[p_3^r] < 3(1 - a) \) and \( L^r < (1 - a) \). Hence (SC) holds for \( r' \). In the remainder of the proof we divide the \((a, b)\) parameter-set into regions and show that (SC) holds for each region.
We start by looking at the last four rounds. In the final round, \( p_1^i = -a \) and \( p_2^i = 0 \). If \( b = 2 \), then \( p_3^i = \min\{b, 1\} \), \( f^1 = 1 \) and \( L^1 < \frac{1}{3}(1-a) \), such that round 1 satisfies (SC). So, let \( b \geq 2 \). Then \( p_3^i = \delta \) and \( f^1 = \frac{2}{3} \), \( D^i = -L^i = \frac{1}{2} a \). Hence, \( \bar{L}_1^2 = \frac{1}{3}(1-2a) < 1-a, \bar{D}_2^2 = \frac{1}{3}(1+a) \) and \( \underline{L}_3^2 = \frac{1}{3}(1-a-b) \). Since \( EU_1^i > 0 \), \( p_3^2 = \text{delay} \) iff \( \bar{D}_2^3 < \underline{L}_3^2 \) iff \( b > 2 \). Let us first consider \( b \leq 2 \). Then \( p_3^1 = -a \), \( p_3^2 = 0 \) and \( p_3^3 = \bar{D}_2^3 \), so that \( f^2 = 1 \) and \( L^2 = \frac{1}{3}(1-2a) < 1-a \). Hence, round 2 meets (SC).

Let us now consider \( b \leq 2 \) and \( a > \frac{1}{2} \). In this case, \( \underline{L}_3^2 \leq \bar{D}_2^2 \leq a \), so that round 2 satisfies (SC’) if \( \bar{D}_2^3 < \underline{L}_3^2 \) or \( \underline{L}_3^2 < 3(1-a) \). So let \( 3(1-a) \leq \underline{L}_3^2 < -\bar{L}_3^2 \). This means that \( 10 - 8a \leq b < 3a \) and \( a > \frac{10}{4} \). Furthermore, \( p_3^1 = -\bar{D}_2^3 \), \( p_3^2 = \underline{L}_3^2 \) and \( p_3^3 = \bar{D}_2^3 \), so that \( f^2 = 1, L^2 = \frac{1}{3} p_3^3 > 1-a \) and \( D^2 = \frac{1}{3} (1 + a + b) \leq a \). Hence,\( p_3^1 = -\bar{D}_2^3 = -\bar{D}_2^3 \), \( p_3^2 = 0 \) and \( p_3^3 = \delta \), so that \( f^3 = \frac{2}{3}, D^3 = -L^3 = \frac{1}{3} D^3 \). From this, \( \bar{L}_4^3 = \frac{1}{27} (8 - 10a - b), \bar{D}_4^3 = \frac{1}{27} (10 + a + b) \) and \( \underline{L}_4^3 = \frac{1}{27} (8b - 10 - a) \). \( 2 < b \leq 2 \) and \( a > \frac{1}{2} \) imply \( \bar{D}_2^4 \leq a \) and \( \bar{L}_4^3 < (1-a) \). \( 10 - 8a \leq b < 3a \) implies \( -\bar{L}_4^3 < \underline{L}_4^3 < \bar{D}_4^3 \). Hence, round 4 satisfies (SC). Thus (SC) holds if

\[
(A) \quad b \leq 2(a + 1)
\]

Let \( b > 2 \) from now on. \( EU_3^r < 0 \) for all \( r \) and player 3 can now delay consecutive rounds and alternatingly delay and not delay. This requires a careful characterization of the dynamic before we proceed. We will call a set of consecutive rounds in which player 3 delays a delaying sequence. We index these sequences by \( s \) (again backwards), with \( s = 1 \) the final delaying sequence, \( s = 2 \) the pre-final delaying sequence etc. Let \( R(s) \) be the set of rounds in the \( s \)-th delaying sequence and define \( \bar{r}(s) = \max R(s) \) and \( r(s) = \min R(s) \). Finally, let \( m(s) = \bar{r}(s) - r(s) + 1 \) be the number of delaying rounds in \( R(s) \).
Let us look at $r(s)-1$. $p_3^{(s)} = \text{delay}$ implies $EU_1^{(s)-1} > 0$ and $f_3^{(s)-1} = \frac{2}{3}$. As player 1 accepts $\delta$ in $r(s)-1$, she also accepts 0 and $p_2^{(s)-1} = 0$. Since $p_1^{(s)-1} = -\overline{D}_2^{(s)-1}$, this means that $D^{(s)-1} = -L^{(s)-1} = -\frac{1}{2}\overline{D}_2^{(s)-1}$. We proceed to rounds $r \in R(s)$. Since player 3 delays, $L_3^{(s)} > \overline{D}_2^{(s)}$ and $p_3^{(s)} = \min\{0, \overline{D}_1^{(s)}\} \leq 0$. As $D^{(s)} = -L^{(s)}$ and $f^{(s)-1} = \frac{2}{3}$, by (5.1) and Lemma 5.1 it must be that $f^1 < 1$, $L = -D < 0$ and $EU_1^r > 0$ for all $r \in R(s)$. Furthermore:

If $D^r = -L^r$, then:

$$\alpha^r_{\text{r}+1} = \overline{D}_2^{r+1} + \overline{L}_1^{r+1} = (1-f^r)(2-a)$$

$$\alpha^r_{\text{r}-1} = \overline{D}_2^{r+1} - \overline{L}_1^{r+1} = (1-f^r)a + 2f^rD^r$$

$$\gamma^r_{\text{r}+1} = \overline{D}_2^{r+1} + \overline{L}_3^{r+1} = (1-f^r)b$$

$$\gamma^r_{\text{r}-1} = \overline{L}_3^{r+1} - \overline{D}_2^{r+1} = (1-f^r)(b-2) - 2f^rD^r$$

In particular, $D^r = -L^r$ holds for $r = r(s)-1, \ldots, r(s)$.

Moreover, as player 3 delays in rounds $r \in R(s)$ and $f^{(s)-1} = \frac{2}{3}$:

$$f^r = \frac{2}{3} + \frac{1}{3} f^{r-1} \forall r \in R(s)$$

$$f^{z(s)+m} = \frac{3^{m+2} - 1}{3^{m+2}} \text{ for } m = -1, 0, 1, \ldots, m(s)-1$$

From (5.4) we get $\alpha^r_{\text{z}(s)} = \frac{1}{3}a + \frac{2}{3}D^{(s)-1}$ and it turns out that $\alpha^r_{\text{r}} = \alpha^r_{\text{z}(s)} \equiv \alpha^r_{\text{z}(s)}$ for $r(s) \leq r \leq \overline{r}(s)+1$. For $s = 1$ it is simple. Suppose $\alpha^r_{\text{r}} = a$ and $D^r = -E^r = \frac{1}{2}a$. Then immediately $\alpha_{\text{r}+1} = a$. Furthermore, due to the symmetry $D^{r+1} = -L^{r+1} = \frac{1}{2}a$. Since $D^{(1)-1} = D^1 = -L^1 = \frac{1}{2}a$ and $\alpha^{(1)-1} = a^1 = a$, by induction it follows that $\alpha^r_{\text{r}} = a$ for $r(1) \leq r \leq \overline{r}(1)+1$. For $s > 1$, we need to assume that $L_1^{(s)} \leq 0$ and $\overline{D}_2^{(s)} \leq a$ and justify it later. Suppose $\overline{L}_1^{(s)} \leq 0$, $\overline{D}_2^{(s)} \leq a$ and $p_3^{(s)} = \text{delay}$. Hence, $p_1^r + p_2^r = -\alpha^r_{\text{r}}$ and, using (5.1), we get $D^r =$
\[ \frac{\frac{1}{2} - \frac{3}{2} (p_r^r - p_j^r) + \frac{1}{2} f^{-1} D_r^{r-1}}{f^r} = \frac{(1 - f^{-1})a + 3f^{-1} D_r^{r-1}}{3f^r}. \]

Substituting this term and using (5.5), we get that
\[ \alpha_r^r - \alpha_{r+1}^r = a (f^r - f^{-1}) + 2 f^{-1} D_r^{r-1} - 2 f^r D^r = 0. \]

Hence, \( \alpha_r^{r+1} = \alpha_r^r \). Furthermore, using the same substitutions, we get
\[ D_r^{r+1} - D_r^r = -\frac{1}{3} (2-a)(1-f^{-1}) < 0 \]
and \( D_r^{r+1} < D_r^r \leq a \). Finally, \( \sigma_1^r = D_r^{r+1} - \sigma_{r+1}^r = D_r^{r+1} - \sigma_r^r < D_r^r - \sigma_r^r = \bar{L}_r^s \leq 0 \). Hence, as \( p_s^r = \text{delay} \) for all \( r \in R(s) \), \( \sigma_r^r = \sigma(r) \) for \( \sigma(r) \leq r \leq \sigma(s) + 1 \).

Using (5.4) and (5.6), we get the following results for \( m = 0, 1, ..., m(s) \):

\[ \begin{align*}
\bar{L}_1^m(s) &= \frac{1}{2} (\alpha_1^m(s) - \alpha_1^{m+1}) = \frac{1}{3} a + \frac{1}{3} a - \frac{1}{2} \alpha(s) \\
\bar{D}_2^m(s) &= \frac{1}{2} (\alpha_2^m(s) + \alpha_2^{m+1}) = \frac{1}{3} a + \frac{1}{3} a + \frac{1}{2} \alpha(s) \\
\bar{L}_3^m(s) &= \gamma_2^m(s) + \frac{1}{3} a + b - \frac{1}{3} \alpha(s) \\
\gamma_3^m(s) &= \gamma_3^m(s) - 2 \bar{D}_2^m(s) = \frac{a + b - 2}{3} - \alpha(s)
\end{align*} \]

Since player 3 only delays in round \( r \) iff \( EU_r^{r-1} > 0 \) and \( \gamma_r^r > 0 \), from (5.7) we get that \( m(s) = \min\{m \in R : \gamma_3^{m(s)+m} \leq 0\} = \text{ceiling} \left[ \frac{\ln(a + b - 2) - \ln(3 \alpha(s))}{\ln(3)} \right] \).

Equivalently, since \( \gamma_3^{m(s)+m(s)-1} > 0 \) and \( \gamma_3^{m(s)+m(s)} \leq 0 \), we get:

\[ 2 - a + 3^{m(s)} \alpha_1(s) < b \leq 2 - a + 3^{m(s)+1} \alpha_1(s) \forall s \]

In round \( \sigma(s) + 1 = \sigma(s) + m(s) \) player 3 will not delay. Since \( EU_1^s > 0 \) and
\[ L_3^{\bar{s}+1} \leq D_2^{\sigma(s)+1} < D_2^{\bar{s}} \leq a, \]

\[ \bar{s} + 1 \text{ satisfies (SC') if } L_3^{\bar{s}+1} < 3(1 - a) \]

\[ L_3^{\bar{s}+1} \leq D_2^{\sigma(s)+1} \leq D_2^{\bar{s}} = \frac{1}{3} (1 + 4a) \text{. Hence, by (5.9) } \bar{s} + 1 \text{ satisfies (SC') if } \]
Let $a > \frac{26}{31}$. It turns out that if $\pi(s) + 1$ does not satisfy (SC), then $\pi(s) + 3 = r(s + 1)$. Let $\pi(s) + 1$ not satisfy (SC). In this case $EU_1^{\pi(s)+1} \leq 0$ and $p_2^{\pi(s)+1} = L_{l_1}^{\pi(s)+1}$. Consequently, $p_2^{\pi(s)+2} = 0$, $p_3^{\pi(s)+2} = \delta$ and $p_1^{\pi(s)+2} = -D_2^{\pi(s)+2} = -D^\pi(s)+1 = -\frac{1}{3} (2D_2^{\pi(s)+1} + L_1^{\pi(s)+1}) = -\frac{1}{6} \left( \frac{2-a+2b}{3^{m(s)+1}} + \alpha(s) \right)$ (using (5.7)). Thus, $f^{\pi(s)+2} = \frac{2}{3}$, $EU_1^{\pi(s)+2} > 0$ and $D^\pi(s)+2 = -E^\pi(s)+2 = -\frac{1}{2} p_1^{\pi(s)+2}$. Using (5.4), we get $\gamma^-_{\pi(s)+3} = \frac{1}{9} \left( 3b - 6 - \alpha(s) - \frac{2-a+2b}{3^{m(s)+1}} \right)$. Furthermore, since $D^{\pi(s)+1} = -\frac{1}{3} p_2^{\pi(s)+1} \leq \frac{1}{2} a$, $\alpha(s) = \frac{1}{3} a + \frac{1}{3} D^{\pi(s)+1} \leq a$ and, by (5.8),

$$\frac{1}{3^{m(s)+1}} \leq \frac{\alpha(s)}{a + b - 2}. \text{ Hence, } \gamma^-_{\pi(s)+3} \geq \frac{1}{9} \left( 3b - 6 - a \left( 1 + \frac{2-a+2b}{a+b-2} \right) \right).$$

Since $b > 2 (a + 1)$ and $a \geq \frac{26}{31}$, $\gamma^-_{\pi(s)+3} > \frac{2}{9} (2a-1) > 0$. Thus, $p_3^{\pi(s)+3} = \text{delay and} \pi(s) + 3 = r(s + 1)$. As a consequence, $\alpha(s + 1) = \frac{1}{3} a + \frac{1}{3} D^{\pi(s)+1} = \frac{1}{3} a + \frac{1}{3} D^{\pi(s)+2}$:

$$\alpha(s + 1) = \frac{1}{3} a + \frac{1}{9} \alpha(s) + \frac{2-a+2b}{3^{m(s)+3}}$$

We conclude our characterization of the delaying sequences by showing we can indeed assume $L_{t_1}^{\pi(s)} \leq 0$ and $D_{t_1}^{\pi(s)} \leq a$ for $s > 1$. Since $\alpha(s) \leq a$ and $a \geq \frac{26}{31}$, $D_{t_1}^{\pi(s)} = \frac{2}{3} (1-a) + \frac{1}{3} \alpha(s) < a$. Showing $L_{t_1}^{\pi(s)} \leq 0$ requires some work. Let $L_{t_1}^{\pi(s)} \leq 0$ or $s = 1$. Using (5.3), $L_{t_1}^{\pi(s)+1} = L_{t_1}^{\pi(s)+3} = -\frac{1}{18} \left( \frac{2-a+2b}{3^{m(s)+1}} + \alpha(s) + 6(a-1) \right)$. Since $\pi(s)+1$ does not satisfy (SC'), $L_{t_1}^{\pi(s)+1} \geq 3(1-a)$ and, using (5.7), this implies $\frac{1}{3^{m(s)+1}} \geq \frac{6-6a+\alpha(s)}{-2+a+2b}$. Furthermore,
(5.10) implies that \( \alpha_-(s) \geq \frac{1}{3}a \) for \( s > 1 \). Hence, \( L_i^{(s+1)} \leq \frac{-2(18 + 9a^2 + a(b - 27))}{27(2b + a - 2)} < 0 \) (as \( b > 2(a + 1) \)). Since in particular \( L_i^{(2)} \leq 0 \), by induction it follows that \( L_i^{(s)} \leq 0 \) for all \( s > 1 \).

We proceed by dividing the parameter-plane not covered by (A) and (B) according to \( m(1) \geq 1 \), the number of rounds player 3 delays in the first delaying cycle, and proof that (SC) holds for some \( \bar{v}(s)+1 \). By \( \alpha_-(1) = a \) and (5.8)

\[
(5.11) \quad 2 - a + 3^{m(1)}a < b \leq 2 - a + 3^{m(1)+1}a
\]

By (5.9), \( \bar{v}(1)+1 \) satisfies (SC') if \( L_3^{(r+1)} = \frac{1}{3}a + b - \frac{1}{2}a < 3(1 - a) \) if

\[
(C) \quad b < 1 + 3^{m(1)+2} - \frac{1}{2}(1 + 5 \cdot 3^{m(1)+1})a
\]

Now, let \( \bar{v}(1)+1 \) not satisfy (SC) and (A) - (C) not hold. Using \( \alpha_-(1) = a \) and (5.10), we get \( \alpha_-(2) = \frac{2 - a + 2b}{3^{m(1)+3}} + \frac{4}{9}a \). Hence, by (5.7), \( \gamma_-^{(2)+m(1)} = \frac{10a + 7b - 20}{3^{m(1)+3}} - \frac{4}{9}a \). This is positive iff:

\[
(D) \quad b > \frac{4}{9}(10 - 5a + 2 \cdot 3^{m(1)+1}a)
\]

This means that if (D) holds \( m(2) > m(1) \) and \( L_3^{(r+1)} \leq L_3^{(2)+m(1)+1} \). Using (5.7) and the upper bound for \( b \) in (5.11), we get \( L_3^{(2)+m(1)+1} = \frac{1}{3} \left( \frac{a + b - 2}{3^{m(1)+1}} - a \right) \leq 0 < 3(1 - a) \). Hence, if (D) is met \( \bar{v}(2)+1 \) satisfies (SC').

Finally, let \( \bar{v}(2)+1 \) not satisfy (SC) and (A) - (D) not hold. As long as \( m(s) = m(1) \), from (5.10) we get

\[
(5.12) \quad \alpha_-(s+1) = \frac{1}{3}a + \frac{1}{9}\alpha_-(s) + \frac{2 - a + 2b}{3^{m(1)+3}}
\]

180
The unique steady state of this difference equation is

\[ \alpha_- = \frac{1}{8} \left( 3a + \frac{2 - a + 2b}{3^{m(1)+1}} \right), \]

which is a global attractor with a monotonic dynamic since \( 0 < \frac{d\alpha_-(s+1)}{d\alpha_-(s)} < 1 \). Using the upper bound for \( b \) in (5.11), \( m(1) \geq 1 \) and \( a \geq \frac{26}{31} \), we get that \( \alpha_- \leq \frac{1}{15}(1 + 7a) \leq a = \alpha_-(0) \). Hence, \( \alpha_-(s) \) decreases monotonically to \( \alpha_- \). Suppose \( m(s) = m(1) \) for all delaying sequences. Using the right-hand side of (C) as lower bound for \( b \) and \( 0 \leq a < 1 \), we get

\[ \frac{b}{3^{m(1)+1}} - \alpha_- \geq \frac{1}{4} \left( 9(1-a) + \frac{2-a}{3^{m(1)+1}} \right) > 0. \]

Hence, if \( m(s) = m(1) \) for all \( s \), there exists an \( \hat{s} \) such that \( L_\delta^{(\hat{s})+1} + L_1^{(\hat{s})+1} = \frac{b}{3^{m(1)+1}} - \alpha_-(\hat{s}) > 0 \) and \( \hat{s} \) meets (SC’).

\( m(s) \) is increasing in \( s \), because \( \alpha_-(s) \) is decreasing in \( s \) and, by (5.8), \( m(s) \) is decreasing in \( \alpha_-(s) \). This means that if \( m(s) \) is not equal to \( m(1) \) for all \( s \), there exists an \( s' \) such that \( m(s') > m(s'-1) = m(1) \). Furthermore, \( \alpha_-(s) > \frac{1}{3}a + \frac{2 - a + 2b}{3^{m(1)+3}} \) (for \( s > 1 \)). Using (5.7), this implies that

\[ L_\delta^{(s')+1} \leq \left( \frac{2(a+b) - 4}{3^{m(1)+3}} - \frac{1}{6}a \right). \]

As a consequence, \( L_\delta^{(s')+1} \leq 0 \) and \( \tau(s'+1) \) satisfies (SC) if

\[ b \leq \frac{1}{4}(8 - 4a + 3^{m(1)+2}a) \]

Hence, if (E) holds, either \( m(s) = m(1) \) for all \( s \) or not, both of which imply that (SC) holds for some \( \tau(s)+1 \). Furthermore, the right-hand side of (E) minus the right-hand side of (D) is \( \frac{1}{38} \left( 4 + 5 \cdot 3^{m(1)}a - 8 \right) \) and this is positive if \( a \geq \frac{26}{31} \) and \( m(1) \geq 1 \). Hence, since (A) – (D) do not hold, (E) must hold.

In conclusion, for each \( (a,b) \in [0,1) \times [a,\infty) \) there exists some round \( r \in \mathbb{N} \) that satisfies (SC) and, hence, the outcome converges to 0 as \( r \) increases.
(i.b) We show that if \( a = b = 1 \), then \( \lim_{r \to \infty} f^r = 1 \) and \( \lim_{r \to \infty} D^r = 0 \).

In round 1, \( p_1^1 = -1 \), \( p_2^1 = 0 \) and \( p_3^1 = 1 \). In round \( r > 1 \), \( p_1^r = -\overline{D}_2 \), \( p_2^r = 0 \) and \( p_3^r = \overline{D}_2 \), with \( f^r = 1 \) and \( D^r = 2 \overline{D}_2 = 2 D^{r-1} \). Consequently, \( \lim_{r \to \infty} f^r = 1 \) and \( \lim_{r \to \infty} D^r = 0 \).

(ii) We show that if \( 1 \leq a < 2 \) and \( b > 1 \), then \( f^* \neq 0 \) and \( f^*, \overline{L} \) or \( D^r \) does not exist, except if \( \frac{11}{7} \leq a = b < 2 \) or \( b = \frac{3}{2}, \frac{5}{4} \leq a < \frac{3}{2} \) or \( \frac{5}{3} \leq b < \frac{11}{17}, a = \frac{5}{7} b - \frac{4}{7} \) or \( \frac{4}{7} \leq b < \frac{7}{4}, \max\{\frac{7}{6} - \frac{1}{10} b, \frac{7}{5} b - \frac{8}{5}\} < a < \frac{6}{5} + \frac{1}{5} b \). In these latter case, the outcome may converge but never to a single outcome in \( Z \).

Let \( 1 \leq a < 2 \) and \( b > 1 \). First, we show that if the outcome converges there exists an \( r^j \) such that \( p_3^i = \text{delay} \) for all \( r > r^j \). Since \( EU_2^r > 0 \) and \( L_1^r > -\overline{D}_2 \), \( \overline{L}_1^r > -1 \) and \( p_3^i \in \mathbb{R} \) for all \( r \geq r^j \). Suppose there exists an \( r' \) such that \( p_3^i = \text{delay} \) for all \( r \geq r' \). This implies that \( EU_1^r > 0 \), \( p_2^r = \overline{L}_1^r < 0 \) and \( p_1^r = -\overline{D}_2 > -a \) for all \( r \geq r' \) and hence by (5.1) that \( f^* = 1 \). Furthermore, using (5.3), (5.4), and the logic behind (5.7), we get that \( \alpha^r = \alpha'^r > 0 \) and, hence, \( D^r = \frac{2}{5} (-p_1^r - p_2^r) + \frac{1}{5} D^{r-1} \geq \frac{1}{5} (-p_1^r - p_2^r) = \frac{1}{5} \alpha'^r \) for all \( r \geq r' \). Now, \( p_3^r = \text{delay} \) only if \( \overline{D}_2 < \overline{L}_3^r \), which implies by (5.4) that \( (1 - f^r)(b - 2) - 2 f^r \)
\[ D^r > 0 \] for all \( r \geq r' \). However, this is not possible, since \( f^* = 1 \) and \( D^r \geq \frac{1}{6} \alpha'^r > 0 \forall r \geq r' \). Hence, there does not exist an \( r' \) such that \( p_3^r = \text{delay} \) for all \( r \geq r' \). Convergence (of \( \overline{L}_3^r - \overline{D}_3 \)) implies the opposite holds:

(5.13) there exists an \( r^j \) such that \( p_3^i = \text{delay} \) for all \( r \geq r^j \)

Second, there can be no convergence to \( \delta \) as \( p_2^r \in \mathbb{R} \) for all \( r \).
Third, there can be no convergence to 0. \( \lim_{r \to \infty} f^r = 1 \) and (5.13) would imply that there exists an \( r' \) such that \( f^r = 1 \) for \( r \geq r' \). This means that \( p_1^r = -\overline{D}_2 \), \( p_2^r = 0 \) and \( p_3^r = \overline{D}_2 \) for all \( r \geq r' + 1 \). Consequently, \( L^r = 0 \) and, hence, \( EU_1^r \leq 0 \) and \( EU_3^r < 0 \) for \( r \geq r' + 1 \). However, if this is the case \( p_3^{r+1} = \delta \), contradicting \( f^r = 1 \) for \( r \geq r' + 1 \).

Finally, we show there can be no convergence to anything else then 0 or \( \delta \), save for four exceptions. Suppose that \( f^*, \overline{L}^* \) and \( \overline{D}^* \) exist, but \( 0 < f^* \leq 1 \) and \( D^* > 0 \). Now, \( f^* = \frac{1}{3}, \frac{2}{3} \) or 1. We have seen above that \( f^* = 1 \) is not possible. Let \( f^* = \frac{1}{3} \). This means that there exists a round \( r' \) such that \( f^r = \frac{1}{3}, \overline{D}_2 \leq a - 1 \) and, thus, \( D^r \leq 3a - 5 \) for \( r \geq r' \). Suppose \( p_2^r = \min\{0, \overline{L}_1^r\} \) for \( r \geq r' \). This implies that \( L^{r+1} = -D^{r+1} \) and \( L^{r+1} < 0 \), such that \( \overline{L}_i < 0 \). Furthermore, \( D^r = -\overline{L}_i = \frac{2}{3}(a-1) + \frac{1}{3}D^{r-1} \). Convergence implies that \( \lim_{r \to \infty} D^r - D^{r-1} = 0 \) and solving for \( D^r = D^{r-1} \) yields \( \overline{D}^* = \overline{L}^* = a - 1 \). However, since \( a < 2 \), \( D^* = a - 1 > 3a - 5 \) and \( D^r \leq 3a - 5 \) cannot hold for all \( r \geq r' \). Hence, \( p_2^r = \min\{\overline{L}_i, 0\} \) and a similar reasoning (with \( a - 1 \leq b - 1 < 3a - 5 \)) shows that \( p_2^r = \max\{0, \overline{L}_1^r\} \). Thus, let \( \pi_2^r(\overline{L}_1) = \pi_2^r(\overline{L}_1^r) = \frac{1}{2} \) for \( r \geq r' \). This means that \( L^r = 0, -\overline{L}_i = \overline{L}_1^r \) and, hence, \( a = b \) for \( r \geq r' \). Furthermore, \( D^r = \frac{1}{2}(\overline{L}_3^r - \overline{L}_1^r) = \frac{2}{3}(a-1) \). Hence, \( \overline{L} = 0 \) and \( \overline{D} = \frac{2}{3}(a-1) \). \( D^r \leq 3a - 5 \) requires

\[
\frac{12}{7} \leq a = b < 2.
\]

Let, ultimately, \( f^* = \frac{2}{3} \). This means that there exists a round \( r' \) such that \( f^r = \frac{2}{3} \), \( 1 - a \leq \overline{L}_i \), \( a - 1 < \overline{D}_2 \leq b - 1 \), \( p_1^r = -\overline{D}_2 \), and \( p_3^r = \delta \) for \( r \geq r' \). In particular, this implies

\[
(5.14) \quad L^r \geq 1 - a \text{ and } \frac{3}{2}a - 2 < D^r \leq \frac{3}{2}b - 2 \forall r \geq r'.
\]
To begin, suppose $p'_2 = 0$. Now, $L' = -D'$ and $D' = \frac{1}{2} \overline{D}_2 = \frac{1}{6} + \frac{1}{3} D'^{-1}$ for $r > r'$. Solving for $D' = D'^{-1}$ yields $D^* = -\overline{L} = \frac{1}{4}$. As $L' = -\frac{1}{4}$, $L'^* < 0$ and $p'_2 = 0$ requires $L'^* < 0$. Together with (5.14), this implies that $b = \frac{3}{7}$ and $\frac{5}{7} \leq a < \frac{3}{7}$. Suppose now that $p'_2 = -\overline{L}_i < 0$ for $r \geq r'$. Hence, $L' = -D'$ and $D' = \frac{1}{2} (\overline{D}_2 - \overline{L}_i) = \frac{1}{6} a + \frac{2}{7} D'^{-1}$. Solving for $D'^{-1}$ yields $D^* = \frac{1}{6} a = -\overline{L}$. However, this is not possible due to (5.14), since $L' \geq 1 - a > -\frac{1}{2} a$. To continue, suppose $\pi'_2(\overline{L}_i) = \pi'_2(L^*_i) = \frac{1}{2}$ for $r > r'$. $L^{r+1}_i + L^{r+1}_i = \frac{1}{2} (4L' + b - a) = 0$ implies $L' = \frac{1}{4} (a - b)$ for $r \geq r'$. Furthermore, $D' = \frac{1}{2} \overline{D}_2 + \frac{1}{6} \overline{L} - \frac{1}{2} \overline{L}_i = \frac{1}{12} (a + b + 4D'^{-1})$ and $L' = -\frac{1}{2} \overline{D}_2 + \frac{1}{6} \overline{L}_i + \frac{1}{4} \overline{L}_i = \frac{1}{12} (b - a + 2 - 4D'^{-1} - 4L'^{-1})$. Solving for $L' = L'^{-1}$ and $D' = D'^{-1}$ and using $L' = \frac{1}{4} (a - b)$, we get that $L' = \frac{1}{4} (a - b)$ and $D^* = \frac{1}{12} (a + b)$. (5.14) implies that $\frac{7}{9} \leq b < \frac{11}{12}$ and $a = \frac{7}{9} b - \frac{4}{7}$.

The last possibility is $p'_2 = L^*_i > 0$. Thus, $L^{r+1}_i + L^{r+1}_i < 0$ and $L' < \frac{1}{4} (a - b)$ for $r > r'$. Furthermore, $D_r = \frac{1}{2} (L^*_r + \overline{L}^*_r) = \frac{1}{2} (b + 2 \overline{D}_r + \overline{D}_r) = \frac{1}{6} \overline{L}_i - \frac{2}{7} \frac{1}{3} \overline{D}$ and $D^* = \frac{3}{12} b - \frac{1}{4}$. (5.14) and $L' < \frac{1}{4} (a - b)$ imply that $\frac{7}{9} \leq b < \frac{11}{12}$ and $a < \frac{11}{12} + \frac{1}{4} b$.

In conclusion, if $1 \leq a < 2$ and $b > 1$, then a necessary condition for convergence is that either of the following holds:

(i) $\frac{11}{12} \leq a = b < 2$ with $f^* = \frac{1}{4}, L' = 0, D^* = \frac{7}{9} (a - 1)$
(ii) $b = \frac{7}{9}, \frac{11}{12} \leq a < \frac{7}{9}$ with $f^* = \frac{2}{3}, D^* = -L' = \frac{1}{4}$
(iii) $\frac{7}{9} \leq b < \frac{11}{12}, a = \frac{7}{9} b - \frac{7}{9}$ with $f' = \frac{2}{3}, L' = \frac{1}{4} (a - b), D' = \frac{1}{12} (a + b)$;
(iv) $\frac{7}{9} \leq b < \frac{11}{12}, \max \{\frac{7}{9} - \frac{1}{10} b, \frac{2}{7} b - \frac{8}{9}\} < a < \frac{6}{7} + \frac{1}{4} b$ with $f' = \frac{2}{3}, L' = \frac{1}{10} b - \frac{2}{7}, D' = \frac{3}{10} b - \frac{1}{3}$.

(Note that these four regions covers a very small part of the parameter plane.)
(iii) If $a \geq 2$, then $f^r = 0$ and $\limsup_{r \to \infty, r \leq r'} D^r \in \mathbb{R}$.

It is immediate that $p_i^r = \delta$ for all $r$ and $i$. Hence, $f^r = 0$ and $D^r = 0$ for all $r$.

Q.E.D.
**Cycles**

To illustrate the cyclic dependence of the equilibrium outcome on $T$ when the core is empty, we provide below the simulation results for $1 < a < 2$ and $a < b < 3$. The color of the area indicates the period of the cycle. White regions indicate that there is a steady state. The darker the color of the area, the higher the period of the cycle. The darkest color indicates the period is equal or higher than 10.
5.7.3 Proofs of Proposition 5.2 and Proposition 5.3 (Informal)

**Proof of Proposition 5.2**

To prove that $\Gamma^I_\tau$ is well-defined, we need to show how the game proceeds given some profile $\sigma \in \Sigma$ and some history $h_\tau$. We first define the first moment of movement:

**Definition 5.5** Given $\sigma \in \Sigma$ and $h_\tau \in H$ and let $R_i(\sigma_i \mid h_\tau) = \{\tau \leq t \leq T: \sigma_i(h_\tau)) = \sigma_i(h_{\tau-1}(h_\tau))\}$. (i) $r_i(\sigma_i \mid h_\tau)$ is the first moment of movement of player $i$. If $R_i(\sigma_i \mid h_\tau) = \emptyset$, then $r_i(\sigma_i \mid h_\tau) \equiv T + \rho$ and player $i$ would not move at any $h_i(h_\tau) \subseteq h_\tau(h_\tau)$. If $R_i(\sigma_i \mid h_\tau) \neq \emptyset$, then $r_i(\sigma_i \mid h_\tau) = \min R_i(\sigma_i \mid h_\tau)$. (ii) We define the first moment of movement $r(\sigma \mid h_\tau) \equiv \min_{i \in X} \{r_i(\sigma_i \mid h_\tau)\}$.

If $R_i \neq \emptyset$, $\min R_i$ must exist, because otherwise (S2) would not hold for $h_{\inf R_i}(h_\tau)$.

Now we can define a function $\gamma$ that returns a history $h'$ as a function of any unresolved history $h_\tau$. The function determines whether the (absence of a) first moment of action directly leads to a resolved history. If this is not the case, it returns another unresolved history with $\tau(h') \geq \tau(h) + \rho$.

**Definition 5.6** Define $\gamma : H \times \Sigma \to \overline{H} \cup \overline{H}, (h_\tau, \sigma) \to h'$, as follows:

- $h_i(h_\tau) \subset h'$, with $r = r(\sigma \mid h_\tau)$.
- If $r > T$, then $h' = h_{T+r}(h_\tau) \in \overline{H}$.
- If $r \leq T$, then $p_i(h') = \sigma_i(h_\tau)) \forall i, j \forall t \in [r, \tau(h')]$
- $h' \in \overline{H}$ if $T - \rho < r \leq T$ or $\sigma_i(h_\tau) = a_j$ for some $i, j$
- $h' \in H$ and $\tau(h') = r + \rho$ if $r \leq T - \rho$ and $\sigma_i(h_\tau) \neq a_j \forall i, j$
Now, it is straightforward to show that that $\Gamma_T^I$ is well-defined.

**Proposition 5.2** The game $\Gamma_T^I$ is a well-defined mapping $\Gamma_T^I(G) : H \times \Sigma \rightarrow \overline{H}_{T+\rho}(h_r, \sigma) \rightarrow \overline{h}$

**Proof:** Consider $\sigma \in \Sigma$ and $h_r \in H$. $\Gamma_T^I$ applies $\gamma$ iteratively. It starts with $\gamma^0(h_r) = h_r$. If $\gamma^k(h_r) \in H$, then $\gamma^{k+1}(h_r) = \gamma(\gamma^k(h_r))$. If $\gamma^k(h_r) \in \overline{H}$, then the procedure stops and $\overline{h} = \gamma^k(h_r)$. Because $\tau(\gamma(h_r)) \geq \tau(h_r) + \rho$ and $T / \rho$ is finite, this procedure will always return a resolved history. Q.E.D.

**Proof of Proposition 5.3**

By $p$ we denote the vector $(p_1, p_2, p_3)$ and by $p_i$ we denote the vector $(p_j, p_k)$. For convenience, we set $u_i(\zeta) \equiv -\infty$.

**Definition 5.7** By $\hat{p} / \hat{z}$ we denote the strategy profile such that the following hold.

1. For each active history $\tilde{h}_r \in \tilde{H}_r$ with $\tau > T - \rho$
   
   (i) $\sigma_i(\tilde{h}_r) = a_j$ iff (a) $u_i(p_j^- (\tilde{h}_r)) \geq u_i(\delta)$ and $u_i(p_k^- (\tilde{h}_r)) > u_i(p_k^- (\tilde{h}_r))$
   or (b) $p_j^- (\tilde{h}_r) = \hat{z}$, $u_i(\hat{z}) \geq u_i(\delta)$ and $u_i(p_k^- (\tilde{h}_r)) \leq u_i(\hat{z})$, or (c) $p_j^- (\tilde{h}_r) \neq \hat{z}$ and $u_i(p_j^- (\tilde{h}_r)) = u_i(p_k^- (\tilde{h}_r)) \geq u_i(\delta)$ and either $p_j^- (\tilde{h}_r) > p_k^- (\tilde{h}_r)$ or $p_j^- (\tilde{h}_r) = \delta$.

   (ii) $\sigma_i(\tilde{h}_r) = \delta$ iff $\sigma_i(\tilde{h}_r) \notin \{a_1, a_2, a_3\}$

2. For each active history $\tilde{h}_r \in \tilde{H}_r$ with $T - \rho < \tau \leq T - \rho$
   
   (i) $\sigma_i(\tilde{h}_r) = a_j$ iff (a) $u_i(p_j^- (\tilde{h}_r)) \geq u_i(\hat{z})$ and $u_i(p_j^- (\tilde{h}_r)) > u_i(p_k^- (\tilde{h}_r))$
   or (b) $p_j^- (\tilde{h}_r) = \hat{z}$ and $u_i(p_j^- (\tilde{h}_r)) \leq u_i(\hat{z})$ or (c) $p_j^- (\tilde{h}_r) \neq \hat{z}$ and $u_i(p_j^- (\tilde{h}_r)) = u_i(p_k^- (\tilde{h}_r)) \geq u_i(\hat{z})$ and either $p_j^- (\tilde{h}_r) > p_k^- (\tilde{h}_r)$ or $p_j^- (\tilde{h}_r) = \delta$.

188
(iii) \( \sigma_i(\tilde{h}_r) = \hat{p}_i \) iff \( \sigma_i(\tilde{h}_r) \notin \{a_1, a_2, a_3\} \)

3. \( z(\Gamma^I_\hat{h}(\hat{p} / \hat{z})) = \hat{z} \) if \( p^r (h_r) = \hat{p} \) and \( \tau \leq T - \rho \).

Such profiles have a special property:

**Lemma 5.2** If \( \sigma = \hat{p} / \hat{z} \) is a SPE of some subgame \( \Gamma^I_{T, h'} \) with \( h^* \in \tilde{H}_{T - \rho} \), then it is a SPE for any subgame \( \Gamma^I_{T, h} \).

**Proof:** Let \( \sigma = \hat{p} / \hat{z} \) be a SPE of \( \Gamma^I_{T, h'} \) and let \( \Gamma^I_{T, h'} \) be a subgame of \( \Gamma^I_T \).

For all \( \hat{h} \) with \( \tau(\hat{h}) > T - \rho \), it is immediate from the definition of \( \hat{p} / \hat{z} \) that \( \sigma \) is a SPE of \( \Gamma^I_{T, \hat{h}} \).

Hence, consider some \( \hat{h} \) with \( \tau(\hat{h}) \leq T - \rho \) and let us look at whether there exists some \( \sigma_i \) such that \( U_i(\sigma_i; \sigma_{-i} | \hat{h}) > U_i(\sigma_i; \sigma_{-i} | \hat{h}) \). If players \( j \) and \( k \) adhere to \( \sigma_{-i} \), then \( p_i^r(h_i) = \hat{p}_{i} \) for all \( h \supset \hat{h}_r \). At \( \hat{h}_r \) each player \( i \) accepts according to \( \sigma \) any proposal yielding a higher payoff than \( u_i(\hat{z}) \). Hence, a necessary condition for a profitable deviation is that one of the following holds:

1. **(i)** a subhistory \( h'_i \supset \hat{h} \) exists under \( \sigma_i \) the most attractive proposal yielding her a higher payoff than \( u_i(\hat{z}) \) (i.e. \( \exists j \neq k : \sigma_i(h'_i) \neq a_j, p_i^r(h'_i) = \zeta, u_i(p_i^r(h'_i)) > u_i(\hat{z}) \) and \( u_i(p_i^r(h'_i)) > u_i(p_i^r(h'_i)) \)).

2. **(ii)** a subhistory \( h'_i \supset \hat{h} \) exists where she can make a deviating proposal with a higher payoff than \( \hat{z} \) that will be accepted in the next active history \( h_{r+\rho} \supset h'_r \). (i.e. for some active history \( h'_r \), \( \exists z : \sigma_i(h'_r) = z \), \( u_i(z) > u_i(\hat{z}) \) and for some \( j \), \( \sigma_j(h_{r+\rho}) = a_i \) for \( h_{r+\rho} \supset h'_r \), \( p_i^{r+\rho}(h'_r, h_{r+\rho}) = z \) and \( p_i^{r+\rho}(h'_r, h_{r+\rho}) = \hat{p}_{i-} \)).

3. **(iii)** \( u_i(\delta) > u_i(\hat{z}) \) and a subhistory \( h'_i \supset \hat{h} \) exists where she can deviate by moving to be silent such that at the next active history no pro-
proposal is accepted. (i.e. \(\sigma_i(h_{t+\rho}), \sigma_i(h_{t+\rho}) \notin (a_1, a_2, a_3)\) for \(h_{t+\rho} \supset h'_t\) with \(p_{t+\rho}^{-}(h'_{t+\rho}) = q\) and \(p_{t+\rho}^{-}(h'_{t+\rho}) = \hat{p}_{t+\rho}^{-}.\)

From the definition of \(\hat{p}/\hat{z}\) it follows that \(\sigma_i(h'_t) = a_j\) iff \(\sigma_i(h'_t) = a_j\) for \(h'_t \supset h^*\) with \(p^{-}(h'_t) = p^{-}(h'_t)\). Hence, if either of aforementioned \((i)-(iii)\) would hold, then player \(i\) could also profitably deviate at either \(\tilde{h}'_t\) or \(h^*\) and \(\sigma\) would not be a SPE of \(\Gamma'_T/_{T,h^*}\).

Hence, no player \(i\) can profitably deviate from \(\sigma_i\), and \(\sigma\) is a SPE of \(\Gamma'_T/_{T,h^*}\).

Q.E.D.

We are now ready to characterize the equilibrium outcomes of \(\Gamma'_T\).

**Proposition 5.3** The set of SPE outcomes is equal to \([c, b] \cup \delta\) for any \(\Gamma'_T\) with \(T \geq \rho\), where \(c = \min\{-a, \max\{-b, b-1\}\}\).

**Proof:** We first show by construction that \([c, b] \cup \delta\) are SPE outcomes of \(\Gamma'_T\) if \(T \geq \rho\). For \(z = 0\), we simply need to observe that \((0, \varsigma, 0) / 0\) is an SPE of any \(\Gamma'_T/_{T,h_{t+\rho}}\) and hence \(\Gamma'_T\). For \(z \in [-a, 0]\) consider the following profile: \(\sigma\) is equal to \(\hat{p}/\hat{z} = (0, \varsigma, 0) / 0\), except that \(\sigma(h'_0) = (\varsigma, z, z)\) and \(\sigma(h''_p) = \sigma(h''_p) = (\varsigma, -a, -a)\) with \(p^{-}(h''_p) = (\varsigma, z, \delta)\) and \(p^{-}(h''_p) = (\varsigma, \delta, z)\). Now, \((0, \varsigma, 0) / 0\) is an SPE of any subgame and \(h_0\) is the only active subhistory of \(h'_p\) and \(h''_p\). Hence, it only remains to be shown that no player can profitably change strategies at \(h'_p, h''_p\) and \(h''_p\). At \(h'_p\), player 1 will obtain her maximal payoff. Furthermore, at \(h''_p\) player 2 nor 3 can profitably deviate: neither of them can accept the other’s proposal and, whatever they propose at \(h''_p\), player 1 will accept \(-a\) at \(h''_p, (h'_p)\) given that the other proposes \(-a\). By the same reasoning, at \(h''_p\) no player can profitably deviate. Finally, no player can profitably deviate at \(h_0\). If player 1 moves away from \(\varsigma\), the outcome will be 0, which is worse for her than \(z\). Players 2 and 3 cannot do better by proposing anything else; in particular, even
if \( 1 - a < z < 1 \) proposing \( \delta \) at \( t=0 \) is not attractive for them, because that will trigger the subgame in which \(-a\) is the outcome (rather than player 1 accepting \( \delta \)). Hence, \( \hat{p} / \hat{z} \) is an SPE of \( \Gamma_T' \).

In a similar way, SPE of \( \Gamma_T' \) can be constructed that support \( z \in (0, b] \) as an outcome. An SPE of \( \Gamma_T' \) that supports \( \delta \) is \( \hat{p} / \hat{z} = (\delta, \delta, \delta) / \delta \), which is obviously an SPE of any \( \Gamma_T' \). Finally, an SPE that supports \( z \in [\max\{-b, b-1\}, -a) \) (if \(-b < b - 1 < -a\)) is the following profile: \( \sigma \) is equal to \( \hat{p} / \hat{z} = (\delta, \delta, \delta) / \delta \), except that \( \sigma(h_0) = (z, z, z) \) and \( \sigma(h_0') = (h, b, z) \) for all \( h_0' \) with \( p_1^{-}(h_0') \neq z \) or \( p_2^{-}(h_0') \neq z \). It is easily verified that no player can profitably deviate from \( \sigma \) at any \( h_0' \) or \( h_0 \).

Second, we show that all points in \( \mathbb{R} \) outside of \( [c, b] \) cannot be equilibrium outcomes. Suppose \( \sigma' \) is an SPE with outcome \( \hat{z} \in \mathbb{R} \setminus [c, b] \). In this case, player 2 can in equilibrium never accept \( x \neq 0 \) at a history \( h_\tau \), because then either player 1 or 3 could profitably deviate by proposing 0 at \( t(h_\tau) \). If 0 is proposed, namely, then in equilibrium either player 2 will accept this, or it will trigger a subgame in which 0 is the outcome under \( \sigma' \).

Player 1 will in equilibrium never accept \( x \) with \( |x| > a \), because then player 2 could profitably deviate by proposing \(-a\) at \( t = t(h_\tau) \) by the same reasoning. Similarly, player 3 will never accept an \( x \) with \( |x| > b \) in equilibrium. This immediately rules out the possibility that \( z \in (-\infty, -b) \cup (b, \infty) \). If \( \hat{z} \in [-b, \min\{-a, \max\{-b, b-1\}\}) \), then it must be accepted by player 3. However, player 3 could then profitably deviate by at no history accepting \( \hat{z} \). Since players 1 and 2 will never accept a proposal outside \([-a, a] \) in equilibrium, the outcome would always be better than \( \hat{z} \) for player 3. Q.E.D.
5.7.4 Experimental Instructions

We present the English translation of the original instructions in Dutch for both treatments.

Instructions Informal Treatment

INSTRUCTIONS

You will initially have fifteen minutes to go through these instructions. When time is up, we will ask whether there is anyone who would like some more time. In case you need more time, please raise your hand and we will simply give you the extra time you need.

Introduction

In a moment you will participate in a decision making experiment. The instructions are simple. If you follow them carefully, you can earn a considerable amount of money. Your earnings will be paid to you individually at the end of the session and separately from the other participants.

You have already received five euros for showing up. In addition, you can earn more money during the experiment. In the experiment the currency is ‘francs.’ At the end of the session, francs will be changed into euros. The exchange rate is 1 euro for each 10 francs.

In this experiment you can also lose money. To prevent that your earnings become negative, you will receive at the beginning of the experiment 75 francs extra. In the unlikely situation that your final earnings will be negative, your earnings will be zero (but you keep the five euros for showing up.)

Your decisions will remain anonymous. They will not be linked in any way to your name. Other participants cannot possibly figure out which decisions you have made. You are not allowed to talk to other participants or communicate with them in any other way. If you have a question, please raise your hand.

Periods and Groups

The experiment consists of 24 periods, each of which will be carried out in groups of three players.

At the beginning of each period, participants will again be randomly divided into groups of three. The chances that you will be with any other participant in the same group for two consecutive periods are therefore very small.
**Choices and Earnings**

In each period, your group of three participants negotiates about choosing a number. The chosen number determines the earnings of each of the three participants for that period. The group can choose any integer between 0 and 100. The group can also choose not to determine any number (the “no number” option).

Hence, the number chosen by the group determines the earnings for each member of the group. These earnings are different per member nevertheless. How much a player earns depends, in addition of the chosen number, also on her ‘ideal value.’ Each player in a group receives an ideal and unique value between 0 and 100. The earnings for a player increase as the outcome lies closer to this ideal value.

If the outcome is exactly equal to the ideal value of a player, then this player receives the maximum earnings of 20 francs. The difference between the ideal value and outcome (if any) decreases the earnings by the same amount in francs. For instance, suppose your ideal value in a certain round is 54. Then you receive 20 francs if the outcome of the period is 54, 19 if the outcome is 53 or 55, 18 if the outcome is 52 or 56 etc. Your earnings may also be negative. If the group, for instance, chooses the number 20, then with an ideal value of 54, your earnings will be equal to -14.

The outcome of a period can also be that the group reaches no agreement. Hence, one chooses “no number.” In this case each member of the group receives 0 francs.

During a round, players are identified by a letter: A, B and C. These are based on their ideal value: the player with the lowest ideal value is A and the player with the highest ideal value is C. For instance, suppose the ideal values of the three players are 16, 54 and 86. Then the player with ideal value 16 is player A, the player with ideal value 54 is player B and, finally, the player with ideal value 56 is player C.

**The negotiations**

The group negotiations on how to choose a number consist of several steps. First, we give an overview. Afterwards, we discuss the separate steps one at a time.

1. Before the negotiations start, each player can send a private message to each other member of the group. A message is a suggestion for the number to choose. Each message from one player to another remains secret for the third player.
2. Then, there will be 2.5 minutes during which participants can make and accept proposals. A proposal is a number between 0 and 100 or a proposal to end the negotiations.
3. As soon as a proposal is accepted by a player other than the proposer, the negotiations end. The accepted proposal is the group’s choice for that period.
4. The period also ends if after two and half minutes no proposal has been accepted. The outcome is then “no number” and all players earn 0 francs.
Information screen

The first screen that you will see in a period, will show which player you are (A, B or C) and the ideal values of you and your group members. Your own letter is marked in red.

If you are ready to proceed, before the time has elapsed, you can press the OK-button.

Sending and receiving messages

Subsequently, you will be able to send a message to each of your two group members and they will be able to send a message to you.

A message is either an integer between 0 and 100 or the word “end.” A number is a suggestion for the group choice. With “end” you tell the two players that you do not want to negotiate (and therefore have earnings 0). You can also choose to send no message by not filling out anything or typing the space bar. To send a message, you fill out a number or “end” in one or both cells and you press OK.

Attention: suggestions you send as a message are not put to a vote and will only be seen by the player who receives the message.

You receive 30 seconds to send messages. If you do not fill out anything and press OK within this time, then no message will be sent. The other players will only see a space at their cell in this case.

After the 30 seconds have elapsed, you will see the messages that the other players sent to you. You will NOT see what the other players sent to each other.

Making and accepting proposals

You are then ready to make and accept proposals. In this phase you will see in the top-left corner of your screen all the necessary information (your identity, the messages,
the ideal values). At the end of these instructions, we will show you the entire screen lay out.

As a group you will have **two and a half minutes** (150 seconds) to accept a proposal (or not accept one). A proposal can once again be **any integer between 0 and 100 or the word “end.”**

During this phase, **you can do three things:** make your own proposal, revise your own proposal or accept a proposal by another player.

To make a proposal, you fill out the number or word you want to propose and press on the “OK” button. This proposal will become immediately visible to the other players in the list “outstanding proposals.” Each of the other two players can accept a proposal you make.

**To revise your proposal,** you simply make another proposal. This must be different from the previous proposal. The old proposal disappears from the list “Outstanding Proposals” (but, as we shall see later, it will remain in the list “Made proposals” on the left of your screen). The new proposal replaces the old one in the list “Outstanding Proposals”

If one of the other layers has made a proposal, then you can **accept a proposal.** You do this by clicking on the proposal you want to accept in the list “Open Proposals” and press the button “Accept this Proposal.”

As soon as a proposal has been accepted by a player, the **period ends.** The choice of the group for this period is then the accepted proposal. If no proposal is accepted within the two and a half minutes, then the group chooses “no number” and all players receive 0 points.
Results

At the end of each period, you will get to see the outcome and the corresponding earnings.

Screens

There is a lot of information you can use while you are making your choices. You can find:

- the player you are
- the ideal values of each player
- the messages you sent and received
- the proposals that have been rejected
- the outcomes of previous periods

At “Previous Periods,” you can find the outcomes of previous periods, together with the ideal values of the player and, between brackets their earnings. The word “You” before the value and payment indicates which player you were.

At the far-left corner below you see in red the total amount of points (Earnings) that you have made across rounds. Because you received 75 francs at beginning, the counter starts at 75. Divide the final score by 10 to determine your earnings in euros. All information about previous periods is shown together on the left side of the screen. On the right side of the screen you will find new information and/or what action you have to take. On top, the ideal values of all players are displayed. Finally, you can find in the far-left corner below a help box with short description of what you have to do.
**Instructions Formal Treatment**

**INSTRUCTIONS**

You will initially have fifteen minutes to go through these instructions. When time is up, we will ask whether there is anyone who would like some more time. In case you need more time, please raise your hand and we will simply give you the extra time you need.

**Introduction**

In a moment you will participate in a decision making experiment. The instructions are simple. If you follow them carefully, you can earn a considerable amount of money. Your earnings will be paid to you individually at the end of the session and separately from the other participants.

You have already received five euros for showing up. In addition, you can earn more money during the experiment. In the experiment the currency is ‘francs.’ At the end of the session, francs will be changed into euros. The exchange rate is **1 euro for each 10 francs**.

In this experiment you can also lose money. To prevent that your earnings become negative, you will receive at the beginning of the experiment 75 francs extra. In the unlikely situation that your final earnings will be negative, your earnings will be zero (but you keep the five euros for showing up.)

Your decisions will remain **anonymous**. They will not be linked in any way to your name. Other participants cannot possibly figure out which decisions you have made.

You are not allowed to talk to other participants or communicate with them in any other way. If you have a question, please raise your hand.

**Periods and Groups**

The experiment consists of **24 periods**, each of which will be carried out in **groups of three players**.

At the beginning of **each period**, participants will again be randomly divided into groups of three. The chances that you will be with any other participant in the same group for two consecutive periods are therefore very small.

**Choices and Earnings**

In each period, your group of three participants negotiates about **choosing a number**. The chosen number **determines the earnings** of each of the three participants for that period. The group can choose **any integer between 0 and 100**. The group can also choose not to determine any number (the “**no number**” option).
Hence, the number chosen by the group determines the earnings for each member of the group. These earnings are different per member nevertheless. How much a player earns depends, in addition of the chosen number, also on her ‘ideal value.’ Each player in a group receives an ideal and unique value between 0 and 100. The earnings for a player increase as the outcome lies closer to this ideal value.

If the outcome is exactly equal to the ideal value of a player, then this player receives the maximum earnings of 20 francs. The difference between the ideal value and outcome (if any) decreases the earnings by the same amount in francs. For instance, suppose your ideal value in a certain round is 54. Then you receive 20 francs if the outcome of the period is 54, 19 if the outcome is 53 or 55, 18 if the outcome is 52 or 56 etc. Your earnings may also be negative. If the group, for instance, chooses the number 20, then with an ideal value of 54, your earnings will be equal to -14.

The outcome of a period can also be that the group reaches no agreement. Hence, one chooses “no number.” In this case each member of the group receives 0 francs.

During a round, players are identified by a letter: A, B and C. These are based on their ideal value: the player with the lowest ideal value is A and the player with the highest ideal value is C. For instance, suppose the ideal values of the three players are 16, 54 and 86. Then the player with ideal value 16 is player A, the player with ideal value 54 is player B and, finally, the player with ideal value 56 is player C.

The negotiations

The group negotiations to choose a number consist of several steps. First, we give an overview. Afterwards, we will discuss the separate steps one at a time.

1. Before the negotiations start, each player can send a separate message to each other member of the group. A message is a suggestion for the number the group can choose. Each message from one player to another remains secret for the third player.
2. Next, at most 10 rounds follow with making proposals and voting.
3. During each round, each of the three participants makes a proposals. This proposal can be any number between 0 and 100 or a proposal to end the negotiations. Subsequently, one of the three proposals is randomly chosen to be put to a vote. The other two participants can then vote “For” or “Against” the chosen proposal (the player who made the chosen proposal automatically votes in favor).
4. If one of these two participants votes “For,” then the proposal is accepted and the period ends. If both participants vote “Against,” then the proposal is rejected and there will be a next round of making proposals and voting. This can continue until nine proposals have been rejected; if also the tenth proposal is rejected, then the period ends and the outcome is ‘no number.’
5. If a proposal to end the negotiations is accepted, then the outcome is “no number” and, consequently, all players receive 0 francs. If a proposed number is accepted, then this number is the choice of the group for that period.
Information screen

The first screen that you will see in a period, will show which player you are (A, B or C) and the ideal values of you and your group members. Your own letter is marked in red.

If you are ready to proceed, before the time has elapsed, you can press the OK-button.

Sending and receiving messages

Subsequently, you will be able to send a message to each of your two group members and they will be able to send a message to you.

A message is either an integer between 0 and 100 or the word “end.” A number is a suggestion for the group choice. With “end” you tell the two players that you do not want to negotiate (and therefore have earnings 0). You can also choose to send no message by not filling out anything or typing the space bar. To send a message, you fill out a number or “end” in one or both cells and you press OK.

Attention: suggestions you send as a message are not put to a vote and will only be seen by the player who receives the message.

You receive 30 seconds to send messages. If you do not fill out anything and press OK within this time, then no message will be sent. The other players will only see a space at their cell in this case.

After the 30 seconds have elapsed, you will see the messages that the other players sent to you. You will NOT see what the other players sent to each other.
Making a proposal

You are then ready to make and accept proposals. In this phase you will see in the top-left corner of your screen all the necessary information (your identity, the messages, the ideal values). At the end of these instructions, we will show you the entire screen lay out.

A proposal can once again be any integer between 0 or 100 or the word “end.” To make a proposal, you fill out this number or word and press “OK.”

To help you calculate quickly which payments belong to which proposal, you also have an earnings-calculator at your disposal. If you fill out any number and press “Calculate” then the earnings will appear that all members would receive should that proposal be accepted. This device is only meant to help you. Nothing that you type there, will be seen by the other players.

**ATTENTION**: In each rounds, everybody fills out a proposal. However, only one of these proposals is (randomly) chosen. This proposal will be revealed to the others and be put to a vote.

You will receive 40 seconds to make your proposal. If you do not type in anything within this time, then ‘0’ will be your proposal.

Voting

After everybody has made a proposal, it will be revealed whose proposal has been chosen. Moreover, the payments everyone would receive if this proposal would be accepted are also shown.

Next, the proposal will be put to a vote. The player who made the proposal, automatically votes “For” and does not press any button.
The other members can vote by simply pressing “For” or “Against.”

If at least one of the two votes is “For”, then the proposal is accepted and it will be the outcome of that period. The group has then made a decision and the period ends.

If both vote “Against,” then the proposal is rejected and you will proceed to a next round of proposing and voting. This can continue until nine proposals have been rejected. If the tenth proposal is also rejected, then the group was not able to reach a decision and the period ends. In this case, the outcome is “no number.”

Results

At the end of each round, you will see how each player voted, whether the proposal has been accepted and whether or not you will go to a next round.

At the end of each period, you will see the outcome and your corresponding earnings.

Screens

There is a lot of information you can use while you are making your choices, You can find:
- the player you are
- the ideal values of each player
- the messages you sent and received
- the proposals that have been rejected
- the outcomes of previous periods

At “Previous Periods,” you can find the outcomes of previous periods, together with the ideal values of the player and, between brackets their earnings. The word “You” before the value and payment indicates which player you were.
At the far-left corner below you see in red the total amount of points (Earnings) that you have made across rounds. Because you received 75 francs at beginning, the counter starts at 75. Divide the final score by 10 to determine your earnings in euros.

All information about previous periods is shown together on the left side of the screen. On the right side of the screen you will find new information and/or what action you have to take. On top, the ideal values of all players are displayed.

Finally, you can find in the far-left corner below a help box with short description of what you have to do.
References


Samenvatting (Dutch Summary)

‘Essays over Onderhandelen en Strategische Communicatie’

Dit proefschrift omvat een reeks van vier essays over onderhandelen en communicatie vanuit een gedragseconomisch perspectief. Onderhandelen en communiceren zijn kenmerkende en nauw verweven onderdelen van sociale interactie. Zodoende kan onderzoek hierover licht werpen op intermenselijke verbanden, ons wellicht helpen om effectiever met elkaar te onderhandelen en misschien zelfs wegen duiden naar beter ingerichte sociale instituties.

De gedragseconomische aanpak van dit proefschrift verschilt van de benadering die gebruikelijk is in ‘standaard’ economisch onderzoek. De standaard economie gaat uit van rationele, egoïstische *hominès economici* die in perfect evenwicht interacteren en exact weten wat zij zelf en anderen denken en doen. De gedragseconomie daarentegen probeert theorieën te ontwikkelen die wortelen in psychologisch realistische aannames om zodoende gedrag goed te kunnen voorspellen. Er zit wel overlap tussen beide aanpakken, aangezien gedragseconomische modellen regelmatig een aanpassing betreffen van de standaard modellen. Gedragseconomen vinden het daarnaast belangrijk om hun theorieën te toetsen aan de hand van laboratorium- en veldexperimenten. De vier essays in dit proefschrift (hoofdstukken 2 tot 5) beogen een bijdrage te zijn aan deze groeiende tak van wetenschap.

Dit proefschrift behoort tot het specifieke onderzoeksveld van de gedragsmatige speltheorie, de tak van de gedragseconomie die zich bezig houdt met strategische interacties. Een rode draad door dit werk is dat alle essays over situaties gaan waarin de standaard speltheoretische aanname dat gedrag rationeel en in evenwicht is niet tot een eenduidige voorspelling leidt. Door minder beperkende aannames te maken en experimenten uit te voeren, kunnen de onderzoeken in dit proefschrift toch tot gerede voorspellingen komen.
Hoofdstukken 3 en 4 bevatten methodologische bijdragen aan de literatuur: hoofdstuk 3 introduceert een nieuwe methode om voorspellingen te maken in specifieke spelsituaties en hoofdstuk 4 stelt deze methode experimenteel op de proef. Hoofdstukken 2 en 5 zijn, daarentegen, toegepaste bijdragen die gedragseconomische methodes gebruiken om openstaande vraagstukken over de economie en politiek op te lossen. Hoofdstukken 2, 3 en 4 bestuderen cheap talk situaties. Cheap talk is de speltheoretische benaming voor communicatie door middel van ‘kosteloze’ boodschappen. Daar bedoelen economen niet alleen Skype mee.. Alle talige uitingen, zoals wat mensen zeggen in een gesprek of toespraak en schrijven in een email of chat, zijn vormen van cheap talk. ‘Cheap,’ omdat het bijvoorbeeld niets kost om te zeggen dat je van iemand houdt of te schrijven dat je een uitermate capabele werknemer zult zijn. (Economen onderscheiden dergelijke kosteloze boodschappen van kostbare signalen, zoals het geven van een bijzonder cadeau om je genegenheid te tonen of het volgen van een moeilijke studie om te laten zien dat je intelligent bent.) Hoofdstukken 2, 3 en 4 kijken specifiek naar elementaire cheap talk situaties waarin een zender met bepaalde informatie een boodschap stuurt aan een ontvanger. De informatie waarover de zender beschikt kan feiten of voorkeuren betreffen en is voor beiden van belang. Het wordt dan interessant wanneer zender en ontvanger deels tegengestelde en deels overeenkomende belangen hebben. We kunnen hierbij denken aan het advies dat een IT-bedrijf aan een klant geeft over de noodzaak van een lucratief IT-project of aan een onderhandeling tussen scheidende ouders wanneer de moeder niet weet hoe vaak de vader de kinderen zou willen zien.

Hoofdstuk 2 geeft een speltheoretische analyse van de invloed van macht op de helderheid van communicatie. Hoewel de relatie tussen macht en communicatie in andere disciplines reeds lang onderkend wordt, is dit een grotendeels onontgonnen gebied in de economie. We richten ons op de observatie dat er aanwijzingen zijn dat personen die behoren tot een machtigere sociale, economische of politieke groep hun wensen duidelijker communiceren. Zo zal een leidinggevende eerder aangeven dat hij of zij een voorstel van een werknemer niet ziet zitten dan omgekeerd. Ook zijn er aanwijzingen dat in patriarchale samenlevingen vrouwen minder helder kunnen zijn in het uiten van hun wensen
dan mannen. In een speltheoretisch model kunnen we een precieze invulling geven aan de begrippen ‘macht’ en ‘helderheid.’ Dit stelt ons in staat om theoretische resultaten te verkrijgen over hoe macht de helderheid van communicatie beïnvloedt. Onze belangrijkste conclusie is dat als een persoon (of instantie) weinig macht heeft, hij vaag moet zijn, omdat de informatie die hij prijsgeeft (ook) tegen hem gebruikt kan worden. Dit betekent tegelijkertijd dat macht een geïnformeerd persoon in staat stelt om duidelijker te zijn, omdat macht hem minder kwetsbaar maakt voor ‘misbruik’ van de informatie die hij verstrekt. Ons model voorspelt bijvoorbeeld dat een vader tijdens onderhandelingen over bezoekregelingen vaag zal zijn als hij juridisch zwak staat of weinig geld heeft voor advocaten. Helderheid is dus een privilege van de machtigen.

Hoofdstuk 3 pakt het probleem van evenwichtselectie op. Speltheorie neemt aan dat gedrag in strategische situaties uiteindelijk in een evenwicht zal geraken. Gedrag wordt in evenwicht beschouwd als niemand er baat bij heeft om zijn of haar gedrag te wijzigen als alle anderen vasthouden aan hun handelwijze. Deze benadering biedt voordelen omdat er in de praktijk maar een aantal evenwichten zijn. Met behulp van zogenaamde selectiecriteria kunnen speltheoretici vervolgens bepalen welk van deze evenwichten het meest plausibel is en een eenduidige voorspelling doen. Voorgaande pogingen om tot algemeen geldende criteria te komen in cheap talk spellen zijn echter niet bijzonder succesvol gebleken. Onze inschatting is dat dit komt omdat er werd uitgegaan van perfect rationale actoren die in een of ander stabiel evenwicht moeten verkeren. Het probleem is dat doorgaans alle cheap talk evenwichten instabiel zijn, omdat ze zogeheten geloofwaardige afwijkingen toelaten. Wij zien een uitweg uit dit probleem in de observatie dat gedrag in de praktijk toch nooit geheel in of uit evenwicht is. Dit betekent wellicht dat de stabiliteit van een evenwicht niet een zwart-wit gegeven is en dat we de mate van (in)stabiliteit moeten meten. Dit idee werken we uit tot het ‘Average Credible Deviation Criterion’ (ACDC). Volgens dit criterium wordt een evenwicht waarschijnlijker naarmate het minder en kleinere geloofwaardige afwijkingen toestaat. ACDC blijkt gedrag minstens zo goed te kunnen voorspellen als reeds bestaande selectiecriteria in eerder
uitgevoerde experimenten. Daarnaast kan ACDC ook voorspellingen doen in situaties waar andere criteria dat niet kunnen.

Hoofdstuk 4 beschrijft een nieuw laboratorium experiment dat we uitvoeren om ACDC te toetsen. In een economisch laboratorium experiment moeten proefpersonen in een abstracte situatie anoniem keuzes maken, waarin hun beslissingen bepalen hoeveel (cash) geld ze aan het eind mee naar huis kunnen nemen. Het voordeel van dit soort experimenten is dat we veel controle kunnen houden op de omgeving, wat ons in staat stelt onze theoretische modellen nauwkeurig te toetsen. In het experiment dat wij uitvoeren vinden we dat onze resultaten de uitkomsten van ACDC schragen. Daarnaast ondersteunen de uitkomsten van het experiment ook de belangrijkste voorspelling uit hoofdstuk 2, namelijk dat meer macht tot meer helderheid leidt.

Hoofdstuk 5 omvat het thema van de politieke besluitvorming. Vanuit een speltheoretisch oogpunt is politieke besluitvorming een groot onderhandelingsspel tussen politieke actoren met verschillende belangen. Ons onderzoek spitst zich toe op de vraag hoe de mate van formaliteit van het politieke besluitvormingsproces de uitkomst beïnvloedt. Politieke besluitvorming vindt namelijk gedeeltelijk plaats in formele situaties, zoals zittingen in de Tweede Kamer en gedeeltelijk in informele situaties, zoals in de wandelgangen, de achterkamertjes en, in Nederland, ‘het Torentje’. Dit hoofdstuk bevat een speltheoretische en experimentele vergelijking van een informeel en een formeel onderhandelingsspel. Het informele onderhandelingsspel biedt zoveel strategische mogelijkheden dat haast alle uitkomsten een evenwicht vormen. Een gevolg hiervan is dat de huidige speltheorie geen voorspelling kan maken over de invloed van formaliteit. Een experiment kan wel licht werpen op deze kwestie. Om deze reden hebben we een experiment uitgevoerd, waarin we deelnemers laten onderhandelen in een formele en een informele situatie. De belangrijkste experimentele bevinding is dat formaliteit van belang is. In het bijzonder zien we dat de speler met de beste uitgangspositie gebaat is bij informele onderhandelingen. Een verklaring hiervoor is dat informele onderhandelingen hem meer ruimte geven om deze uitgangspositie uit te buiten. Dit impliceert in concreto dat een middenpartij, zoals het CDA, liever achterkamertjes politiek bedrijft dan onderhandelt in de Tweede Kamer, omdat zij zo beter de partijen
aan de flanken tegen elkaar kan uitspelen. Een extreme partij, zoals de PVV of de SP, is juist gebaat bij formele procedures, omdat dit haar verzekert dat haar meningen en voorstellen gehoord worden.