Doctoral Thesis

Competition and cooperation
Sociological studies in experimental game theory

Author(s):
Berger, Joël

Publication Date:
2015

Permanent Link:
https://doi.org/10.3929/ethz-a-010568360

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
COMPETITION AND COOPERATION
SOCIOLOGICAL STUDIES IN
EXPERIMENTAL GAME THEORY

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

JOËL BERGER
MA Soc, University of Bern
born on October 12, 1982
citizen of Reiden (LU)

Prof. Dr. Andreas Diekmann (examiner)
Prof. Dr. Siegwart Lindenber (co-examiner)
Prof. Dr. Thomas Voss (co-examiner)

2015
Summary

This dissertation is a collection of five studies on competition (chapters 2 and 3) and cooperation (chapters 4, 5 and 6) – two societal constants. Methodologically, all the contributions are based on or, at least inspired by, game theory, and the hypotheses are tested with experimental and quasi-experimental laboratory or field studies.

Chapter 2, the first substantive chapter, outlines a competition model proposed by the French sociologist Raymond Boudon (1982 [1977]) and presents experimental results. The model explains a long-standing sociological puzzle, namely that an improvement in the availability of opportunities for actors in a social system (e.g. a society or a firm) can coincide with a growing rate of frustrated individuals. For instance, uprisings have repeatedly been preceded by forms of political liberalization that have actually provided greater opportunities. In organizations, satisfaction with regard to promotion opportunities can be negatively associated with objective chances of promotion. We conducted three laboratory experiments to test the model’s predictions. The results are mixed: when opportunities increase, frustration remains constant or increases only slightly. Based on the results, we suggest a modification to the original formulation of the model.

The second substantive chapter, chapter 3, addresses whether a competitive human trait, the desire for social status, can explain cooperative behavior in social dilemmas, as predicted by explanatory models based on signaling theory. For example, when people buy conspicuous environmentally friendly cars for status reasons, at the same time, they contribute to a healthier environment – that is, they cooperate in a social dilemma. My study is an extended replication of Nelissen and Meijers (2011). In a series of experiments they found that people who display luxury goods are perceived as higher status and thus gain advantages in social interaction, which is a motive for buying luxury goods. I test the hypothesis that the public display of costly environmentally friendly goods brings about the same effects as the display of luxury goods – an increase in social status and benefits in social interactions – in the first test of this so-called green signaling hypothesis in
a natural field setting. In contrast to the original study, I did not observe positive effects of either the luxury treatment or the green treatment in any of my five experiments.

The third and fourth substantive chapters are devoted to social norm enforcement – a classic cooperation problem. In laboratory experiments, people are willing to sanction norms at a cost – a tendency called altruistic punishment. However, the degree to which these findings can be generalized to natural interactions is still debated. In the third substantive chapter 4, results from a field experiment of norm enforcement are reported. We investigated the enforcement of the anti-littering norm in Bern, Zurich, and New York City and its sensitivity to cost and benefit manipulations. Findings indicate that norms are universally enforced, although significantly less than in the standard laboratory experiment, and that enforcement is significantly more common in Switzerland than in New York. We also find that enforcement drops as the benefits of norm enforcement decrease or the costs increase. Moreover, individuals prefer more subtle forms of enforcement to direct punishment, likely because they reduce costs and the risk of retaliation.

Chapter 5 is concerned with an alternative explanation of norm enforcement. While the previous chapter builds on a strand of research that ascribes norm enforcement to pro-social preferences, according to the hypothesis under study, self-interest is a sufficient motive for norm enforcement in many instances. Further, individual heterogeneity in the costs and benefits of enforcement raises the likelihood of its occurrence, and individuals with a high net benefit are more likely to enforce norms – a hypothesis that has already been corroborated in laboratory experiments. In order to validate these findings we provide evidence from the field. We let a confederate play loud music on their mobile phone in an open-plan train car and measure the time until a negative sanction occurs (if any). In accordance with our prediction, we find that people sitting near the norm violator, who thus have a greater net benefit from enforcement, are more inclined to do so.

The last substantive chapter, chapter 6, focuses on the individual motives behind cooperative behavior. The evolutionary legacy hypothesis states that the altruistic behavior regularly observed in contemporary, anonymous, one-time-only interactions could be due to evolutionary maladaptation rather than to other-regarding preferences. According to this hypothesis, humans evolved in small groups and virtually all interactions were repeated and non-anonymous. Uncooperative behavior was punished severely, e.g. by ostracism, and it was thus in everybody’s self-interest to cooperate. Human psychology is therefore adapted to non-anonymous, repeated interactions. In an anonymous, one-time-only interaction, people are thus not capable to adjust their behavior and behave as if they were in a repeated, non-anonymous situation. The empirical support for this
claim rests on studies showing that altruism among adults increases in the presence of cues that imply that one’s behavior is observable. Stylized eye spots and faces are prominent examples of the cues researchers have used. Overall, the empirical results are thoroughly mixed. To address this problem, we conducted an eye spot experiment with children. Children are especially informative subjects for this kind of study because altruism develops rapidly during childhood with potential differences between girls and boys. This means that, if altruism in anonymous, one-time-only interactions depends primarily on payoff-irrelevant social cues, then variation in altruism by age or sex should be closely linked to variation in sensitivity to eye spots. Although we did find variation in altruism among children, we found no variation in sensitivity to eye spots. More generally, we found no eye spot effects of any kind. This result contradicts the evolutionary legacy hypothesis.
Kurzfassung

Die vorliegende Dissertation ist eine Kollektion von fünf Studien zu den thematischen Schwerpunkten 'Wettbewerb' (Kapitel 2 und 3) und 'Kooperation' (Kapitel 4-6). Methodische Grundlage aller Beiträge sind ihre spieltheoretische Anbindung und die experimentelle Prüfung sämtlicher Hypothesen entweder im Labor oder im Feld.


Das letzte inhaltliche Kapitel 6 untersucht die individuellen Motive hinter kooperativem Verhalten. Eine einflussreiche Hypothese, die sogenannte evolutionary legacy hypothesis, postuliert, dass altruistisches Verhalten, das regelmässig in nicht wiederholten anonymen Interaktionen auftritt, nicht etwa durch prosoziale Präferenzen, sondern durch einen psychologischen „Bias“ hervorgerufen wird.
Contents

Summary iii

Kurzfassung vii

1 Introduction 1

2 The logic of relative frustration: Boudon’s competition model and experimental evidence 7
   2.1 Introduction ................................................. 8
   2.2 Boudon’s competition model and hypotheses ................. 10
       2.2.1 The model .............................................. 10
       2.2.2 Model implications and hypotheses ................... 13
   2.3 Experimental methods ....................................... 17
   2.4 Relative frustration in the lab ............................. 19
       2.4.1 Test of hypotheses ................................. 19
       2.4.2 Beyond the loser rate: Gini coefficient .......... 21
   2.5 Discussion and conclusion .................................. 23
   2.A Appendix .................................................. 26
       2.A.1 Central statistical analyses .......................... 26
       2.A.2 Instructions ......................................... 27
       2.A.3 Additional information on the model .............. 31
       2.A.4 Additional information on the experimental procedures and design ........................................... 33
       2.A.5 Additional statistical analyses ..................... 35

3 Are luxury brand labels and ‘green’ labels costly signals of social status? 45
   3.1 Introduction ................................................ 46
       3.1.1 Conspicuous consumption as a signal of social status 46
3.1.2 Public generosity and pro-environmental behavior as signals of status and prosociality 46
3.1.3 The present study 48

3.2 Experiment 1 49
3.2.1 Materials and method 49
3.2.2 Results 50

3.3 Experiment 2 51
3.3.1 Materials and method 51
3.3.2 Results 52

3.4 Experiment 3 53
3.4.1 Materials and method 53
3.4.2 Results 54

3.5 Experiment 4 55
3.5.1 Materials and method 55
3.5.2 Results 56

3.6 Experiment 5 57
3.6.1 Materials and method 58
3.6.2 Results 58

3.7 Discussion and conclusion 58

3.A Appendix 62
3.A.1 Shirts and caps used as experimental treatments 62
3.A.2 Regression estimates for experiments 2-5 63
3.A.3 Regression estimates with respect to sex for experiments 1-5 67

4 An international field experiment of norm enforcement in anonymous interactions: How costs and benefits impact altruistic punishment 73
4.1 Introduction 74
4.2 Theoretical background 75
4.3 Experimental design and evidence 79
4.4 Discussion and conclusion 90

4.A Appendix 93
4.A.1 Additional tables 93
4.A.2 Additional figures 97

5 The sanctioning dilemma: A quasi-experiment on social norm enforcement on the train 99
5.1 Introduction 100
5.2 Theory and hypotheses 102
CONTENTS

5.2.1 Diffusion of responsibility in the symmetric volunteer’s dilemma ........................................ 102
5.2.2 Tacit coordination in the asymmetric volunteer’s dilemma .................................................. 104
5.2.3 Decreasing benefits in the volunteer’s timing dilemma ......................................................... 105
5.2.4 Norm salience ................................................................................................................... 106
5.3 Materials and methods ............................................................................................................. 106
5.3.1 Procedures and design ........................................................................................................ 106
5.3.2 Data ................................................................................................................................... 109
5.4 Results ....................................................................................................................................... 110
5.5 Discussion and conclusion ....................................................................................................... 114
5.6 Appendix .................................................................................................................................. 117

6 Eye spots do not increase altruism in children .............................................................................. 119
6.1 Introduction .................................................................................................................................. 121
6.2 Experimental methods ............................................................................................................... 124
6.3 Results ......................................................................................................................................... 127
6.4 Discussion and conclusion ......................................................................................................... 133
6.A Appendix .................................................................................................................................... 136
6.A.1 Experimental methods .......................................................................................................... 136
6.A.2 Data analysis ........................................................................................................................ 137

7 Summary and conclusion .............................................................................................................. 143

References .......................................................................................................................................... 151

Acknowledgments .............................................................................................................................. 171

Curriculum vitae .................................................................................................................................. 173
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Parameter values and model predictions in experiments 1-3</td>
<td>18</td>
</tr>
<tr>
<td>2.A.1</td>
<td>Probability of investing and losing, respectively</td>
<td>26</td>
</tr>
<tr>
<td>2.A.2</td>
<td>Gini coefficients, variances and results from statistical tests</td>
<td>27</td>
</tr>
<tr>
<td>2.A.3</td>
<td>Effect of design factors on individual probability of investing and losing (experiment 1)</td>
<td>36</td>
</tr>
<tr>
<td>2.A.4</td>
<td>Individual satisfaction with outcome of the competition (experiments 1-3)</td>
<td>37</td>
</tr>
<tr>
<td>2.A.5</td>
<td>Satisfaction by competition outcome (loser, sustainer, winner) and positions ($k = 1, k = 2, k = 5$) for all experiments (predictions)</td>
<td>42</td>
</tr>
<tr>
<td>2.A.6</td>
<td>Satisfaction by competition outcome (loser, sustainer, winner) and positions ($k = 1, k = 2, k = 5$) for all experiments: Tests for statistical differences</td>
<td>43</td>
</tr>
<tr>
<td>3.1</td>
<td>Perception ratings (experiment 1)</td>
<td>50</td>
</tr>
<tr>
<td>3.A.1</td>
<td>Regression estimates for experiment 2</td>
<td>63</td>
</tr>
<tr>
<td>3.A.2</td>
<td>Regression estimates for experiment 3</td>
<td>64</td>
</tr>
<tr>
<td>3.A.3</td>
<td>Regression estimates for experiment 4</td>
<td>65</td>
</tr>
<tr>
<td>3.A.4</td>
<td>Regression estimates for experiment 5</td>
<td>66</td>
</tr>
<tr>
<td>3.A.5</td>
<td>Regression estimates with respect to sex for experiment 1</td>
<td>67</td>
</tr>
<tr>
<td>3.A.6</td>
<td>Regression estimates with respect to sex for experiment 2</td>
<td>69</td>
</tr>
<tr>
<td>3.A.7</td>
<td>Regression estimates with respect to sex for experiment 3</td>
<td>70</td>
</tr>
<tr>
<td>3.A.8</td>
<td>Regression estimates with respect to sex for experiment 4</td>
<td>71</td>
</tr>
<tr>
<td>3.A.9</td>
<td>Regression estimates with respect to sex for experiment 5</td>
<td>71</td>
</tr>
<tr>
<td>4.1</td>
<td>Rates of norm enforcement in Bern, New York City, and Zurich.</td>
<td>89</td>
</tr>
<tr>
<td>4.A.1</td>
<td>Logit regression estimates for Bern (coefficients)</td>
<td>93</td>
</tr>
<tr>
<td>4.A.3</td>
<td>Logit regression estimates for Zurich</td>
<td>95</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.A.4</td>
<td>Logit regression, pooling cities</td>
<td>96</td>
</tr>
<tr>
<td>5.1</td>
<td>The volunteer’s dilemma (VOD)</td>
<td>103</td>
</tr>
<tr>
<td>5.2</td>
<td>Summary of study designs</td>
<td>109</td>
</tr>
<tr>
<td>5.3</td>
<td>Descriptive statistics of car- and passenger-level variables</td>
<td>110</td>
</tr>
<tr>
<td>5.4</td>
<td>Regression models</td>
<td>112</td>
</tr>
<tr>
<td>6.1</td>
<td>Ordinal logistic model of transfers</td>
<td>131</td>
</tr>
<tr>
<td>6.2</td>
<td>Logistic model of transfers</td>
<td>132</td>
</tr>
<tr>
<td>6.A.1</td>
<td>The payment scheme used to exchange gold coins from the dictator game into toys.</td>
<td>138</td>
</tr>
<tr>
<td>6.A.2</td>
<td>Model selection, ordered logit models of dictator transfers</td>
<td>139</td>
</tr>
<tr>
<td>6.A.3</td>
<td>Model selection, logit models of dictator transfers treated dichotomously as zero (0) or positive (1).</td>
<td>140</td>
</tr>
</tbody>
</table>
# List of Figures

2.1 Individual decision situation ................................................. 11  
2.2 Game matrix from the perspective of player $i$ ............................... 12  
2.3 $i$’s expected payoff for strategy ‘invest’ ........................................ 14  
2.4 Model predictions .......................................................... 15  
2.5 Predicted and observed rates of investors and losers ......................... 20  
2.6 Predicted and observed path of the variance and the Gini coefficient .......................... 22  
2.A.1 Two extreme cases: Low frustration case and high frustration case 33  
2.A.2 Satisfaction by competition outcome and positions for experiment 1 ........................................ 38  
2.A.3 Satisfaction by competition outcome and positions for experiment 2 ........................................ 39  
2.A.4 Satisfaction by competition outcome and positions for experiment 3 ........................................ 40  
2.A.5 Predicted and observed path of the variance for all experiments 41  
3.1 Relative frequency of compliance by experimental condition in experiment 2 ........................................ 52  
3.2 Average donation by experimental condition (experiment 3) .................. 54  
3.3 Relative frequency of compliance by experimental condition in experiment 3 ........................................ 56  
3.4 Average donation by experimental condition (experiment 5) .................. 57  
3.A.1 Pictures used for perception rating in experiment 1 .......................... 62  
3.A.2 Shirts and caps used in experiments 2-5 ..................................... 62  
4.1 Self-reported reason for not directly and indirectly enforcing by experimental condition ........................................ 82  
4.2 Bern. Relative frequency of total, indirect and direct norm enforcement, dependent on experimental condition ........................................ 84
4.3 New York City. Relative frequency of total, indirect and direct
norm enforcement, dependent on experimental condition . . . 86
4.4 Zurich. Relative frequency of total, indirect and direct norm
enforcement, dependent on experimental condition . . . . . 88
4.A.1 Control condition and littered condition in Bern. . . . . . . 97

5.1 Example of passenger seating in an intercity (IC) train car . . 108
5.2 Proportion of sanctions across car types . . . . . . . . . . 111
5.A.1 Form used to record the setting of an intervention in an IC car . 117

6.1 The distribution of dictator transfers . . . . . . . . . . . . 128
6.A.1 The stimulus used for the control treatment. . . . . . . . . 141
6.A.2 The stimulus used for the eye spot treatment . . . . . . . . 141
6.A.3 The spatial arrangement of the experimental arena . . . . . . 142
Chapter 1

Introduction

This dissertation is a collection of five studies on competition and cooperation – two societal constants. In modern, meritocratic societies, scarce and valuable resources are allocated through competition (Young, 1958).¹ For example, prospective students compete for scarce slots at the best universities, university graduates for lucrative jobs on the labor market and employees for promotion. But competition is not only a core principle of resource allocation in contemporary society; it seems that competitive traits are deeply rooted in human psychology (Griskevičius, Cantù, and Vugt, 2012). That is, humans have a taste for relative status: not only do they want to do well, but they want to keep up with, or even to do better than, their neighbors and colleagues (Frank, 1985). Therefore, not only do people compete when it is in their material self-interest, they are even willing to sacrifice resources in order to outcompete others on the status ladder – a behavioral pattern that manifests itself differently, depending on the societal context. In modern societies people demonstrate status with new shiny cars (Veblen, 1965 [1899]) or an exceptionally fine taste for good wines and exotic holiday destinations (Bourdieu, 1977; Bourdieu, 1984). In tribal societies people demonstrate status with risky hunting maneuvers or by growing the largest yams in the village instead of maximizing their overall harvest (Bliege-Bird and Smith, 2005). Surprisingly, such competitive traits might even foster cooperative behavior, as we will learn later on. Conversely, when it is too fierce, competition can also generate frustration and lead to quarrels and even revolts (Boudon, 1982 [1977]).

Despite of its omnipresence in societal life, competition is just one side of the coin. A society can only work smoothly when its members cooperate – namely,

¹ Here, competition is defined as a social interaction in which several actors wish to obtain a certain good or achieve a certain goal. Yet, these goods or goals are scarce and it is therefore not possible for all actors to obtain a good or reach their goal (Lautmann, 1995).
Introduction

when taxpayers declare their assets honestly, officers resist the temptation of taking a bribe and bystanders intervene when someone is destroying public property (Coleman, 1990; Putnam, 2001). Even markets – arenas of competition – perform better when actors follow certain cooperation norms, such as when they adhere to contracts (Knack and Keefer, 1997) or reciprocate good services with favorable ratings in online auctions (Diekmann, Jann, et al., 2014). Not only competitive, but also cooperative, behavior seems to be a standard feature of human psychology: in no other species than humans have such complex cooperation systems among genetically unrelated individuals evolved (Gil-White and Richerson, 2003). Humans regularly cooperate for the sake of others, or the group as a whole, even when it is costly for them, whether in terms of money, time, or effort (Bowles and Gintis, 2002). Still, cooperation problems are also widespread and can have devastating consequences (Hardin, 1968). For example, not everywhere is tax morale or the willingness to fight corruption equally strong (see Alm and Torgler, 2006 and Mocan, 2008, respectively), and despite the threats of climate change, up until now, it has not been possible to achieve a worldwide reduction in CO2 emissions (Milinski et al., 2006).

Given their significance in social life, it is not surprising that competition and cooperation have been two core themes in the social sciences since their beginning. Early on, social philosophers debated the effects of competition on society. While Adam Smith (2003 [1776]) has pointed out the positive consequences of market competition on the wealth of nations, Karl Marx (1969 [1890]) stressed issues such as social inequality and poverty. Similarly, the question of how stable cooperation could be possible in light of humans’ self-interested nature was reflected on by Hobbes (2003 [1651]). While the early economists since Mill answered this question by referring to the purely self-regarding *homo economicus* (Mill, 1874; see Persky, 1995), early sociologists suggested a converse human model of *homo sociologicus* (Durkheim, 1982 [1895]; Parsons, 1937). Whereas the former model of man describes an actor that cooperates for the greater good if it is in his or her self-interest only, the latter model describes a creature who always cooperates, as long as internalized norms and values require cooperation in a given situation. Both models of men, however, have difficulties in explaining certain phenomena. One the one hand, humans donate blood, anonymously give to charity or jump into dangerous rivers to save lives – phenomena that are hardly in line with the *homo economicus* assumption. On the other hand, widespread cooperation problems such as corruption, tax fraud, or climate change would rarely

---

2 In accordance with Diekmann and Lindenberg (2001, p. 2752) I define cooperation as a behavior in an interaction “that will lead to a collectively rational outcome when other actors behave cooperatively as well”.
occur if human action were largely determined by social norms and moral values (Fehr and Gintis, 2007; Lindenberg, 1990).

In recent decades, the interdisciplinary field of experimental game theory has picked up these basic questions on the individual motives driving competitive and cooperative behavior and their societal consequences (Diekmann, 2008). The modern tools of game theory and experimentation allow the development of more rigorous theoretical explanations than pure verbal description and cleaner empirical tests than most observational data, respectively. Game theory on the one hand is not only a common language for the social sciences that enables categorizing and precisely describing social interaction structures; it also provides tools for model building and deducing new hypotheses (Camerer and Fehr, 2004). Laboratory experiments, on the other hand, allow testing of model predictions with great rigor (Guala, 2005; Morton and Williams, 2010). Differences between behavior predicted by a model and observed in experiments then inspire model adaption (Camerer, 2006; Sugden, 2005), for instance models of men that incorporate features of both of their ancestors, *homo economicus* and *homo sociologicus* (e.g. Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999). Complements to laboratory experiments are experimental and quasi-experimental field studies, which allow validation of results from the laboratory. While the field environment allows less control than a laboratory setting, it is also less artificial. The ‘participants’ do not know that they are taking part in a study and behave in their ‘natural habitat’ as they always do (Levitt and List, 2007).

While the iterative process of rigorous model building and experimental testing – the main characteristic of experimental game theory – is a major driver of scientific progress not only in neighboring disciplines such as economics but also in the natural sciences, both formal modeling and experimental methodology are still rare in sociology (Diekmann, 2008; Willer and Walker, 2007). This dissertation is a small contribution to fill this ‘methodical lacuna’. It contributes to sociological, game theoretically founded or, at least, game theoretically inspired experimental research in competition (chapters 2 and 3) and cooperation (chapters 4 to 6). Chapter 2 presents a competition model that explains a long-standing sociological puzzle, namely that social improvements can coincide with a larger number of frustrated individuals, and puts the model to a first empirical test. In chapter 3 the hypothesis that a competitive human trait, namely the desire to demonstrate social status, can promote environmentally friendly behavior, is tested for the first time by means of a field experiment. Chapters 4 and 5 examine social norm enforcement – a classic cooperation problem – in the field. These studies complement previous research on norm enforcement in laboratory settings. Chapter 6 contributes to the question of the motivation behind coopera-
tive behavior: are humans solely driven by self-interest or also by other-regarding preferences?

In the remainder I give an overview of the five studies that constitute this dissertation, their central question, methodological background, and main results. The first substantive chapter 2 reports on a study that for the first time tests predictions from a competition model outlined by the French sociologist Raymond Boudon (1982 [1977]). The model explains a finding that has puzzled social scientists for a long time, namely that an improvement in the availability of opportunities for actors in a social system (e.g. a society or a firm) can coincide with a growing rate of frustrated individuals. For instance, uprisings have repeatedly been preceded by forms of political liberalization that have actually provided greater opportunities (Brinton, 1965; Tocqueville, 1952 [1856]). In organizations, satisfaction with regard to promotion opportunities can be negatively associated with objective chances of promotion (Stouffer et al., 1965 [1949]). Together with Andreas Diekmann, I conducted three laboratory experiments to test the model’s predictions. The results are mixed: when opportunities increase, frustration remains constant or increases only slightly. Based on the results, we suggest a modification to the original formulation of the model. This study can be seen as typical for research in experimental game theory. The hypotheses are derived from a mathematical model and tested in the laboratory.

Conversely, the study on status signaling, presented in chapter 3, is merely inspired by game theoretic reasoning. This contribution bridges the two poles, competition and cooperation. One of the main questions is whether a competitive human trait, the desire for social status, can explain cooperative behavior in social dilemmas as predicted by explanatory models based on signaling theory (Spence, 1973; Zahavi, 1975; Zahavi, 1977). For example, when people buy conspicuous environmentally friendly cars for status reasons, at the same time, they contribute to a healthier environment, that is, they cooperate in a social dilemma (Whitfield, 2011). My study is an extended replication of Nelissen and Meijers (2011). In a series of experiments they found that people who display luxury goods are perceived as higher status and thus gain advantages in social interaction, which is a motive for buying luxury goods. I also test the hypothesis that the display of costly environmentally friendly goods brings about benefits in social interactions, and thus, an incentive to buy green products – the first test of this so-called green signaling hypothesis in a natural field setting. In contrast to the

---

3 A social dilemma is a situation where, should everybody follow their self-interest, a situation results that is worse than if everybody cooperated (Dawes, 1980). An example is climate change: For most people it is more convenient to drive a car or take a flight than walking or taking a long train ride, respectively – with negative consequences for the climate and thus for everyone.
original study, I did not observe positive effects of either the luxury treatment or the green treatment in any of five experiments.

The two following chapters examine classic cooperation problems. Both studies are on social norm enforcement, which is, in essence, a social dilemma. Everybody profits from social norms and their enforcement, but everyone also has an incentive to leave enforcement to anyone else. For instance, most people can hardly relax or work in a silence compartment of a train, when someone is playing punk rock at a high volume nearby. Still, it is tempting to leave the awkward and probably even risky job of enforcing the silence norm and asking the norm violator to turn down the music to someone else. In laboratory experiments, people are willing to sanction norms at a cost – a behavioral tendency called altruistic punishment (Fehr and Gächter, 2002). However, the degree to which these findings can be generalized to natural interactions is still debated (Guala, 2012). In chapter 4, results from a field experiment of norm enforcement are reported. Together with Debra Hevenstone, I investigated the enforcement of the anti-littering norm in Bern, Zurich, and New York City and its sensitivity to cost and benefit manipulations. Findings indicate that norms are universally enforced, although significantly less than in a standard laboratory experiment, and that enforcement is significantly more common in Switzerland than in New York. We also find that enforcement drops as the benefits of norm enforcement decrease or the costs increase. Moreover, individuals prefer more subtle forms of enforcement to direct punishment, likely because they reduce costs and the risk of retaliation.

In a fourth study, outlined in chapter 5, together with Wojtek Przepiorka, I examine a new explanation of norm enforcement. While the previous chapter builds on a strand of research that ascribes norm enforcement to pro-social preferences, according to the hypothesis under study, self-interest is a sufficient motive for norm enforcement in many instances. Further, individual heterogeneity in the costs and benefits of enforcement raises the likelihood of its occurrence, and individuals with a high net benefit are more likely to enforce a norm – a hypothesis that has already been corroborated in a laboratory experiment (Diekmann and Przepiorka, 2015; Przepiorka and Diekmann, 2013). In order to provide complementary evidence from the field, we let a confederate play loud music on their mobile phone in an open-plan train car and measure the time until a negative sanction occurs (if any). In accordance with our prediction, we find that people sitting near the norm violator, who thus have a greater benefit from enforcement, are more inclined to do so.

The last study, reported on in chapter 6, is joint work with Sonja Vogt, Charles Efferson and Ernst Fehr. This contribution focuses on the individual motives behind cooperative behavior. It is still debated whether humans only cooperate when it is in their self-interest or whether they are additionally motivated by
other-regarding preferences. A test bed for this question is the dictator game, a simple sharing task. One person, the sender, divides a sum, usually money, between himself/herself and a receiver. The sender and receiver are mutually anonymous and will never knowingly interact again. The receiver can thus never reward the sender’s generosity or punish stinginess. In principle, self-interest can therefore be excluded as a motive for the sender to donate in the dictator game (Camerer, 2006). However, the evolutionary legacy hypothesis posits that the altruistic behavior regularly observed in the dictator game could be due to a psychological bias rather than to other-regarding preferences. According to this hypothesis, humans evolved in small groups and virtually all interactions were repeated and non-anonymous. Uncooperative behavior was punished severely, e.g. by ostracism, and it was thus in everybody’s self-interest to cooperate. Human psychology is therefore adapted to non-anonymous, repeated interactions. In an anonymous, one-time-only interaction, people are thus not able to adjust their behavior and behave as if they were in a repeated, non-anonymous situation (Haley and Fessler, 2005). The empirical support for this claim rests on studies showing that altruism among adults increases in the presence of cues that imply that one’s behavior is observable. Stylized eye spots and faces are prominent examples of the cues researchers have used. As meta-analyses show, overall the empirical results are thoroughly mixed (Nettle et al., 2013; Sparks and Barclay, 2013). To address this problem, we conducted an eye spot experiment with children. Children are especially informative subjects for this kind of study because altruism develops rapidly during childhood with potential differences between girls and boys. This means that, if altruism in anonymous one-time only interactions depends primarily on payoff-irrelevant social cues, then variation in altruism by age or sex should be closely linked to variation in sensitivity to eye spots. Although we did find variation in altruism among children, we found no variation in sensitivity to eye spots. More generally, we found no eye spot effects of any kind. This result contradicts the evolutionary legacy hypothesis.

Chapter 7 then summarizes the main findings and presents some concluding remarks.
Chapter 2

The logic of relative frustration: Boudon’s competition model and experimental evidence

Abstract An improvement in the availability of opportunities for actors in a social system (e.g. a society or a firm) can coincide with a growing rate of frustrated individuals. For instance, uprisings have repeatedly been preceded by forms of political liberalisation that have actually provided greater opportunities (the so-called Tocqueville paradox). In organisations, satisfaction with regard to promotion opportunities can be negatively associated with objective chances of promotion. Raymond Boudon has proposed a game-theoretic competition model which specifies the micro-mechanisms that produce these puzzling phenomena at the aggregate level, and clarifies the conditions under which they emerge. We conducted three laboratory experiments to test the model’s predictions, making our study the first empirical test of Boudon’s model. The results are mixed: when opportunities increased, the rate of the relatively frustrated losers in the group remained constant, or increased only slightly. However, when applying another aggregation rule, which accounts for all social comparison processes and does not merely focus on the losers an increase in relative frustration under improved conditions was observed. Our results imply that under specific conditions there is a trade-off between opportunities and social mobility, on the one hand, and social inequality and relative frustration, on the other.

---

1 This chapter is an edited preprint version of the following paper: Berger, Joël and Diekmann, Andreas (2015). “The logic of relative frustration: Boudon’s competition model and experimental evidence.” European Sociological Review. Online first.
2.1 Introduction

Surprisingly, an improvement in a society’s opportunities can coincide with a larger share of frustrated individuals. The first to discuss this was Tocqueville, who claimed that the outbreak of the French Revolution was triggered by economic and social improvements (Tocqueville, 1952 [1856]). The factual truth of his historical account is not an issue here, but his thesis is of general interest in the social sciences. The phenomenon that political reforms or economic improvements can lead to an increase in the rate of the frustrated and, with a certain probability, to uprisings is called Tocqueville’s paradox (Neckel, 2010). Brinton (1965) and Coleman (1990) discuss several historical examples of Tocqueville’s paradox. A recent example is China, where rapid economic growth coincides with growing dissatisfaction (Brockmann et al., 2009; Ishida, Kosaka, and Hamada, 2014); as a consequence, Tocqueville’s oeuvre has become popular among the party elite (Pei, 2013).

Related to the Tocqueville paradox is Durkheim’s observation of rising suicide rates during rapid economic growth (Durkheim, 1952 [1897]). The effect of more overall frustration under improved circumstances does not only appear on the societal level but also within organisations. In their study on social mobility in the US army, Stouffer et al. (1965 [1949]) report the classic finding that satisfaction regarding promotion chances was lower in branches with high objective opportunities. Related and no less interesting examples are discussed by Gladwell (2013).

There are two widespread explanations for the phenomenon of greater aggregate frustration under improved conditions (Esser, 2001). The first explanation assumes frustration to arise from intra-individual comparisons of an individual’s expectations regarding a situation with the actual state of affairs. For instance, if, due to social improvements, expectations grow faster than the actual improvement, frustration may arise from this gap (e.g. Brinton 1965; Durkheim 1952 [1897]). The second approach derives frustration from inter-individual comparison. Stouffer et al. (1965 [1949]), as an example, supposed soldiers to evaluate their situation in relation to that of a reference group. A non-promoted soldier among a majority of non-promoted soldiers felt less dissatisfied than a non-promoted individual among a group with a large share of promoted.

In the following, for frustration arising from comparison with either of the reference points (i.e. an individual’s expectations or an individual’s reference
group), we speak of relative frustration. We use the terms intra-individual and inter-individual relative frustration when referring to either case separately. To avoid confusion, we will not use the term relative deprivation, which usually refers to frustration arising from inter-individual comparison (Runciman, 1966).

Based on verbally formulated relative frustration theory (e.g. Merton and Rossi 1957; Runciman 1966), attempts to quantify the theory were made. Several authors suggested the Gini coefficient or a derivation thereof as a measure of inter-individual relative frustration at the aggregate level (e.g. Kakwani 1984; Yitzhaki 1979; see D’Ambrosio and Frick 2007 for a brief literature review and an empirical application). Boudon (1982 [1977], ch. 5),\(^3\) going beyond a mere measure of relative frustration, introduced a game-theoretic model, explaining the puzzle of more frustration under improving (or better) conditions as the consequence of rational individual decisions in interdependent competition situations.\(^4\) As the opportunities in a social system improve, competition becomes fiercer, leading to more overall frustration.\(^5\) At a certain point, the direction of the correlation changes and more opportunities go along with a decreased frustration level. In short, the model predicts an inversely U-shaped path of aggregate relative frustration when opportunities increase. This, however, only happens under specific conditions. In other regions of the parameter space, a negative association between opportunities and relative frustration is possible, too.

While Boudon outlined the model in a rather sketchy manner, Raub (1984) formalised it rigorously and analysed it game theoretically. Raub demonstrated the core prediction of an inversely U-shaped rate of relative frustration to hold under different behavioural assumptions and he specified the boundary conditions of this phenomenon precisely. Kosaka (1986) relaxed the model assumption of homogeneous actors and found that social inequality between competitors dampens the effect of more frustration under improving conditions to some degree. Yamaguchi (1998) demonstrated that this only holds true when contest mobility exceeds sponsored mobility (as defined by Turner 1960), while the opposite is the case when sponsored mobility prevails. Manzo (2009) implemented the competition model as an agent based simulation. With this different approach, he replicated his forerunners’ findings that the effect of greater frustration under improved conditions only emerges in certain regions of the parameter space.

---

\(^3\) For reasons of simplicity, we use the most parsimonious version of the model, where actors decide between investing and not investing. In Boudon’s notation: \(B2 = C2 = 0\) (Boudon, 1982 [1977], pp. 116–117).

\(^4\) Alternative explanations and models can be found in Davies (1962), Davis (1959), Elster (1991), Gambetta (2005), and Hirschman and Rothschild (2011).

\(^5\) The model explains variation in relative frustration, with variation in the opportunities of a social system. Whether opportunities differ due to changes over time or are cross-sectional is irrelevant.
Moreover, Manzo (2011) showed that network topology might impact both the spread and intensity of relative frustration. Specifically, social comparison in local networks as compared to global comparison fosters relative frustration.

While the competition model has been analysed and developed theoretically, to our knowledge there has never been an empirical test of the model’s central implications. To fill this gap, we conducted three laboratory experiments testing the central hypotheses and an inversely U-shaped path of relative frustration under improving conditions.

2.2 Boudon’s competition model and hypotheses

2.2.1 The model

We start with $N$ actors facing a decision on whether or not to invest resources $C$ such as time, effort, or money in a competition for a scarce and highly valued good, for instance a high-prestige position within a firm. There are $k$ lots (e.g. positions) and $n$ investors (competitors). Because the good is scarce, it must hold that $k < N$. While $k$ is common knowledge, the number of investors can range from 0 to $N$. Successful investors get access to the scarce good (e.g. promotion) and therefore receive the high payoff $\alpha$, which is given by gross benefit $B$ (e.g. prestige, power or money) minus investment costs $C$ (see figure 2.1; see also Hedström 2005, p. 57).

Since the probability of success depends on the total number of investors, the investment decision is strategic in nature. If the number of investors $n$ exceeds the amount of free positions $k$, some investors necessarily fail to obtain a desired position. The losers are relatively frustrated due to both intra- and inter-individual comparison: First, their expectation of getting promoted by investing is disappointed. Second, the losers adopt the successful investors as their reference group: While for the same stake $C$ the successful *winners* receive the high payoff $\alpha$, the *losers* receive nothing in return. This is represented by the low payoff $\gamma$. *Sustainers* (non-investors), e.g. individuals deciding not to spend time and money training in order to qualify for promotion, will neither get promoted nor

---

6 See also Appendix 2.A, paragraph 2.A.5 on intensity of relative frustration.
7 The two extreme cases are trivial: If $k = 0$, there is no competition at all; ‘not invest’ is the dominant strategy. If $k = N$, every investor wins by definition and, again, there is no competition.
Boudon's competition model and hypotheses

Figure 2.1: Individual decision situation: The strategy ‘invest’ is risky and leads to the high payoff \( \alpha \) with probability \( s \) and to the low payoff \( \gamma \) with probability \( 1 - s \). The strategy ‘not invest’ leads to the medium payoff \( \beta \) with certainty.

will they lose any resources. They consequently end up with the medium payoff \( \beta \). Formally, the payoffs satisfy the inequalities \( \alpha > \beta > \gamma \).

Individual investment decisions depend on the opportunities provided by a given competition, which is determined by the number of scarce positions \( k \) and by the cost-benefit ratio \( Q \):

\[
Q = \frac{(B - \beta)}{C}.
\]

\( Q \) expresses how lucrative the winners’ payoff \( \alpha \) is in comparison to the sustainers’ payoff \( \beta \), considering the investment costs \( C \).

The probability of success \( s \) for a player choosing the strategy ‘invest’ is given by the ratio of the number of scarce positions \( k \) to the number of investors \( n \). This is provided that the number of investors exceeds the number of positions and 1, otherwise:

\[
s = \begin{cases} 
\frac{k}{n} & \text{for } k < n \\
1 & \text{for } k \geq n.
\end{cases}
\]

Note that all \( N \) players decide simultaneously whether or not to invest. Hence, before the decisions are made, the actual number of investors \( n \) is unknown, while

---

\( ^8 \) We tested the micro assumption that the losers are frustrated, i.e. less satisfied than the winners and the non-investors. The results suggest that this assumption holds. More specifically, the winners report the highest and the losers the lowest satisfaction, with the non-investors in between (see Appendix 2.A, table 2.A.4).
The logic of relative frustration

Figure 2.2: Game matrix from the perspective of player $i$. The expected payoff $E(k, n)$ is determined by equation (2.1).

the number of positions $k$ is common knowledge. The higher the number of investors, the lower the chances of success for each player.

Given $\alpha$, $\gamma$, and $k$ the expected payoff $E(k, n)$ for a specific number of investors $n$ is provided by:

$$E(k, n) = \begin{cases} k\alpha + \frac{n-k}{n}\gamma & \text{for } k < n \\ \alpha & \text{for } k \geq n. \end{cases} \quad (2.1)$$

With this information, the game matrix from the perspective of any player $i$ can be constructed (see figure 2.2). In the matrix, the expected payoff of the strategy ‘invest’ for each possible number of player $i$’s competitors $n - 1$ is depicted:

The model assumptions are in accordance with classical game theory. The rules of the game are common knowledge and each player knows that all players are fully rational and maximise their expected utility.

So, what will a rational actor do in this situation? If the expected payoff of the strategy ‘invest’ $E(k, n)$ exceeds the sustainers’ payoff $\beta$ independently of the actual number of investors, ‘invest’ is the dominant strategy, which means that all players will invest.\(^9\) For the case that no dominant strategy exists, there is a threshold $n^*$ with the property that, as long as maximally $n^*$ players choose to invest, the expectation of an investment exceeds the payoff $\beta$. So there are $\binom{n}{n^*}$ asymmetrical Nash equilibria in pure strategies, in which $n^*$ players choose ‘invest’ and $N - n^*$ players choose ‘not invest’. Nevertheless, since homogeneous actors are assumed and communication is not possible, none of these equilibria is likely to be realised.

Rather, the rational solution lies in mixed strategies, with an optimal investment probability $p^*$. The mixed strategy solution of this game is in accordance with the rationality theory developed by Harsanyi and Selten (1988). By the axioms of the Harsanyi-Selten theory, in a symmetric game the solution must be a

\(^9\) For all of his co-players’ strategy choices, a players’ strictly dominant strategy leads to a higher pay-off than any other strategy available to the player (e.g. Rasmusen, 2001).
symmetric Nash equilibrium. The mixed strategy equilibrium is the unique symmetric Nash equilibrium of the game and, therefore, it is the rational solution according to the Harsanyi-Selten theory.

Of course, there are other decision principles, with a particularly simple one being the maximin strategy. Here, the maximin strategy is ‘not invest’ yielding payoff $\beta$; the same payoff a player expects when he employs the mixed equilibrium strategy. However, maximin is not an equilibrium strategy. If other players choose maximin, it is advantageous to deviate and to choose the strategy ‘invest’.

Both the maximin strategy and the mixed equilibrium strategy give rise to an efficiency problem. In a cooperative game with binding agreements actors would agree on an asymmetric equilibrium with exactly $k$ investors who are, for example, determined by the drawing of lots. Then, investors receive $\alpha$ and non-investors receive $\beta$. This is the Pareto optimal payoff vector with the welfare gain: $k\alpha + (N - k)\beta - N\beta > 0$. However, without the possibility to agree on a binding contract the Pareto optimal solution is not attainable.

The counterintuitive phenomenon that additional positions lead to an increase in relative frustration occurs if a small increase in $k$ tempts an overly large number of players to choose to invest and, as a result, the increase in the number of losers exceeds the number of additional positions substantially. Roughly speaking, this is the case if the benefit of the strategy ‘invest’ is significantly superior to the one of ‘not invest’ while the costs of investing are rather low; that is, if $Q$ is sufficiently high.

### 2.2.2 Model implications and hypotheses

We illustrate the discussed mechanisms using two numerical examples. Both of these examples have been implemented experimentally to test model predictions.

**Mixed strategy case** Let there be a social system of $N = 6$ players choosing whether or not to compete for one of $k$ scarce positions. Sustainers receive a medium-level payoff $\beta = 6.5$ with certainty. Successful investors receive the high payoff $\alpha = 9$, which is given by gross benefit $B = 10.5$ minus investment costs $C = 1.5$. Losers invest their resources in vain and end up with the low payoff, in this case $\gamma = \beta - C = 5$. The number of positions $k$ varies from 1 to 5. In figure 2.3, the game matrix for these parameter values and the case of 1 position is displayed. The payoff of the strategy ‘not invest’ is fixed at 6.5, while for a given player $i$ the expectation regarding ‘invest’ depends on the total number of her competitors $n - 1$ and can be calculated using equation (2.1). If player $i$ is the only investor, she will get 9 for sure. Of course, the more players enter the competition, the smaller is the expectation of the strategy ‘invest’ for any of the
The logic of relative frustration

Figure 2.3: Game matrix from the perspective of player \( i \). The expected payoff \( E(k, n) \) for the strategy ‘invest’, given the number of player \( i \)’s competitors \( (n-1) \), is determined by equation (2.1). Parameter values: \( k = 1, \alpha = 9, \gamma = 5, N = 6 \). The payoff \( \beta \) for ‘not invest’ is fixed at 6.5.

<table>
<thead>
<tr>
<th>Player ( i )</th>
<th>Invest</th>
<th>Not invest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( 0 )</td>
<td>( 1 )</td>
</tr>
<tr>
<td>Invest</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Not invest</td>
<td>6.5</td>
<td>6.5</td>
</tr>
</tbody>
</table>

Evidently, ‘invest’ is not a dominant strategy: As soon as two or more other players than \( i \) choose to invest, the expected payoff of ‘invest’ is worse than the payoff of the strategy ‘not invest’. Hence, a rational actor will apply a mixed strategy.

The optimal mixture of both strategies can be gained by exploiting the principle of indifference. If all players choose to invest with the optimal probability \( p^* \), all players are indifferent between their pure strategies. This results in a Nash equilibrium in mixed strategies (Nash, 1950). To derive \( p^* \), the overall expected payoff of the strategy ‘invest’ \( E(k, \bullet) \) for a given value of \( k \) and all possible permutations of investors is equated with the payoff of the strategy ‘not invest’ \( \beta \) (see equation (2.2)). Solving for \( p \) yields the optimal investment probability \( p^* \) (Raub, 1984).

\[
E(k, \bullet) = \sum_{n=1}^{N} \left( \frac{N - 1}{n - 1} \right) p^{n-1} (1 - p)^{N-n} \cdot E(k, n) = \beta \tag{2.2}
\]

Given one position, for the parameter values specified above, the rational solution is to invest with a probability of \( p^* = .429 \) and to sustain with \( 1 - p^* = .571 \). The expected relative frequency of investors then equals \( p^* \).\(^{10} \) Provided there are more investors than positions, the rate of winners in a social system of \( N \) is given by the ratio of free positions to the total number of individuals \( (k/N) \), and the loser rate is determined by subtracting the winner rate from the investor rate \( (p^* - k/N) \). Should the share of investors undershoot the share of positions, every investor wins and there are no losers. Importantly, Boudon defines the rate of relative frustration in a social system as the loser rate, abstracting from the non-investors and winners.

\(^{10}\) For a detailed illustration of how to derive \( p^* \), see Appendix 2.A, paragraph 2.A.3.
For the discussed parameter constellation, the expected rates of investors, winners and losers as a function of $k$ are plotted in figure 2.4a. If $k = 1$, as derived from equation (2.2), 42.9% of all players are expected to choose the strategy ‘invest’. Since there are $(k/N) 100 = 16.7\%$ winners, there will be $42.9\% - 16.7\% = 26.2\%$ losers. When $k$ is doubled from 1 to 2 positions, the investor rate grows sharply by 46.0 percentage points: the increase in positions leads to an overly large increase in investors and thus to fiercer competition. As a result, the rate of frustrated losers increases by 29.3 percentage points. This increase is sharper than the increase in the winner rate of 16.7 percentage points. Obviously, there are more additional losers than additional winners, even though the number of positions has doubled from one to two.

If $k$ keeps rising, the loser rate starts decreasing: As soon as 100% of the population is competing, additional positions can only diminish the rate of the relatively frustrated.

In short, the frustration rate follows an inversely U-shaped trajectory with a maximum at $k = 2$. We call the discussed situation the *mixed strategy case*.
because the paradoxical effect occurs at $k = 2$, where an equilibrium in mixed strategies exists.

**Dominant strategy case**  In our second example, the winners’ payoff $\alpha$ is enhanced from 9 to 10, such that ‘invest’ is the dominant strategy if there are two or more positions. In this case, weaker rationality assumptions – namely no assumption of actors playing mixed strategies – are required in order to deduce the hypothesis of an inversely U-shaped trajectory of relative frustration. Let us call this situation the *dominant strategy case*.

As depicted in figure 2.4b, if there is only one prestigious position 55.1% of all players are predicted to choose ‘invest’, assuming players apply mixed strategies. Whichever decision rule actors actually follow in the case of one position, if there are two positions ‘invest’ becomes the dominant strategy. This means that for every player it is best to invest, independently of what the other players do. Even if all competitors of player $i$ choose to invest, $i$’s expected gain of ‘invest’ is still greater than the payoff of ‘not invest’. Hence, all 6 players will compete for only two positions and, as a consequence, the level of frustrated losers reaches its maximum (100% - 33.3% = 66.7% compared to 38.4% in the low-mobility system with 1 position). If $k$ keeps rising, more and more investors are promoted, while there are no additional losers. Again, the rate of relative frustration follows an inversely U-shaped functional form with its maximum at $k = 2$.

Importantly, the model highlights that the increase in relative frustration under improved chances of upward mobility does not necessarily occur. For instance, if ‘invest’ is the dominant strategy and all players invest anyway, additional positions $k$ coincide with fewer frustrated losers and more satisfied winners. Furthermore, if the expected payoff of an investment is not too tempting, additional positions lead to an according investment rate, and relative frustration remains constant (see figure 2.A.1 in the appendix). Hence, the model implies that, at the aggregate level, there is no general law connecting opportunities to frustration rates. Depending on the parameter constellation, social improvements can lead to a higher as well as a lower frustration rate or leave it unchanged.

In our experiments, we chose parameters in a way that maximises investment, and, thus, frustration when opportunities increase (for details, see Appendix 2.A.3), in order to test the main model prediction of an inversely U-shaped rate of relative frustration. More specifically, we aimed at testing the following hypotheses:

**H1:** The higher the number of positions, the higher the number of investors according to the predictions of the game-theoretic model.
**H2: The rate of the frustrated losers is an inversely U-shaped function of the number of positions, given the parameter values specified.**

### 2.3 Experimental methods

We conducted a series of three laboratory experiments. In the first experiment, we implemented the parameters of the mixed strategy case (the first of the two numerical examples discussed in the preceding session) in a within-subjects design.\(^{11}\) Since the results deviated somewhat from predictions (see below), we wanted to minimise the chances that the design was responsible for this. Specifically, giving feedback after each round of the experiment might have affected the subjects’ investing behaviour. In experiment 2, in order to exclude this we replicated the experiment in a between-subjects design. Additionally, we projected the payoffs into the positive domain to eliminate loss-aversion (Kahneman and Tversky, 1979). The third experiment is a replica of the second case discussed in the preceding section; that is, the dominant strategy case. In this case, weaker rationality assumptions are necessary in order to derive the inversely U-shaped path of relative frustration from the model.

In experiment 1, we compared competitions with one position (*low-mobility condition*), two positions (*medium-mobility condition*), and five positions (*high-mobility condition*) in order to test the hypothesis of an inversely U-shaped rate of relative frustration. In experiments 2 and 3, we restricted ourselves to the comparison of low-mobility systems to medium-mobility systems and omitted the high-mobility system so as to focus on the effect of an increase in the rate of relative frustration under improved conditions. An overview of all parameter values and predictions is given in table 2.1.

72 students participated in experiment 1, and 60 each in experiments 2 and 3. In each experiment, subjects were randomly assigned to groups of six; separate competitions were held for each group. When all the individuals of a given group had decided whether or not to invest, \(k\) winners were chosen randomly from among the investors of the corresponding group. In the first experiment, the participants played the competition game for 6 rounds, while only one round was played in experiments 2 and 3.\(^{12}\)

\(^{11}\) As is common in experiments with monetary incentives we assume that, approximately, individuals’ utility increases linearly with money in the payoff range of the experiment.

\(^{12}\) See Appendix 2.A for a detailed description of the experimental design and procedures.
Table 2.1: Parameter values and model predictions in experiments 1-3

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameters</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winners’ payoff $\alpha$</td>
<td>7</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Sustainers’ payoff $\beta$</td>
<td>1</td>
<td>6.5</td>
<td>6.5</td>
</tr>
<tr>
<td>Losers’ payoff $\gamma$</td>
<td>-3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Number of positions $k$</td>
<td>1, 2, 5</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>Group size</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><strong>Predictions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investors (%)</td>
<td>k = 1</td>
<td>k = 1</td>
<td>k = 1</td>
</tr>
<tr>
<td></td>
<td>k = 2</td>
<td>k = 2</td>
<td>k = 2</td>
</tr>
<tr>
<td></td>
<td>k = 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losers (%)</td>
<td>39.7</td>
<td>42.9</td>
<td>55.1</td>
</tr>
<tr>
<td></td>
<td>83.3</td>
<td>88.9</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.0</td>
<td>26.2</td>
<td>38.4</td>
</tr>
<tr>
<td></td>
<td>50.0</td>
<td>55.5</td>
<td>66.7</td>
</tr>
<tr>
<td></td>
<td>16.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: payoffs in CHF.
2.4 Relative frustration in the lab

2.4.1 Test of hypotheses

For all three experiments, we report descriptive statistics that are also visualised in figure 2.5. Tests for statistical significance were conducted by means of logit models. The corresponding estimations can be found in table 2.A.1 in the appendix.

In the first experiment, as expected, subjects invested with a higher probability as the number of positions was increased (figure 2.5a). The investor rate rose from 36.1% in the low-mobility condition \((k = 1)\), to 54.9% in the medium-mobility condition \((k = 2)\), and reached 90.3% in the high-mobility condition \((k = 5)\). Both the investor rate of the low-mobility condition and the investor rate of the high-mobility condition differ significantly from the reference category \((k = 2)\) at least at the 1% significance level. Nevertheless, while predicted and observed values correspond quite well in the cases of one (39.7% predicted, 36.1% observed) and five positions (100% predicted, 90.3% observed), there were notably fewer investors than predicted in the medium-mobility system (83.3% predicted, 54.9% observed). Consequently, as depicted in figure 2.5b, the rate of the relatively frustrated losers remained roughly constant as the number of positions was doubled from one position (20.8% losers) to two positions (22.9% losers, \(p = .748\)). Finally, and in line with the model, the frustration level of 10.4% was lower in the high-mobility system in comparison to the medium-mobility system \((p = .025)\). In short, no inversely U-shaped rate of relative frustration was observed. Rather, the loser rate remained constant when the number of positions was doubled and then decreased in the high-mobility condition.

In the second experiment, in the low-mobility condition 50.0% of the subjects chose to invest, while 80.0% entered the competition in the medium-mobility condition. This difference is statistically significant \((p = .007)\) and the observed values approximate the predictions of 42.9% \((k = 1)\) and 88.9% \((k = 2)\) quite well (see figure 2.5c). Furthermore, in correspondence with the model, the loser rate of 46.7% in the medium-mobility condition exceeds the loser rate in the low-mobility condition (33.3%, see figure 2.5d). However, since more losers than expected were generated in the low-mobility and fewer in the medium-mobility condition, the difference in the loser rate amounts only to 13.3 percentage points, while the model predicts a 29.3 percentage point difference. Because the difference is not even half as large as predicted, it does not reach statistical significance.

In the third experiment, ‘invest’ is the dominant strategy in the medium-mobility system and therefore all six subjects of a given group were predicted to compete for only two positions. Despite this, as depicted in figure 2.5e, only
73.3% of all participants entered the competition. Due to this discrepancy, the loser rates did not differ much between the low-mobility condition (36.7%, predicted: 38.4%) and the high-mobility condition (40.0%, predicted: 66.7%, see figure 2.5f, \( p = .731 \)).

To sum up, in all three experiments investment increased with the number of positions, which means that hypothesis 1 is supported from a qualitative point of view. Nevertheless, while investment behaviour corresponds neatly with the model at the extreme points of one and five positions, participants invested more cautiously than expected in the medium-mobility condition with two positions. Consequently, frustration remained constant when the number of positions was doubled from one to two (experiments 1 and 3) or increased only slightly (experiment 2). Hypothesis 2 is thus rejected: The rate of the frustrated losers does not follow an inversely U-shaped trajectory but remains constant as positions are enhanced on a low level and decreases in the high-mobility system – a result that lies between model predictions and the naive view of decreasing frustration under improving conditions.
2.4.2 Beyond the loser rate: Gini coefficient

Drawing on the literature on relative frustration discussed in the introduction, we questioned Boudon’s narrow focus on the loser rate. While we agree with Boudon that the losers are the only ones being frustrated due to disappointed expectations, the sustainers might still be relatively frustrated due to a comparison with the winners. Even Boudon acknowledged that even though the sustainers might not be as frustrated as the losers, they most probably are not as satisfied as the winners either (Boudon, 1982 [1977], p. 115).

As a measure for relative frustration in a group, the Gini coefficient was suggested (Kakwani, 1984). The Gini accounts for all the three types of actors (i.e. losers, winners and sustainers) as well as for the different degree of relative frustration perceived by losers and non-investors by summing up the absolute differences in the payoffs in a group. That is:

\[
\sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij},
\]

where \(d_{ij}\) is the absolute value of the difference in the payoffs between actors \(i\) and \(j\). This summation procedure yields, after proper normalisation, the desired measure.\(^{14}\)

We computed the Gini coefficient predicted by the model and the corresponding empirical values for each experimental condition of experiments 2 and 3 (see figure 2.6 and figure 2.A.5, Appendix 2.A). There are problems with the interpretation of the Gini coefficient in the presence of negative values, however. In order to resolve this, in experiment 1 we calculated the variance as a substitute.\(^{15}\)

Let us look at the predicted Gini coefficients and variances. In the beginning \((k = 1)\) there are few winners and few losers, and the sustainers prevail – the group composition is largely equal. When positions increase sharply on a low level \((k = 2)\), the formerly equal sustainers are divided into winners and losers. The dispersion in the payoffs and therefore the number of comparison processes increases. Thereafter, the sustainers vanish, while the losers are diminished. Hence, after a peak, the Gini coefficient declines with increasing positions.

\(^{13}\) See also footnote 4.

\(^{14}\) The Gini coefficient is defined as: \(\frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij}\), where \(d_{ij}\) is the absolute value of the difference between person \(i\) and \(j\). The Gini is 0 when all payoffs are equal and approximates 1 with increasing inequality. A meaningful interpretation of the Gini coefficient requires a ratio scale with a meaningful zero value as with monetary payoffs.

\(^{15}\) See table 2.A.2 and paragraph 2.A.5 in the appendix for further details of our analyses of Gini coefficients and the variance.
Figure 2.6: Predicted and observed path of the variance (experiment 1) and the Gini coefficient (experiments 2 and 3).

Not only the predicted, but also the empirical path of the Gini coefficient and the variance is inversely U-shaped – at least in tendency. In experiment 1, the variance increases significantly \((p = .024)\) from 9.18 (predicted: 10.44) to 13.05 (predicted: 21.11) and finally decreases to 11.44 (predicted: 14.12), although not statistically significantly \((p = .593)\). Note that even though the increase in the variance from the low-mobility to the medium-mobility condition is significant, this increase is not as steep as predicted. In experiment 2, the Gini rises significantly \((p = .006)\) from .106 (predicted: .098) in the low-mobility condition to .140 (predicted: .142) in the medium-mobility condition. Similar results are observed in experiment 3, where the Gini rises from .128 (predicted: .129) to .161 (predicted: .167) when the positions are doubled \((p = .025)\).

In all three experiments, the measures of dispersion increase when the number of positions is doubled and, in tendency, decrease thereafter. It is worth mention-

---

\(^{16}\) Here we report results from the high stake condition only. In the low stake condition, the decrease in the variance from the high to the low stake condition was more pronounced \((p = .102)\). Note that splitting the sample into a high and a low stake condition implies a decrease in statistical power.
ing that this pattern is in line with the *Kuznets curve*. Kuznets (1955) argued that in the course of the economic growth occurring during a society’s transition from rural to industrial, the trend of income inequality is inversely U-shaped.

Taken together, our findings suggest that Boudon’s explanation of the puzzling phenomena reported by Tocqueville, Durkheim and Stouffer et al. might need revision. While it is true that additional opportunities tempt an increasing share of actors to compete, in none of the three experiments was this increase as sharp as predicted. Consequently, relative frustration, defined as the loser rate, remained constant or increased only slightly when conditions improved on a low level. However, when not merely focusing on the losers and accounting for all comparison processes between losers, sustainers and winners, an increase in relative frustration is observed.

### 2.5 Discussion and conclusion

An improvement in the opportunities offered within a society (Tocqueville, 1952 [1856]) or an organisation (Stouffer et al., 1965 [1949]) can lead to an increase in frustration at the aggregate level. Boudon (1982 [1977]) suggests a game-theoretic model explaining these counterintuitive effects as the unintended consequence of strategic individual decisions. As in a social system the chances of getting access to a scarce and lucrative good (e.g. a prestigious position within a firm) increase, then under specific conditions the additional investors considerably outnumber the additional positions. As a consequence of this fierce competition, the rate of the relatively frustrated losers increases and Tocqueville’s paradox emerges. When the number of positions is further enhanced, more and more investors achieve promotion and aggregate frustration diminishes again. In short, the model predicts an inversely U-shaped association between opportunities and relative frustration.

To test the main model predictions, we conducted three laboratory experiments with different parameter values – the first empirical test, as far as we are aware. In accordance with the model, in all three experiments participants invested with a higher probability as the number of opportunities was enhanced. However, especially in the medium-mobility system, where we expected the rate of the frustrated losers to peak, participants invested more cautiously than predicted. As a consequence, the loser rate remained constant. When opportunities further improved, as expected, the loser rate decreased again. This result lies in between model predictions of an inversely U-shaped rate of frustration and the intuitive belief of a decreasing frustration level under improving opportunities.
Post-hoc we questioned Boudon’s conception of relative frustration narrowly focusing only on the losers of a competition. Following the received conception of relative frustration resulting from inter-individual comparison, we operationalised relative frustration as the Gini coefficient (Kakwani, 1984). In its essence, this measure sums over all differences in the players’ payoffs and, in doing so, captures all intra-group comparison processes. Summing over all differences is an alternative aggregation rule and, applying this rule, an inversely U-shaped rate of relative frustration is not only predicted by the model but also found empirically. The observed increase in the Gini coefficient (or the variance, which we used as a substitute in experiment 1) when chances improve was observed in all three experiments. This result is also in line with Kuznets’ (1955) thesis of an inversely U-shaped rate of social inequality under improving economic conditions. Note, however, that we derived the inversely U-shaped path of the Gini coefficient from Boudon’s model. Concerning the micro-macro problem (Coleman, 1990), our modification demonstrates the importance of the question of how to transform individual decisions into macro-level effects. The implications of the model on the macro level crucially depend on the proper choice of aggregation rule.

Taken together, our experiments demonstrate that a more favourable opportunity structure can indeed generate more social inequality, perpetuating relative frustration. However, as derived from the model, this is only likely to occur under specific conditions (Manzo, 2009; Raub, 1984), e.g. when the expected benefit of investing into a competition for upward mobility is considerably larger than the benefit of not doing so and when the increase in opportunities occurs on a low level. Then, there can be a trade-off between social mobility on the one side and social inequality and relative frustration on the other.

Our study has some limitations. We tested model predictions with a specific set of parameter values and with Swiss students in a laboratory setting. It is possible that the predicted effect of an increasing loser rate would still occur when varying any of these features of the design, which we suggest for future studies. Further, we focused on whether investing behaviour changes with opportunities as predicted by the model and elicited frustration only by a verbal question. Introducing a behavioural measure of frustration would thus be a natural next step. Moreover, testing hypotheses derived from an expanded version of the model might be of interest. Given our results, mechanisms reinforcing frustration, such as comparison in local networks (Manzo, 2011), would be of special interest.

Finally, we would like to emphasise that Boudon’s competition model is not restricted to the investigation of relative frustration. Rather, it provides a conceptual basis for theoretical and empirical investigations of the interconnection of competition structures, social mobility and inequality, status processes, and in-
The model generalises “winner-take-all markets” (Frank and Cook, 1995; Lutter, 2013) and is a fruitful expansion of classic market-entry models that have been previously used to describe situations such as these (Fischbacher and Thöni, 2008). It is worth mentioning that Boudon applied a variant of his model to analyse the effects of educational institutions on individual decisions (Boudon, Cibois, and Lagneau 1975, see also Raub 1984, ch. 5). We believe that in the sociology of education the competition model could be useful as an interdependent counterpart of parametric models (e.g. Breen and Goldthorpe, 1997). Given the actuality and sociological relevance of these phenomena, it seems worthwhile to further investigate the competition model and exploit it as a theoretical framework for empirical research.
## 2.A Appendix

### 2.A.1 Central statistical analyses

Table 2.A.1: Probability of investing and losing, respectively

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th></th>
<th>Experiment 2</th>
<th></th>
<th>Experiment 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Invest</td>
<td>Lose</td>
<td>Invest</td>
<td>Lose</td>
<td>Invest</td>
<td>Lose</td>
</tr>
<tr>
<td>1 position</td>
<td>-.188**</td>
<td>-.021</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td></td>
<td>(-2.76)</td>
<td>(-.32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 positions</td>
<td>Ref.</td>
<td>Ref.</td>
<td>.300**</td>
<td>.133</td>
<td>.200**</td>
<td>.033</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.05)</td>
<td>(1.29)</td>
<td>(2.41)</td>
<td>(.34)</td>
</tr>
<tr>
<td>5 positions</td>
<td>.354***</td>
<td>-.125*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(5.76)</td>
<td>(-2.24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Invest</th>
<th>Lose</th>
<th>Invest</th>
<th>Lose</th>
<th>Invest</th>
<th>Lose</th>
</tr>
</thead>
</table>

| N groups       | 12           | 12       | 10           | 10       | 10           | 10       |
| N individuals  | 72           | 72       | 60           | 60       | 60           | 60       |
| N decisions    | 432          | 432      | 60           | 60       | 60           | 60       |
| R²             | .18          | .03      | .08          | .02      | .12          | .08      |

Notes: Discrete change effects derived from logit models. Cluster-robust standard errors. z statistics in parentheses. Controlled for session (not shown). * p < .05, ** p < .01, ***p < .001.
Table 2.A.2: Gini coefficients, variances and results from statistical tests

<table>
<thead>
<tr>
<th>Experiment</th>
<th>k</th>
<th>G or V</th>
<th>S.E.</th>
<th>Difference</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>V = 9.18</td>
<td>.070</td>
<td>k = 1 - k = 2</td>
<td>z = −2.25, p = .024</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>V = 13.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>V = 11.44</td>
<td>.078</td>
<td>k = 5 - k = 2</td>
<td>z = −0.53, p = .593</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>G = .106</td>
<td>.010</td>
<td>k = 1 - k = 2</td>
<td>t_{(df=8)} = −3.19, p = .006</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>G = .114</td>
<td>.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>G = .128</td>
<td>.014</td>
<td>k = 1 - k = 2</td>
<td>t_{(df=8)} = −2.30, p = .025</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>G = .161</td>
<td>.003</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


2.A.2 Instructions

Notes: The instructions for experiments 2 and 3 are provided. The instructions for the first experiment differ only slightly, in that it was mentioned that the game was played for several rounds. English translations are presented while the original language is German. We implemented Boudon’s competition model as a laboratory experiment using z-Tree (Fischbacher, 2007).

a) Instruction sheet:

Guide to the competition  In the course of this experiment you will decide whether or not to take part in a competition.

- If you do not take part in the competition, you will receive a medium-level payoff.
- If you take part in the competition, you will either gain a high payoff with a certain probability or you will only get a low payoff.

The rules of the competition are the following:

1. You and all other subjects will be randomly assigned to a group of six. A separate competition is held for each group.
2. You will receive information about the chances of winning and the amount of all three payoffs.
3. You must decide whether or not to take part in the competition. If you decide to take part, you will have to pay an investment fee.

4. After all the individuals within a group have taken their decisions, the winners will be chosen randomly from among the competitors in a given group.

5. You will receive information about your payoff. There are three possible payoffs:

   - Medium payoff: If you do not take part in the competition, you receive a medium payoff.

   - High payoff (winners’ payoff): If you take part in the competition and win, you receive a winner’s payoff. Since you will have to pay an investment fee if you choose to compete, this fee is subtracted from your gross benefit, i.e. winners’ payoff = gross benefit - investment fee.

   - Low payoff: If you take part in the competition and you lose, you receive the very same medium payoff as the sustainers. However, you will lose your investment fee: Low payoff = medium payoff - investment fee.

In short, you have the choice between definitely receiving a medium payoff and taking part in the competition. The competition either leads to a high payoff or to a low payoff. Before you make your decision, you will be told the amount of all three payoffs. Chance of winning:

   - You will be randomly assigned to a group of six. A separate competition is held for each group. What happens in the other groups has no influence on your chances or payoffs.

   - In a given competition, up to six individuals compete for a given number of high payoffs.

   - Before the decision, neither you nor the other five individuals in your group know how many of the group members will actually decide to compete.

   - Before the decision, you and all other persons are informed how many high payoffs there are.

   - Thus, if you take part in the competition, the chance of winning is influenced by two factors:
     - Number of high payoffs (common knowledge); and
     - Number of competitors in a group of 6 (unknown).
Selection of winners:

- The high payoffs are randomly assigned to the competitors within a group.
- If in a given competition the number of competitors equals the number of high payoffs or if there are fewer competitors than high payoffs, then all competitors will receive a high payoff.

**Example 1 (fictive numbers):** If in a given competition there is one high payoff and five out of six subjects of a given group take part in the competition, then...

- the person who has not taken part in the competition gets a medium payoff;
- one of the five competitors gets a high payoff (gross payoff minus investment fee); and
- four of the five competitors get a low payoff (medium payoff minus investment fee).

**Example 2 (fictive numbers):** If in a given competition there is one high payoff and one out of the six persons of a given group takes part in the competition, then...

- the five who have not taken part in the competition get a medium payoff; and
- the winner gets a high payoff (gross payoff minus investment fee).

**b) Instructions on screen**

**Screen 1:** (Welcome, general rules of the lab)

**Screen 2:** Control questions

1. How many additional persons are there in your group? (5; 7; 9)

2. Which of the following does not influence my chance of winning (number of competitors in my group; number of payoffs in the competition; persons in the other groups of six)

3. Which of the following three statements are correct?
   - It does not cost me any money to compete.
- I have to compete.
- It is up to me whether or not to compete.

4. If I do not take part in the competition, then...
   - I will probably receive a high payoff.
   - I will definitely receive a low payoff.
   - I will definitely receive a medium payoff.

5. How many CHF are 100 MP? (1; 0.1; 0 since there is no real money to gain)

6. Before the decision, who knows how many subjects will actually compete? (only the other people; only myself; nobody)

7. If I take part in the competition, then...
   - I will receive a low payoff.
   - I will receive a low or a high payoff.
   - I will receive a high payoff.

Screen 3: (Correct answers are shown)

Screen 4: As soon as all subjects are ready, we will clarify those questions that were answered incorrectly. Should anyone have a question after this, please raise your hand. Someone will come and answer your question. Please do not put up your hand before the microphone announcement has ended.
   (Questions that were answered incorrectly were explained through the microphone)

Screen 5: Information about the competition
- Number of high payoffs: 1, 2, 5
- Amount of high payoff: $\alpha$ MP
- Amount of low payoff: $\gamma$ MP
- Amount of medium payoff for sustainers: $\beta$ MP
- Do you want to take part in the competition? (Yes; No)
Screen 6:

- You did not take part in the competition. You will receive $\beta$ MP.
- You have won. You will receive $\alpha$ MP.
- You have lost. You will receive $\gamma$ MP.

How satisfied are you with this result? (0 not at all satisfied . . . 10 fully satisfied).

After a short questionnaire, subjects disbanded.

2.A.3 Additional information on the model

Derivation of mixed equilibria

As discussed in the main article, when no dominant strategy is available, players are assumed to apply mixed strategies. As an illustration, the numerical example from the main article is discussed in more detail in the following.

Parameter values (mixed strategy case as implemented in experiment 2):

- Group size: $N = 6$
- Number of positions: $k = 1$
- Winners’ payoff: $\alpha = 9$
- Sustainers’ payoff: $\beta = 6.5$
- Losers’ payoff: $\gamma = 5.$

Applying equation (2.1) (see main article), the expected payoff $E(k, n)$ for one position and all possible numbers of competitors $n - 1$ can be calculated. As an example, when in addition to $i$ there are two other investors, it is $E(k, n) = E(1, 3) = 1/3 \cdot 9 + 2/3 \cdot 5 = 6.3$.

Once the expected payoffs of the strategy ‘invest’ for all possible numbers of investors have been determined (see figure 2.3 in the main article), we can fill in equation (2.2) (see main article):

$$E(k, \bullet) = (1 - p)^5 \cdot 9 + \binom{5}{1} p(1 - p)^4 \cdot 7 + \binom{5}{2} p^2(1 - p)^3 \cdot 6.3 +$$
$$\binom{5}{3} p^3(1 - p)^2 \cdot 6 + \binom{5}{4} p^4(1 - p) \cdot 5.8 + p^5 \cdot 5.7 = 6.5$$

Because the group size is 6, besides focal player $i$, there can be up to 5 other investors. The binomial coefficient $\binom{N-1}{n-1}$ accounts for the fact that there are, for instance, $\binom{5}{2} = 10$ possibilities that besides $i$ 2 players invest and 3 refrain from
doing so. Thus, the probability that there are \( p^{(n-1)} \) additional investors and \((1 - p)^{(1-(n-1))}\) non-investors must be weighted by the according binomial coefficient.

The resulting overall expected payoff of the strategy ‘invest’ \( E(k, \bullet) \) is then equated with the sustainer’s payoff \( \beta \). Solving for \( p \) yields the optimal probability of investing. As stated in the main article, in this example \( p^* \) equals .429 and, thus, the investor rate amounts to 42.9%.

When ‘invest’ is a dominant strategy (which is the case when \( k > 2 \) in the numerical example at hand), the investor rate is 100%.

**Opportunities and frustration rates: No universal law**

As mentioned in the main article, there is no universal law connecting opportunities to frustration rates on the aggregate level. Put differently, the effect of a larger frustration rate under improved conditions emerges only in certain regions of the parameter space. Here we further elaborate on this in order to give the interested reader a deeper understanding of the model. We do not, however, deliver thorough game theoretic analyses of the boundary conditions of the paradoxical effect. Such analyses already exist and can be found in Raub (1984). For similar analyses based on agent based models, see Manzo (2009).

**Low frustration case**

If the winners’ payoff is not that lucrative (\( \alpha = 6.6, \beta = 6.5, \gamma = 5 \)) only a minor number of players choose to invest, the frustration rate remains relatively flat and, in the extreme case where the investor rate grows in proportion to the available positions, is even zero (see figure 2.A.1, panel (a)).

**High frustration case**

Figure 2.A.1b demonstrates the opposite extreme. A position is so lucrative (\( \alpha = 14, \beta = 6.5, \gamma = 5 \)) that all players enter the competition even if there is only one single position. If this is the case, an increase in the number of positions necessarily diminishes the frustration rate – a negative correlation between the number of positions and frustration over the whole range of positions becomes evident.

**Inversely U-shaped frustration case**

The paradoxical effect of an increasing loser rate under improving conditions (and, more generally, an inversely U-shaped path of frustration) emerges in between these two extremes.
2.A.4 Additional information on the experimental procedures and design

In this section, the experimental methods are discussed more thoroughly.

**Group size and parameter values**

The *group size* of six was chosen for pragmatic reasons. While leaving room for the phenomenon of interest to emerge (i.e. an increasing loser rate under improved conditions), group size was kept small in order to maximise the number of groups and thus the statistical power.

The *payoffs* were chosen such that the chances of observing the phenomenon of interest are optimal according to the model, while also allowing the investigation of both a situation where the phenomenon of interest should emerge when actors play mixed strategies (‘mixed strategy case’, experiments 1 and 2) and a situation where the assumption of mixed strategies is not necessary (‘dominant strategy case’, experiment 3). Further increasing $\alpha$ enhances the frustration level.
in the situation with one position more than in the situation with two positions and thus diminishes the difference in the loser rate between both situations since the system converges towards the high frustration case depicted in figure 2.A.1b. On the other side, decreasing $\alpha$ shifts the peak in the loser rate to the right-hand side of the x-axis, then getting flatter and finally turning into the low frustration case depicted in figure 2.A.1a.

**Experiment 1**

As stated in the main article, 72 subjects participated in the *first study*. Each one took part in six competitions – namely, in two competitions with one position, two competitions with two positions and two competitions with five positions. Each competition (1 position, 2 positions, 5 positions) was conducted once in a *high stake* ($\alpha = 7, \beta = 1$ and $\gamma = -3$) and once in a *low stake* ($\alpha = 1.75, \beta = .25$ and $\gamma = -.75$) condition. Half of the groups started with the high-stake and the other half with the low-stake conditions. In both the high- and the low-stake models, the predictions are identical and as provided in the main article. This procedure was implemented as a robustness check. According to the theory, subjects should not be influenced by the absolute level of payoffs but only by the relation of the payoffs. Varying stakes allowed us to observe the sensitivity of investment behaviour towards the absolute level of payoffs.

Additionally, the competitions were presented once in an *ascending order* (1 position, 2 positions, 5 positions) and once in a *descending order* (5 positions, 2 positions, 1 position). Half of the groups started with the competition with one position (ascending order) and the other half started with the competition with 5 positions (descending order).

Analyses on the impact of both design factors (stake and order of competitions) are presented in the first paragraph of the subsequent section.

After each of the 6 competitions, the outcome of the actual competition was reported (i.e. investors learned whether or not they were successful) and individual *satisfaction* was measured with the question “How satisfied are you with this result?” Participants rated their satisfaction on a scale ranging from 0 (not at all satisfied) to 10 (fully satisfied). This allowed us to test the micro assumption that losers are frustrated, i.e. less satisfied than the winners and the non-investors. This, however, comes with two problems. First, it is a matter of controversy as to whether satisfaction measures can be compared between individuals and should be aggregated (see Johnson and Fornell 1991). Second, the model is based on the

---

17 In the first experiment, the game was played repeatedly in order to maximise the number of observations. As described in the main article, there were some issues related to this and we thus continued with a between-subjects design.
assumption that the losers of a competition are dissatisfied (that is, relatively frustrated or, equivalently, relatively deprived) with the outcome of the competition because they have invested the same amount of resources as the winners but do not get anything in return (dissatisfaction with a state). As a consequence, their scepticism towards the system grows (dissatisfaction with a process). Hence, it remains unclear which of the two satisfaction components should be measured. We decided to only measure satisfaction with the outcome, because a cognitive pre-test revealed that people were irritated by the question on their satisfaction with the competition as a whole. Subjects claimed that their answers to both questions would not differ anyway. Despite these difficulties, we want to provide the interested reader with the results on testing this micro assumption. The results are reported in the second paragraph of the subsequent section.

Experiments 2 and 3

As described in the main article, experiments 2 and 3 were between-subjects (only one decision per subject). There was no high- and low-stake condition either. In short, the design was much simpler than that of experiment 1 and is thus sufficiently described in the main article.

2.A.5 Additional statistical analyses

Effects of design factors (experiment 1)

We conducted statistical tests by regressing the dependent variables investor (= 1, 0 else) and loser (= 1, 0 else) on a set of predictors by means of logit models. Discrete change effects are reported in table 2.A.3.

When controlling for the design factors stake and order, the results do not differ substantially from those reported in the main article. Turning to the design factors, the regression estimations suggest that a fourfold increase in the payoffs (the high-stake condition) does not affect the investor rate. Presenting the number of positions in an increasing rather than in a decreasing order (1, 2, 5 vs. 5, 2, 1) does not influence investment behaviour either. Results are pretty robust to variation in design factors disregarded by the model.

Test of micro assumptions (experiments 1-3)

In order to test the model’s micro assumption that losers are frustrated, we conducted an OLS regression analysis with self-reported satisfaction as the dependent variable and dummies for losers and winners as predictors, with ‘sustainer’ as the reference category (see table 2.A.4).
Clearly, the losers are the least satisfied group while the winners are the most satisfied group, with the non-investors in between. All differences are statistically significant at least at the 1% level. Evidently, the model’s micro assumptions hold. However, it also becomes clear that the non-investors are less satisfied than the winners. This justifies our alternative conception of relative frustration not merely focusing on the losers (see main article).

When plotting the losers’, sustainers’, and winner’ satisfaction in each competition (low-mobility, medium-mobility and high-mobility) separately, the micro assumption still holds. Moreover, losers and sustainers seem even to be more dissatisfied in competitions with two and five positions respectively, than in competitions with one (see figure 2.A.2). In other words, there seems to be a negative correlation between opportunities and the intensity of individual relative frustration. The basic form of Boudon’s model abstracts from frustration intensity, but see Manzo (2011) for a discussion and an agent-based model of both frequency and intensity of relative frustration.
## Table 2.A.4: Individual satisfaction with outcome of the competition (experiments 1-3)

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Losers</td>
<td>-4.05***</td>
<td>-4.05**</td>
<td>-5.07***</td>
</tr>
<tr>
<td></td>
<td>(-11.07)</td>
<td>(-4.62)</td>
<td>(-8.31)</td>
</tr>
<tr>
<td>Sustainers</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>Winners</td>
<td>3.65***</td>
<td>3.02**</td>
<td>2.32***</td>
</tr>
<tr>
<td></td>
<td>(7.98)</td>
<td>(4.17)</td>
<td>(6.51)</td>
</tr>
<tr>
<td>Constant</td>
<td>5.23***</td>
<td>6.71***</td>
<td>7.44***</td>
</tr>
<tr>
<td></td>
<td>(15.28)</td>
<td>(10.04)</td>
<td>(21.23)</td>
</tr>
</tbody>
</table>

| N_{individuals} | 72           | 60           | 60           |
| N_{decisions}    | 432          | 60           | 60           |
| R^2              | .60          | .65          | .72          |

Note: Coefficients (OLS regression). Standard errors clustered on groups. \( t \) statistics in parentheses. *\( p < .05 \), **\( p < .01 \), ***\( p < .001 \).

The tendency toward higher frustration intensity in medium and high-mobility systems might contribute to Tocqueville’s paradox. Nevertheless, at least in our experiments, intensity cannot contribute much to overall satisfaction. First, the composition of winners, losers and sustainers impacts overall satisfaction more than slight differences in individual frustration intensity. For instance, in a system with five positions, the five winners’ satisfaction weighs much more than the fact that the only loser of a group is more frustrated than a loser in a more competitive system.

Second, the negative correlation between opportunities and frustration intensity is not very robust. While there is a similar tendency (which is not statistically significant though) in experiment 3 (see figure 2.A.4), no such correlation is observed in experiment 2 (see figure 2.A.3).\(^{18}\)

\(^{18}\) All satisfaction scores (predictions from OLS regression) and statistical tests (F-tests) are listed in table 2.A.5 and 2.A.6 respectively. Note that our experiments were not designed to measure frustration intensity and the number of cases in each cell is quite small in experiments 2 and 3.
Figure 2.A.2: Satisfaction by competition outcome (loser, sustainer, winner) and positions ($k = 1, k = 2, k = 5$) for experiment 1.

Gini coefficients and variance

The interpretation of Gini coefficients in the presence of negative values is problematic (negative values can lead to coefficients greater than 1). For this reason, we calculated the variance as a substitute for the Gini in experiment 1.

Predicting Gini coefficients

In order to predict Gini coefficients in experiments 2 and 3, the following modification was necessary. In our experimental groups of six, a loser or winner rate of, for example, 12% would produce a fractional number of individuals with the according payoffs. When calculating the Gini coefficient, however, we cannot work with fractions of individuals and in a group of six individuals rounding would have a substantial impact on the predictions. Due to this, we calculated the predicted Gini for groups of 60 for all three experiments.

Furthermore, it should be noted that according to the model the effect of an increase in the Gini under improving conditions emerges under a broader range
of conditions than the increase in the loser rate and is thus harder to falsify, empirically. Only when the investor rate does not increase notably, despite the better opportunities or when the winners’ payoff is substantially greater than the other payoffs, does the Gini not follow an inversely U-shaped path but rather remains flat (first case) or falls monotonously (second case).\footnote{We thank Anshul Kumar Sing for an analysis of the path of the Gini coefficient under different conditions.}

**Testing for statistical differences between Gini coefficients**

In experiments 2 and 3, the Gini coefficients and their standard errors were calculated using Stata (i.e. the ‘svylorenz’ procedure by Stephen P. Jenkins).\footnote{See \url{http://www.stata.com/meeting/12uk/uksug2006_jenkins.pdf} (retrieved 05 Feb. 2014).} This procedure exploits all observations and thus is not just an approximation. After having derived the Gini coefficients and the adjusted standard errors, simple un-

---

**Figure 2.A.3**: Satisfaction by competition outcome (loser, sustainer, winner) and positions ($k = 1, k = 2$) for experiment 2.
The logic of relative frustration

Figure 2.A.4: Satisfaction by competition outcome (loser, sustainer, winner) and positions ($k = 1, k = 2$) for experiment 3.

Paired t-tests were conducted ($df = 8$), which was possible because observations are statistically independent between groups.21

Testing for statistical differences between variances

Figure 2.A.5 displays the path of the variance for all three experiments. The path of the variance tends to be inversely U-shaped theoretically and empirically for all three experiments. While in experiments 2 and 3 predictions and observed values match neatly, in experiment 1 the observed variance is somewhat lower than predicted.

When testing the hypothesis of variance homogeneity in the data of experiment 1, we accounted for the complex data structure applying Stata’s intreg command. We modelled the variance using the experimental groups and the treatments (number of positions) as predictors (dummy variables).

21 The number of groups (and not individual observations) was used for determining the degrees of freedom ($df$).
In the main article, for reasons of simplicity, we report results from the high stake condition only. When calculating the variances for the low stake condition, results are similar. The variance increases statistically significantly ($p = .015$) from .545 ($k = 1$) to .949 ($k = 2$), and then falls in tendency ($p = .102$) to .713 ($k = 5$). Note that calculating the variance separately for the high stake and the low stake condition implies a reduction in statistical power due to the smaller sample size.

For experiments 2 and 3, we conducted Levene’s test of variance homogeneity, using Stata’s robvar command. The hypothesis of variance homogeneity in the payoffs between the experimental conditions (low-mobility condition and moderate-mobility condition) is clearly rejected for both experiment 1 ($V_{k1} = 1.84, V_{k2} = 3.22, p = .007$) and experiment 2 ($V_{k1} = 2.96, V_{k2} = 4.82, p = .006$).
Table 2.A.5: Satisfaction by competition outcome (loser, sustainer, winner) and positions ($k = 1, k = 2, k = 5$) for all experiments (predictions).

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Positions $k$</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loser</td>
<td>1</td>
<td>1.77</td>
<td>2.40</td>
<td>3.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.486)</td>
<td>(.589)</td>
<td>(.530)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.25</td>
<td>2.86</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.288)</td>
<td>(.592)</td>
<td>(.634)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.69</td>
<td>-</td>
<td>.260</td>
</tr>
<tr>
<td>Sustainer</td>
<td>1</td>
<td>5.37</td>
<td>6.67</td>
<td>7.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.248)</td>
<td>(.937)</td>
<td>(.359)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5.30</td>
<td>6.83</td>
<td>7.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.330)</td>
<td>(.484)</td>
<td>(.712)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.65</td>
<td>-</td>
<td>.667</td>
</tr>
<tr>
<td>Winner</td>
<td>1</td>
<td>8.94</td>
<td>10.0</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.512)</td>
<td>-</td>
<td>(.197)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8.88</td>
<td>9.60</td>
<td>9.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.211)</td>
<td>(.287)</td>
<td>(.099)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>8.94</td>
<td>-</td>
<td>.181</td>
</tr>
</tbody>
</table>

| $N_{individuals}$ | 72 | 60 | 60 |
| $N_{decisions}$    | 432| 60 | 60 |
| $R^2$              | .65| .65| .73|

Note: Predictions (OLS regression). Cluster-robust S.E. in parentheses.
Table 2.A.6: Satisfaction by competition outcome (loser, sustainer, winner) and positions ($k = 1, k = 2, k = 5$) for all experiments: Tests for statistical differences.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Pos. $k$</th>
<th>$k = 1$ vs. $k = 2$</th>
<th>$k = 1$ vs. $k = 5$</th>
<th>$k = 2$ vs. $k = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Losers</td>
<td>1</td>
<td>$F = 1.42, p = .259$</td>
<td>$F = 10.2, p = .009$</td>
<td>$F = 4.47, p = .058$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$F = .30, p = .597$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$F = 2.02, p = .189$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$F = .02, p = .878$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$F = .069, p = .429$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Winners</td>
<td>1</td>
<td>$F = .00, p = .967$</td>
<td>$F = .01, p = .912$</td>
<td>$F = .15, p = .703$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$F = 1.94, p = .197$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$F = .21, p = .661$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Chapter 3

Are luxury brand labels and ‘green’ labels costly signals of social status?

Abstract

Costly signaling theory provides an explanation for why humans are willing to pay a premium for conspicuous products such as luxury brand-labeled clothing or conspicuous environmentally friendly cars. According to the theory, the extra cost of such products is a signal of social status and wealth and leads to advantages in social interactions for the signaler. A previous study found positive evidence for the case of luxury brand labels. However, an issue of this study was that some of the experiments were not conducted in a perfectly double-blind manner. We resolved this by replicating variations of the original design with a double-blind procedure. Additionally, besides the luxury label condition, we introduced a ‘green’ label condition. Thus, the hypothesis that signaling theory is able to explain pro-environmental behavior was tested for the first time in a natural field setting. Further, we conducted experiments in both average and below-average socioeconomic neighborhoods, where, according to signaling theory, the effects of luxury signals should be even stronger. In contrast to the original study, we did not find positive effects of the luxury brand label in any of five experiments. Nor did we find evidence for a green-signaling effect. Moreover, in below-average socioeconomic neighborhoods a slight negative tendency of the luxury label actually became evident.

---

1 This chapter is an edited version of the following paper: Berger, Joël (2015). “Are luxury brand labels and ‘green’ labels costly signals of social status?” Unpublished working paper.
3.1 Introduction

3.1.1 Conspicuous consumption as a signal of social status

Sociologist and economist Thorstein Veblen (1965 [1899]) coined the term ‘conspicuous consumption’, referring to the display of easily recognizable expensive goods when cheaper functional equivalents exist. Examples are sports cars, expensive watches, and luxury clothes. Why are humans willing to pay a luxury premium for such items?

Costly signaling theory (CST) (Grafen, 1990a; Grafen, 1990b; Spence, 1973; Zahavi, 1975; Zahavi, 1977) offers a parsimonious explanation. CST is based on the following tenets: First, members of a group vary with respect to a desirable yet not directly observable quality. Second, this quality is correlated with the signal in a reliable way, which means that lower-quality individuals are not able or cannot afford to emit the signal. Third, an observer derives some benefit from the possibility to discriminate between individuals with different levels of this quality. Fourth, signaling allows the receiver of the signal to make an inference about the sender’s quality (Bliege-Bird and Smith, 2005; Gambetta, 2009).

This implies that, according to CST, an individual displaying luxury items signals that she or he is able to ‘waste’ money and is thus wealthy and of a high social status, which is the unobservable yet desirable quality. Humans have a preference for bonding with wealthy and high-status individuals and thus treat them more favorably in social interactions (Miller, 2009; Saad, 2007). Since neither wealth nor social status is directly observable, when bonding, humans rely on signals for these traits (Bourdieu, 1984; Bouska and Beatty, 1968; Doob and Gross, 1968). Obviously, according to the theory, the luxury signals provide benefits to both signalers and observers.

Nelissen and Meijers chose the example of luxury brand-labeled clothes to test the CST explanation for conspicuous consumption (Nelissen and Meijers, 2011). They are the first to present evidence suggesting that individuals displaying luxury brands are indeed treated more favorably during social interactions.

3.1.2 Public generosity and pro-environmental behavior as signals of status and prosociality

It has been suggested that unconditional prosocial behavior in public (e.g., public generosity) is an even more efficient investment in social capital (i.e., bonding with desirable allies) than the mere display of luxury goods. As is the case with the consumption of expensive goods, public generosity imposes costs on the signaler and is thus a signal of wealth and social status. In addition, potential allies
may infer that the signaler will spread his or her benevolence to them too. Hence, the desire to associate with prosocial signalers might even be greater than the wish to affiliate with luxury signalers (Gintis, Smith, and Bowles, 2001).

As an example of prosocial signaling, among the Meriam, a small-scale tribal society in the South Pacific, young men hunt turtles and present them to the community in a public feast. This gift is not reciprocated by the other group members. However, the turtle hunters gain in status; as a result, they have the opportunity to mate with one of the most desired women and have a higher number of offspring than average (Smith and Bliege-Bird, 2000). Similar phenomena have been reported for other tribal societies (Boone, 1998; Hawkes, 1993; Kaplan and Hill, 1985; Marlowe, 2010).

Analogous to turtle hunting in modern societies are non-anonymous donations (Glazer and Konrad, 1996; Harbaugh, 1998; Sieg and Zhang, 2012). It has been shown experimentally that individuals prefer to engage publicly rather than privately in prosocial behavior (Vugt and Hardy, 2009), and that publicly generous individuals gain in social status (Anderson and Kilduff, 2009; Ellingsen and Johannesson, 2011) and are perceived as more cooperative and trustworthy (Bereczkei, Birkas, and Kerek, 2010; Fehrler and Przepiorka, 2013; Klapwijk and van Lange, 2009). Consequently, they are more often chosen as allies, economic exchange partners (Barclay and Willer, 2007; Chiang, 2010; Gambetta and Przepiorka, 2014), and group leaders (Milinski et al., 2002; Willer, 2009). In some instances, they can even expect higher returns from economic transactions in the long run (Sylwester and Roberts, 2010). Probably in anticipation of such advantages, under conditions of partner selection, individuals even compete to be the most altruistic, sending a signal of prosociality and status in order to maximize their chances of securing the most desired allies. This behavioral pattern of “competitive altruism” (Roberts, 1998) was first detected in animals (Zahavi, 1989; Zahavi, 1990) but has also been observed in humans (Barclay, 2004; Hardy and van Vugt, 2006).

More recently, it has been pointed out that conspicuous environmentally friendly behavior is an instance of conspicuous prosocial behavior and thus can be explained within the framework of CST (Griskevicius, Cantù, and Vugt, 2012; Whitfield, 2011). A typical example of a ‘green signal’ is the hybrid car ‘Toyota Prius’. Compared to fuel-powered models in the same price segment it offers less comfort. This is the cost that makes the signal reliable: Only an individual who truly cares about the environment, and thus the public, is willing to forgo comfort for the greater good. This interpretation is strengthened by the fact that the Prius has no conventional counterpart. It has been argued that this explains the huge success of the Prius relative to other hybrids: The former is unique and therefore can serve as a signal, while the latter cannot easily be distinguished
from their fuel-powered counterparts (Thaler and Sunstein, 2008). Indeed, recent findings corroborate this hypothesis (Delgado, Harriger, and Khanna, 2015; Furlong, 2011; Sexton and Sexton, 2014). Similarly, in fictive purchase situations, subjects favor the ecological product over its conventional equivalent if the ecological product is costlier and conspicuous (Griskevicius, van den Bergh, and Tybur, 2010). This is consistent with the hypothesis that ‘green’ products serve as a signal of both status and prosocial attitudes and so increase people’s attractiveness as allies and exchange partners. Indeed, in experimental games people donating to a climate fund gain in reputation (Milinski et al., 2006).

In a nutshell there might be two different processes from which green signalers benefit. First, as with luxury consumption, they signal their ability to ‘waste’ money and are thus perceived as higher in status and wealthier. Second, since they buy not merely luxury products but products benefiting the environment (and thus, the public), they are also perceived as more cooperative and trustworthy. Both mechanisms enhance their desirability as interaction partners.

### 3.1.3 The present study

The purpose of the present study is twofold. First, we aimed to reproduce the findings by Nelissen and Meijers, which suggest that the display of luxury brand yields benefits in social interactions. In doing so, we expanded on the original design by adding a perfectly double-blind procedure, suggested by Nelissen and Meijers for further research. Second, we conducted a first test of the hypothesis that green signals have the same favorable impact on social relations as luxury brands in a natural field setting. Additionally, since we could not replicate the original findings, we conducted the experiments in neighborhoods with average and below-average socioeconomic status. This is of interest because CST predicts the effects of luxury brand labels to be stronger in neighborhoods with a below-average socioeconomic status since the costs of a luxury brand are higher for less wealthy individuals, and therefore the signal is perceived as more reliable.

The remainder of this article is structured as follows. First, we investigate how the labels exploited in the study at hand are perceived (experiment 1). Thereafter, we test whether people are more likely to comply with our confederate’s request to take part in a short survey when wearing a luxury brand-labeled shirt or a green (environmentally friendly) shirt as compared to the control condition without a label. In experiment 3, we test the hypothesis that displaying luxury labels or green labels yields financial benefits. Experiment 4 is a replication of the compliance study in a neighborhood with low socioeconomic status, while in experiment 5 money is collected in a below-average socioeconomic neighbor-
hood. All experiments but one were conducted with both the original procedure and a new, double-blind variant.

3.2 Experiment 1: status perception

As a precondition for a brand label to work as a signal it must be associated with a desirable trait such as status, wealth, or cooperativeness. Hence, we measured the perception of the same individual when wearing a luxury brand-labeled shirt, an organic-labeled shirt and a baseline shirt without a label.

3.2.1 Materials and method

150 passersby ($M_{age} = 40.0$ years, S.D. $= 14.83$, 59.33% female) where surveyed at several spots in the inner city of Zurich. Participants were randomly assigned to one of three conditions, namely no label ($n = 50$), luxury label ($n = 50$), and green label ($n = 50$). Lacoste is the better known of the two luxury brands used in the original study; thus we used Lacoste as a luxury brand label. For the green-signaling condition, we used the Bio label. Bio is a well-known Swiss label indicating organic products that are more expensive than equivalent non-organic products, and so fulfills a precondition for working as a costly signal. The photographs used in experiment 1 can be found in Appendix 3.A.1.

Respondents were invited to take part in a short study on the formation of first impressions and to spontaneously answer the questions without further reflection. The questionnaire consisted of a photograph of a female confederate wearing a polo shirt on which one of the two above-mentioned labels was digitally superimposed, or a polo shirt without a label. Participants rated on a scale from 1 (not at all) to 5 (completely) the person’s status (this person has a high status), wealth (this person has money), attractiveness (this person is attractive), kindness (this person is kind), trustworthiness (this person is trustworthy), prosociality (this person is prosocial), and environmental consciousness (this person is environmentally conscious). For each dependent variable (i.e., each rating), tests of significance were conducted by means of an OLS regression model, with both labels (luxury and green) as dummy variables and the control condition as a reference category. Cluster-robust standard errors were used because an individual’s ratings for the traits are not statistically independent.
Table 3.1: Perception ratings (experiment 1)

<table>
<thead>
<tr>
<th></th>
<th>Status</th>
<th>Wealth</th>
<th>Attractiveness</th>
<th>Kindness</th>
<th>Trustworthiness</th>
<th>Prosociality</th>
<th>Pro-environm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>0.320+</td>
<td>0.286*</td>
<td>0.080</td>
<td>-0.060</td>
<td>-0.340+</td>
<td>-0.460**</td>
<td>-0.393+</td>
</tr>
<tr>
<td></td>
<td>(1.84)</td>
<td>(2.02)</td>
<td>(0.37)</td>
<td>(-0.39)</td>
<td>(-1.68)</td>
<td>(-2.87)</td>
<td>(-1.90)</td>
</tr>
<tr>
<td>Green label</td>
<td>0.120</td>
<td>-0.214</td>
<td>0.240</td>
<td>0.080</td>
<td>0.020</td>
<td>-0.079</td>
<td>0.776***</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(-1.51)</td>
<td>(1.28)</td>
<td>(0.61)</td>
<td>(0.11)</td>
<td>(-0.47)</td>
<td>(3.82)</td>
</tr>
<tr>
<td>Const.</td>
<td>3.400***</td>
<td>2.714***</td>
<td>3.360***</td>
<td>4.280***</td>
<td>3.940***</td>
<td>3.939***</td>
<td>3.653***</td>
</tr>
<tr>
<td></td>
<td>(26.61)</td>
<td>(29.43)</td>
<td>(24.13)</td>
<td>(47.28)</td>
<td>(30.52)</td>
<td>(36.89)</td>
<td>(25.42)</td>
</tr>
</tbody>
</table>

Notes: A confederate approached passersby in the city and asked them to rate a person on a photograph (another confederate) on a scale from 1 to 5 with respect to status, wealth, attractiveness, kindness, trustworthiness, prosociality, and environmental consciousness when wearing a shirt with no label (control conditions), a Lacoste label (luxury label condition) or a Bio label (green label condition). The tables list coefficients from OLS regression models (t statistics in parentheses; + < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001; cluster-robust standard errors). ‘No label’ is the reference category.

3.2.2 Results

Table 3.1 summarizes the results. In line with CTS and the original findings, the individual is perceived as higher status when wearing the shirt with the Lacoste brand label, (t = 1.84, p = .068) and, more clearly, as wealthier (t = 2.02, p = .045) when compared to the control condition. Interestingly, the Lacoste label lowers the ratings on the dimensions ‘trustworthiness’ (t = −1.68, p = .096), ‘prosociality’ (t = −2.87, p = .005, not measured in the original study) and ‘pro-environmental attitude’ (t = −1.90, p = .059), while it does not influence perception of ‘attractiveness’ (t = 0.37, p = .709) and ‘kindness’ (t = −.39, p = .699). This is in contrast to the original study, where the Lacoste brand was associated with higher ratings on the trustworthiness dimension. Evidently, even though people wearing luxury brand labels are perceived as higher status they are not necessarily seen as more trustworthy and prosocial. The only clearly statistically significant effect of the green label is a substantially higher average rating on the environmentally consciousness dimension (t = 3.82, p = .000). Obviously, the higher rating on the pro-environment dimension was not transferred to other dimensions, such as prosociality. Also, even though Bio products are more expensive, the person in the picture was not perceived as richer in the green condition compared to the control condition. This might be due to the fact that
pro-environmental attitudes are generally not associated with wealth in the public perception.

### 3.3 Experiment 2: compliance

To test whether individuals wearing brand-labeled shirts are not only perceived differently but also treated more favorably in social interactions, experiments 2 and 3 were conducted. Specifically, in experiment 2 we test the hypothesis that passersby more often comply with a request of a luxury signaler and a green signaler because they are perceived as higher status.

#### 3.3.1 Materials and method

720 passersby ($M_{age} = 36.4$ years, S.D. = 11.8, 61.11% female) were approached by a female confederate (as in the original study by Nelissen and Meijers) in summer 2013 at different spots in the inner city of Zurich. Following the procedure of the original study, the confederate approached with a questionnaire and clipboard and asked the targeted individual to answer a few short questions. Only unaccompanied individuals were chosen. In experiment 1 as in all other experiments, the age and sex of the participants were estimated by the respective confederates. There were three experimental conditions and two different experimental procedures. Under the standard procedure, as in the original study, the confederate wore a polo shirt without a label ($n = 120$), a polo shirt with a luxury brand label ($n = 120$), or a shirt with a green label ($n = 120$). For experiment 2 and for all other experiments we used the labels introduced in the section on experiment 1. Under the alternative procedure, instead of shirts, caps with the respective labels were used (no label $n = 111$, luxury label $n = 129$, green label $n = 120$). The cap procedure was suggested by Nelissen and Meijers for future studies in order to eliminate potential experimenter effects in their own experiments. In the original procedure with polo shirts, experimenter effects cannot be ruled out even if the confederate is blind to the hypothesis under study since the treatment might be obvious. However, in the present study, the confederate was not aware of which cap she wore because she drew it blindly from a bag and put it back before drawing anew. This procedure, of course, leads to a somewhat uneven distribution of cases between the treatments of each cap experiment. The author supervised the confederate every day at varying points in time to ensure that this double-blind procedure was applied properly.

Importantly, the same confederate conducted all trials of experiment 2. Moreover, the shirts as well as the caps were identical except for the labels that were
taken from original items (photographs of all items are presented in Appendix 3.A.1). In order to rule out effects of potential time-dependent confounding effects, the confederate chose the first experimental condition randomly and then repeatedly changed the cap or shirt after a predefined number of trials. However, due to practical reasons, the cap and shirt trials were conducted on different days. Pictures of the prepared shirts and caps used in experiments 2-5 are presented in Appendix 3.A.1.

### 3.3.2 Results

![Graph showing relative frequency of compliance by experimental condition in experiment 2.](image)

Figure 3.1: Relative frequency of compliance by experimental condition in experiment 2. A confederate approached passersby in the city and asked them to take part in a short survey. Experimental conditions: No label (control condition), Lacoste label (luxury label condition), Bio label (green label condition), printed on shirts or caps (double-blind procedure). 95% ci. Regression outputs can be found in Appendix 3.A.2.

Statistical tests were conducted by means of logit regression models with compliance (yes = 1, 0 else) as the dependent variable and dummies for the lux-
Experiment 3

ury label condition and the green label conditions as predictors, with the control condition as a reference category. Separate models were specified for the shirt and for the cap experiments. For a test of the effect of wearing a shirt or a cap, a third model containing only the dummy variable “shirt” (1 = yes, 0 else) was specified. In all of the subsequent experiments, these three steps of model specification were applied (except for experiment 5, where the third step was omitted because only the shirt conditions were conducted). Outputs of these regression models are listed in Appendix 3.A.2.

Figure 3.1 summarizes the results. No statistical difference between the control condition and the luxury condition becomes evident, either in the case of shirts ($z = .84, p = .401$) or in the case of caps ($z = -.57, p = .571$). The same holds true for the green label (shirt: $z = .35, p = .729$; cap: $z = .95, p = .341$). Contrarily, passersby complied with a significantly lower probability with the request when the confederate wore a cap ($z = 3.92, p = .000$), independent of the label. Evidently, the type of clothing worn was more important for the perception of a person than labels.

3.4 Experiment 3: charity donations

While the former experiment was designed to investigate whether people are more likely to comply with a request stemming from an individual wearing a luxury brand label or a green label, the third experiment investigates whether wearing labeled clothes yields financial benefits.

Again this might be the case because signalers are perceived as higher status and thus observers comply with their request. For green labels, a second process might reinforce this tendency. Green signalers might be perceived as more cooperative and trustworthy, which might enhance the chances that people donate, believing that their donations would indeed reach the target instead of being misused.

For this study, we cooperated with Kinderkrebshilfe, a well-known charity that supports children suffering from cancer as well as their parents (e.g., by providing counseling and financial aid).

3.4.1 Materials and method

A female confederate (as in the original study) who had not taken part in experiment 1 approached 880 passersby ($M_{age} = 41.7$ years, S.D. = 13.6, 61.7% female) at different spots in the city of Zurich. The locations were allocated by the city to charity campaigns.
The confederate approached passersby with a moneybox and information on the charity. She explained that she was collecting money for Kinderkrebshilfe and handed out literature. The confederate wore the same shirts or caps as those used in experiment 2. Each time after receiving a donation, the confederate put the money in a treatment-specific envelope. The dependent measure was the average amount of money collected.

### 3.4.2 Results

![Graph showing average donation by experimental condition](image)

Figure 3.2: Average donation by experimental condition (experiment 3). A confederate approached passersby in the city and asked them to donate to a charity. Experimental conditions: No label (control condition), Lacoste label (luxury label condition), Bio label (green label condition), printed on shirts or caps (double-blind procedure). 95% ci. The corresponding regression tables can be found in Appendix 3.A.2.

Statistical tests were conducted by means of OLS regression models. The models were specified as described before. As evident from figure 3.2, neither the Lacoste label (shirt: \( t = .64, p = .519 \); cap: \( t = -.28, p = .776 \) ) nor the Bio
Experiment 4

3.5 Experiment 4: compliance in a neighborhood with low socioeconomic status

In contrast to the original study, and in spite of the considerably larger number of subjects, a positive effect of labels on either compliance or on charitable donations could not be detected. Nor could we find a positive effect of the green label. This holds true for both shirts and caps.

A precondition for brands to function as a signal is that they are perceived as costly. Although study 1 demonstrates that this condition is met at least for the Lacoste label, we replicated the compliance experiment in a neighborhood with a low socioeconomic status in Zurich since this should raise the relative cost of the brands from the perspective of people on a low income.

3.5.1 Materials and method

720 subjects were approached (\(M_{\text{age}} = 45.0, \text{ S.D.} = 14.9, 58.9\% \text{ female}\)), either by the female confederate (shirt experiment) from experiment 4 or by a new, male confederate (cap experiment), and asked if they were ready to take part in a short survey (see the section on experiment 2 for further details). A male confederate was introduced in addition to the female confederate, because luxury signals might work differently depending on the sex of the signaler and of the observer (Buss, 1989; Saad and Gill, 2001), although Nelissen and Meijers did find beneficial effects of luxury brands for both women and men.

While the shirt experiment was conducted at a street corner in the city center, the cap experiments was conducted with people waiting at a bus stop. This implies that the baseline cooperation cannot be directly compared between the shirt and the cap procedure. However, if CTS applies, the postulated effects should be observable under both procedures.
Figure 3.3: Relative frequency of compliance by experimental condition in experiment 3. A confederate approached passersby in the city (neighborhood with low socioeconomic status) and asked them to take part in a short survey. No label (control condition), Lacoste label (luxury label condition), Bio label (green label condition), printed on shirts or caps (double-blind procedure). 95% ci. The corresponding regression tables can be found in Appendix 3.A.2.

### 3.5.2 Results

As in study 2, neither the luxury label (shirt: \( z = .44, p = .663 \); cap: \( z = -1.71, p = .087 \)) nor the green label (shirt: \( z = 1.13, p = .257 \); cap: \( z = -1.18, p = .239 \)) affected the rate of compliance (see figure 3.3 for an overview), even though there is a slight negative tendency for the luxury brand for the cap – but not for the shirt – at the 10% level of significance. The fact that this time the baseline compliance is higher in the cap experiment (\( z = 2.60, p = .009 \)) is most probably due to the fact that the cap experiment was conducted with participants waiting at a bus stop, while in the shirt experiment, passersby were approached. All the same, in either case labels have no positive effect on compliance.
3.6 Experiment 5: charity donations in a neighborhood with low socioeconomic status

In study 4, which was conducted in a neighborhood with low socioeconomic status, for the first time an effect at least reached the 10% significance level (even though in the opposite direction than hypothesized). For this reason, we also replicated study 3 (i.e., collecting money for a charity) in a neighborhood with below-average socioeconomic status. Instead of Zurich the experiment was conducted in the Swiss capital Bern to ensure that it is not an unknown idiosyncratic trait of the inhabitants of Zurich that is responsible for the lacking effect of the brand labels.

![Bar chart showing average donation by experimental condition (experiment 5). A confederate approached passersby in the city (neighborhood with low socioeconomic status) and asked them to donate to a charity. Experimental conditions: No label (control condition), Lacoste label (luxury label condition), Bio label (green label condition), printed on shirts. 95% ci. The corresponding regression tables can be found in Appendix 3.A.2.](image-url)
3.6.1 Materials and method

The same male confederate as in experiment 4 approached 360 individuals near a shopping center in a neighborhood with a low socioeconomic status (\(M_{\text{age}} = 48.9, \ S.D. = 13.8, \ 53.3\% \ \text{female}\)). The procedure was equivalent to the one of experiment 3, except that only the shirt procedure was realized. The beneficiary of the collected money was the same charity as before.

3.6.2 Results

Figure 3.4 displays the main results. In comparison to the shirt without a label, we do not find a significant effect either of the luxury brand label (\(t = -1.50, \ p = .135\)) or of the green brand label (\(t = .74, \ p = .458\)) – a result that is consistent with our former study on donations. As in experiment 4, however, there is a slight but statistically insignificant negative effect of wearing a luxury brand-labeled shirt on the amount of donations.

3.7 Discussion and conclusion

Nelissen and Meijers argue that CST explains the human taste for luxury goods in general and for luxury brand labels in particular. According to the theory, an individual displaying luxury items signals that she or he can afford to ‘waste’ money and is thus wealthy and of a high social status. This leads to an advantageous treatment of the signaler in social interactions because humans have a preference for associating with wealthy and high-status people and are ready to pay a premium to affiliate with them. Because neither wealth nor social status is directly observable, when bonding, humans rely on signals, inferring these traits.

As an extension of the luxury-signaling thesis it has been suggested that public prosociality, such as conspicuous generosity or environmentally friendly behavior, works in an analogous way. Public prosociality may function as a signal of wealth and, additionally, of prosocial values and cooperativeness. Hence, people sending prosocial or pro-environmental signals, like people sending luxury signals, should be treated advantageously in social interactions.

The present study aimed to replicate the original study by Nelissen and Meijers. Expanding on their design, we implemented a perfectly double-blind procedure proposed by the authors of the original study and, additionally, a green-signaling condition. The latter provides a first test of the hypothesis that conspicuous pro-environmental behavior works in an analogous way to conspicuous consumption in a natural field setting. Further, we conducted experiments in neighborhoods of average as well as below-average social status. In the latter,
effects of costly labels are predicted to be even stronger since the signal should be perceived as more reliable from the perspective of less wealthy observers.

While we replicate the finding of the original study of an individual wearing a luxury brand-labeled shirt being rated as higher status and wealthier than an individual wearing a shirt without a label, in contrast to the original study, the individual was also perceived as less trustworthy, less prosocial and less environmentally friendly when wearing a luxury brand-labeled shirt. In the green label condition, even though the wearer was perceived as more environmentally friendly, there were no spillover effects from this to other unobservable but positive traits, such as prosociality, trustworthiness, or wealth, as predicted by the green-signaling hypothesis (experiment 1).

Moreover, neither the confederate asking people to take part in a survey (experiment 2) nor the confederate collecting money for a charity (experiment 3) was treated more favorably when wearing a luxury brand-labeled shirt or cap. Nor did the green label have an effect in experiment 2 or 3. However, wearing a cap was disadvantageous in both experiments, independent of the brand.

Subsequently, we replicated experiments 2 and 3 in two neighborhoods of low socioeconomic status in order to make sure that the cost condition of a CTS explanation was met. In both studies, no positive effect of any label became evident. On the contrary, in studies 4 (cap) and 5 (shirt), a negative tendency of the luxury brand label was observed. Note, however, that this tendency was not statistically significant at the 5% level ($p = 0.087$ in study 4 and $p = 0.135$ in study 5).

Taken together, we were not able to reproduce the findings by Nelissen and Meijers. We do not find any positive effects of luxury brand labels on the cooperative behavior of observers.

A possible explanation for these findings would be that brands are costly signals in the sense of CST, but that being wealthy is not perceived as positive throughout. For instance, inequality aversion – and especially aversion towards disadvantageous inequality (Fehr and Schmidt, 1999) – might lead observers to evaluate people displaying wealth rather negatively. This would explain the negative attitudes toward luxury brand labels in neighborhoods with a low socioeconomic status. More generally, Nelissen and Meijers’ premise that people have a preference for bonding with high-status individuals in general might be wrong. Rather, people might prefer interacting and bonding with people of similar social status. Such tendencies of homophily have been observed not only in humans (McPherson, Smith-Lovin, and Cook, 2001) but also in other primates (Massen and Koski, 2014). This implies that, depending on preferences dominant in a given subgroup, positive or negative features might be associated
with a given brand. On average, such effects might cancel each other out, resulting in zero effect.

A tendency for homophily might also explain our finding that caps lower the chances of cooperation. Caps are more often worn by teenagers than by the average Swiss. Hence, wearing a cap might be perceived as a signal; the wearers may pay a cost in the sense that they lower their chances of bonding with average people, but at the same time enhance their chances of associating with members of certain subgroups (e.g., younger people).

We also tested for the possibility that there are signaling effects for certain sex configurations only, as reported elsewhere (Buss, 1989; Saad, 2007). There is no consistent pattern, although females perceive both luxury and green labels somewhat more favorably than males (see Appendix 3.A.3 for details). When it comes to behavior, however, in experiments 2 and 3 there is no effect of any label for either male or female subjects (and a female confederate, as in the original study). In experiment 4, males comply somewhat more with the request to take part in a survey \( p = 0.058 \) when the (female) confederate wears a green label rather than a neutral one. In contrast, male subjects react somewhat more adversely to the male confederate’s request to give money to a charity when he wears a luxury brand-labeled shirt as compared to a neutral shirt \( p = 0.025 \), while there is no significant effect for female subjects. However, these findings might also be a result of repeated testing.

The important point is that we did not find any positive effect of luxury labels in behavioral experiments 2-5, for either the sample as a whole or for the male or female subgroups. Hence, potential interaction effects between the sexes of the confederates and the subjects cannot explain our failure to replicate the results of the original study.

Nevertheless, future studies should systematically vary the confederates’ sexes in order to clarify whether or not there are interaction effects between the signaler’s and the observer’s sex because such effects have been found in similar settings (Milinski et al., 2006; Saad, 2007). Future research should also test the homophily hypothesis of signaling. For instance, labels with a specific connotation could be tested among different subgroups. Luxury brand experiments could be conducted in neighborhoods with high socioeconomic status, with average or even low socioeconomic status neighborhoods as a control group. Accordingly, green brands could be tested in neighborhoods where people with a high ecological awareness dominate.

As a last point, our results cast some doubt on whether luxury brands and green brands are costly enough in relation to conventional analogs to serve as signaling devices. Should further research again fail to replicate the findings of
the original study, this would call for another explanation for the human taste for expensive luxury brands and environmentally friendly products.
3.A Appendix

3.A.1 Shirts and caps used as experimental treatments

Figure 3.A.1: Pictures used for perception rating in experiment 1. The photograph was not blackened in the study but only for the purpose of anonymity in this publication.

Figure 3.A.2: Shirts and caps used in experiments 2-5.
### 3.A.2 Regression estimates for experiments 2-5

Table 3.A.1: Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Comply shirts</th>
<th>Comply caps</th>
<th>Comply shirts vs. caps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>0.284</td>
<td>-0.303</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.84)</td>
<td>(-0.57)</td>
<td></td>
</tr>
<tr>
<td>Green label</td>
<td>0.120</td>
<td>0.447</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.95)</td>
<td></td>
</tr>
<tr>
<td>No label</td>
<td>Ref.</td>
<td>Ref.</td>
<td></td>
</tr>
<tr>
<td>Shirt</td>
<td></td>
<td></td>
<td>0.941***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.92)</td>
</tr>
<tr>
<td>Cap</td>
<td></td>
<td></td>
<td>Ref.</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.671***</td>
<td>-2.555***</td>
<td>-2.473***</td>
</tr>
<tr>
<td></td>
<td>(-6.68)</td>
<td>(-6.96)</td>
<td>(-12.57)</td>
</tr>
<tr>
<td>N</td>
<td>360</td>
<td>360</td>
<td>720</td>
</tr>
</tbody>
</table>

Notes: Coefficients from logit regression models. \( z \) statistics in parentheses. \( ^+ p < 0.10, ^* p < 0.05, ^** p < 0.01, ^*** p < 0.001 \).
Table 3.A.2: Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Donation shirts</th>
<th>Donation caps</th>
<th>Donation shirts vs. caps</th>
<th>Stop shirts</th>
<th>Stop caps</th>
<th>Stop shirts vs. caps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>0.2</td>
<td>-0.09</td>
<td>-0.215</td>
<td>0.162</td>
<td></td>
<td>-0.65</td>
</tr>
<tr>
<td></td>
<td>-0.64</td>
<td>(-0.28)</td>
<td>(-0.87)</td>
<td>-0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green label</td>
<td>-0.348</td>
<td>0.121</td>
<td>-0.273</td>
<td>0.261</td>
<td></td>
<td>-1.08</td>
</tr>
<tr>
<td></td>
<td>(-1.12)</td>
<td>-0.39</td>
<td>(-1.11)</td>
<td>-1.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No label</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shirt</td>
<td></td>
<td></td>
<td></td>
<td>-0.009</td>
<td>0.857***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.05)</td>
<td></td>
<td></td>
<td>(-0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cap</td>
<td></td>
<td></td>
<td></td>
<td>Ref.</td>
<td>Ref.</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.696**</td>
<td>0.642**</td>
<td>0.655***</td>
<td>1.033***</td>
<td>-0.13</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>-3.17</td>
<td>-2.95</td>
<td>-4.89</td>
<td>-5.75</td>
<td>(-0.76)</td>
<td>-0.1</td>
</tr>
<tr>
<td>N</td>
<td>480</td>
<td>400</td>
<td>880</td>
<td>480</td>
<td>400</td>
<td>880</td>
</tr>
</tbody>
</table>

Notes: (1-3): Coefficients from OLS regression models. $t$ statistics in parentheses. (4-6): Coefficients from logit regression models (coefficients). $z$ statistics in parentheses. $^+p < 0.10, ^* p < 0.05, ^** p < 0.01, ^*** p < 0.001.$
## Table 3.A.3: Experiment 4

<table>
<thead>
<tr>
<th>Label Type</th>
<th>Comply shirts</th>
<th>Comply caps</th>
<th>Comply shirts vs. caps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>0.127</td>
<td>-0.452*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.44</td>
<td>(-1.71)</td>
<td></td>
</tr>
<tr>
<td>Green label</td>
<td>0.324</td>
<td>-0.312</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.13</td>
<td>(-1.18)</td>
<td></td>
</tr>
<tr>
<td>No label</td>
<td>Ref.</td>
<td>Ref.</td>
<td></td>
</tr>
<tr>
<td>Shirt</td>
<td></td>
<td></td>
<td>-0.414**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-2.60)</td>
</tr>
<tr>
<td>Cap</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.055***</td>
<td>-0.241</td>
<td>-0.487***</td>
</tr>
<tr>
<td></td>
<td>(-5.06)</td>
<td>(-1.34)</td>
<td>(-4.49)</td>
</tr>
<tr>
<td>N</td>
<td>360</td>
<td>360</td>
<td>720</td>
</tr>
</tbody>
</table>

Notes: Coefficients from logit regression models (coefficients). \( z \) statistics in parentheses. \( ^* p < 0.10, \ ^*\!\! p < 0.05, \ ^*\!\!\! p < 0.01, \ ^*\!\!\!\! p < 0.001. \)
Table 3.A.4: Experiment 5

<table>
<thead>
<tr>
<th></th>
<th>Donation shirts</th>
<th>Stop shirts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>-0.446 (-1.50)</td>
<td>-0.415 (-1.11)</td>
</tr>
<tr>
<td>Green label</td>
<td>0.222</td>
<td>0.466</td>
</tr>
<tr>
<td></td>
<td>-0.74</td>
<td>-1.43</td>
</tr>
<tr>
<td>No label</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>Constant</td>
<td>0.911***</td>
<td>-1.609***</td>
</tr>
<tr>
<td></td>
<td>-4.32</td>
<td>(-6.57)</td>
</tr>
<tr>
<td>N</td>
<td>360</td>
<td>360</td>
</tr>
</tbody>
</table>

Notes: (1): Coefficients from OLS regression model. t statistics in parentheses. (2): Coefficients from logit regression model. Coefficients. z statistics in parentheses. + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001.
### Table 3.A.5: Experiment 1

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Status (male)</th>
<th>Status (female)</th>
<th>Money (male)</th>
<th>Money (female)</th>
<th>Attractiveness (male)</th>
<th>Attractiveness (female)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury</td>
<td>0.158</td>
<td>0.460*</td>
<td>0.342</td>
<td>0.253</td>
<td>-0.051</td>
<td>0.204</td>
</tr>
<tr>
<td>label</td>
<td>-0.55</td>
<td>-2.14</td>
<td>-1.42</td>
<td>-1.52</td>
<td>(-0.14)</td>
<td>-0.88</td>
</tr>
<tr>
<td>Green</td>
<td>0.074</td>
<td>0.13</td>
<td>-0.286</td>
<td>-0.233</td>
<td>-0.057</td>
<td>0.324*</td>
</tr>
<tr>
<td>label</td>
<td>-0.24</td>
<td>-0.65</td>
<td>(-1.13)</td>
<td>(-1.42)</td>
<td>(-0.16)</td>
<td>-1.75</td>
</tr>
<tr>
<td>Constant</td>
<td>3.364***</td>
<td>3.429***</td>
<td>2.571***</td>
<td>2.821***</td>
<td>3.182***</td>
<td>3.500***</td>
</tr>
<tr>
<td><em>N</em></td>
<td>61</td>
<td>89</td>
<td>58</td>
<td>89</td>
<td>61</td>
<td>89</td>
</tr>
</tbody>
</table>
Table 3.A.5 continued

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Kindness (male)</th>
<th>Kindness (female)</th>
<th>Trustworthiness (male)</th>
<th>Trustworthiness (female)</th>
<th>Pro-sociality (male)</th>
<th>Pro-sociality (female)</th>
<th>Environmental consciousness (male)</th>
<th>Environmental consciousness (female)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>-0.099 (-0.39)</td>
<td>-0.027 (-0.14)</td>
<td>-0.524 (-1.64)</td>
<td>-0.19 (-0.71)</td>
<td>-0.686** (-2.87)</td>
<td>-0.283 (-1.31)</td>
<td>-0.389 (-1.08)</td>
<td>-0.378 (-1.59)</td>
</tr>
<tr>
<td>Green label</td>
<td>-0.085 (-0.42)</td>
<td>0.155 (-1.51)</td>
<td>-0.42 (-1.54)</td>
<td>0.261 (-1.23)</td>
<td>-0.47 (-0.75)</td>
<td>0.149 (-0.75)</td>
<td>0.857* (-2.28)</td>
<td>0.685** (-3.08)</td>
</tr>
<tr>
<td>N</td>
<td>61</td>
<td>89</td>
<td>61</td>
<td>89</td>
<td>59</td>
<td>88</td>
<td>59</td>
<td>89</td>
</tr>
</tbody>
</table>

Notes: (1-3): Coefficients from OLS regression models. *t* statistics in parentheses, cluster-robust standard errors. *p < 0.10, **p < 0.05, ***p < 0.01, ****p < 0.001. The confederate and the person in the picture were female.
### Table 3.A.6: Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Comply (male subjects)</th>
<th>Comply (female subjects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>0.262</td>
<td>-0.0293</td>
</tr>
<tr>
<td></td>
<td>(0.62)</td>
<td>(-0.08)</td>
</tr>
<tr>
<td>Green label</td>
<td>0.155</td>
<td>0.274</td>
</tr>
<tr>
<td></td>
<td>(0.37)</td>
<td>(0.72)</td>
</tr>
<tr>
<td>Control</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>Shirt</td>
<td>1.053**</td>
<td>0.833*</td>
</tr>
<tr>
<td></td>
<td>(2.90)</td>
<td>(2.57)</td>
</tr>
<tr>
<td>Cap</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.452***</td>
<td>-2.671***</td>
</tr>
<tr>
<td></td>
<td>(-5.87)</td>
<td>(-7.72)</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>280</td>
<td>440</td>
</tr>
</tbody>
</table>

Notes: (1-3): Coefficients from logit regression models (z statistics in parentheses). $^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$. Female confederate.
Table 3.A.7: Experiment 3

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Donation (male)</th>
<th>Donation (female)</th>
<th>Stop (male)</th>
<th>Stop (female)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>-0.201 (-0.74)</td>
<td>0.249 -0.78 (-0.49)</td>
<td>-0.137 -0.22</td>
<td>0.0486</td>
</tr>
<tr>
<td>Green label</td>
<td>-0.125 (-0.45)</td>
<td>-0.14 (-0.45)</td>
<td>-0.0907 -0.27</td>
<td>0.0582</td>
</tr>
<tr>
<td>Control</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>Shirt</td>
<td>0.0106 -0.05</td>
<td>-0.0429 (-0.17)</td>
<td>0.749** -3.27</td>
<td>0.926*** -5.13</td>
</tr>
<tr>
<td>Constant</td>
<td>0.611** -2.7</td>
<td>0.730** -2.78</td>
<td>0.2 -0.89</td>
<td>-0.101</td>
</tr>
</tbody>
</table>

| N              | 337 543        | 337 543           |

Notes: (1-2): Coefficients from OLS regression models (t statistics in parentheses). (3-4): Coefficients logit regression models (z statistics in parentheses). *p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001. Female confederate.
### Table 3.A.8: Experiment 4

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Comply Shirts (male)</th>
<th>Comply Shirts (female)</th>
<th>Comply Caps (male)</th>
<th>Comply Caps (female)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>0.575</td>
<td>-0.091</td>
<td>-0.476</td>
<td>-0.47</td>
</tr>
<tr>
<td></td>
<td>-1.12</td>
<td>(-0.25)</td>
<td>(-1.18)</td>
<td>(-1.33)</td>
</tr>
<tr>
<td>Green label</td>
<td>0.932*</td>
<td>-0.014</td>
<td>-0.427</td>
<td>-0.252</td>
</tr>
<tr>
<td></td>
<td>-1.9</td>
<td>(-0.04)</td>
<td>(-1.07)</td>
<td>(-0.71)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.476***</td>
<td>-0.853***</td>
<td>-0.0834</td>
<td>-0.341</td>
</tr>
<tr>
<td></td>
<td>(-3.77)</td>
<td>(-3.43)</td>
<td>(-0.29)</td>
<td>(-1.47)</td>
</tr>
</tbody>
</table>

**N** 137 223 159 201

Notes: (1-3): Coefficients from logit regression models (z statistics in parentheses). *p < 0.10, **p < 0.05, ***p < 0.01. Shirts: Female confederate. Caps: Male confederate.

### Table 3.A.9: Experiment 5

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Donation (male)</th>
<th>Donation (female)</th>
<th>Stop (male)</th>
<th>Stop (female)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury label</td>
<td>-0.920*</td>
<td>-0.061</td>
<td>-0.931</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>(-2.27)</td>
<td>(-0.14)</td>
<td>(-1.72)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Green label</td>
<td>-0.243</td>
<td>0.597</td>
<td>0.288</td>
<td>0.651</td>
</tr>
<tr>
<td></td>
<td>(-0.60)</td>
<td>(1.39)</td>
<td>(0.65)</td>
<td>(1.33)</td>
</tr>
<tr>
<td>Control</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
<td>Ref.</td>
</tr>
<tr>
<td>Constant</td>
<td>1.308***</td>
<td>0.597*</td>
<td>-1.229***</td>
<td>-1.998***</td>
</tr>
<tr>
<td></td>
<td>(4.46)</td>
<td>(1.99)</td>
<td>(-3.74)</td>
<td>(-5.30)</td>
</tr>
</tbody>
</table>

**N** 168 192 168 192

Notes: (1): Coefficients from OLS regression model. t statistics in parentheses. (2): Coefficients from logit regression model. Coefficients. z statistics in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01, ****p < 0.001. Male confederate.
Chapter 4

An international field experiment of norm enforcement in anonymous interactions: How costs and benefits impact altruistic punishment

Abstract In laboratory experiments people are willing to sanction norms at a cost – a behavioral tendency called altruistic punishment. However, the degree to which these findings can be generalized to natural interactions is still debated. Only a small number of field experiments have been conducted and initial results suggest that punishment is less frequent in natural interactions than in the lab. This study focuses on norm enforcement in Bern, Zurich, and New York City and its sensitivity to cost and benefit manipulations. Findings indicate that norms are universally enforced, although significantly less than in the standard laboratory experiment, and that enforcement is significantly more common in Switzerland than in New York. We also find that enforcement drops as benefits decrease or costs increase. Moreover, individuals prefer more subtle forms of enforcement to direct punishment, likely because they reduce costs and the risk of retaliation.

1 This chapter is an edited version of the following paper: Berger, Joël and Hevenstone, Debra (2015). “An international field experiment of norm enforcement in anonymous interactions: How costs and benefits impact altruistic punishment.” Unpublished working paper.
4.1 Introduction

Social norms regulate interactions between individuals in all known societies (Fehr and Fischbacher, 2004b; Hechter and Opp, 2001a). In many instances social norms constrain individual behaviors that have negative externalities (Ullmann-Margalit, 1977) and, as such, help stabilize social order (Hardin, 2013).

To ensure that social norms are followed, they must be enforced (Coleman, 1990; Oliver, 1980). As enforcement is costly (e.g. in terms of effort and/or time) there is an incentive to free ride, leaving enforcement to others (Heckathorn, 1989; Yamagishi, 1986). Sociological research has identified mechanisms that overcome this “second-order free rider problem.”² For instance, enforcers might indirectly profit via reputation, status, or social capital or could even be directly rewarded by other group members (Axelrod, 1986; Horne, 2004; Horne, 2007; Willer, 2009). But these mechanisms mainly work in repeated, non-anonymous interactions. It is still debated whether there is norm enforcement in the anonymous one-shot encounters that are common in contemporary societies (Rauhut and Krumpal, 2008).

Evidence from laboratory experiments indicates that people are willing to pay to punish individuals who violate social norms even in anonymous one-shot interactions – a behavioral tendency called altruistic (or costly) punishment (Fehr and Gächter, 2002). However, further laboratory studies have challenged the idea that altruistic punishment is the main mechanism sustaining social norms in the real world. In the lab, subjects are more reluctant to punish violators when the cost of sanctions is high (Egas and Riedl, 2008), or when retaliation is possible (Janssen and Bushman, 2008) – two key factors in many real world-interactions. As such, Guala (2012) objects that it cannot be concluded from the existing evidence that punishment is a frequent and efficient way of maintaining social norms in society. Instead, he suggests conducting field experiments to study norm enforcement in natural interactions, to develop a deeper understanding of the processes maintaining social norms.

To date only a small number of field experiments have been conducted. Initial results suggest that in natural interactions, people are less willing to enforce norms than in the laboratory (Balafoutas and Nikiforakis, 2012), and that people who feel personally affected by norm violations are more likely to sanction (Brauer and Chekroun, 2005).

Understanding why and when norms are enforced is of interest not only to basic social science, but also to social policy. The enforcement of some norms,

---
² Breaking a norm for one own’s benefit, and in the process hurting the collective (e.g. littering), is called the “first-order free-rider problem.” In contrast, the “second-order free-rider problem” refers to the question of who enforces the norm, thus resolving the original free-rider problem.
like littering, is a public good, and insofar as conditions for enforcement can be duplicated, these conditions are also sometimes viewed as a potential public policy tool. It is for this reason that the New York City Parks and Recreation website states, with respect to the new ban on smoking in parks, “the new law will be enforced mostly be New Yorkers themselves. We expect that New Yorkers will ask people to follow the law and stop smoking.”

This study extends our knowledge of norm enforcement in anonymous real-life interactions. First, we conducted field experiments in Bern, Zurich and New York, which allowed us to assess the prevalence and generality of norm enforcement as well as differences across locations. Second, we measure the impact of cost-benefit manipulations on norm enforcement. If altruistic punishment is robust to cost-benefit manipulations, it is more likely to be an important mechanism stabilizing norms under a broad range of conditions (Egas and Riedl, 2008). Third, we report characteristics of enforcers. Finally, we report quantitative and qualitative observations of alternative reactions beyond direct punishment.

The remainder of the article is organized as follows: After a discussion of the theoretical background, we describe the experimental design and present results from Bern, New York, and Zurich. We then summarize the main findings and their implications for norm enforcement and altruistic punishment. Finally, we suggest further research.

### 4.2 Theoretical background

#### Social norms and their enforcement

While there are many definitions of a social norm, every definition contains at least one of the following three aspects: First, a social norm is a rule about desirable or undesirable conduct that usually is contingent on a specific social context. Second, if a norm is being followed, this generates a behavioral regularity on the aggregate level. Third, social norms are enforced by sanctions that increase the probability of compliance (Bicchieri, 2006; Hechter and Opp, 2001b; Voss, 2001). Sanctions can be positive, i.e. reward norm compliance, or negative, i.e. punish deviance. For example, parents might praise their children for throwing a piece of paper in the garbage (positive) while a passerby might reprimand someone for littering (negative) (Coleman, 1990). Our study focuses on negative

---

sanctions, as does most of the experimental literature on norm enforcement in anonymous one-shot interactions.\textsuperscript{4}

Many social norms limit behaviors that have negative externalities and, as such, bystanders have an interest in enforcing norms (Heckathorn, 1988; Ullmann-Margalit, 1977). For example, smoking might harm (or bother) surrounding individuals, and as such, those individuals have an interest in stopping smokers (Opp, 2002). However, norms are not universally enforced because enforcement can be costly, e.g. it might be awkward to ask a smoker to put out his or her cigarette or might instigate a confrontation. As such, even if the benefit of enforcing a norm is greater than the costs, everyone has an incentive to leave enforcement to others – this is called the “second-order free rider problem” (Heckathorn, 1989; Olson, 1965; Yamagishi, 1986).

**Norm enforcement in repeated non-anonymous interactions**

Various solutions for the second-order free rider problem have been proposed. Some scholars propose the concept of meta-norms prescribing norm-enforcement, i.e. individuals are rewarded for enforcing a norm or punished for not doing so (Axelrod, 1986; Ellickson, 1991). This hypothesis is supported by laboratory experiments showing that peers reward those who punish norm violators and that rewards increase with growing sanctioning costs (Horne and Cutlip, 2002). Research shows that prosocial acts engender benefits such as gains in status, power, and desirable allies. This means that enforcing social norms might be in one’s self-interest, as long as norm enforcement is observable by group members (Bliege-Bird and Smith, 2005; Posner, 2000; Willer, 2009). Horne (2004; 2007) demonstrates that people also enforce norms in hope of engaging in profitable exchange thereafter.

An alternative to the rather costly form of norm-enforcement via punishment is gossip and social isolation (Guala 2012; Hirschmann 1970; Nowak and Sigmund 1998a; see also Coleman 1990). There is evidence that this social pressure enforces norm compliance in contexts ranging from the modern suburbs (Baumgartner, 1984) to tribal societies (Boehm, 1999; Marshall, 1961), that social pressure facilitates compliance with fishing quotas in Norway (Gezelius, 2002), and that it deters white-collar crime (Paternoster and Simpson, 1996).

These mechanisms only work in environments with stable inter-personal relations (Granovetter, 1985). In contrast, in anonymous one-shot interactions, meta-norms, gossip, and indirect benefits do not work. In an anonymous context a

\textsuperscript{4} For more information on positive sanctions in anonymous one-shot interactions see Berger (2011) and Diekmann, Jann, et al. (2014).
transgressor’s reputation is not at stake and enforcers cannot improve their reputations when no group member is present to observe their behavior either. It is for this very reason that social capital theorists (Putnam, 2001), exchange theorists (Homans, 1950), and criminologists (Sampson, Raudenbusch, and Earls, 1997) all highlight the importance of inter-personal relations for norm enforcement and social order (for a counterargument see Flache and Macy 1996). But in the modern world many interactions are more or less anonymous (Rauhut and Krumpal, 2008).

**Norm enforcement in anonymous one-shot interactions: altruistic punishment**

While experimental economists have studied norm enforcement in anonymous interactions, they have done so almost exclusively using laboratory experiments. Results indicate that a large proportion of individuals are willing to pay money to punish non-cooperation (Fehr and Gächter, 2002) and to enforce distribution norms (e.g. Camerer 2006; Cameron 1999). Even unaffected third parties sanction norm violators at a cost (Fehr and Fischbacher, 2004b). Just as the use of social pressure to enforce norms in non-anonymous repeated interactions has been observed in tribal to modern societies, so has norm enforcement in anonymous non-repeated interactions in the laboratory.\footnote{Specifically, there are “lab experiments” conducted outside of the lab with non-student subjects. In these experiments standard laboratory conditions were applied: subjects knew that they took part in an experiment, knew the rules of the game and payoffs, and understood that all decisions were anonymous. This is called a framed field experiment whereas our study is a natural field experiment, which is conducted in an everyday situation and where the subjects do not know that they are taking part in an experiment (Harrison and List, 2004).} However, these experiments have found differences in the degree of enforcement across cultures (Henrich, Elreath, et al., 2006).

Sanctioning in the absence of indirect benefits is called altruistic (or costly) punishment (Fehr and Gächter, 2002). The term “altruistic” highlights that those who punish are willing to pay for the sake of the group. In other words, people are motivated not only by narrow self-interest, but also by internalized prosocial values. Violations that contradict internalized values trigger emotions that could motivate sanctioning (Fehr and Gintis 2007, for a similar argument: Elster 1989). However, these emotions do not necessarily lead to sanctioning because there are costs associated with sanctioning. As these costs increase, prosocial behavior is less likely to occur (e.g. Carpenter 2007; Fehr and Gintis 2007) – a pattern that concurs with rational-choice models that assume individual actions are driven
An international field experiment

by both selfish and normative motives (e.g. Braun and Gautschi 2014, Coleman 1990, pp. 243-44, Opp 2013).

The problem of external validity

Some scholars argue that altruistic punishment is the primary mechanism sustaining social norms and is of particular importance in those interactions where other enforcement mechanisms are absent (e.g. Fehr and Fischbacher 2004b; Fehr and Gächter 2002). However, laboratory evidence alone is insufficient to justify this claim. Even if laboratory experiments measure a real behavioral tendency, it does not follow that people manifest this tendency in natural interactions (Harrison and List, 2004; Jackson and Cox, 2013). For instance, while in the typical laboratory study individuals have the choice between punishment and passivity, in quotidian life there are many more options (Guala, 2012). External validity is further threatened by the fact that more cooperative individuals might volunteer for laboratory experiments or because participants are eager to please the experimenter (Levitt and List, 2007; Orne, 1962). Finally, experiments show that punishment decreases as sanction efficacy decreases, and when there is the possibility of retaliation (Anderson and Putterman, 2006; Egas and Riedl, 2008; Janssen and Bushman, 2008). All together these factors could be so strong as to entirely discourage sanctioning or at least enough to push individuals towards less costly (non-confrontational) means of norm enforcement than punishment.

Field experiments on norm enforcement in anonymous interactions

Despite these issues, very few field experiments on norm enforcement in anonymous interactions have been conducted. Balafoutas and Nikiforakis (2012) investigated two norms: the norm against littering and the norm that on an escalator those not wishing to walk should stand on the right-hand side, allowing others to pass on the left. They found that those violating the anti-littering norm were punished 4.0% of the time, while those violating the escalator norm were punished 19.3% of the time. Enforcing the anti-littering norm is a classic second-order social dilemma: every observer has an incentive to leave enforcement to someone else. For the escalator norm, however, it is questionable whether this is an instance of norm enforcement, since people might simply want to pass. The same issue holds for a replication of the escalators study by Wolbring, Bozoyan, and Langner (2013). Further, it is known that people are more willing to sanction norms when there are self-interested motives for doing so. For instance, when someone violates the queuing norm, the person behind them has a large incentive
to sanction due to self-interest, beyond prosocial motives (Milgram et al. 1986; Schmitt, Dubé, and Leclerc 1992, see also Brauer and Chekroun 2005; Przepiorka and Diekmann 2013).

This study extends our knowledge about norm enforcement in anonymous real-life interactions where the individual benefit from norm enforcement is rather small, thus focusing on altruistic motives for sanctioning. As there is little individual benefit from enforcing the anti-littering norm, we focus on transgressions of the littering norm rather than the escalator or queuing norms. We conduct natural field experiments in multiple locations, comparing the prevalence of altruistic punishment and testing for its generality and for differences across locations. Second, we measure the impact of cost-benefit manipulations on norm enforcement. This is important as altruistic punishment that is robust to cost-benefit manipulations is more likely to successfully maintain social norms. It is known that in the laboratory norms can break down when the costs of punishment are too high or benefits too small (e.g. Egas and Riedl 2008). However, we do not yet know whether in the field, norm enforcement is also sensitive to costs and benefits. In addition, we consider the characteristics of enforcers, which differ considerably from the student population used in laboratory experiments. Finally, we describe quantitative evidence and qualitative observations of other types of norm enforcement, beyond direct punishment.

4.3 Experimental design and evidence

Operationalization

We chose the anti-littering norm because it is a generally accepted norm that creates a typical second-order free rider problem, and because it was tested in a prior field experiment on norm enforcement (Balafoutas and Nikiforakis, 2012).

We define norm enforcement as either confronting a violator (direct enforcement) or picking up the litter (a plastic bottle) (indirect enforcement). Direct enforcement is clearly an instance of altruistic punishment and negative sanctioning. In contrast, indirect enforcement is more ambiguous. An enforcer who picks up a bottle might be observed by the violator, in which case picking up the bottle functions as a direct reprimand, or it might be observed by bystanders, in which

---

6 Game theoretically speaking, in a volunteer’s dilemma, where the individual benefit of norm enforcement exceeds the individual benefit, the assumption of self-regarding preferences can be sufficient to explain enforcement. Conversely, in the so-called missing hero dilemma, the individual cost exceeds the individual benefit, and thus, only altruistic motives can explain norm enforcement (Diekmann and Przepiorka, 2015). Here, we intended to create a missing hero dilemma.

7 See footnote 5.
An international field experiment

case it functions as a form of norm reinforcement in the community. Picking up a dropped bottle also directly contributes to the first-order public good of a clean environment (thus confounding the two motives of cleaning up and norm enforcement). Given this ambiguity, picking up the bottle is considered “indirect” norm enforcement. Both direct and indirect enforcement are costly in terms of time and effort. Confronting a violator also has psychological costs (Adams and Mullen, 2012) and engenders the possibility of retaliation. A third possible reaction is no enforcement. During our experiment there were a substantial number of bystanders who shook their heads, looked angrily at the violator, or talked to each other about the incident (potentially bonding or reinforcing the norm). While these reactions could be considered a sort of norm enforcement, they are subtle and subjective to assess. As such, we took a conservative approach and considered them not to be enforcement in the quantitative analyses.

For the cost-benefit manipulation, the experiment was conducted in clean and littered settings (experiments 1 and 2 in Bern and NYC). In the clean control condition we removed surrounding litter if necessary, while in the littered condition we placed a bag of garbage and several pieces of litter in the experiment setting. (A picture of the clean/littered conditions in Bern is displayed in figure 4.A.1 in the appendix.) We hypothesize that when there is already litter, observers consider one additional piece of garbage as marginally less harmful to the public good, thus decreasing the value of norm enforcement.

Data was collected on two types of people: “targets” and “enforcers.” The target was the first individual between ages 18 through 70 who stood at a pre-defined spot after a train or bus left the experimental area. The norm violation was conducted directly in front of these “targets” with the expectation that they would enforce. However, despite the targets’ proximity to the violation, other bystanders often enforced. In fact, targets seldom enforced. “Enforcers” were, of course, those individuals who enforced the norm. We collected measures of

---

8 We asked norm-enforcers for their motives on the last day of experimentation in Zurich using a standardized questionnaire. Individuals picking up the bottle reported both a desire to demonstrate their concern about littering (5 of 12 cases) and a desire to keep the station clean (7 cases).

9 There are multiple ways one could manipulate the costs and benefits of enforcing the anti-littering norm, such as using a more threatening looking violator. However, changing the violator would likely include a change of various traits simultaneously, making it difficult to know which exact characteristic generated the result. In contrast, we chose to change the amount of litter in the surrounding area, as it is a clearer manipulation (see also Shadish, Cook, and Campbell (2002) on construct validity in field experiments). That said, there is also another interpretation of how littering the environment might impact normative behavior (see Keizer, Lindenberg, and Steg (2008) and Lindenberg and Steg (2007)).

10 While we only chose targets that were not in groups, other enforcers were occasionally in groups. When enforcers were in a group, we could not disentangle altruistic motives from indirect benefits. However, in very few cases were enforcers in a group.
individual characteristics (estimated age, sex, and presumed race or ethnicity) for both targets and enforcers. As targets are a random sample of adults, they can be used as a comparison group for the enforcers. For each experiment we also counted the number of people in the 3 x 7 meter area in which the violation was easily visible, to measure the bystander density.

Manipulation check

To test whether littering the experimental environment successfully altered bystanders’ cost-benefit calculations, we conducted a vignette survey of randomly selected bystanders in Zurich (n = 137). There were two vignettes, one describing an unknown passerby who dropped a plastic bottle in a clean environment, and another describing an unknown passerby who dropped a plastic bottle in a littered environment. Respondents saw just one of the two types of vignettes and were asked whether they would directly sanction the transgressor and if they would pick up the bottle. If their answer was “no” for either question, they were asked to choose the most important of 6 potential reasons for not taking action (see figure 4.1). The reason that direct enforcement yields too small a benefit was selected roughly three times as often in the littered condition (21.43%) as in the control condition (7.46%, χ² = 5.36, p = .021). The same tendency was found for picking up the bottle: 14.29% reported that picking up the bottle provides too small a benefit in the littered condition compared to 1.49% in the clean condition (χ² = 7.59, p = .006). The vignette study suggests that people consider the surrounding conditions when choosing whether to enforce an anti-littering norm.

Experiment 1: Bern

The first experiment was conducted in Bern, the capital of Switzerland (population 138,000).11 The experiment was conducted at a tram stop in front of the main rail station. The stop, Hirschengraben, is used by approximately 40,000 passengers a day.12 We selected a central crowded stop to maximize anonymity and population diversity. On the platform we chose the area around a trashcan as the setting. The individual standing nearest to the trashcan was defined as the target. One of the experimenters was assigned to the role of the norm violator and the other to the role of observer. Roles were exchanged after every 5 trials. Both experimenters (female, white, 36 years old, 1.73 m tall and male, white, 30 years old, and 1.83 m tall) wore blue jeans and a black shirt.

12 Estimate by the public transit authority for 2012.
Figure 4.1: Self-reported reason for not directly and indirectly enforcing by experimental condition (control vs. littered).

The observer, who coded the variables of interest, was discreetly seated on a bench behind a bike stand 7 meters from the trashcan. The violator slowly walked towards the garbage can and threw an empty plastic bottle at the trashcan from about 2 meters away. The bottle, missing the trashcan, fell to the floor near the target, and the violator continued walking without picking it up. To bystanders, this looked as if the violator was too lazy to pick up the bottle after having missed the trashcan. The experiment was never conducted when a tram was at the stop, as the noise would have made it difficult to hear the dropping bottle. The next trial was not conducted until all bystanders from the prior experiment had left. The experiment was conducted on four working days in July 2013 between 1.30 pm and 7.30 pm with the intention of encompassing both non-rush hour and rush hour traffic. The first experimental condition was determined randomly, and then was switched every 2.5 hours.

For each violation the target, someone else or both, might have enforced the anti-littering norm. In only 6 of the 33 violations where a norm was enforced, did the target intervene. To analyze the likelihood of overall enforcement, the
violation, not the enforcer, is the observation of interest. As such, in the statistics that follow, the frequency of sanctioning per violation is reported.$^{13}$

In 13.58% ($n = 11$) of the violations under the control condition (a clean environment) there was direct norm enforcement (i.e. verbal reprimands or a request to pick up the bottle) compared to 23.46% ($n = 14$) indirect norm enforcement (a bystander picking up the bottle), while in 5 (3.14%) of the cases both direct and indirect enforcement occurred. All together, in almost a third of all cases (30.86%, 25 of 81) the anti littering norm was enforced.

Manipulating the benefits (i.e. littering the area) had a strong impact on the probability of enforcement (see figure 4.2). Direct enforcement decreased by 10.00 percentage points from 13.58% to 3.58% ($\chi^2 = 4.69, p = 0.030$). Indirect enforcement dropped even more by 17.05 percentage points to 6.41% ($\chi^2 = 9.01, p = 0.003$). Total enforcement (direct or indirect) dropped from 30.86% to just 10.26% ($\chi^2 = 10.26, p = 0.001$) of all trials.

When analyzing the individual characteristics that might be correlated with norm enforcement, the enforcer, not the violation, is a single observation. This is a key distinction, as multiple people could react to a single violation. If we were to only report the characteristics of enforcers, we would not know whether these characteristics motivated enforcement or were over-represented at the tram stop. Originally a “target” individual was incorporated into the experimental design so that we could estimate the effect of individual characteristics on enforcement among the pool of targets. However, few targets sanctioned while many bystanders did. As such, we use the target population, a random sample of people waiting at the stop, as a comparison group for enforcers. This design also allows us to examine the correlation of individual characteristics with enforcement.

Results show that discernible gender and ethnicity (as coded by the observer) were not correlated with enforcement. Forty-five percent of targets were male compared to 50% of enforcers ($\chi^2 = 0.319, p = .572$) and 89% of targets were Swiss compared to 91% of enforcers ($\chi^2 = 0.2156, p = .642$). Estimated age was also uncorrelated, with an average target’s age of 39.5 compared to 40.5 for enforcers (one tail $t$ test, $t = .413, p = .340$).

Finally, in order to control for setting characteristics, we conducted logit regression analyses with norm enforcement (direct, indirect, or both) as the dependent variable, first using only the treatment variable as an independent variable and then adding in setting variables, such as time, weather, and violator (see table 4.A.1 in the appendix). None of these variables had a significant effect on norm

---

$^{13}$ This design is a slight deviation from the Athens experiment, where the target and the littering incident were they same. In that experiment the authors managed to isolate an individual and then drop a plastic bottle in front of them. After a few pilots we found it impossible to isolate individuals and to prevent other observers from becoming enforcers, and thus had to change the design.
enforcement except that people were less likely to pick up the bottle when it was sunny and hot. Most importantly, the coefficient for the treatment variable (litter) is stable after introducing controls. Notably, the density of bystanders has no significant effect on norm enforcement (using a continuous variable, as shown in the appendix, or using dummies, not shown). If people had the same likelihood of responding, regardless of the number of bystanders, we would anticipate a higher likelihood of enforcement with more people. This might suggest that diffusion of responsibility lowers the individual propensity of engaging in norm enforcement (see Chekroun and Brauer 2002).

**Experiment 2: New York**

We replicated the experiment in New York City on two working days in August 2013 under similar conditions as to those in Bern. We selected two platforms at the Union Square subway station. The intention of selecting a central, but not the most central station (e.g. Grand Central or Penn Station), as we did in Bern, was
to generate a representative sample of the population, but with lower volumes of pedestrian traffic, closer to the Bern case. However, even as a secondary station, Union Square is still frequented about 2.5 times more than Hirschengraben; in 2012, 108,000 people per day used Union Square station.\textsuperscript{14} In New York the experiment was conducted in exactly the same way as in Bern, except race was coded differently to reflect the local context.\textsuperscript{15}

We conducted 155 trials, 78 in the control condition, and 77 in the littered condition. As in Bern, norms were enforced significantly more often in the control condition (14.10%, 11 cases) than in the littered condition (2.60%, 2 cases, \(\chi^2 = 6.68, p = .010\), see figure 4.3). This effect is mainly driven by indirect norm enforcement (12.82% in the control condition vs. zero cases in the littered condition). Direct enforcement was observed so rarely under both conditions (three times in the control and twice in the littered condition (\(\chi^2 = .19, p = .660\)) that we cannot ascertain whether litter played a role.

In NYC there were significant demographic differences between targets and enforcers. Sixty percent of the targets were men compared to 91% of enforcers \((\chi^2 = 4.167, p = .041)\). With respect to race, 50% of the targets were non-white, 72% of the enforcers were \((\chi^2 = 2.181, p = .139)\). Finally, with an average age of 35 among both targets and enforcers, there was no age effect \((t = .127, p = .450)\).

In sum, it seems that men are more likely to enforce the anti-littering norm in New York. We would conjecture that this difference might be due to less fear of confrontation.

The experimental context (rush hour, density, and norm violator)\textsuperscript{16} had no effect on the probability of enforcement. More importantly, the effect of the treatment variable remains approximately constant when controlling for context (see table 4.A.2 in the appendix).

**Experiment 3: Zurich**

Results from the first two experiments suggested that both direct and indirect norm enforcement is considerably more frequent in Bern than in New York and – insofar as our design is comparable to Balafoutas and Nikiforakis (2012) – than in Athens. We conducted a third experiment for two reasons. First, we wanted to

\textsuperscript{14}http://www.mta.info/nyct/facts/ridership/ridership_sub.htm, retrieved December 28, 2013.

\textsuperscript{15}Namely Caucasian, African-American, Asian, Hispanic, and other rather than Swiss, white foreigner, and non-white foreigner.

\textsuperscript{16}As no one picked up bottles under the littered condition, there is no model predicting indirect norm enforcement by experimental condition. The weather in NYC was only cloudy or sunny (never rainy) and on the cloudy day there were no direct sanctions in the control condition. As such, the effect of weather, independent of the disorder condition cannot be estimated.
replicate the high sanctioning rate as found in Bern. Second, conducting a third experiment allowed us to add a comparison treatment focusing on how increasing costs (rather than decreasing benefits) impacts norm enforcement.

The third experiment was conducted in Zurich’s main train station on two underground platforms for the local commuter train, where about 400,000 passengers pass through daily.\textsuperscript{17} Zurich has about 400,000 inhabitants (over 1.9 million, when counting the whole metropolitan area of Zurich).

While in the first two experiments the benefit of enforcing the anti-littering norm was varied (the clean vs. littered environment), in the third experiment, we also manipulated the costs of norm enforcement by varying the location of the littering – first adjacent to a trashcan (control condition), and then 12 meters away (no trashcan condition). Under the no trashcan condition, a person wishing to throw away the dropped bottle would have to either walk to the trashcan or hold the bottle until the train arrived, increasing the cost of indirect enforcement.

\textsuperscript{17} Estimate from Swiss Federal Railways.
Because under the new condition, bystanders could interpret the litterer’s intent differently than in the original experiment (purposeful littering versus laziness) a third treatment was introduced: the trashcan condition. Under this condition a bottle was dropped near the trashcan rather than aimed at the trashcan. In sum, the control condition serves as a basis for comparison to the experiments in Bern and New York, while comparing the trashcan and no trashcan conditions tests whether norm enforcement responds to cost manipulations.

The Zurich experiment was conducted on three working days in September and October 2013. One hundred and seventy four trials were conducted, 58 in each of the three conditions (control, trashcan, no trashcan), changing conditions after each trial.

Figure 4.4 depicts the main results. In the control condition, there was direct enforcement in 12.07% ($n = 7$) of all trials compared to indirect enforcement in 37.93% ($n = 22$) of all trials. The total enforcement rate was 39.66% of all trials. Direct enforcement (reprimands) did not differ between conditions (12.07% control, 8.62% trashcan, 6.90% no trashcan, $\chi^2 = .964, p = .618$). In contrast, indirect enforcement occurred substantially less frequently when the bottle was dropped far from the trashcan (6.90%) compared to the control treatment (37.93%, $\chi^2 = 6.06, p = .000$) or the trashcan treatment (27.59%, $\chi^2 = 8.70, p = .003$). The control condition and the trashcan condition were indistinguishable with respect to direct, indirect, and total enforcement ($\chi^2 = .372, p = .542$; $\chi^2 = 1.409, p = .235$; and $\chi^2 = .333, p = .564$, respectively). In sum, targeting the trashcan and just dropping the bottle next to the trashcan are no different, while dropping the bottle far from the trashcan discouraged people from picking it up.

In Zurich, as in Bern, there were no significant differences between enforcers and targets. Fifty-six percent of targets were men compared to 54% of enforcers ($\chi^2 = .086, p = .770$), 85% of targets were Swiss compared to 92% of enforcers ($\chi^2 = .863, p = .353$), and the average target was about 37 compared to 37.5 among enforcers ($t = .173, p = .413$). Setting variables also did not influence enforcement probability and the treatment was stable when controlling for setting (see table 4.A.3 in the appendix).

### Three-city comparison

Table 4.1 summarizes the rates of direct, indirect, and total norm enforcement in Bern, New York and Zurich. Total enforcement rates do not differ between Zurich (37.07%, control and near bin conditions pooled) and Bern (30.86%, control condition) ($\chi^2 = .812, p = .367$) but do between Zurich and New York (14.10%, $\chi^2 = 12.25, p = .000$) and between Bern and New York ($\chi^2 = 6.374, p = .012$).
Figure 4.4: Zurich. Relative frequency of total, indirect and direct norm enforcement, dependent on experimental condition (control vs. low littered), 95%-confidence intervals.

(both control condition). Direct enforcement rates (reprimands) also do not differ between Zurich (10.35%, pooling the control and near bin conditions) and Bern (13.58%, control condition) ($\chi^2 = .484$, $p = .487$) but do between Zurich and New York (3.85%, $\chi^2 = 2.761$, $p = .097$) and between Bern and New York ($\chi^2 = 4.689$, $p = .030$). Finally, indirect enforcement rates (picking up the bottle) also do not differ between Zurich (32.76%, control and near bin condition pooled) and Bern (23.46%, control condition, $\chi^2 = 2.007$, $p = .157$) but do between Zurich and New York (12.82%, $\chi^2 = 9.957$, $p = .002$) and between Bern and New York ($\chi^2 = 3.015$, $p = .083$). In sum, there is consistently no difference between the two Swiss cities, but a significant difference between New York and the two Swiss cities.

Looking at table 4.A.4 in the appendix, to compare norm enforcement between cities, we again consider only the experiments run in a clean environment, pooling those experiments where the bottle was thrown next to the trashcan with those where it was dropped next to the trashcan. Using Bern as the contrast group,
Table 4.1: Rates of norm enforcement in Bern, New York City, and Zurich.

<table>
<thead>
<tr>
<th>Norm enforcement (%)</th>
<th>Bern</th>
<th>New York</th>
<th>Zurich</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct</td>
<td>13.58</td>
<td>3.85</td>
<td>10.34</td>
</tr>
<tr>
<td>indirect</td>
<td>23.46</td>
<td>12.82</td>
<td>32.76</td>
</tr>
<tr>
<td>total</td>
<td>30.86</td>
<td>14.10</td>
<td>37.07</td>
</tr>
</tbody>
</table>

Notes: BE/NYC: Control condition. ZH: Control condition and trashcan condition (pooled data). Total: combination of direct and indirect enforcement strategies (if both strategies occurred in one trial, this was treated as one case of norm enforcement).

there is a significant difference between NYC and Bern in terms of direct and total sanctioning, such that the predicted probability of a direct sanction in NYC is 3% compared to 12% in Bern. Using Zurich as the contrast group, NYC is also significantly different such that the predicted probability of a bottle being picked up in NYC is only 11% compared to 28% in Zurich. Controlling for contextual variables does not change the effects.

In sum, the two Swiss cities were indistinguishable, while NYC had systematically lower direct and indirect enforcement. There were also qualitative differences between New York and Switzerland. In NYC there were fewer subtle reactions, where people shook their heads or spoke to one another about the litterer. In Bern we even had trouble maintaining the littered condition, as people repeatedly removed the litter we had purposely strayed about. Two times by-standers actually removed an entire bag of garbage from the scene. In Zurich, our female experimenter was even pushed by an elderly lady walking by.

Summary

In a series of three experiments we demonstrate that bystanders enforce the anti-littering norm in anonymous one-shot interactions in Bern, New York and Zurich. In Bern and Zurich there is no difference between enforcers and a random sample of the population in terms of gender, age, or race, while in NYC men are more likely to enforce the norm. This corroborates the thesis that people are generally motivated to enforce social norms and we conjecture the gender difference in New York might be due to different groups’ relative fears of reprisal. This was confirmed by the vignette study, which found that the main motivation for not sanctioning was the fear of conflict (see figure 4.1).
We found that increasing the cost-benefit ratio, through a manipulation of either costs or benefits, leads to a substantial drop in norm enforcement. Both direct and indirect enforcement (reprimands and picking up the bottle, respectively) occur less frequently in the littered condition, where the benefit of norm enforcement is diminished (experiments 1 and 2). Analogously, increasing the costs of indirect enforcement, by dropping the bottle further from the trashcan, decreases the probability of indirect enforcement (experiment 3). Predicting the odds of any sort of enforcement (either direct or indirect), and pooling the increased cost and decreased benefit manipulations, the predicted probability of enforcement falls from 29% to just 8% when either manipulation is implemented. This suggests that, although altruistic punishment is a general phenomenon, it is sensitive to the cost-benefit ratio. The fact that the number of bystanders does not correlate with the probability of norm enforcement suggests that norm enforcement is very sensitive: the more bystanders there are, the less likely an individual will be to enforce, i.e. the second-order free rider problem is greater in larger groups.

In sum, yes, people enforce, but they are easily dissuaded from doing so. Moreover, indirect and more subtle forms of enforcement such as picking up litter or talking to bystanders about the transgression seem to be at least, if not more, important than confronting a transgressor.

A further result is the significant difference between enforcement in Switzerland and New York City. The rate of direct enforcement (confrontation) was 3.85% in New York, almost equivalent to the 4% rate in Athens, while the rates of 13.58% in Bern and 10.34% in Zurich are substantially higher.

4.4 Discussion and conclusion

Social norms regulate interactions between individuals and play a key role in stabilizing social order. Norms are largely maintained through enforcement, but since enforcement is costly, everyone has an incentive to leave it to others. Sociological research has identified mechanisms that overcome this “second-order free rider problem” (e.g. Axelrod 1986; Coleman 1990; Horne 2004). However, these mechanisms mainly work in repeated, non-anonymous interactions. But many interactions in modern societies are anonymous with a low chance of repeated encounters. It is still debated whether people are willing to enforce norms in these types of interactions (Rauhut and Krumpal, 2008).

In laboratory experiments individuals are willing to enforce social norms in anonymous non-repeated interactions at a net cost – a phenomenon called altruistic or costly punishment (Fehr and Gächter, 2002). However, it is still debated how far these findings can be transferred to natural interactions and more field
evidence is needed to answer this question conclusively (Guala, 2012). Yet up to now, only a small number of field experiments have been conducted, with results suggesting considerably less willingness to enforce norms in the field than in the standard laboratory experiment (Balafoutas and Nikiforakis, 2012), except for norms that have a strong direct impact on bystanders (Brauer and Chekroun, 2005).

We conducted field experiments in Bern, Zurich, and New York City. Norm enforcement was observed in all three cities, with higher rates in Bern (direct: 13.58% of all trials, indirect: 23.46%) and Zurich (direct: 10.34%, indirect: 32.76%) than in New York (direct: 3.85%, indirect: 12.82%). Age, gender and ethnic or racial background were uncorrelated with enforcement except in New York, where men were more likely to enforce. In all three cities enforcement was sensitive to cost-benefit manipulation. In sum, altruistic punishment specifically (i.e. confronting transgressors), and norm enforcement generally, occur in natural anonymous one-shot interactions and are conducted by a broad range of individuals across societies. However, enforcement is sensitive, depending on costs, benefits, and location.

Despite the universality of altruistic punishment, concerns (e.g. Guala 2012) seem justified. The level of punishment in the field is considerably lower than in the standard laboratory experiment. This is likely because in field experiments, the costs of punishment are typically higher and the benefits lower. This hypothesis is confirmed by the fact that altruistic punishment is sensitive to cost-benefit manipulations both in the lab (e.g. Egas and Riedl 2008) and, as shown here, in the field. As such, punishment (in the sense of directly confronting a transgressor) might not be as widespread or important as proposed. Instead, individuals seem to prefer less costly alternatives. In the case of our experiment this included picking up litter or talking to other bystanders about the litterer (bonding about the norm). This sort of enforcement reduces the risk of retaliation, strengthens the norm among bystanders, and signals to the transgressor that their behavior is unacceptable. Considering such subtle forms of norm enforcement, the total enforcement rate would be even higher in all cities.

Bonding against norm violations might be the mechanism that ultimately leads to the emergence of sanctioning institutions. Anthropological field evidence suggests that a group or an institution usually executes norm enforcement rather than single individuals (Ostrom, 1990; Wiessner, 2002). In laboratory experiments, subjects also prefer institutionalized punishment mechanisms to individual peer punishment (Traulsen, Röhler, and Milinski, 2013). This makes sense, as institutions distribute the high costs of norm enforcement equally across members of the group and overcome the second-order free rider problem (Boyd, Gintis, and Bowles, 2010; Coleman, 1990). Bystanders complaining to each other about
people littering are the early interactions that can lead a group to develop new institutions, such as a police-enforced fine for littering.

Our results also suggest dynamics that could ultimately lead to the deterioration of norms. We found that observers were less willing to take action when the benefits of a sanction decreased. The more people who litter, fail to clean up after their dogs or smoke in a non-smoking area, the smaller the marginal benefit of norm enforcement, the lower the enforcement rate, and the higher the chance of transgression. Thus we see that norm transgressions can generate a sort of positive feedback loop where one transgression generates future transgressions because the probability of being sanctioned decreases. This might well be a further mechanism fueling the spreading of disorder as described elsewhere (Cialdini, Reno, and Kallgren, 1990; Keizer, Lindenberg, and Steg, 2008).

Finally, the considerable differences in the level of norm enforcement between New York City and Swiss cities should be addressed. Our design does not allow us to identify the reason for this difference, although we would conjecture that the difference is at least partly cultural. Qualitative evidence (namely the stories that ex-pats in Switzerland like to tell one another) also suggests that social control is greater in urban Switzerland than in the urban US (Hevenstone, 2015). A survey study by Brauer and Chaurand (2010) indicates that the level of social control in a society is negatively related to the degree of individualism as conceptualized by Hofstede (2001). As such, it might be the US’s greater individualism that leads to less enforcement. It might also be that differences in beliefs and values influence which norms are seen as important in a society (Wildavsky, 1993). Beyond culture, there are, of course, other potential reasons for the difference. For example, people in New York City might have a greater fear of reprisal – which would also explain why men are more likely to enforce than women in New York, but not in Switzerland. Also, a factor related to city size might account for these differences – although Zurich and Athens are comparable with respect to city size while they differ substantially in punishment rates.

In conclusion, this study demonstrates that people are willing to enforce social norms in natural anonymous situations using punishment as well as more subtle strategies; that the costs and benefits impact the probability of enforcement; and that enforcement varies considerably across contexts. Although altruistic punishment seems to be a general mechanism and might be able to sustain social norms under favorable conditions, our results suggest that it is possibly not robust under a broad range of conditions and that more subtle mechanisms of norm enforcement seem to play a major role as well.
### Table 4.A.1: Logit regression estimates for Bern (coefficients)

<table>
<thead>
<tr>
<th></th>
<th>(1a) direct</th>
<th>(1b) direct</th>
<th>(1c) indirect</th>
<th>(1d) indirect</th>
<th>(1e) total</th>
<th>(1f) total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Littered (ref: control)</td>
<td>-1.368* (2.04)</td>
<td>-1.434* (2.05)</td>
<td>-1.498** (2.82)</td>
<td>-1.920** (-3.22)</td>
<td>-1.363** (-3.07)</td>
<td>-1.642*** (-3.37)</td>
</tr>
<tr>
<td>Rush hour (not rush h.)</td>
<td>0.262 (0.41)</td>
<td>0.032 (0.06)</td>
<td>0.098 (0.21)</td>
<td></td>
<td>0.098 (0.21)</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>0.031 (0.21)</td>
<td>-0.224 (-1.59)</td>
<td>-0.117 (-1.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad Weather (Ref.: sunny)</td>
<td>0.594 (0.91)</td>
<td>1.370* (2.39)</td>
<td>1.157* (-2.41)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violator male (ref.: female)</td>
<td>0.436 (0.73)</td>
<td>-0.700 (-1.44)</td>
<td>-0.188 (-0.45)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.851*** (-5.71)</td>
<td>-2.651* (-2.80)</td>
<td>-1.183*** (-4.51)</td>
<td>-0.621 (-0.82)</td>
<td>-0.806*** (-3.35)</td>
<td>-0.822 (-1.25)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>159</th>
<th>159</th>
<th>159</th>
<th>159</th>
<th>159</th>
<th>159</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>.05</td>
<td>.07</td>
<td>.07</td>
<td>.15</td>
<td>.07</td>
<td>.11</td>
</tr>
</tbody>
</table>

Notes: $t$ statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent variable: norm enforcement (4.A.1a/b: direct, 4.A.1c/d: indirect, 4.A.1e/f: total).
### Table 4.A.2: Logit regression estimates for New York

<table>
<thead>
<tr>
<th></th>
<th>(2a) direct</th>
<th>(2b) direct</th>
<th>(2c) total</th>
<th>(2d) total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Littered (ref: control)</td>
<td>-0.405 (-0.44)</td>
<td>-0.378 (-0.40)</td>
<td>-1.818* (-2.31)</td>
<td>-1.875* (-2.36)</td>
</tr>
<tr>
<td>Rush hour</td>
<td>0.911 (0.94)</td>
<td>-0.343 (-0.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>0.010 (0.11)</td>
<td>-0.023 (-0.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violator male</td>
<td>0.352 (0.38)</td>
<td>0.873 (1.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-3.219*** (-5.47)</td>
<td>-3.937*** (-3.63)</td>
<td>-1.807*** (-5.55)</td>
<td>-2.042** (-2.85)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>155</th>
<th>155</th>
<th>155</th>
<th>155</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>.00</td>
<td>.03</td>
<td>.08</td>
<td>.11</td>
</tr>
</tbody>
</table>

Notes: $t$ statistics in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Dependent variable: norm enforcement (4.A.2a/b: direct, 4.A.2c/d: indirect). Models for indirect norm enforcement not estimable due to zero cases of indirect norm enforcement in the littered condition. Effect of weather not estimable (constant across days of experimentation).
### Table 4.A.3: Logit regression estimates for Zurich

<table>
<thead>
<tr>
<th></th>
<th>(3a) direct</th>
<th>(3b) direct</th>
<th>(3c) indirect</th>
<th>(3d) indirect</th>
<th>(3e) total</th>
<th>(3f) total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin</td>
<td>-0.375</td>
<td>-0.367</td>
<td>-0.473</td>
<td>-0.463</td>
<td>-0.222</td>
<td>-0.217</td>
</tr>
<tr>
<td>(Ref.: Cont)</td>
<td>(-0.61)</td>
<td>(-0.59)</td>
<td>(-1.18)</td>
<td>(-1.14)</td>
<td>(-0.58)</td>
<td>(-0.56)</td>
</tr>
<tr>
<td>No bin</td>
<td>-0.617</td>
<td>-0.665</td>
<td>-2.110***</td>
<td>-2.191***</td>
<td>-1.141**</td>
<td>-1.426**</td>
</tr>
<tr>
<td>(Ref.: Cont)</td>
<td>(-0.94)</td>
<td>(-0.97)</td>
<td>(-3.61)</td>
<td>(-3.65)</td>
<td>(-3.03)</td>
<td>(-2.97)</td>
</tr>
<tr>
<td>Rush hour</td>
<td>-0.759</td>
<td></td>
<td>0.059</td>
<td></td>
<td>-0.073</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.00)</td>
<td></td>
<td>(0.13)</td>
<td></td>
<td>(-0.17)</td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>-0.013</td>
<td></td>
<td>-0.019</td>
<td></td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.23)</td>
<td></td>
<td>(-0.49)</td>
<td></td>
<td>(-0.02)</td>
<td></td>
</tr>
<tr>
<td>Bad Weather</td>
<td>0.594</td>
<td></td>
<td>-0.269</td>
<td></td>
<td>-0.310</td>
<td></td>
</tr>
<tr>
<td>(Ref.: sunny)</td>
<td>(0.88)</td>
<td></td>
<td>(-0.65)</td>
<td></td>
<td>(-0.81)</td>
<td></td>
</tr>
<tr>
<td>Violator male</td>
<td>-0.297</td>
<td></td>
<td>-0.643</td>
<td></td>
<td>-0.479</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.55)</td>
<td></td>
<td>(-1.67)</td>
<td></td>
<td>(-1.36)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.986***</td>
<td>-1.970</td>
<td>-0.492*</td>
<td>0.163</td>
<td>-0.420</td>
<td>-0.479</td>
</tr>
<tr>
<td></td>
<td>(-4.93)</td>
<td>(-2.39)</td>
<td>(-1.82)</td>
<td>(0.13)</td>
<td>(-1.56)</td>
<td>(-1.36)</td>
</tr>
<tr>
<td>N</td>
<td>174</td>
<td>174</td>
<td>174</td>
<td>174</td>
<td>174</td>
<td>174</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.01</td>
<td>.04</td>
<td>.09</td>
<td>.11</td>
<td>.05</td>
<td>.07</td>
</tr>
</tbody>
</table>

Notes: $t$ statistics in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Dependent variable: norm enforcement. (4.A.3a/b: direct, 4.A.3c/d: indirect, 4.A.3e/f: total).
Table 4.A.4: Logit regression, pooling cities

<table>
<thead>
<tr>
<th></th>
<th>(4a) direct</th>
<th>(4b) direct</th>
<th>(4c) indirect</th>
<th>(4d) indirect</th>
<th>(4e) total</th>
<th>(4f) total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYC (Ref.: Bern)</td>
<td>-1.368*</td>
<td>-1.499*</td>
<td>-.734</td>
<td>-.844</td>
<td>-1.000*</td>
<td>-1.087*</td>
</tr>
<tr>
<td></td>
<td>(-2.04)</td>
<td>(-2.14)</td>
<td>(-1.71)</td>
<td>(-1.87)</td>
<td>(-2.47)</td>
<td>(-2.54)</td>
</tr>
<tr>
<td>Zurich (Ref.: Bern)</td>
<td>-0.309</td>
<td>-0.326</td>
<td>.464</td>
<td>.446</td>
<td>.277</td>
<td>.190</td>
</tr>
<tr>
<td></td>
<td>(-0.69)</td>
<td>(-0.62)</td>
<td>(1.41)</td>
<td>(1.17)</td>
<td>(.90)</td>
<td>(.53)</td>
</tr>
<tr>
<td>Rush hour</td>
<td>-0.139</td>
<td>-.239</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.28)</td>
<td>(-.69)</td>
<td>(-1.03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>-0.0209</td>
<td>-0.0183</td>
<td>0.00237</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.40)</td>
<td>(-0.55)</td>
<td>(0.07)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad weather (Ref.: sunny)</td>
<td>0.450</td>
<td>0.309</td>
<td>0.112</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.86)</td>
<td>(0.87)</td>
<td>(0.34)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violator male</td>
<td>0.217</td>
<td>-0.165</td>
<td>0.618</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
<td>(-0.57)</td>
<td>(0.22)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.851***</td>
<td>-2.040***</td>
<td>-1.183***</td>
<td>-1.048**</td>
<td>-0.806***</td>
<td>-0.741*</td>
</tr>
<tr>
<td></td>
<td>(-5.71)</td>
<td>(-3.62)</td>
<td>(-4.51)</td>
<td>(-2.66)</td>
<td>(-3.35)</td>
<td>(-2.01)</td>
</tr>
<tr>
<td>N</td>
<td>275</td>
<td>275</td>
<td>275</td>
<td>275</td>
<td>275</td>
<td>275</td>
</tr>
<tr>
<td>R²</td>
<td>.03</td>
<td>.04</td>
<td>.03</td>
<td>.04</td>
<td>.04</td>
<td>.05</td>
</tr>
</tbody>
</table>

Notes: $t$ statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dependent variable: norm enforcement (4.A.4a/b: direct, 4.A.4c/d: indirect, 4.A.4e/f: total).
4.A.2 Additional figures

Figure 4.A.1: Control condition and littered condition in Bern.
Chapter 5

The sanctioning dilemma: A quasi-experiment on social norm enforcement on the train

Abstract
Numerous laboratory experiments have established peer-sanctioning as an important driver of norm compliance and cooperation in human groups. However, systematic evidence of peer-sanctioning occurring in the field is still rare. Here we present results from a field experiment investigating the enforcement of the silence norm on the train. We let a confederate play loud music on their mobile phone in an open-plan train car and measure the time until a negative sanction occurs (if any). The silence norm is enforced in 45 of 90 cases, enforcement rates do not differ across silent and non-silent area cars, and the more passengers are in a car, the more likely the silence norm is to be enforced. Passengers’ propensities to enforce the silence norm are in line with predictions derived from the asymmetric volunteer’s dilemma (VOD). The higher a passenger’s net benefit from enforcing the silence norm, the more likely the passenger is to negatively sanction the norm breaker. Our findings extend the validity of results from lab experiments which conceive of the second-order free-rider problem as a VOD.

1 This chapter is an edited version of the following paper: Przepiorka, Wojtek and Berger, Joël (2015). “The sanctioning dilemma: A quasi-experiment on social norm enforcement on the train.” Unpublished working paper.
5.1 Introduction

Social norms and their enforcement through positive and negative sanctions are indispensable for social cohesion and the functioning of societies. While many social norms are formalized in terms of paragraphs in legal codes, far from all social norms can be formally defined and even fewer enforced at all times by a legitimate authority. Although not unheard of, most people are reluctant to instantly call the police if someone is smoking in a non-smoking area, listening to loud music in public transport, jumping the queue in the supermarket, dressing inappropriately at a wedding, or free riding on a group project at work. Such norm violations are, if at all, negatively sanctioned (i.e. punished) by the norm breakers’ peers. Peer-punishment comes in many forms; everyday norm violations can be punished by disapproving words, looks or gestures, by negative gossip and ostracism, or acts of aggression (Brauer and Chekroun, 2005; Feinberg, Willer, and Schultz, 2014; Gordon, Madden, and Lea, 2014; Guala, 2012). But why do people engage in the informal punishment of their deviant peers – even at the risk of suffering retaliation?

In most instances, peer-punishment can be explained by the fact that the norm breaker and the punisher are going to meet again in the future (e.g. next-door neighbors, colleagues at work, individuals and organizations with a reputation to lose, etc.); that is when the benefits of establishing a cooperative environment outweigh the costs of upholding the punishment threat in the long run (Gächter, Renner, and Sefton, 2008). However, in many instances of everyday norm violations it is a priori unlikely that the norm breaker and the affected parties will encounter each other again. Yet, many of those who are affected by the norm violation object and punish the norm breaker. Punitive preferences have been suggested as a proximate explanation for the punishment of norm breakers in one-time-only encounters. Theoretical considerations (Gintis, Smith, and Bowles, 2001) as well as empirical findings from laboratory experiments (Fehr and Gächter, 2002; Ostrom, Walker, and Gardner, 1992) corroborate that many humans are indeed inclined to ‘sacrifice resources for rewarding fair and punishing unfair behavior even if this is costly and provides neither present nor future material rewards’ (Fehr and Gächter, 2002, p. 3, emphasis in original).

---

2 In the literature on peer punishment as a means of enforcing cooperation, by punishing, an actor incurs a cost to inflict a cost on a deviant peer. Although such punishment does not always bring about a higher level of cooperation (e.g. Herrmann, Thöni, and Gächter 2008), it communicates disapproval regarding free-riding behavior and, as such, is often understood as enforcing a cooperation norm. Although they do not need to mean the same thing in general, the terms peer-punishment, negative sanctions and norm enforcement are often used as synonyms.
These findings have been contested as to their empirical and methodological validity (Guala, 2012; Henrich, Elreath, et al., 2006; Herrmann, Thöni, and Gächter, 2008; Krasnow et al., 2012; Levitt and List, 2007; Nikiforakis, 2008; West, Mouden, and Gardner, 2011). However, only relatively recently have experimental social psychologists (e.g., Brauer and Chaurand 2010; Brauer and Chekroun 2005; Chekroun and Brauer 2002; Diekmann, Jungbauer-Gans, et al. 1996) and experimental economists started addressing issues of external validity by investigating peer-punishment in the field (Balafoutas and Nikiforakis, 2012; Balafoutas, Nikiforakis, and Rockenbach, 2014). For example, Balafoutas and Nikiforakis (2012) conducted a field experiment to study the punishment of norm breakers in a large subway station in Athens. A confederate violated the “do not litter in public places” norm and the “stand right walk left on the escalator” norm 150 times each. Their results show that the no-litter norm was enforced in 4% of the cases and the escalator norm was enforced in 19.3% of the cases (for replications regarding the no-litter norm and the escalator norm in other cities see Berger and Hevenstone 2015, and Wolbring, Bozoyan, and Langner 2013, respectively). Based on potential punishers’ statements in a follow-up questionnaire, Balafoutas and Nikiforakis (2012) conjecture that the low enforcement rate of the no-litter norm is due to potential punishers being afraid of retaliation. That is, since the no-litter norm is commonly better known than the escalator norm, the violation of the no-litter norm is a credible sign of the norm breaker’s intentionality, anti-social preferences and thus, his or her higher propensity to retaliate punishment. Ultimately, the relatively low punishment rates reported in these field studies cast further doubt on the external validity of results regarding peer-punishment obtained in laboratory experiments.

We contribute to this literature by pointing out an understudied aspect of peer-punishment. Most studies investigating peer-punishment in the lab or in the field have neglected the strategic nature of many punishment situations. Like many scholars theorizing about social norms did before us (Axelrod, 1986; Coleman, 1990; Demsetz, 1967; Horne, 2001; Ostrom, 2000; Posner, 2000; Ullmann-Margalit, 1977), we start from the assertion that the violation of a social norm creates negative externalities for a group of people and thus, the enforcement of the norm creates a (second-order) public good (Heckathorn, 1989; Oliver, 1980; Yamagishi, 1986). Moreover, in many instances of every-day norm violations, a collective demand for negative sanctions is created that can be satisfied by one actor alone. And if the benefits outweigh the costs of producing the second-order public good, it can be in the actor’s self-interest to produce it for the entire group (Olson, 1965; Roberts, 2013).

In the next section, we detail our theoretical argument and derive our hypotheses. In Section 5.3, we describe two quasi-experimental field studies we
conducted to test our hypotheses. We investigate the enforcement of the silence norm among passengers who happen to sit in the same open-plan train car. In 90 instances, we let a confederate play annoying music on their mobile phone and measure the time until a sanction occurs (if any) as well as relevant contextual variables and passengers’ characteristics. In the first study, comprising 31 interventions with a male confederate, we merely investigate the prevalence of norm enforcement. In the second study, comprising 59 interventions with female confederates, we also systematically vary whether the intervention takes place in a silent-area car or in a non-silent-area car. We present the results of our two studies in Section 5.4. In the last section, we discuss the potential of game theoretic models to explain norm enforcement, the limitations of our study and possibilities for future research.

5.2 Theory and hypotheses

The violation of the silence norm in an open-plan train car constitutes the first order free-rider problem. The quiet environment can be conceived of as a common pool resource (CPR), and a person behaving loudly over extracts the CPR at a cost for those who do not. The negative externality produced by the norm breaker manifests itself in the increased costs of appropriating one’s fair share of the CPR (e.g. Ostrom, Walker, and Gardner 1992). Thus punishing the norm breaker re-establishes a quiet environment for the remaining duration of the train ride and, as such, produces a public good. However, the punishment of the norm breaker is costly and therefore subject to a (second-order) free-rider problem (Heckathorn, 1989; Oliver, 1980; Yamagishi, 1986). This second-order free rider problem can be modelled with the volunteer’s dilemma (Diekmann, 1985; Raihani and Bshary, 2011).

5.2.1 Diffusion of responsibility in the symmetric volunteer’s dilemma

The volunteer’s dilemma (VOD) is a step-level public good game where only one actor’s contribution is necessary and sufficient to produce the public good for the entire group (Diekmann, 1985; Palfrey and Rosenthal, 1984). In our case, punishing the norm breaker to re-establish silence for all passengers in the train car constitutes the (second-order) public good. More formally, a public good of value $\sum U_i$ for a group of size $n \geq 2$ is produced by a single actor $i$ choosing to punish the norm breaker at a cost $K_i$ where $U_i > K_i > 0 \forall i$. The public good is not provided if all actors choose not to punish the norm breaker, and there is
Table 5.1: The volunteer’s dilemma (VOD)

<table>
<thead>
<tr>
<th>Passenger i’s choice</th>
<th>0</th>
<th>1</th>
<th>...</th>
<th>n − 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>punish norm breaker</td>
<td>$U_i - K_i$</td>
<td>$U_i - K_i$</td>
<td>$U_i - K_i$</td>
<td>$U_i - K_i$</td>
</tr>
<tr>
<td>don’t punish norm breaker</td>
<td>0</td>
<td>$U_i$</td>
<td>$U_i$</td>
<td>$U_i$</td>
</tr>
</tbody>
</table>

a welfare loss if more than one actor punishes the norm breaker. The VOD thus has $n$ welfare maximizing, pure strategy Nash equilibria, in which one passenger punishes the norm breaker while all other $n − 1$ passengers do not. Table 5.1 presents the payoff structure of the VOD from passenger $i$’s perspective.

The social dilemma comprised in the VOD arises from the fact that, without communication, it is difficult for a group of passengers to tacitly agree on which one of them should punish the norm breaker. Although the benefits outweigh the costs of punishment (i.e. $U_i > K_i > 0 \ \forall \ i$), free riding on another passenger’s punishment is even more beneficial. As a consequence, the entire group may end up suffering from the negative externalities produced by the norm violation, while waiting for someone else to punish the norm breaker. Assuming the symmetric VOD, where all passengers have the same benefits from and costs of punishing the norm breaker (i.e. $U_i = U_j$ and $K_i = K_j \ \forall \ i \neq j$), this diffusion of responsibility effect (Darley and Latané, 1968; Diekmann, 1985) can be derived from the mixed strategy equilibrium (MSE).

$$p_i^* = 1 - \frac{1}{n-1}\sqrt{K_i/U_i} \quad (5.1)$$

In the MSE, each passenger $i$ punishes the norm breaker with a certain probability $p_i^*$. Furthermore, with $q_i^* = 1 - p_i^*$, we can calculate the probability $p^*$ that at least one passenger will punish the norm breaker and the second-order public good will be produced (see Diekmann 1985 or Palfrey and Rosenthal 1984 for the derivation of equations 5.1 and 5.2).

$$p^* = 1 - \prod_{i=1}^{n} q_i^* \quad (5.2)$$

Consistent with the diffusion of responsibility effect, both $p_i^*$ and $p^*$ are decreasing in $n$, the size of the group. Based on our theoretical argument thus far, we can derive our first hypothesis:

**H1:** The larger a group of passengers sitting in the same train car is, the less likely it is that the silence norm will be enforced.
5.2.2 Tacit coordination in the asymmetric volunteer’s dilemma

Recall, however, that H1 derives from the MSE of the symmetric VOD, which assumes the same benefits and costs for all passengers. In the punishment situation under scrutiny, this assumption is likely to be violated. First, some passengers try to make the best of the time they spend on a train by, for instance, working or reading, whereas other passengers simply enjoy looking out of the window, dozing, consuming media (usually using their earphones) or chatting with others (possibly on the phone). Thus the punishment of the norm breaker will produce a greater benefit for those passengers who require a quiet environment to carry on their activities. Second, the costs arise from the actual act of punishing the norm breaker. That is, getting up, approaching the norm breaker, making an assertive statement that the music should be turned down, and possibly facing an aggressive response all sum up to an individual’s total cost of punishment. While male passengers derive on average the same benefits from enforcing the silence norm as female passengers, they are likely to have lower costs. Men are on average taller and more aggressive than women; these factors have and are believed to have a deterring effect on the norm breaker, making an aggressive response by the norm breaker less likely. Third, irrespective of their activity and gender, passengers sitting closer to the source of noise will be more disturbed and, therefore, the punishment of the norm breaker will generate a larger benefit for them. Moreover, sitting closer to the norm breaker also reduces the costs of punishment as getting up to approach the norm breaker might not be necessary.

This set of observable factors, i.e. passengers’ distance to the source of noise, their activity and gender, generate heterogeneity in passengers’ net benefits from punishing the norm breaker. This heterogeneity can be accounted for with the asymmetric VOD (Diekmann, 1993), where \( U_i \neq U_j \) and/or \( K_i \neq K_j \) \( \exists i \neq j \). Based on the asymmetric VOD, it has been shown theoretically (Diekmann, 1993; He, Wang, and Li, 2014) as well as empirically (Diekmann and Przepiorka, 2015; Przepiorka and Diekmann, 2013) that the person with the largest net benefit from norm enforcement will be the most likely to sanction the norm breaker (see also Brauer and Chekroun 2005). Correspondingly, our next hypotheses are as follows:

**H2-1:** The closer a passenger sits to the source of noise, the more likely it is that this passenger will enforce the silence norm.

**H2-2:** The more silence a passenger’s activity requires, the more likely it is that this passenger will enforce the silence norm.
**H2-3:** A male passenger will be more likely to enforce the silence norm than a female passenger.

Under asymmetric conditions, the diffusion of responsibility effect hypothesized under H1 will be less likely to occur. Passengers’ perceivable differences in net benefits from norm enforcement facilitate the group’s tacit coordination on the passenger with the highest net benefit to sanction the norm breaker (Przepiorka and Diekmann, 2013). Thus in asymmetric situations, group size will have less bearing on the probability that a sanction occurs. Hypotheses H1 and H2 can thus be seen as testing two alternative models – the symmetric and the asymmetric VOD, respectively.

### 5.2.3 Decreasing benefits in the volunteer’s timing dilemma

Both the symmetric and asymmetric VOD are simultaneous move games, where all actors make their decisions at the same time, without knowing what other group members decide. However, the punishment situation on the train is dynamic in that all passengers can observe all other passengers’ moves and refrain from punishing the norm breaker after someone else does. A dynamic version of the VOD was first described by Bliss and Barry (1984). Later, Weesie (1993) suggested a similar dynamic conceptualization of the VOD and called it the volunteer’s timing dilemma (VTD). In both strands of literature, it is assumed that the benefits of establishing the (second-order) public good decrease over time. In the punishment situation on the train, the benefits of enforcing the silence norm decrease as the train approaches its destination. At the same time, the costs of enforcing the silence norm stay constant. We can therefore state our next hypothesis.

**H3:** The closer the train is to its destination, the less likely will the silence norm be enforced.

Moreover, hypotheses H1 and H2 can also be derived from the symmetric and asymmetric VTD, respectively, but now with the timing of norm enforcement as the dependent concept (Weesie, 1993). For instance, H1 and H2-1 can be rephrased as follows: The larger a group of passengers sitting in the same train car is, the *later* the silence norm will be enforced; the closer a passenger sits to the source of noise, the *earlier* this passenger will enforce the silence norm. Given the short time frame of an intervention (the music stops after 4 minutes and 48 seconds if no punishment occurs) we treat the two types of hypotheses H1 and H2 as equivalent. In the results section, we estimate both models with the probability, and timing of norm enforcement as dependent variables.
5.2.4 Norm salience

Most people, at least in Switzerland, where the two studies were conducted, would expect most others to agree that overly loud behavior in an open-plan train car is inappropriate and should be negatively sanctioned. However, passengers’ beliefs might be different on trains in which there is an explicit distinction between silent-area cars and non-silent-area cars. In these trains, passengers in a silent-area car should feel more entitled to reprimand someone breaking the silence norm than passengers in a non-silent-area car.

**H4:** The silence norm will be more likely to be enforced in a silent-area car than in a non-silent-area car.

However, based on what Balafoutas and Nikiforakis (2012) conjecture based on their findings (see above), we might also expect the opposite. Since the breaking of the silence norm in the silent-area car, where this norm is made explicit by visible signs in the car, would be indicative of the norm breaker’s retaliation potential, passengers in a silent-area car should feel more reluctant to enforce the silence norm than passengers in a non-silent-area car.

**H4a:** The silence norm will be less likely to be enforced in a silent-area car than in a non-silent-area car.

5.3 Materials and methods

Trains provide an ideal setting for conducting quasi-experiments as they restrict passengers in their action space for some time and, at the same time, allow for careful measurement and controlled intervention (Levitt and List, 2007). In this section, we detail the design and procedures of the two studies we conducted to test our hypotheses. Since the two studies differ in a few but important respects, we start with an in-depth description of study 1 and then only point out the differences between the first and second study. In the last part of this section, we give a brief description of the data that we collected.

5.3.1 Procedures and design

The first study was conducted on the intercity train (IC) between Zurich and Bern. This train comprises non-silent-area cars only, takes 58 minutes between destinations and has no stops in between. In total 14 train rides were taken, seven from Zurich to Bern and seven in the opposite direction. Between one and four interventions per train ride and 35 interventions in total were recorded. The first
two train rides were used to test the intervention. We discarded four out of six interventions recorded on these two rides because they differ from subsequent interventions in the volume at which the music was played. In total, 31 valid interventions were conducted and data on 204 passengers were collected. The number of interventions that could be recorded on one ride depended on how crowded the train was and therefore on whether the experimenters could find an opportunity to carry out the intervention (see below). As mentioned earlier, the benefits from sanctioning a norm violation decrease as the train approaches its destination, whereas the costs stay constant. To avoid the costs of enforcing the silence norm exceeding the benefits and thus the punishment situation failing to comprise a VOD, no interventions were started in the last 15 minutes of the ride. The interventions were conducted by two experimenters: one norm breaker, who violated the silence norm by playing a loud song on his mobile phone, and one observer, who recorded the contextual variables, passenger characteristics and passengers’ reactions to the norm violation. It was always the same person who was the norm breaker or the observer.

The following procedures were followed (also see figure 5.1): The norm breaker (X) enters the train car and takes a seat in an empty compartment. Then the observer (O) enters the car and takes a seat sufficiently distant from the norm breaker in order not to interfere with the intervention. The observer takes about five minutes to record the observable characteristics of the other passengers present in the car (e.g., passengers P1 through P6 in figure 5.1) on a prepared form (one version of the form is presented in figure 5.A.1 in the appendix; translated from German by the authors). The other passengers’ gender, estimated age, activity and their position in the car are registered by the observer. Then, the observer indicates to the norm breaker inconspicuously that he has finished recording the situation data and the norm breaker can start the intervention. Then the norm breaker starts playing the song “Robot Rock” by Daft Punk on his mobile phone at an annoying volume. The intervention ends if a passenger in the car sanctions the norm breaker or when the song ends after 4 minutes and 48 seconds. The norm breaker stops the music only if another passenger asks him directly to do so; mere gestures of disapproval or exclamations not directed at the norm breaker or not comprising a request to stop the music are not counted as sanctions. During the intervention, the norm breaker avoids eye contact with the other passengers in order not to make it too easy for them to inflict negative sanctions on him. After the intervention, the norm breaker and observer leave the car one after the other. Before the next intervention, the norm breaker helps the observer to complete

3 The song can be listened to here: http://www.youtube.com/watch?v=r4rhX6iB-2o (retrieved December 15, 2014). Within each of the two studies, the volume at which the song is played is kept constant across interventions.
The sanctioning dilemma

Figure 5.1: Example of passenger seating in an intercity (IC) train car

The form with the other passengers’ characteristics and reactions that the observer might have missed from his position during the intervention.

The second study differs from the first study in three important respects. First, all interventions were conducted on the direct line between Zurich and Basel, which takes 53 minutes without stops. Second, the norm breaker was always female. However, unlike in our first study, five different pairs of norm breaker and observer conducted the interventions. Third, we systematically varied whether the silence norm was violated in a silent-area car or in a non-silent-area car. It was only possible to vary the car type in intercity express (ICE) trains, in which silent-area cars are marked as such with conspicuous signs and non-silent-area cars lack such signs. Sometimes, taking the ICE was not possible. In these cases, interventions were conducted in the same IC trains (comprising non-silent-area cars only) as in study 1. In total, 59 interventions were conducted and data on
823 passengers were collected. Everything else was done in the same way as in study 1. Table 5.2 summarizes the most important aspects of the two studies.

### 5.3.2 Data

The data we have collected have multiple levels. The two levels that are relevant for our data analysis are the train car level, at which the $N_1 = 90$ interventions took place, and the (individual) passenger level, with a total of $N_2 = 1027$ cases. The most important car-level variables that we recorded are the car type, the number of passengers in the car (including the observer but not the norm breaker), and the minutes left at the start of an intervention until the train reaches its destination. Recall that no interventions were started in the last 15 minutes of the train ride. The most important passenger-level variables that we recorded are the passengers’ activities, their discernible gender and estimated age, and their location (i.e. seat) in the car. Based on each passenger’s seat number and the seat number of the norm breaker (both the norm breaker’s and the observer’s seats were also recorded), each passenger’s distance to the norm breaker (in terms of number of seats) was calculated. Table 5.3 lists the descriptive statistics of these variables. The data are available from the authors on request.

The main outcome variables that we recorded at the car level are whether or not the norm breaker was negatively sanctioned by another passenger and, if he or she was sanctioned, the time in seconds until the sanction occurred. At the passenger level, we also recorded the passenger who sanctioned the norm breaker. That is, we marked this passenger as “punisher” on the data recording form (see figure 5.A.1 in the appendix). In study 1, we also recorded passengers’ reactions which did not qualify as sanctions. However, recording these reactions accurately was not always possible from the observer’s location in the car. In order for
Table 5.3: Descriptive statistics of car- and passenger-level variables

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>missing</th>
<th>mean</th>
<th>median</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Car-level variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>car type</td>
<td>90</td>
<td>0</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-silent (ICE)</td>
<td>28</td>
<td>0</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>silent (ICE)</td>
<td>20</td>
<td>0</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-silent (IC)</td>
<td>42</td>
<td>0</td>
<td>0.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of passengers in car</td>
<td>90</td>
<td>0</td>
<td>12.39</td>
<td>10</td>
<td>7.10</td>
</tr>
<tr>
<td>min. to destination</td>
<td>90</td>
<td>0</td>
<td>34.82</td>
<td>35.5</td>
<td>12.41</td>
</tr>
<tr>
<td><strong>Passenger-level variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>passenger’s activity</td>
<td>1027</td>
<td>0</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wears earphones on</td>
<td>71</td>
<td>0</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>talks to others or phone</td>
<td>251</td>
<td>0</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reads or works</td>
<td>386</td>
<td>0</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dozes, eats, looks out window</td>
<td>219</td>
<td>0</td>
<td>0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other unknown</td>
<td>100</td>
<td>0</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>is female</td>
<td>1024</td>
<td>3</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>estimated age</td>
<td>1024</td>
<td>3</td>
<td>39.55</td>
<td>35</td>
<td>16.80</td>
</tr>
<tr>
<td>distance to norm breaker</td>
<td>1027</td>
<td>0</td>
<td>13.57</td>
<td>12</td>
<td>9.24</td>
</tr>
</tbody>
</table>

the observer to be better able to focus on recording the contextual variables and passenger characteristics, we decided to discontinue recording passengers’ other reactions in study 2. We therefore do not report results on passengers’ other reactions in this paper.

### 5.4 Results

Punishment occurs in 50% of our 90 interventions. The punishment rate is lower in study 1 (29%) than in study 2 (61%). This statistically significantly difference ($\chi^2_{(1)} = 8.32$, $p = 0.004$) could be due to the fact that the norm breaker is male in study 1 and female in study 2. However, since it is not only the norm breaker’s gender that differs across the two studies (see table 5.2), alternative explanations cannot be ruled out.

Surprisingly, as is apparent from figure 5.2, the sanctioning rate does not differ significantly across car types ($\chi^2_{(2)} = 1.63$, $p = 0.443$). The silence norm is enforced at a slightly higher rate in ICE trains in non-silent-area cars (57%) than
in silent-area cars (55%), and norm enforcement is lowest in IC trains in non-
silent-area cars (43%). The latter rate, although not significantly different from
the other two, might be lower because most interventions in IC trains were con-
ducted in study 1, where the norm breaker was male. In any case, this evidence
does neither support our hypothesis H4 nor hypothesis H4a; the two counteracting
mechanisms hypothesized under H4 and H4a might be at work at the same
time or not at all. We will come back to this point in the discussion section.

We now turn to multiple regression analyses to further test our hypotheses
(table 5.4). We continue with testing our car-level hypotheses (H1, H3, H4 and
H4a); in the second part of this section, we will test our passenger-level hypothe-
ses (H2-1 through H2-3). Apart from the variables measuring and operational-
zizing the concepts in our hypotheses, all models in table 5.4 also account for
time-constant unobserved (and observed) differences across the two studies.

The first model in table 5.4 is a logistic regression of the probability that the
silence norm will be enforced. Accounting for other factors, the evidence of no
difference in enforcement rates between silent and non-silent area cars in ICE
trains does not change; hypotheses H4 and H4a remain unsupported. Moreover,
we do not find support for the diffusion of responsibility effect hypothesized under
H1. In fact, the statistically significant coefficient estimate suggests that the more
other passengers are in a car, the more likely it is that the silence norm will be
enforced. However, recall that a test of H1 assumes passengers to be homogenous

Figure 5.2: Proportion of sanctions across car types
### Table 5.4: Regression models

<table>
<thead>
<tr>
<th></th>
<th>Sanction occurred (0/1)</th>
<th>Time to sanction (cens. at 288&quot;)</th>
<th>Passenger sanctioned norm breaker (0/1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Logit 1</td>
<td>Cox PHM</td>
<td>Logit 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FE logit</td>
</tr>
<tr>
<td><strong>H1</strong>: number of passengers in car</td>
<td>0.141**</td>
<td>0.071**</td>
<td>-0.033</td>
</tr>
<tr>
<td></td>
<td>-0.047</td>
<td>-0.023</td>
<td>-0.024</td>
</tr>
<tr>
<td><strong>H3</strong>: minutes to destination</td>
<td>0.050*</td>
<td>0.021+</td>
<td>0.026**</td>
</tr>
<tr>
<td></td>
<td>-0.021</td>
<td>-0.012</td>
<td>-0.009</td>
</tr>
<tr>
<td><strong>H4, H4a</strong>: car type (non-silent ICE car is ref. cat.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>silent (ICE)</td>
<td>-0.26</td>
<td>-0.092</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>-0.617</td>
<td>-0.371</td>
<td>-0.245</td>
</tr>
<tr>
<td>non-silent (IC)</td>
<td>2.424*</td>
<td>1.207**</td>
<td>0.498</td>
</tr>
<tr>
<td></td>
<td>-0.93</td>
<td>-0.424</td>
<td>-0.328</td>
</tr>
<tr>
<td><strong>H2-1</strong>: passenger’s distance to the norm breaker</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.137***</td>
<td>-0.128***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.03</td>
<td>-0.029</td>
<td></td>
</tr>
<tr>
<td><strong>H2-2</strong>: passenger’s activity (reads or works is ref. cat.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wears earphones</td>
<td>-1.154</td>
<td>-1.915+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.706</td>
<td>-1.07</td>
<td></td>
</tr>
<tr>
<td>talks with others or on the phone</td>
<td>-0.943*</td>
<td>-0.812</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.4</td>
<td>-0.501</td>
<td></td>
</tr>
<tr>
<td>dozes, eats, looks out of window, etc.</td>
<td>-0.953*</td>
<td>-0.973+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.476</td>
<td>-0.51</td>
<td></td>
</tr>
<tr>
<td>other / unknown</td>
<td>-0.445</td>
<td>-0.527</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.572</td>
<td>-0.7</td>
<td></td>
</tr>
<tr>
<td><strong>H2-3</strong>: passenger’s gender (male is ref. cat.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>-0.127</td>
<td>-0.378</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.328</td>
<td>-0.343</td>
<td></td>
</tr>
<tr>
<td><strong>Control (study 1 is ref. cat.)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>study 2</td>
<td>2.985**</td>
<td>1.682***</td>
<td>1.369***</td>
</tr>
<tr>
<td></td>
<td>-0.96</td>
<td>-0.456</td>
<td>-0.376</td>
</tr>
<tr>
<td>Const.</td>
<td>-6.516***</td>
<td>-2.845***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.618</td>
<td>-0.792</td>
<td></td>
</tr>
<tr>
<td><strong>N1</strong></td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td><strong>N2</strong></td>
<td></td>
<td>1024</td>
<td>619</td>
</tr>
<tr>
<td><strong>pseudo R²</strong></td>
<td>0.21</td>
<td>0.06</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>χ² (df)</strong></td>
<td>17.64(5)**</td>
<td>24.21(5)**</td>
<td>55.84(11)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35.82(6)**</td>
</tr>
</tbody>
</table>

Notes: The table lists coefficient estimates with standard errors in parentheses (+\(p < 0.1\), *\(p < 0.05\), **\(p < 0.01\), ***\(p < 0.001\), for two-sided tests). The first two models are based on car-level data only and estimate the probability (Logit 1) and speed (Cox PHM) with which a sanction occurs, conditional on car-level characteristics. The two models are estimated with heteroscedasticity robust standard errors. The last two models are based on passenger-level data and estimate the probability of a passenger to enforce the silence norm. The third model (Logit 2) accounts for car-level clustering and the fifth model (FE Logit) accounts for car fixed effects.
in terms of their costs of and benefits from enforcing the silence norm. In our train setting, this assumption is likely violated, in which case other predictions suggest themselves. We will come back to this point in the second part of this section, where we test our passenger-level hypotheses. Finally, we find support for hypothesis H3. The closer the train is to its destination, the less likely it is that the silence norm will be enforced. However, using models for binary outcome variables, such as logit, does not entirely suit the process that generated our data. In particular, these models do not account for the fact that our dependent variable not only tells us whether a sanction occurred or not, but also when it occurred if it did. If the norm breaker was negatively sanctioned, the sanction occurred after 124 seconds on average (median = 120 sec.; sd = 87.1). What is more, our dependent variable is censored at 288 seconds – the length of the song played to provoke negative sanctions. Our second model accounts for these characteristics of our dependent variable.

The second model in table 5.4 is a proportional-hazard model (PHM) for continuous time-to-event data, aka Cox regression (e.g. Hosmer, Lemeshow, and May 2008). Here too we find a higher number of passengers to be positively related to an earlier occurrence of sanctions, no difference between silent and non-silent-area cars in ICE trains, and a decreasing propensity to enforce the silence norm over time. Although the coefficient estimate of the “minutes to destination” variable is now only significant at the 10%-level (p = 0.074), we can conclude that our results thus far are robust across different model specifications.

Before we turn to testing our passenger-level hypotheses, we will first discuss the results of our third model with regard to the three car-level variables. Model three in table 5.4 is a logistic regression of the probability that the silence norm will be enforced by a particular passenger. Note first that the coefficient estimate for the number of passengers, although statistically insignificant, is now negative. This does not mean that once we control for possible heterogeneity among the passengers in a car, we might be able to identify a diffusion of responsibility effect (hypothesis H1). If anything, the negative coefficient estimate is an artifact. Since there can be only one punisher (recall that an intervention ends as soon as a negative sanction occurs), larger group sizes decrease the relative frequency of a single punisher. Moreover, model four does not give new evidence regarding hypotheses H4 and H4a, and it confirms that it is indeed the individual passengers’ propensity to enforce the silence norm that decreases as the train approaches its destination (H3).

As mentioned earlier, our train setting is more likely to resemble an asymmetric than a symmetric volunteer’s dilemma. Asymmetry implies that passengers differ in the costs of and benefits from enforcing the silence norm. Hypotheses H2-1 through H2-3 are based on this assumption and predict that passengers with
a higher net benefit from enforcing the silence norm will be more likely to punish the norm breaker. We find good evidence in support of two of the three hypotheses. First, passengers who sit closer to the norm breaker are significantly more likely to enforce the silence norm (H2-1). Second, passengers who are engaged in activities that require silence are significantly more likely to enforce the silence norm (H2-2). That is, passengers who read or work are more likely to enforce the silence norm than passengers who wear earphones (although $p = 0.102$), who talk with others or on the phone, or who doze, eat or look out of the window. Finally, we find no support for hypothesis H2-3. Although female passengers tend to be less likely to punish the norm breaker, the negative coefficient estimate is statistically insignificant. We also estimated models interacting the norm breaker’s gender with the norm enforcer’s gender, and we estimated a mixed-effects model with random intercepts at the car level, but these analyses did not produce any new insights (not shown).

One could object that something else, e.g. something related to the norm breakers or the train cars, that we failed to record or keep constant when conducting our studies, might be causing the individual-level effects. After all, our data do not stem from a controlled experiment in which subjects are randomly assigned to experimental conditions. To address some of the objections that could arise in this regard, we also estimate a logistic regression with car fixed effects (aka conditional logit) in our fourth and last model (see e.g. Angrist and Pischke 2009; Halaby 2004; Snijders and Bosker 2012). Although, by doing this, we lose data on all passengers who sat in cars in which no sanction occurred (there is no variation in the dependent variable), we find our previous results hardly affected. First, there is still a substantial and statistically significant negative effect of a passenger’s distance to the source of noise on the passenger’s propensity to enforce the silence norm. Second, passengers who read or work are still more likely to punish the norm breaker than passengers wearing earphones, talking with others or on the phone (although $p = 0.105$), or dozing, eating or looking out of the window. Finally, although now substantially stronger, the lower propensity of female passengers to punish norm breakers remains statistically insignificant.

5.5 Discussion and conclusion

Our paper contributes to the research on how peer-punishment promotes cooperation and enforces social norms in two important respects. First, we argue that most research on peer-punishment has by and large neglected the strategic nature of the punishment situation. Many everyday norm violations produce a demand for negative sanctions in a group of people, which can be satisfied by
one person alone. In these cases, a coordination problem can arise with regard to which group member should punish the norm breaker. This second-order free-rider problem can be formally described by a step-level public good game such as the VOD (Diekmann, 1985; Diekmann, 1993; Raihani and Bshary, 2011). Previous research using computerized laboratory experiments has shown that conceptualizing the second-order free-rider problem as a VOD can lead to new insights into the mechanisms of second-order public good provision. In particular, these studies have demonstrated that punishment cost heterogeneity has a major impact on the efficiency and effectiveness with which peer-punishment promotes (first-order) cooperation (Diekmann and Przepiorka, 2015; Przepiorka and Diekmann, 2013).

Our second major contribution to the literature lies in our extending the validity of these previous findings by showing that similar results can be obtained in domains of social life which resemble the set-ups created in the experimental lab. We conduct two field studies to investigate the enforcement of the silence norm by groups of passengers sitting in open-plan train cars. In 90 instances, we let a confederate play annoying music on their mobile phone and measure the time until a sanction occurs. Playing loud music in an open-plan train car produces negative externalities and creates a second-order cooperation problem that resembles the VOD. In our study, the norm breaker was punished and the silence norm was enforced in 45 out of 90 instances.

Furthermore, our findings support hypotheses derived from the asymmetric VOD (H2-1 through H2-3), in which the interacting parties are assumed to differ in their net benefits from producing the second-order public good. We find that passengers with lower costs of and/or higher benefits from punishing the norm breaker are more likely to do so. At the same time, our evidence clearly speaks against the diffusion of responsibility hypothesis (H1) – the hypothesis that the sanctioning rate will decrease with group size. We find – to the contrary – that the more passengers are in a car, the more likely it is that the silence norm will be enforced. It is important to stress that the diffusion of responsibility hypothesis can be derived from the symmetric VOD, which assumes homogenous actors. Future studies testing the diffusion of responsibility hypotheses should take care when establishing the conditions under which the assumption of homogenous actors is justified. Individuals differ in many significant ways. In most social interactions occurring naturally, there will be at least one actor who has different preferences, endowments and/or constraints than his or her peers.

More obvious expectations find no support in our data. Norm enforcement rates do not differ across train cars with and without a sign making a silence norm being in effect explicit. On the one hand, we expected that passengers sitting in a silent-area car would feel more entitled to punish the norm breaker and therefore
would do so more than passengers sitting in a non-silent-area car (H4). On the other hand, breaking a norm that is made explicit by clearly visible signs may induce fear of retaliation in passengers as the norm breaker will be more likely to be perceived as someone seeking a quarrel. Fear of retaliation may reduce the enforcement rate in silent-area cars as compared to non-silent area cars (H4a). Since the two mechanisms hypothesized under H4 and H4a work in opposite directions, our results are inconclusive inasmuch as they leave open whether the two mechanisms neutralize each other or are not at work at all.

This is also where the limitations of our study become apparent. As the title of this paper reveals already, ours is not an experiment in which subjects are randomly assigned to experimental conditions. However, random assignment would have greatly facilitated the exclusion of potential confounders. For example, some passengers choose silent-area cars exactly because they embrace a general reluctance to reprimand others for their rude behavior, possibly wrongly believing that others sitting in silent-area cars with them will do it on their behalf. Future studies aiming at disentangling the two mechanisms hypothesized under H4 and H4a will need to find (or create) a setting in which the norm breaker can be negatively sanctioned but cannot respond to the sanction. Only then will it be ruled out that fear of retaliation might curb the entitlement to negatively sanction a norm breaker.

Despite its limitations and unresolved issues, the study at hand shows that it can be fruitful to model social interactions in general, and situations of social norm enforcement in particular, in game theoretic terms. Individuals’ considerations regarding whether to punish norm breakers do not evade strategic thinking, which accounts for the costs and benefits of their and other bystanders’ potential actions. Also in line with this idea are our findings in support of hypothesis H3, suggesting that passengers indeed acknowledge the decreasing benefit of enforcing the silence norm as the train approaches its destination. Finally, confirming one of Olson’s (1965) early intuitions, our study shows that as long as the benefits outweigh the costs of producing the (second-order) public good, it can be in an actor’s self-interest to produce it for the entire group by him or herself.
Figure 5.A.1: Form used to record the setting of an intervention in an IC car
Chapter 6

Eye spots do not increase altruism in children

Abstract\textsuperscript{1} Identifying the evolutionary mechanisms that support human altruism in anonymous one-shot interactions has proven to be difficult because straightforward explanations like kinship and repeated interactions do not naturally apply. Nonetheless, an influential hypothesis argues that humans retain a reciprocity-based psychology from the late Pleistocene, and this psychology produces the apparent altruism observed in contemporary settings where interactions are not repeated. The empirical support for this claim rests on studies showing that altruism among adults increases in the presence of payoff-irrelevant cues that imply one’s behavior is observable. Stylized eye spots and faces are prominent examples of the cues researchers have used. Overall the empirical results are thoroughly mixed. To address this problem, we conducted an eye spot experiment with children. Children are especially informative subjects for this kind of study because altruism develops rapidly in early childhood with potential differences between girls and boys. This means that, if anonymous one-shot altruism depends primarily on payoff-irrelevant social cues, then variation in altruism by age or sex should be closely linked to variation in sensitivity to eye spots. Children are additionally informative as subjects because they do not have strong executive control over decision making. If the response to eye spots is automatic, the behavioral effects should be unusually pronounced in children because they cannot employ executive control to override the automatic response. Although we did find variation in altruism among children, with boys significantly less generous than girls, we found no variation in sensitivity to eye spots. More generally, we found no eye spot effects of any kind. Although

\textsuperscript{1} This chapter is an edited preprint version of the following paper: Vogt, Sonja, Efferson, Charles, Berger, Joëlle, and Fehr, Ernst (2015). “Eye spots do not increase altruism in children.” In: Evolution and Human Behavior 87, pp. 224-231.
payoff-irrelevant social stimuli can increase prosocial behavior under the right conditions, these conditions appear to be quite restrictive. Consequently, an ancestral, reciprocity-based psychology is increasingly unlikely to be the principal evolutionary mechanism responsible for anonymous one-shot altruism.
6.1 Introduction

Altruistic cooperation and the prosocial preferences on which it depends play a crucial role in human societies (Camerer, 2006; Fehr and Fischbacher, 2003; Fehr and Fischbacher, 2004a). When social interactions take place in their usual setting, a setting characterized by incomplete contracts, social norms, and informal institutions, prosocial behavior can dramatically attenuate the inefficiencies that follow from strictly self-regarding behavior (Bowles, 2004). Nonetheless, in spite of the crucial role prosocial behavior has in human societies, the evolutionary mechanisms responsible for such behavior remain a highly contentious matter. This is especially true in the special and important case of altruism in anonymous one-shot interactions with genetically unrelated partners (Bowles and Gintis, 2011; Henrich, 2004; Pennisi, 2005).

Anonymous one-shot interactions are special because ethnographic data suggest that, although people in ancestral societies may have had one-shot interactions, anonymous interactions were probably quite rare (Fehr and Henrich, 2003). Moreover, some scientists argue even further that ancestral social interactions were typically repeated. From this perspective, one-shot interactions, anonymous and otherwise, were also rare (Hagen and Hammerstein, 2006; Haley and Fessler, 2005). The nature of social life in the distant past will always involve a certain amount of speculation, but all in all the evidence suggests that anonymous one-shot interactions constitute a special class of evolutionarily recent phenomena.

Anonymous one-shot interactions are additionally important for the following reasons. First, much of the experimental research clearly documenting human altruism is based on interactions of this sort (Camerer, 2003; Fehr and Gächter, 2002; Henrich, Elreath, et al., 2006; Henrich, Ensminger, et al., 2010; Henrich, Robert, et al., 2004; Kagel and Roth, 1995). Second, even if nameless and ephemeral interactions were rare in the past, they are presumably quite common now, and for this reason they matter in contemporary human societies. Finally, the evolutionary basis for anonymous one-shot altruism among unrelated strangers is especially hard to identify and explain. In particular, the only evolutionary explanations for prosocial behavior that are widely regarded as unproblematic are kin-based altruism and the enlightened material self-interest of reciprocity in its various forms (Axelrod and Hamilton, 1981; Bowles and Gintis, 2011; Hamilton, 1964; Henrich, 2004; Nowak and Sigmund, 1998b; Panchanathan and Boyd, 2004; Trivers, 1971). With respect to anonymous one-shot altruism, however, neither kinship nor reciprocity provides a ready explanation. An influential approach to this seeming impasse is to transfer unproblematic explanations based on reciprocity and reputation to situations in which they do not explicitly apply. This is the evolutionary legacy hypothesis for the evolution of human altruism.
Eye spots do not increase altruism in children (Hagen and Hammerstein, 2006; Haley and Fessler, 2005). The hypothesis posits that contemporary human psychology reflects the ancestral conditions that held in the late Pleistocene. Supporters of the hypothesis claim that in the late Pleistocene human social groups were small and cohesive, social contact was intense, interactions were always repeated, and one’s reputation as a more or less cooperative individual was always at stake. Life was effectively like a “camping trip that lasted a lifetime” (Cosmides and Tooby, 2013, p. 203). Under circumstances of this sort, a person would have cooperated to protect her reputation as a prosocial individual, and she would have done so to gain the benefits that followed when members of her group reciprocated in the future.

The link to anonymous one-shot interactions is based on an assumed evolutionary lag. Specifically, the key assumption is that contemporary prosocial behavior observed in anonymous one-shot settings stems from a mismatch between the ancestral, reputation-based psychology of the actors and the actual anonymity of the contemporary social interaction. Put differently, even if the explicit structure of the interaction is anonymous and one-shot, the implicit structure in the hunter-gatherer recesses of an actor’s psyche is such that she will behave in a way that bolsters her good reputation. The discrepancy between the explicit structure of the interaction and the implicit structure persists simply because selection on human psychology, hypothetically, has not caught up with the novel features of contemporary social life. In this sense, the evolutionary legacy hypothesis for the evolution of human altruism is fully consistent with the larger claim in evolutionary psychology that human cognition is adapted to life as hunter-gatherers in the late Pleistocene (Cosmides and Tooby, 2013; Smith, 2000).

The empirical evidence for human altruism as a kind of anachronism is typically experimental. In particular, a number of studies have shown that exposure to face-like stimuli increases altruistic choices in incentivized economic games (Haley and Fessler, 2005; Nettle et al., 2013; Sparks and Barclay, 2013). Importantly, this increase occurs even though the face-like stimuli are sometimes quite abstract and never affect material payoffs. The resulting interpretation of the data stipulates that face-like stimuli, relative to control stimuli, activate the ancestral, reputation-based psychology of participants, and this leads to the observed increase in altruism. Simply put, if people think they are being watched, whether consciously or not, they will behave themselves because any future benefits received from those who might reciprocate depend on it. An especially strong interpretation of this idea argues that every anonymous one-shot interaction involves a variety of uncontrolled cues (e.g. other people in the room) that imply one’s reputation is at stake (Haley and Fessler, 2005). If this is true, an important logical possibility is that these cues, via the ancestral psychology they regulate, are
producing all of the ostensibly other-regarding behavior observed by researchers in anonymous one-shot interactions.

The evolutionary legacy hypothesis has inspired a number of compelling studies, and many of them provide empirical support for the effects of payoff-irrelevant social cues. Importantly, however, current research has two problems. First, existing evidence does not allow us to evaluate the strong interpretation that all observed human altruism stems from a disparity between ancestral and contemporary social conditions. Showing that one can modulate anonymous one-shot altruism by varying payoff-irrelevant stimuli neither implies nor excludes the possibility that all such behavior is due to payoff-irrelevant stimuli. A residual measure of altruism, large or small, may or may not exist. Assessing the strong version of the evolutionary legacy hypothesis requires an approach with predictions more specific than simply saying payoff-irrelevant cues should increase altruism if they imply one’s reputation is at stake. Second, though a number of studies have provided evidence for the importance of payoff-irrelevant social cues (Bateson, Nettle, and Roberts, 2006; Burnham and Hare, 2007; Ernest-Jones, Nettle, and Bateson, 2011; Francey and Bergmüller, 2012; Haley and Fessler, 2005; Mifune, Hashimoto, and Yamagishi, 2010; Milinski and Rockenbach, 2007; Oda et al., 2011; Rigdon et al., 2009), the evidence overall is decidedly mixed. A number of studies have also failed to find an effect (Carbon and Hesslinger, 2011; Fehr and Schneider, 2010; Lamba and Mace, 2010; Raihani and Bshary, 2012; Tane and Takezawa, 2011).

To address these problems, we conducted a dictator game study with payoff-irrelevant stimuli and a distinctive but informative subject pool. In particular, like previous studies, our experimental treatments involved either an asocial control stimulus or a social, face-like stimulus. Unlike previous studies, however, we recruited children of ages five and eight of both sexes to participate in our study. Children represent an informative subject pool because previous research suggests that altruistic behavior in anonymous one-shot encounters develops rapidly between the ages of five and eight, and this developmental process is potentially different for girls and boys (Benenson, Pascoe, and Radmore, 2007; Blake and McAuliffe, 2011; Fehr, Bernhard, and Rockenbach, 2008; House et al., 2013). As we now explain, this means that children offer especially interesting subjects for studies of altruism that use payoff-irrelevant cues.

Recall that the strong interpretation of the evolutionary legacy hypothesis posits that all anonymous one-shot altruism in humans is due to an active reputational psychology. If this is true, any group of people who are systematically more altruistic than others can only be this way because they are more sensitive to the payoff-irrelevant social cues in their environment. For our purposes, this implies the following. If any group of children, defined by age, sex, or both, is
Eye spots do not increase altruism in children

more prosocial than other children, it can only be because this relatively altruistic
group is more sensitive to payoff-irrelevant social stimuli. In this case we should
observe a positive interaction between the relatively altruistic group, whatever it
may be, and exposure to a payoff-irrelevant stimulus that triggers one’s reputa-
tional psychology. Because our design varies payoff-irrelevant stimuli for both
girls and boys of different ages, it allows us to test for this kind of interaction.

In addition, pre-adolescent children have less executive control over decision
making than young adults. In particular, children are less able than adults to se-
lectively focus attention on the key features of a situation, and they are less able to
inhibit prepotent responses (Diamond, 2013). If responding to payoff-irrelevant
social stimuli represents a kind of automatic or unconscious process that inter-
feres with volitional goals (Cosmides and Tooby, 2013; Haley and Fessler, 2005),
then strong executive functioning should dampen any effects associated with the
stimuli in question. Executive control, in short, should improve the ability of in-
dividuals to focus on and respond to the explicit, payoff-relevant structure of the
situation rather than the implicit, payoff-irrelevant structure. More to the point, if
payoff-irrelevant social cues can increase altruistic behavior by evoking a reputa-
tional psychology, and if this is an automatic process in conflict with materially
beneficial decisions under executive control, then effects associated with payoff-
irrelevant social stimuli are most likely to surface in children. This follows be-
cause children should be relatively unable to inhibit their automatic, prosocial re-
sponse to the stimuli. Under this logic, because all of our subjects were children,
our study is especially favorable for detecting a difference between our control
stimulus and our face-like stimulus.

6.2 Experimental methods

We conducted our experiment in the spring and fall of 2012 in 27 kindergarten
and primary schools throughout the German-speaking part of Switzerland. The
study was approved by the Department of Education of the Canton of Zurich and
the Human Subjects Committee of the Faculty of Economics, Business Admin-
istration, and Information Technology at the University of Zurich. Finally, we
received informed consent from the parents of all participating children.

After arrival, while the children were in another area with their teacher, at least
two experimenters constructed four arenas for the experimental session. These
arenas consisted of two waiting areas, an area for the dictator game itself, and a
toy store. After the experimental arenas were built, the experimenters brought the
children together and introduced themselves. The experimenters explained that
they came from the University of Zurich, and they were going to let each child
Experimental methods

play a game for the next hour or so. During the game, the child would have the opportunity to earn gold coins based on her choices in the game. In actuality, the gold coins were simply stylized gold coins printed on small pieces of paper, and the experimenter took great care to explain the meaning and value of these gold coins.

Specifically, one of the experimenters held up a single gold coin and explained their value. At the end of the session, each child would have a chance to turn in her gold coins for a toy from the toy store. More coins would provide a child with the option to choose a toy from a larger selection that included larger and more valuable toys. At this point, the experimenter led the children to the toy store, where they observed different groups of toys separated both by value and whether they were targeted at girls or boys. After the children saw the toy store, the experimenters covered all toys with a blanket so the children would not be distracted during the experiment itself.

After this brief introduction, the experimenters led participants to the first waiting area. As mentioned above, we used two waiting areas, and we always placed them as far apart as possible. One waiting area was for participants waiting to play the game, while the other was for participants who had already played. Children solved puzzles in both waiting areas, and together the experimenters and the class teacher ensured that each child was in the appropriate waiting area. Importantly, the use of two waiting areas allowed us to separate experienced participants from naive participants. This in turn meant that the children who had played could not distort the play of the children still waiting to play by providing uncontrolled information about the game or recommendations about what to choose.

A single experimenter had primary contact with all participants, and so we can exclude experimenter effects. In addition, participants were randomly assigned to treatment within each session, and so we can also exclude the possibility of some unknown systematic relationship between session and treatment. For the dictator game itself, the principal experimenter led a randomly selected child to a small tent that was constructed away from the waiting areas. The child’s teacher was not present at this point. The child sat inside the tent, while the experimenter sat in the doorway of the tent. The child sat in front of a laminated pad approximately 30 x 42 cm. The lower half of the pad was blank, and 10 gold coins were placed on this blank area. The upper half of the pad consisted entirely of a graphical figure. In the control treatment, the figure was a meaningless but regular set of perpendicular lines (appendix). In the eye-spot treatment, the figure was the same stylized eye spots used in Haley and Fessler (2005).

To the left of the pad was a sealed box with a small opening. Two envelopes were to the right of the pad. One envelope was white, and it was for the par-
Eye spots do not increase altruism in children

Participant’s own coins. The other envelope was one of four colors, and it was for the recipient’s coins. We used the colored envelopes to record the gender and treatment for each participant while still maintaining the participants’s complete anonymity. Specifically, the recipient’s envelope was either red, blue, green, or yellow. For each participant, the experimenter used an envelope of a specific color to record the appropriate gender-treatment combination for that participant. To avoid systematic effects associated with color preferences among the children, the meaning of the four colors was randomly determined for each session. As explained below, the child left the envelope behind after making her decision and leaving the tent. Our use of colored envelopes allowed us to later identify the gender and treatment for each participant without the need to assign an identification card, ask the participant her name, or take any other action that would reduce the explicit anonymity of the child.

The experimenter explained the dictator game as follows. Specifically, the experimenter explained that the dictator’s task was to divide the 10 gold coins between herself and a recipient. At the end of the session, the dictator would have the opportunity to turn in her gold coins for a toy from the toy store, and more gold coins would allow her to choose from a larger set of toys that included more valuable toys. The recipient was another child in a different kindergarten or primary school, and neither the identity of the dictator nor the identity of the recipient would ever be revealed to the other. The experimenter explained that the recipient had no coins herself. Any coins received from the dictator, however, would be turned in for a toy, and recipients receiving more gold coins would get more valuable toys. The dictator was told that, when making her decision, she should place her coins in the white envelope and the recipient’s coins in the colored envelope. The experimenter explained that the dictator would be alone in the tent at this point, and the dictator should wear a nearby set of sound-proof earmuffs (Haley and Fessler, 2005). These earmuffs were especially designed for children, and our intention was to use both the tent and the earmuffs to eliminate uncontrolled social stimuli and so isolate the effect of our treatment variation. After distributing the coins, the experimenter told the dictator that she should deposit the recipient’s envelope in the sealed box in the tent and then exit with her own envelope. In addition, the experimenter encouraged the dictator to put her envelope in her pocket, if possible, before proceeding to the second waiting area. Altogether, our experimental design achieved a degree of anonymity well beyond other studies conducted with children of this young age (cf. Fehr, Bernhard, and Rockenbach 2008).

Before leaving the child alone to make a choice, the experimenter asked the child several control questions to test understanding. Specifically, the experimenter asked the child where to put her own gold coins, where to put the gold
coins for the recipient, what to do with each of the two envelopes once the coins were distributed, whether the child knew or ever would know the recipient, and whether the recipient initially had any of her own gold coins. If the child answered any question incorrectly, the experimenter explained the task again and only continued when the child answered all questions correctly. Once the child had answered all control questions correctly, the experimenter helped the child put on her sound-proof earmuffs, closed the tent door so that the tent was opaque from all directions, and left the child to make her choice in private.

Altogether, we conducted the experiment with 201 children. 24 of them left the tent without dividing the coins, which leaves a total of 177 observations (94 boys, 83 girls). If the participant distributed the coins but did not take her envelope when exiting the tent, we recorded this fact and control for it in regression analyses below. This occurrence was rare and only happened with 12 dictators. If the participant distributed the coins but did not deposit the recipient’s envelope in the sealed box, we also recorded this information and control for it in our analyses. This happened 44 times. After a particular participant was finished, the experimenter ensured the participant went to the appropriate waiting area, and then the experimenter returned to the tent to prepare the task for the next participant.

At the end of each session, dictators went to the toy store one by one. Each dictator then turned in her gold coins for the toy of her choice from the toys she could afford with the number of coins she had. For each recipient, we translated the coins received into a toy from the appropriate price category. We then delivered the toys to a Swiss non-governmental organization that had previously agreed to distribute the toys to specific children of the appropriate ages from families in need.

6.3 Results

The mean transfer over all participants was 2.418, which is significantly positive (Wilcoxon signed rank test with continuity correction, $N = 177$, $p < 0.001$). Children randomly assigned to the control transferred an average of 2.552 coins, while children assigned to the eye spot treatment transferred an average of 2.289 gold coins. The difference is not significant (Wilcoxon signed rank test with continuity correction, $N = 177$, $p = 0.5152$). More generally, the transfer distributions are extremely similar for the treatment and control (figure 6.1a).

If we dichotomize transfers as zero or positive (Nettle et al., 2013), 64.37% of participants in the control transferred some positive amount, while in the eye spot treatment 62.22% transferred a positive amount. Again, the difference is not
Eye spots do not increase altruism in children

Figure 6.1: The distribution of dictator transfers. Dictator transfers are separated by treatment (a) and by the sex of the dictator (b). Model selection (appendix) and associated regression results show that the sex of the dictator is an important variable for explaining variation in dictator transfers, while experimental treatment is not. In particular, boys transferred significantly less than girls (table 6.1), and they transferred positive amounts at a significantly lower rate than girls (table 6.2). The data consist of 177 observations.

Figure 6.1: The distribution of dictator transfers. Dictator transfers are separated by treatment (a) and by the sex of the dictator (b). Model selection (appendix) and associated regression results show that the sex of the dictator is an important variable for explaining variation in dictator transfers, while experimental treatment is not. In particular, boys transferred significantly less than girls (table 6.1), and they transferred positive amounts at a significantly lower rate than girls (table 6.2). The data consist of 177 observations.

significant (Fisher’s exact test, $p = 0.876$). These results show that we have no pure treatment effect. Moreover, this is true even though we took great care to randomize assignment to treatment within sessions, to randomize the meaning of our colored envelopes in each session, to prevent any contamination of par-
Results

Participants waiting to play, and to reduce uncontrolled social stimuli as much as possible in order to isolate the effect of our treatment variation.

Although we find no treatment effect over all participants, treatment effects could nevertheless be present but vary by the sex and age of the dictator. For example, perhaps only dictators in a particular age-sex category respond to eye spots. When all dictators are pooled, the effect becomes extremely difficult to detect. As an even more complex possibility, perhaps one age-sex combination increases transfers when assigned to the eye spot treatment, while another age-sex combination decreases transfers. We conducted our experiment with children of ages five and eight to evaluate exactly this kind of possibility. More to the point, existing behavioral evidence suggests that prosocial behavior develops rapidly between the ages of five and eight, and the developmental process could be different for boys and girls (Benenson, Pascoe, and Radmore, 2007; Blake and McAuliffe, 2011; Fehr, Bernhard, and Rockenbach, 2008; House et al., 2013). If this is true, insofar as prosocial behavior depends on payoff-irrelevant social cues, children in different age-sex categories should respond differently to these cues.

To evaluate this idea, we modeled transfers in two ways. First, we modeled player transfers using ordinal logistic regressions. Ordinal logistic regressions assume that the response variable falls into a relatively small number of discrete, ordered categories (Verbeek, 2008). In our case, ordinal logistic models are especially appropriate both because they do not assume that transfers represent a cardinal measure of preferences and because we have a number of observations at one of the boundaries of the action space (i.e. 0). Second, we also coded transfers as either zero (0) or positive (1) and modeled them using simple logistic regressions. This binary treatment of transfers allows us to identify variation in the probability of transferring any positive amount without regard for the magnitude of positive transfers (Haley and Fessler, 2005; Nettle et al., 2013).

Under both treatments of the response variable, we specified six regression models. The models vary in terms of the independent variables they include (appendix, tables 6.A.2 and 6.A.3). All models control for whether the dictator dropped the recipient’s envelope in the sealed box in the tent and whether the dictator exited the tent with or without her own envelope. The set of independent variables additionally includes the age, sex, and treatment for each dictator. Altogether we specified models without interaction terms as well as models that include second-order and third-order interactions. Our complete set of models allows us to identify, with complete flexibility, any differential response to the eye spots based on the age of the dictator, the sex of the dictator, or both. Furthermore, this is true for both the regressions that model the magnitude of dictator transfers (i.e. ordinal logistic) and the regressions that simply treat transfers as zero or positive (i.e. simple logistic).
To identify the best fitting models, we used AICc as a criterion to identify the best models in the set (Burnham and Anderson, 2002). AICc is an information theoretic model selection criterion that optimizes the trade-off between the omitted-variable biases associated with underfitting and the lack of generalizable conclusions associated with overfitting. It is a derivative form of Akaike’s original asymptotic criterion (Akaike, 1973) that corrects for finite samples (Burnham and Anderson, 2002). We implemented the model selection exercise twice, once using ordinal logistic regressions and once using simple logistic regressions.

Both model selection exercises produced a robust result. Namely, the critical independent variable is the sex of the dictator (figure 6.1b). Boys transferred significantly less than girls (table 6.1), and they transferred positive amounts with a significantly lower probability than girls (table 6.2). Neither model selection nor associated regression results indicate that experimental treatment is an important predictor of dictator behavior. For both ordinal logistic regressions and simple logistic regressions, the best set of predictor variables includes the age and sex of the dictator, but no interactions and no variable for experimental treatment. Under this specification, whether using an ordinal logistic or a simple logistic regression, the sex of the dictator is highly significant, and it is the only significant predictor (appendix). In the present paper, we do not present these best-fitting models. Simply for the sake of generality, we present models that additionally include a treatment dummy (tables 6.1 and 6.2). With or without the treatment dummy, the conclusions are the same. Boys gave less than girls, and they gave positive amounts with lower probability than girls.
Table 6.1: Ordinal logistic model of transfers.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Odds ratio</th>
<th>Robust std. error</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Dropped</td>
<td>2.344</td>
<td>1.159</td>
<td>1.72</td>
<td>0.085</td>
</tr>
<tr>
<td>Not Taken</td>
<td>1.614</td>
<td>0.903</td>
<td>0.86</td>
<td>0.392</td>
</tr>
<tr>
<td>Primary</td>
<td>0.598</td>
<td>0.194</td>
<td>-1.59</td>
<td>0.112</td>
</tr>
<tr>
<td>Male</td>
<td>0.421</td>
<td>0.119</td>
<td>-3.06</td>
<td>0.002</td>
</tr>
<tr>
<td>Eye Spot</td>
<td>0.791</td>
<td>0.226</td>
<td>-0.82</td>
<td>0.412</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intercepts</th>
<th>Estimate</th>
<th>Robust std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 0/1</td>
<td>-1.227</td>
<td>0.406</td>
</tr>
<tr>
<td>Intercept 1/2</td>
<td>-0.662</td>
<td>0.392</td>
</tr>
<tr>
<td>Intercept 2/3</td>
<td>-0.314</td>
<td>0.391</td>
</tr>
<tr>
<td>Intercept 3/4</td>
<td>0.222</td>
<td>0.392</td>
</tr>
<tr>
<td>Intercept 4/5</td>
<td>0.846</td>
<td>0.401</td>
</tr>
<tr>
<td>Intercept 5/6</td>
<td>1.544</td>
<td>0.412</td>
</tr>
<tr>
<td>Intercept 6/7</td>
<td>1.872</td>
<td>0.429</td>
</tr>
<tr>
<td>Intercept 7/8</td>
<td>2.112</td>
<td>0.435</td>
</tr>
<tr>
<td>Intercept 8/9</td>
<td>2.300</td>
<td>0.442</td>
</tr>
</tbody>
</table>

Notes: Independent variables indicate if the dictator dropped the recipient’s envelope in the sealed box and also if she took her own envelope when exiting the tent. The independent variables additionally indicate the age (Primary), sex (Male), and treatment (Eye Spot) for each dictator. The best fitting model (appendix) does not include the treatment dummy. The number of observations is 177.
Table 6.2: Logistic model of transfers.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Odds ratio</th>
<th>Robust std. error</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.086</td>
<td>1.269</td>
<td>2.74</td>
<td>0.006</td>
</tr>
<tr>
<td>Not Dropped</td>
<td>1.156</td>
<td>0.505</td>
<td>0.33</td>
<td>0.739</td>
</tr>
<tr>
<td>Not Taken</td>
<td>2.756</td>
<td>2.311</td>
<td>1.21</td>
<td>0.227</td>
</tr>
<tr>
<td>Primary</td>
<td>0.931</td>
<td>0.349</td>
<td>-0.19</td>
<td>0.849</td>
</tr>
<tr>
<td>Male</td>
<td>0.372</td>
<td>0.126</td>
<td>-2.91</td>
<td>0.004</td>
</tr>
<tr>
<td>Eye Spot</td>
<td>0.851</td>
<td>0.277</td>
<td>-0.50</td>
<td>0.619</td>
</tr>
</tbody>
</table>

Notes: Independent variables indicate if the dictator dropped the recipient’s envelope in the sealed box and also if she took her own envelope when exiting the tent. The independent variables additionally indicate the age (Primary), sex (Male), and treatment (Eye Spot) for each dictator. The best fitting model (appendix) does not include the treatment dummy. The total number of observations is 177.
6.4 Discussion and conclusion

In the first eye spot study, Haley and Fessler (2005) provided a bold explanation for anonymous one-shot altruism. They essentially argued that altruism of this sort, under an appropriate model of human psychology, is neither anonymous nor one-shot. It arises instead from an evolutionary history of reciprocity and reputation. In particular, the proximate psychology behind the observed altruistic behavior is not the outcome of an evolutionary history that selected for anonymous one-shot altruism. It is instead the outcome of an evolutionary history that selected for prosocial behavior that was ultimately in the material self-interest of the prosocial individual.

Haley and Fessler (2005) further suggested that ostensibly anonymous one-shot social interactions actually involve a rich array of social cues that should activate our ancestral, reciprocity-based psychology. Because we can modulate altruistic behavior with cues of this sort, we know they matter. Moreover, because we cannot eliminate all payoff-irrelevant social cues, we cannot conclude that any additional explanation for anonymous one-shot altruism is necessary. This logic places us on the brink of accepting the strong interpretation of the evolutionary legacy hypothesis. Although Haley and Fessler (2005) do not use this same terminology, they do argue that we should preferentially favor explanations for anonymous one-shot altruism that rely on an ancestral psychology shaped by an evolutionary history of kinship and repeated interactions. This is equivalent to arguing that, in the absence of a clear reason to reject the strong version of the evolutionary legacy hypothesis, we should accept it given the evidence we have. We should do so, in short, because the underlying theory is conventional, and our only alternative is to rely on the extraordinary claims offered by proponents of strong reciprocity.

Support for the evolutionary legacy hypothesis rests on empirical results that show how payoff-irrelevant social cues can increase altruism. A number of studies provide this kind of evidence, but a small and growing number of studies find no effect. All in all, the literature is thoroughly inconclusive. We conducted our study with children of ages five and eight because existing research suggests this is an important developmental period for prosocial behavior. Under the strong interpretation of the evolutionary legacy hypothesis, variation in prosocial behavior among different categories of children should correspond closely to variation in their sensitivity to payoff-irrelevant social cues. We found no support for this prediction. Specifically, although we found systematic variation in prosocial behavior, with males being considerably less altruistic than females, we did not find any associated variation in sensitivity to eye spots. Indeed, we found no eye spot
Eye spots do not increase altruism in children

Effects of any kind, which means that across studies the empirical findings are even more mixed than before.

Two recent papers (Nettle et al., 2013; Sparks and Barclay, 2013) presented meta-analyses of eye spot studies to clarify the reasons for the mixed empirical results. Two conclusions emerged. First, eye spots do not reliably affect the average degree of generosity, but they do seem to increase the probability that a person gives some positive amount (Nettle et al., 2013). In contrast to this finding, our data show no effect of either sort. Whether we analyze the magnitude of dictator transfers or the probability of giving any positive amount, eye spots had no effect on dictator choices in our experiment.

The second conclusion to surface in recent meta-analyses is that eye spots tend to have a reliable effect only when exposure to the payoff-irrelevant stimulus is sufficiently brief and just before the decision maker chooses (Sparks and Barclay, 2013). In our case, the experiment was not computerized because we were working with young children. Consequently, for decisions to be anonymous, the payoff-irrelevant stimuli had to be continuously present inside the tent where the experiment took place. Exposure necessarily lasted the entire 5-10 minutes required for each participant. Whatever the exposure threshold separating a response to eye spots from no response, the threshold seems to be much less than this (Sparks and Barclay, 2013). Thus, we cannot exclude the possibility that the children in our study did not respond to eye spots because exposure time was beyond the critical threshold, whatever this threshold may prove to be.

Importantly, however, if exposure to eye spots for more than a minute or two eliminates any effect, then eye spots cannot do the work originally asked of them. Eye spot results were originally put forward as a way to explain anonymous one-shot altruism by way of conventional, unproblematic explanations like reciprocity under repeated interactions (Haley and Fessler, 2005). In particular, the economic experiments providing much of the evidence for anonymous one-shot altruism always involve some kind of social stimuli that should trigger our ancestral, reciprocity-based psychology. Subjects often mingle outside the lab before entering, they participate in a lab full with other people, they often interact with the experimenters, and they know they might see at least one or two of their fellow participants on the way out the door. Because these kinds of payoff-irrelevant social stimuli produce altruism, and because arguably they can never be fully eliminated, an ancestral psychology evoked by payoff-irrelevant social stimuli is all we need to explain anonymous one-shot altruism.

Importantly, this reasoning is not plausible if the relevant stimuli have no effect after a minute or two. The vast majority of behavioral experiments yielding evidence for one-shot anonymous altruism involve exposure to uncontrolled social stimuli for a much longer period of time. In our own lab in Zurich, for ex-
ample, which has produced some share of the evidence for anonymous one-shot altruism, subjects convene outside the lab before an experiment, they interact with the experimenter as they read instructions and answer questions to test their understanding of the game, and they are typically in the lab for one to two hours with up to 35 additional people. If experimental subjects habituate to payoff-irrelevant social stimuli after a minute or two, then such stimuli cannot explain the altruism that appears in this kind of setting.

They also cannot explain the acts of anonymous one-shot altruism that occur in contemporary large-scale societies, societies where such acts take place amid an unceasing barrage of stimuli that imply others could be watching. We have never counted how many times a day Julia Roberts and Brad Pitt look at us from the many billboards that garnish the streets of Zurich, but we suspect the number is astonishingly high. If we habituate to such stimuli quickly, then the stimuli cannot explain the altruism it is being asked to explain. We are left with our original problem, namely a form of altruism that cannot be explained in terms of conventional evolutionary mechanisms. Paradoxically, by identifying the importance of limited exposure time, Sparks and Barclay (2013) argue that they reconcile disparate results from eye spot studies and thus vindicate the importance of reciprocity and reputation in anonymous one-shot settings. In reality, they do exactly the opposite. Rapid habituation, if robust, ensures that eye spots may be an empirical curiosity, but they cannot be a general explanation for anonymous one-shot altruism in everyday social life. Payoff-irrelevant social stimuli are simply too pervasive.

Ultimately, we suspect that the challenges facing eye spot research will be interpretive. The empirical results are subject to refinement (Nettle et al., 2013; Sparks and Barclay, 2013). Some studies will produce positive results; some will produce negative results. Ongoing research will sort out the differences. Whatever these differences turn out to be, the number of existing studies with positive results suggest that eye spot effects in some form are here to stay. Nevertheless, as the domain for payoff-irrelevant social stimuli becomes increasingly restricted, any associated ancestral psychology will be increasingly unable to explain altruism to the exclusion of other mechanisms. This is the crux of the interpretive problem. Haley and Fessler (2005) granted eye spots a privileged explanatory status because eye spots allow us to retain conventional evolutionary explanations for observed behavior. However, as the limitations of an evoked reputational psychology become increasingly clear, we are left with our original problem of how to explain altruism in situations that really are anonymous and one-shot, both in explicit material terms and in implicit psychological terms.
6.A Appendix

6.A.1 Experimental methods

Instructions

Now I will explain to you how the game works. Here are 10 gold coins. You can buy toys with these gold coins at the end of this game. The more gold coins you have, the bigger the toys you can buy. When I am done explaining the game, you will decide on your own if you want to give some of the gold coins to another child or not. Only do this after I have explained the entire game to you.

The other child goes to another school or kindergarten. You will never know who the other child is, and the other child will never know who you are. The other child has no gold coins. But the other child can also get a toy. The more gold coins the other child has, the bigger the toys the child can get.

Please wait until I am gone to make your decision. As soon as you are alone in the tent, you can decide. Put the gold coins you want to keep in the WHITE envelope. [The experimenter points on the white envelope]. You can then put the white envelope in your trouser pocket.

[Counterbalance the ordering of the following two statements across participants.]

[Giving Statement] If you want to give gold coins to the other child, please put them in the [specify color] envelope. [The experimenter points to the colored envelope.] Then put the [specify color] envelope in the box.

[Keeping Statement] If you want to keep all gold coins for yourself, please just put the empty [specify color] envelope in the box. [The experiment points to the colored envelope.]

After you make your decision, you can come out of the tent and continue playing. At the end of the class you can buy toys with the gold coins in your white envelope. Do you have any questions?

[After answering any questions] I have some questions in order to see if you understood the game correctly.

• In which envelope do you put the gold coins that you want to keep?
• In which envelope do you put the gold coins that you want to give to the other child?
• What do you do with the white envelope?
• What do you do with the [specify color] envelope?
• How many gold coins does the other child have?
Do you know the other child?

Does the other child know you?

[If necessary, explain the instructions again. When the child has answered all questions correctly, proceed.] So, now I am leaving the tent and you can make your decision. Afterwards, please come out of the tent. Please put on these ear muffs before you make your decision. It will help you to concentrate. [Experimenter waits until the child is wearing the ear muffs. Then the experimenter leaves the tent. The experimenter waits at a distance. When the child is done, the experimenter must make sure that the child goes to the second waiting area. The experimenter then prepares the experiment for the next child.]

Stimuli and experimental arena

Figures 6.A.1 – 6.A.3 show the stimuli used in our two treatments and a schematic of the experimental arena.

Payment scheme

We purchased several different toys targeted at both boys and girls of ages five and eight. We divided toys into three categories according to their local market price. Toys in the low category ranged in price from approximately eight to 12 Swiss Francs. Toys in the middle category ranged from approximately 15 to 20 Swiss Francs. Finally, toys in the high category ranged in price from approximately 25 to 30 Swiss Francs. Table 6.A.1 shows the payment scheme used to exchange gold coins from the dictator game into toys.

6.A.2 Data analysis

To identify the best way to model dictators’ transfers, we conducted two model selection exercises. In the first case (Table 6.A.2), we modeled transfers using ordinal logistic regressions. Models of this sort do not assume that the cardinal response variable represents an underlying cardinal preference relation. Moreover, because ordinal logit models assume an unobserved latent variable, they are also appropriate when modeling choices, as in our experiment, in a bounded action space (Verbeek, 2008). In the second case (Table 6.A.3), we modeled transfers using logistic regressions. Specifically, as suggested by the meta-analysis of Nettle et al. (2013), we modeled transfers as either zero (0) or positive (1).

As we argue in the main paper, the strong version of the evolutionary legacy hypothesis implies the following. If any category of child is relatively altruistic,
this can only be because this category of child is relatively sensitive to payoff-
irrelevant social cues. This implies a positive interaction between being a member
of the relevant category and participating in our eye spot treatment.

To test this idea, we categorize the children who participated in our experi-
ment in terms of age and sex. In some models, we interact the associated dummy
variables with experimental treatment. In the most general case (Models 1 in
Tables 6.A.2 and 6.A.3), we categorize children jointly in terms of age and sex.
Thus, altogether we have four categories. They include girls in kindergarten,
boys in kindergarten, girls in primary school, and boys in primary school. To
avoid cumbersome three-way interactions, we simply create three separate dum-
mies to categorize children. The omitted category is girls in kindergarten, and so
the three dummies are boys in kindergarten (M·K), girls in primary school (F·Pr),
and boys in primary school (M·Pr). In the remaining models (Models 2 – 6 in
Tables 6.A.2 and 6.A.3), we do not interact experimental treatment with the joint
categorization of children by age and sex, and so we simply work with a dummy
for primary school (Pr) and a dummy for boys (M).

Table 6.A.1: The payment scheme used to exchange gold coins from the dictator
game into toys.

<table>
<thead>
<tr>
<th>Gold coins</th>
<th>Category of toy</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 5</td>
<td>Low</td>
</tr>
<tr>
<td>&gt; 5 and &lt; 9</td>
<td>Middle</td>
</tr>
<tr>
<td>≥ 9</td>
<td>High</td>
</tr>
</tbody>
</table>
Table 6.A.2: Model selection, ordered logit models of dictator transfers.

<table>
<thead>
<tr>
<th>#</th>
<th>ND</th>
<th>NT</th>
<th>Pr</th>
<th>M</th>
<th>E</th>
<th>Pr-E</th>
<th>M-E</th>
<th>F-K</th>
<th>M-K</th>
<th>F-Pr</th>
<th>M-Pr</th>
<th>F-K-E</th>
<th>M-K-E</th>
<th>F-Pr-E</th>
<th>M-Pr-E</th>
<th>Par.</th>
<th>AIC&lt;sub&gt;c&lt;/sub&gt;</th>
<th>w&lt;sub&gt;i&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>18</td>
<td>694.114</td>
<td>0.036</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>17</td>
<td>696.262</td>
<td>0.012</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>14</td>
<td>690.497</td>
<td>0.22</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>12</td>
<td>696.688</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>14</td>
<td>690.493</td>
<td>0.221</td>
</tr>
<tr>
<td>6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>13</td>
<td>688.858</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Notes: Independent variables include (i) a dummy (ND) indicating that the dictator did not drop the recipient’s envelope in the sealed box, (ii) a dummy (NT) indicating that the dictator did not take her own envelope when leaving the tent, (iii) a dummy (Pr) for an approximately eight-year-old dictator in primary school, (iv) a dummy (M) for a male dictator, (v) a dummy (E) for participation in the eye-spot treatment, and various interactions between the age, sex, and treatment of the dictator. For ease of interpretation, some of these interactions in some models use a dummy (K) indicating an approximately five-year-old dictator in kindergarten instead of using the primary school dummy (Pr). A check mark indicates that the variable is included in the associated regression. The last three columns of the table show the number of parameters estimated for each model, the AIC<sub>c</sub> value, and the Akaike weight (w<sub>i</sub>). The Akaike weights follow from a simple normalization of AIC<sub>c</sub> values across all six models, indexed by i, under consideration (Burnham and Anderson, 2002). The normalization is such that ∑<sub>i</sub> w<sub>i</sub> = 1, and higher w<sub>i</sub> values indicate a better model. The model selection exercise shows that models including the sex and age (i.e. kindergarten versus primary) of the dictator without interactions involving experimental treatment are the best models.
### Table 6.A.3: Model selection, logit models of dictator transfers treated dichotomously as zero (0) or positive (1).

<table>
<thead>
<tr>
<th>#</th>
<th>ND</th>
<th>NT</th>
<th>Pr</th>
<th>M</th>
<th>E</th>
<th>Pr-E</th>
<th>M-E</th>
<th>F-K</th>
<th>M-K</th>
<th>F-Pr</th>
<th>M-Pr</th>
<th>F-K-E</th>
<th>M-K-E</th>
<th>F-Pr-E</th>
<th>M-Pr-E</th>
<th>Par.</th>
<th>AIC&lt;sub&gt;c&lt;/sub&gt;</th>
<th>wi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>10</td>
<td>235.045</td>
<td>0.095</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>9</td>
<td>238.717</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6</td>
<td>233.704</td>
<td>0.186</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>4</td>
<td>238.247</td>
<td>0.019</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6</td>
<td>233.512</td>
<td>0.205</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5</td>
<td>231.81</td>
<td>0.48</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Independent variables include (i) a dummy (ND) indicating that the dictator did not drop the recipient’s envelope in the sealed box, (ii) a dummy (NT) indicating that the dictator did not take her own envelope when leaving the tent, (iii) a dummy (Pr) for an approximately eight-year-old dictator in primary school, (iv) a dummy (M) for a male dictator, (v) a dummy (E) for participation in the eye-spot treatment, and various interactions between the age, sex, and treatment of the dictator. For ease of interpretation, some of these interactions in some models use a dummy (K) indicating an approximately five-year-old dictator in kindergarten instead of using the primary school dummy (Pr). A check mark indicates that the variable is included in the associated regression. The last three columns of the table show the number of parameters estimated for each model, the AIC<sub>c</sub> value, and the Akaike weight (wi). The Akaike weights follow from a simple normalization of AIC<sub>c</sub> values across all six models, indexed by i, under consideration (Burnham and Anderson, 2002). The normalization is such that \( \sum_i w_i = 1 \), and higher wi values indicate a better model. The model selection exercise shows that models including the sex and age (i.e. kindergarten versus primary) of the dictator without interactions involving experimental treatment are the best models.
Figure 6.A.1: The stimulus used for the control treatment.

Figure 6.A.2: The stimulus used for the eye spot treatment (see Haley and Fessler 2005).
Figure 6.A.3: The spatial arrangement of the experimental arena. The schematic shows the position of the child, the door of the tent, the pad with the coins and stimulus, the two envelopes for the dictator (D) and recipient (R), and the box for the recipient’s envelope. As the figure shows, the spatial orientation of the child, the box, and the envelopes were counterbalanced across sessions.
Chapter 7

Summary and conclusion

This dissertation contains five studies on competition and cooperation. Not only are competition and cooperation two societal constants, but competitive and cooperative traits also seem deep-rooted in human psychology (Gil-White and Richardson, 2003; Griskevicius, Cantù, and Vugt, 2012). Classical social scientists and social philosophers debated positive and negative effects of competition on society (Marx, 1969 [1890]; Smith, 2003 [1776]) or pondered what makes enduring social cooperation possible (Hobbes, 2003 [1651]). While such key sociological problems have been discussed since the dawn of the discipline, in recent decades the interdisciplinary field of experimental game theory has generated profound new insights. The quick development in the field is driven by the iterative process of rigorous model building and experimentation. Unfortunately, sociology has largely neglected both game theory and experimentation (Diekmann, 2008; Fehr and Gintis, 2007). This dissertation is a small contribution to fill this ‘methodical lacuna’. Inspired by game theoretic reasoning and based on experimental methods, it contributes to sociologically relevant questions regarding competition and cooperation: How is it possible that competitive structures generate more frustration when they provide more opportunities? Can a competitive human trait, the desire for social status, promote cooperation, namely environmentally friendly behavior? Are people in real-world settings willing to enforce social norms for the sake of the group even when this is costly, as they do in the laboratory? Do children respond to cues that they are being watched, and what does the result tell us about the debate on whether humans are only motivated by self-regarding preferences or also by other-regarding preferences?

In what follows, the main results of each study are summarized and thereby answers to the above-mentioned questions are given. The chapter then closes with an outlook for a new research direction. Chapter 2 investigated a long-standing
sociological puzzle, already discussed by Tocqueville (1952 [1856]), Durkheim (1952 [1897]) and Stouffer et al. (1965 [1949]): How is it possible that in competitive situations more opportunities can generate more frustration? We discussed a game theoretic competition model, introduced by Boudon (1982 [1977]), and experimentally tested its main implications. The model explains the phenomenon of more frustration under better conditions as a consequence of rational, strategic individual decisions in competitive interaction structures. The main model prediction is an inversely U-shaped path of frustration: Under certain conditions, specified by the model, a small increase in the opportunities leads to an overly large increase in the number of actors entering the competition. As a consequence, the number of additional frustrated losers exceeds the number of additional satisfied winners. At a certain point, the direction of the correlation changes and more opportunities are accompanied by a decreasing frustration level. We tested the prediction of an inversely U-shaped rate of relative frustration in three laboratory experiments. The results are mixed. In contradiction to the model predictions, when opportunities increase, the rate of the frustrated losers remains constant, or increases only slightly. In accordance with the model’s predictions, thereafter, the loser rate decreases again. Post hoc, we questioned Boudon’s narrow focus on the losers of a competition. Not only the losers, but also the actors who did not enter the competition, may feel relatively frustrated. Drawing on received relative frustration theory (Kakwani, 1984) we measured relative frustration using the Gini coefficient, which accounts for all social comparison processes in a group. Indeed, the observed Gini coefficient follows an inversely U-shaped path. Taken together, the results suggest that when opportunities increase, more actors feel inclined to enter the competition and hence the disparity in the payoffs between all actors grows. This increase in social inequality fosters overall frustration. Importantly, the effect of more frustration under improved conditions only emerges under certain conditions – for example, when the opportunities are enhanced on a low level and when the payoff for a successful competitor is considerably more lucrative than that for an actor refraining from competition.

In chapter 3, I examined whether a competitive human trait, the taste for social status, can explain why humans are willing to pay a premium for products such as luxury clothes or environmentally friendly cars. According to signaling theory (Spence, 1973; Zahavi, 1975; Zahavi, 1977), the extra costs of conspicuous luxury products or expensive environmentally friendly products are signals of social status and wealth that lead to advantages for signalers in social interactions. For example, luxury signalers are selected in situations of partner choice because most people prefer bonding with wealthy and high status individuals (Nelissen and Meijers, 2011). Such advantages in social interactions are, in addition to a taste for status, a motive for buying luxury products. This argument extends to
‘green’ signalers – that is, people who buy conspicuous and expensive environmentally friendly products (Griskevicius, Cantù, and Vugt, 2012). Additionally, those paying a premium for environmentally friendly products, instead of buying less expensive but environmentally more harmful ones, signal a willingness to engage for the public (Whitfield, 2011). Because people hope to profit from the cooperativeness of green signalers, they might even be more preferable allies than luxury signalers (Gintis, Smith, and Bowles, 2001). As with luxury products, these advantages in social interactions could be a motive for buying expensive and conspicuous green products. Nelissen and Meijers (2011) found positive evidence for the luxury-signaling hypothesis. People wearing luxury-labeled shirts were treated more favorably in social interactions, as predicted. However, an issue of this study was that some of the experiments were not double blind. I thus replicated these experiments with a double-blind procedure. And in addition to the luxury label condition, I introduced a green label condition. Therefore, the hypothesis that signaling theory can explain environmentally friendly behavior was tested for the first time in a natural field setting. Further, I conducted experiments in both socioeconomically average and below-average neighborhoods, where, according to signaling theory, the effects of luxury signals should be even stronger. In contrast to the original study, I did not find positive effects of the luxury brand label in any of the five experiments. Nor did I find evidence for a green-signaling effect. Moreover, in socioeconomically below-average neighborhoods a slightly negative tendency of the luxury label actually became evident. This finding sheds some doubt on a signaling theory explanation of both the consumption of luxury products and environmentally friendly products. However, I used only one specific label for each of the treatments. Hence, before the signaling explanation of luxury product consumption and green product consumption is abandoned, further studies with a broader range of labels should be conducted.

The previous chapter connected competition and cooperation. In chapters 4 and 5, we examined a classic cooperation problem: the enforcement of social norms. In principle, everybody profits from clean cities or silent train compartments. Still, enforcement requires engagement and is potentially risky because the sanctioned norm violators could retaliate. Hence, when someone litters or plays annoying music in a silence compartment of a train, everybody also has an incentive to leave enforcement to anyone else (Heckathorn, 1989). Nevertheless, evidence from laboratory experiments suggests that a surprisingly large proportion of people is willing to enforce norms at a cost – a behavioral pattern called altruistic punishment (Fehr and Gächter, 2002). It is still debated to what extent these findings can be generalized to natural interactions (Guala, 2012). To extend our knowledge of norm enforcement in natural situations, in chapter 4, we systematically violated the anti-littering norm at subway and bus stations in Bern,
Summary and conclusion

Zurich, and New York City. Moreover we manipulated the costs and benefits of enforcement. We find that norms are universally enforced, although significantly less than in the standard laboratory experiment, and that enforcement is significantly more common in Switzerland than in New York and than found in a previous study in Athens (Balafoutas and Nikiforakis, 2012). We also observe that enforcement drops as benefits decrease or costs increase. Moreover, individuals prefer more subtle forms of enforcement to direct punishment. A survey among bystanders suggests that they are concerned with the possibility of conflict when directly enforcing a norm. This means that the costlier and riskier norm enforcement is, the less likely altruistic punishment is to stabilize social norms. This points to a potentially significant strand of further research. Researchers should go beyond the basic question of whether humans are willing to enforce norms to the question of how frequent enforcement must be in order to stabilize a given norm.

While the previous chapter builds on a strand of research that ascribes norm enforcement to pro-social preferences, in chapter 5 we focused on an alternative explanation of norm enforcement. This explanation stresses that in many situations, self-interest is sufficient to motivate people to enforce a norm, and that heterogeneity in the costs and benefits of norm enforcement among bystanders increases the probability that a norm will be enforced. Among homogeneous actors, it is unclear who should volunteer and enforce the norm. Conversely, among heterogeneous actors, it is possible to coordinate on the actor with the highest net benefit of enforcement. While this hypothesis has already been corroborated in a laboratory experiment (Diekmann and Przepiorka, 2015; Przepiorka and Diekmann, 2013), we are the first to validate this hypothesis in a natural field setting. We let a confederate violate the silence norm in a train by playing loud music on a mobile phone. In accordance with our hypothesis, a passenger with a high net benefit from enforcing the silence norm (e.g. those sitting near the confederate) asked the violator to turn down the music more frequently. Overall, the silence norm was enforced in approximately 50 percent of all trials. This prevalence of enforcement is considerably higher than the prevalence we found in the littering experiments discussed in chapter 4. We suggest that this is because the externality of playing loud music in a train car is greater than the externality of dropping one piece of litter. More generally, in the first case (i.e. enforcement of the silence norm), the individual benefit of enforcement outweighs the individual cost, while in the second case (i.e. enforcement of the anti-littering norm), the individual cost outweighs the individual benefit. Game theoretically, the first situation can be conceived of as a volunteer’s dilemma, while the second can be viewed as a missing hero dilemma. In a volunteer’s dilemma, in principle, norm enforcement can be explained by self-interest. Conversely, in a missing hero dilemma it is
a dominant strategy not to cooperate and thus the assumption of other-regarding preferences is necessary to explain norm enforcement (Diekmann and Przepiorka, 2015). Obviously, the two interaction structures differ strikingly. Hence, in future studies on norm enforcement researchers should clearly state which of the two interaction structures they are studying.

Finally, in chapter 6 we investigated the motives behind cooperative behavior and thereby contribute to the debate on whether humans only cooperate when it is in their self-interest or whether they are additionally motivated by other-regarding preferences. The *evolutionary legacy hypothesis* states that the altruistic behavior regularly observed in the dictator game could be due to an evolutionary mal-adaptation rather than to other-regarding preferences. According to this hypothesis, human psychology is adapted to interactions in small groups, which were predominantly non-anonymous and repeated. Un-cooperative individuals were punished severely, for example by ostracism, and it was thus in everybody’s self-interest to cooperate. As a consequence of this evolutionary legacy, when in an anonymous one-time only interaction, people are hardly capable of adapting their behavior and therefore behave cooperatively even if it is not in their self-interest (Haley and Fessler, 2005). The empirical support for the evolutionary legacy hypothesis rests on studies showing that altruism among adults increases in the presence of cues that imply one’s behavior is observable (e.g. stylized eye spots and faces). However, the empirical support for the hypothesis is mixed (Nettle et al., 2013; Sparks and Barclay, 2013). To address this problem, we conducted an eye spot experiment with children. Children are especially informative subjects for this kind of study because altruism develops rapidly during childhood with differences between girls and boys. This means that, if altruism in anonymous one-time only interactions depends primarily on social cues, then variation in altruism by age or sex should be closely linked to variation in sensitivity to eye spot cues. Although we did find variation in altruism among children, we found no variation in sensitivity to eye spots. More generally, we found no eye spot effects of any kind. This result contradicts the evolutionary legacy hypothesis and suggests that other-regarding preferences also motivate human behavior. Of course, one study can hardly end the long-standing debate on human motivation. One caveat of this study is the long time span our subjects were exposed to the eye cues. A meta-study suggests that eye cues are only capable of evoking altruism when shown briefly (Sparks and Barclay, 2013). Hence, a subsequent study should shorten exposure time.

In closing this dissertation, I suggest thorough examination of the effects of competition on cooperation as a new direction of research. The question of how competition affects cooperation has not only been key in the social sciences since their origin (e.g. Marx, 1969 [1890]; Polanyi, 1957; Smith, 2003 [1776]) but
is also of practical relevance. For example, an optimal balance between competition and cooperation among the employees of a firm is crucial for its performance (Abell, 1996). Similarly, competitive and cooperative institutional features should be well-balanced at the societal level in order to avoid a sub-optimal allocation of limited resources on the one hand and the crowding out of trust and social capital on the other (Bowles, 1998; Stiglitz, 2000). Still, there is very little experimental evidence on the effects of competition on cooperation, and it seems that depending on the structure of interaction, competition can either promote or dampen cooperation. Here, we first briefly review studies that highlight positive effects of competition on cooperation. We then turn to another strand of research identifying negative effects of competition on cooperation.

A first strand of research highlights positive effects of competition on cooperation. A prominent example is the behavioral pattern of competitive altruism (Roberts, 1998), first detected in the Arabian babbler, a bird species. Individual babblers try to be the most altruistic of a group in order to climb the status latter. For example, they compete over feeding the nestlings or the weakest individuals of a group (Zahavi, 1989). As discussed in chapter 3, the preference for social status can also promote cooperation in humans. For instance, among the Meriam, a tribal small-scale society in the South Pacific, young men hunt turtles and offer them in public feasts as an investment in their social status (Bliege Bird and Smith 2005). Another example of how competition promotes cooperation becomes evident in economic exchange. Competition for customers among sellers can promote cooperative behavior in sellers toward their costumers. To illustrate this, suppose you are the only seller of secondhand cars in town. It may then pay off to sell mediocre cars for high-pitched prizes. As soon as there are other secondhand car dealers, however, you have an incentive to behave more cooperatively and offer fair deals, or otherwise, no customer will buy twice at your place (Huck, Luenser, and Tyran, 2012).

A second strand of studies stresses the negative effects of competition on cooperation that can emerge when actors compete for a limited number of lucrative goods (e.g. scarce and well-paid positions on the labor market). There are two different paths connecting competition and cooperation problems. The first path is that competition can negatively impact cooperativeness mediated through relative frustration and trust. Specifically, competition fosters relative frustration, as demonstrated in chapter 2. Further, relatively frustrated individuals tend to be less trustful of other society members as well as toward societal institutions (Fischer and Torgler, 2013). Trust, in turn, is a precondition for cooperation (Yamagishi, 1986). For instance, a police officer who thinks that most of her colleagues are corrupt is more inclined to take a bribe and a citizen who has no trust in societal institutions is more inclined to cheat on taxes (Frey and Torgler, 2007; Helliwell
The second path is that relative frustration can trigger plain anti-social behavior, which is, of course, devastating for cooperation. To be more precise, it has been found that the relatively frustrated tend to compensate their disadvantage with illegal means, for example with stealing. On top of that, they also tend to retaliate against more privileged individuals or the agency responsible for their underprivileged status (Kim and Glomb, 2014; Skarlicki and Folger, 1997; Wilensky, 1963).

Although there is evidence on negative effects of competition on cooperation, such studies are not only rare, but are also scattered across disciplines and are mostly limited in the extent to which their findings can be interpreted as causal effects. To resolve this, I suggest combining the competition model I discussed in chapter 2 with a social dilemma game (e.g. a public goods game). Based on such a design, questions such as the following could be investigated: Is there a negative causal effect of fierce competition and relative frustration on trust and cooperation? If so, what would then be the optimal trade-off between competition and cooperation such that limited societal resources could still be distributed in accordance with meritocratic principles, but with a minimum of negative side effects? A deeper understanding of how competition impacts trust and cooperation would therefore be of interest not only for basic social theory but also for more applied questions concerning institutional design.
References


REFERENCES


REFERENCES


Fehr, E. and F. Schneider (2010). “Eyes are on us, but nobody cares: are eye cues relevant for strong reciprocity?” In: Proceedings of the Royal Society B: Biological Sciences 277, pp. 1315–1323.


REFERENCES


REFERENCES

Gordon, D. S., J. R. Madden, and S. E. G. Lea (2014). “Both loved and feared: third party punishers are viewed as formidable and likeable, but these reputational benefits may only be open to dominant individuals.” In: *PLoS ONE* 9, e110045.


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Acknowledgments

Special thanks are due to my supervisor, Andreas Diekmann, for making this dissertation possible as well as for his continuous guidance and advice. I am also indebted to my co-supervisors, Siegwart Lindenberg and Thomas Voss, for their engagement and helpful suggestions, as well as to my co-authors, Charles Efferson, Ernst Fehr, Debra Hevenstone, Sonja Vogt, and Wojtek Przepiorka, who contributed substantially to making this dissertation successful, too. Further, I want to thank my colleagues Heidi Bruderer Enzler, Marc Höglinger, Julia Jerke, Matthias Näf, and Stefan Wehrli, for inspiring discussions that helped developing parts of this dissertation. Deserving of specific mention is Jonas Blatter, who excellently assisted me with both collecting data and formatting this dissertation. Another big thank you goes to Benita Combet, who greatly supported me with her helpful comments on earlier drafts of this dissertation. Last but not least, I wish to thank my family for their support.
Curriculum vitae

Personal
Name Joël BERGER
Date of birth 12 October 1982
Nationality Swiss

Education
02/2012 – present PhD candidate, ETH Zurich, Chair of Sociology, Prof. Dr. A. Diekmann
02/2009 – 08/2011 MA in sociology (summa cum laude), University of Bern
10/2006 – 12/2010 BA in education and sociology (summa cum laude), University of Bern
10/2003 – 07/2006 BA in primary school education, University of Applied Sciences and Arts Northwestern Switzerland (FHNW)
08/1998 – 07/2002 Swiss Baccalaureate (Maturität), secondary school (Kantonsschule), Wohlen
08/1989 – 07/1998 Elementary school, Muri
Curriculum vitae

Positions
10/2011 – present  
Research Assistant, 
ETH Zurich, Chair of Sociology, Prof. Dr. A. Diekmann

02/2013 – 07/2013  
Lecturer, 
University of Bern, Institute of Sociology, Prof. Dr. B. Jann

09/2012 – 05/2013  
Research Associate, 
University of Zurich, Chair of Micro Economics and Experimental Economic Research, Prof. Dr. E. Fehr

Lecturer, 
University of Bern, Institute of Sociology, Prof. Dr. B. Jann

09/2010 – 09/2011  
Undergraduate Research Assistant, 
University of Bern, Institute of Sociology, Prof. Dr. B. Jann

04/2009 – 08/2010  
Undergraduate Research Assistant, 
University of Bern, Institute of Education, Prof. Dr. R. Becker

08/2005 – 07/2009  
Teacher, 
Elementary school, Villmergen

International short visits
06/2014  
University of Leipzig, Institute of Sociology

06/2013  
University of Luxembourg, Institute for Education and Society

Teaching
Spring semester 2015  
Seminar ‘Empirical Research Methods’ (ETH Zurich)

Spring semester 2014  
Seminar ‘Empirical Research Methods’ (ETH Zurich)

Spring semester 2013  
Seminar ‘Values and Norms’ (University of Bern)

Fall semester 2011  
Seminar ‘Experimental Sociology’ (University of Bern)
Awards, grants and funding

2013  Doc.CH grant (Excellence instrument of the Swiss National Science Foundation): CHF 162,308

2011  Award for the best Master’s degree in fall 2011 (grade point average: 6.0 out of 6.0) at the Faculty of Business, Economics and Social Sciences, University of Bern: CHF 1,000

Research interests

• Experimental sociology
• Cooperation and social order
• Competition and social inequality
• Applied game theory

Other activities

• Member of the Swiss Sociological Association (SGS)
• Member of the American Sociological Association (ASA)
• Reviewer for:
  – Games
  – Kölner Zeitschrift für Soziologie und Sozialpsychologie
  – Swiss Journal of Sociology
  – Conference on Norms and Institutions, Ascona, 10-15 May 2015
  – 15th International Conference on Social Dilemmas, ETH Zurich, 10-13 July 2013
Publications

Articles in peer-reviewed journals


Contributions in edited books


**Articles under review**

Berger, Joël; Hevenstone, Debra: “An international field experiment of norm enforcement in anonymous interactions: How costs and benefits impact altruistic punishment.”


Przepiorka, Wojtek; Berger, Joël: “The sanctioning dilemma: A quasi-experiment on social norm enforcement on the train.”

**Work in progress**

Berger, Joël: “Is Willingness to Pay for Environmentally Friendly Products a Signal for Trustworthiness?”


Efferson, Charles; Vogt, Sonja; Berger, Joël; Fehr, Ernst: “The Evolution and Generalization of Social Norms in Children and Grown Ups.”

**Talks and conferences¹**

**Invited Talks**


**Conference Organisation**

Conference on the Legacy of Raymond Boudon, 29-30 May 2014, ETH Zurich. Organised with Andreas Diekmann.

¹ * = Presenter
Conf}erence Talks

Berger, Joël*; Przepiorka, Wojtek; Hevenstone, Debra: “Social Norm Enforcement in the City and on the Train. Evidence from (Quasi-) Experimental Field Studies.” International Conference on Norms and Institutions, Ascona, 13 May 2015.


Berger, Joël*; Diekmann, Andreas: “Are Luxury Brand Labels and Green Labels Costly Signals? Evidence from a Field Experiment.” 15th International Conference on Social Dilemmas, ETH Zurich, 10-13 July 2013.


