

Understanding and Increasing Ethical Behaviour Through Mechanism Design

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2015

vorgelegt von

M.Sc. Janna Ter Meer

aus

Leiderdorp

Referent: Prof. Bernd Irlenbusch

Korreferent: Prof. Bettina Rockenbach

Tag der Promotion: 28.01.2015

Declaration of Authorship

I, Janna TER MEER, hereby declare that I have completed this thesis titled, 'Understanding and Increasing Ethical Behaviour Through Mechanism Design' and the work presented in it without help from third parties and without means of assistance, apart from those indicated. I have cited the sources of all direct and indirect quotations, dates and ideas that are not my own. No other persons were involved in preparing the contents of this work, except for the contribution of listed co-authors to the respective chapters. I certify that I have not used the paid services of consultation firms, and that I have paid no one, directly or indirectly, for tasks connected to the contents of this dissertation. The work has not yet been submitted in the same or similar form to another institution in Germany or abroad. I certify that this statement is true and complete to the best of my knowledge.

Signed:

Date:

“The teacher can lead the student only as far as she has gone herself.”

Donna Farhi

Acknowledgements

This dissertation marks the end of a four-year period of intellectual as well as personal development. Looking back to the start of this journey I see myself as ambitious though naive about my own skill set and what it means to engage in and succeed in academia.

Throughout the last four years I had the pleasure to work in an environment with incredible thinkers, many of whom were friendly and supportive, an enormous amount of freedom to pursue ideas, discuss them with others and strike up collaborations based on joint interests. At the same time, I also learned that academia is an arena with a strong and steep hierarchy, that to most people writing top publications matters more than anything else, including teaching and collaboration, that ideas are sometimes better kept to oneself to avoid others taking off with them and that most of the work is done alone behind your desk. I learned that academia is ruthlessly competitive, that PhD students can have their motivation crushed by supervisors who do not allow them to attend conferences or do a research exchange, that it is considered ‘normal’ for months to go by without meeting with your supervisor.

Navigating this landscape and working towards the dissertation has been both challenging and rewarding. For the end result I’m indebted to a number of people.

First I would like to thank my supervisor, Bernd Irlenbusch, for recruiting me into a PhD after my Master’s program at the London School of Economics, his support, involvement in our joint projects and encouraging me to start attending conferences with a project that was only half finished at the time. My thanks also go to Bettina Rockenbach, who accepted to be my second supervisor and whose profound expertise across a wide range of topics is a continuous inspiration. I would like to thank Uri Gneezy for allowing me to study at the University of California, San Diego, on two research visits and taking so much time for one-on-one meetings, challenging me to formulate my overall research agenda and develop new ideas, dismissing so many of them as not interesting or relevant and sharing your insights about what it takes to succeed in academia. I fondly look back on our interesting, yet at times nervewracking Monday walks across campus where I was given one minute to pitch research ideas. I would also like to thank Michèle Belot for many engaging discussions on our joint project and for making our work relationship feel like one from peer-to-peer, rather than professor to PhD student. Our collaboration has taught me more about doing behavioural economic research than any course I have ever taken. I would also like to thank Jeroen van de Ven, Anna Dreber Almenberg, On Amir, Thomas Buser and Roberto Weber for their interest in and support of my work and Theo Offerman for providing me with a temporary office at the University of Amsterdam. Finally, I have been very fortunate to share offices with Daniela Iosub in

Cologne, Silvia Saccardo in San Diego and Eszter Czibor in San Diego and Amsterdam. Thank you for countless hours in which we were each other's academic soundboard for ideas as well as sharing in each other's successes and frustrations. Your ambition, skill and kindness that I have come to discover over the past years has been truly inspiring.

Lastly, I would like to mention my family and friends for supporting me in my academic endeavours as well as other aspects of life. Saying 'thanks' here does not do justice to the profound gratitude and love I feel towards each of you. Thanks to my mother and Mart for for your endless support and for making home always feel close despite being miles apart. Forfeiting a paid contract to pursue a second visit to UCSD would not have been possible without your emotional and financial support. Thanks to my father and Etel for so much advice on life as well as pushing me on through difficult decisions. Heartfelt thanks to Shian, Katie, Hanna, Amanda, Eric and Yiwen for continuously teaching me courage, love, respect and patience. You help me recognize how powerful it is to pursue your passions and it is beautiful to see you flourish after having made such a life choice. Thank you for so fully enjoying life and those around you. I feel incredibly lucky to be a part of it. Thank you to my brother, Clien, Roos, Sophie, Dave, Rachel, Andrew, Megan, Prabha, John, Jack, Felix, Julia, Willem, Leo and Mau for your smiles and support that make me the person I am today. Thanks to Tim, whose words about my choices and life's truths are somehow never far beneath the surface.

Contents

Declaration of Authorship	i
Acknowledgements	iii
Contents	v
List of Figures	viii
List of Tables	ix
Introduction	xii
1 Lying in public good games with and without punishment	1
1.1 Introduction	2
1.2 Literature review	4
1.3 Hypotheses	6
1.3.1 Incentives for lying in the public good game	7
1.3.2 Treatment-specific hypotheses	9
1.4 Experimental Design	12
1.5 Results	14
1.5.1 Overall contributions and earnings	14
1.5.2 Lying and beliefs	15
1.5.3 The role of punishment	20
1.5.3.1 Punishment assigned	21
1.5.3.2 Reactions to punishment	24
1.6 Discussion	25
1.7 Conclusion	26
2 Fooling the Nice Guys: Explaining receiver credulity in a public good game with lying and punishment	28
2.1 Introduction	29
2.2 Method	30
2.3 Results	32
2.3.1 SVO classification and general patterns	32
2.3.2 The effect of announcements on beliefs	34

2.3.3	The effect on contributions and punishment	37
2.4	Conclusion	38
3	The indirect effect of monetary incentives on deception	39
3.1	Introduction	40
3.2	Literature review and hypotheses	41
3.3	Experimental design and procedures	44
3.4	Results	46
3.4.1	Incentive effects	47
3.4.2	Relative performance under feedback	49
3.4.3	Efficiency	51
3.5	Conclusion	51
4	Are social investments rewarded? A Pay-What-You-Want field experiment with Fair Trade products	53
4.1	Introduction	54
4.2	Literature review	55
4.2.1	Mechanisms	55
4.2.2	The Pay-What-You-Want literature	56
4.2.3	Literature on WTP and ethical consumption	57
4.3	Hypotheses	58
4.4	Experimental procedures	60
4.4.1	The Fair Trade label	60
4.4.2	General procedures	61
4.4.3	Treatments and randomization	62
4.5	Results	63
4.5.1	Checking randomization	63
4.5.2	Profile of the customer	63
4.5.3	Amount paid	64
4.5.4	Purchase rates	67
4.5.5	Profits	68
4.6	Discussion and conclusion	69
A	Appendix Chapter 1: Lying and Public Goods	71
A.1	Additional regression results	72
A.2	Instructions public good game (P-ACT/ANN treatment)	74
B	Appendix Chapter 2: Fooling the Nice Guys	79
B.1	The 32 allocation decision tasks of the ring measure (Liebrand, 1984)	80
B.2	SVO angles and corresponding classifications based on our 25% and Liebrand and McClintock (1988)	81
B.3	Tobit regression: Drivers of the contribution decision	82
B.4	Robustness checks	83
B.4.1	Analysis according to classification of cooperative and individualistic types	83

B.4.2	SVO degree angle as an independent variable in the belief formation regression	84
B.5	Instructions public good game (ANNOUNCE treatment)	85
B.6	Instructions ring measure, translated from the original German	89
C	Appendix Chapter 3: The indirect effect of monetary incentives on deception	91
C.1	Additional regression results	92
C.2	Experimental instructions	95
C.2.1	Instructions	95
C.2.2	Part 1	95
C.2.3	Part 2	97
C.2.4	Private instructions for player A in part 2	98
D	Appendix Chapter 4: Are social investments rewarded?	101
D.1	Photos of stand materials	102
D.2	Randomization	103
D.3	Details of the markets	105
D.4	Script	106
D.4.1	Main interaction	106
D.4.2	Suggested answers to questions from customers	106
D.5	Additional regression results	109
	Bibliography	111

List of Figures

1.1	Contributions to the public good over time across treatments	15
1.2	Beliefs and displayed contributions across treatments.	18
1.3	Errors in belief adjustment across treatments.	18
1.4	Punishment assigned for (perceived) low and high contributions across treatments	22
2.1	Average contributions, announcements and beliefs across periods in ANNOUNCE	35
2.2	Punishment assigned by different types in the two treatments conditional on beliefs (ANNOUNCE) and actual contributions (STANDARD)	38
3.1	Message sent across the incentive treatments, without and with feedback	48
4.1	Sticker for the regular (non-Fair Trade) product	62
4.2	Sticker for the Fair Trade certified product	62
4.3	Average amount paid by condition	64
D.1	Stand display, separate Fair Trade condition	102
D.2	Stand display, separate regular condition	102
D.3	Detail of product and sign, separate regular condition	102
D.4	The stand from a distance with banner and research assistant	102

List of Tables

1.1	Overview of the different treatments	13
1.2	General descriptive statistics	16
1.3	Tobit regression: the effect of reports on subject's contribution decision	20
1.4	Tobit regression: the effect of actual and perceived deviations from the social optimum on punishment assigned	23
1.5	Tobit regression: the effect of received punishment on contribution	24
2.1	General descriptive statistics	34
2.2	Tobit regressions - belief formation in ANNOUNCE	36
3.1	Payoff matrix Y, effective when the receiver does not choose the actual performance level	45
3.2	General descriptive statistics	47
3.3	Probit and OLS regressions: the effect of incentives in the work task on subsequent honesty	49
3.4	Probit regressions: the effect of average and relative performance in the work task on subsequent honesty	50
4.1	OLS regression: Drivers of the amount paid	66
4.2	Marginal and estimated average profit (per 10.000 traffic) per condition	68
A.1	Tobit regression: the effect of beliefs on the contribution decision across treatments	72
A.2	Tobit regression: the role of lies on punishment assigned	73
B.1	The 32 allocation decision tasks comprising the ring measure (Liebrand, 1984)	80
B.2	SVO angles of all experimental subjects and corresponding classifications based on our 25% and Liebrand and McClintock (1988)	81
B.3	Tobit regression: Drivers of the contribution decision	82
B.4	General descriptive statistics for alternative classification	83
B.5	Tobit regression: the effect of the SVO degree angle on belief formation	84
C.1	OLS regressions: the effect of average and relative performance in the work task on the message sent	92
C.2	Probit regression: the effect of relative performance on honesty across incentive conditions	93
C.3	Probit regression: the effect of average earnings on honesty across incentive conditions	94

D.1	Overview of dates and randomization for each market	103
D.2	Demographics of purchaser by condition	104
D.3	General descriptive statistics by market	105
D.4	Purchase rates and total traffic by market	105
D.5	OLS regression: Drivers of the amount paid	109
D.6	OLS regression: Drivers of the amount paid	110

To my grandmother

Introduction

On a typical day, most of us engage in a considerable number of behaviours that could be classified as ‘unethical’. At work, we might be tempted to tell our colleagues that we are on schedule for that joint project, despite not having worked on it productively for days. To the person soliciting donations for Greenpeace outside the neighbourhood supermarket, we lie that we are already a member of their organization so that we will be left alone. While doing our shopping we might purchase non-certified coffee, even when we realize that the farmer growing the beans does not earn enough to make a proper living. And what about stealing money or equipment from the workplace, misrepresenting sales figures, bribing medical staff to receive better health care or outright violent behaviour towards others? While most of us will never be involved in such larger scandals, they are a frequent occurrence in many parts of the world today. Additionally, even more mundane forms of unethical behaviour can have notable repercussions if followed by enough people.

Unethical behaviour has caught the attention of behavioural economists for two main reasons. The first is that at the collective level these behaviours have large negative pay-off consequences to another party or lead to a large redistribution of resources between individuals or entities. For example, the Association of Certified Examiners estimates that occupational fraud accounts for a loss of 5% of revenues, or \$3.5 trillion dollars, at the global level every single year (ACFE, 2012). The second reason for the interest in unethical behaviour is that individuals do not seem to do enough of it. Until relatively recently, the main framework used to understand behaviour such as stealing, lying and bribery was that of Becker (1968)’s model of criminal behaviour. Individuals engage in such behaviours if the benefits of doing so outweigh the costs, which depend on the imposed punishment (eg. a fine or prison sentence) and the probability of getting caught. For many situations in which unethical behaviour takes place, the benefits are considerable and the costs small or zero. Consider the decision to rob someone’s house. While some individuals protect their houses with sophisticated security equipment, many of us simply lock the front door and, if we have been unattentive, left a window open on the first floor. It would be relatively trivial to enter the premise while the stakes (eg. laptop computers, jewelry) are quite large. Furthermore, taking only a few items would not warrant the time and effort of the police for a full scale investigation, making the cost of robbery minimal.

Two main insights from behavioural economics contributed to understanding this puzzle. First, many individuals exhibit social preferences (Fehr and Schmidt, 1999; Andreoni and Miller, 2002; Charness and Rabin, 2002), meaning that they generally care for the

welfare of others and thus may choose to forgo benefits so as not to impose costs on other people. The second insight is that a substantial proportion of individuals are averse to engaging in unethical behaviour. Even in a setting with considerable stakes and no probability of getting caught many individuals do not lie (Gneezy, 2005; Fischbacher and Föllmi-Heusi, 2013; Abeler et al., 2012), steal (Belot and Schröder, 2013) or bribe (Gneezy et al., 2013b).

The four chapters in this dissertation refine and apply insights from this literature through the lens of mechanism design. The study of mechanism design has at its core the design of economic institutions that achieve some predetermined behavioural outcome, such as efficiency, revenue, profits, as well as cooperation or honesty. Incorporating behavioural insights on unethical behaviour into the study of mechanism design is important for several reasons. First, it is possible that the occurrence of unethical behaviour changes the effectiveness of mechanisms that are considered optimal in a more abstract environment (Chapters 1 and 2). Second, mechanism designers, such as employers and policy makers, may consider ethical behaviour a desirable objective in itself. From this perspective, it is important to consider the interaction between chosen incentives, such as revenue-sharing and tournament schemes, and unethical behaviour (Chapter 3). Finally, different mechanisms can be used to determine whether key behavioural assumptions, such as social preferences, have predictive power in actual ethical behaviour (Chapter 4).

Controlled experiments are key in establishing causal relationships between economic institutions and unethical behaviour. As such the chapters in this dissertation rely on both laboratory and field experiments. The two main advantages of studying unethical behaviour in the laboratory are the possibility of measuring its occurrence and quantifying the payoff consequences. To illustrate, consider a salesperson tasked with submitting a subjective review report on customer satisfaction. To determine whether information in the report has been inflated, it is necessary to ascertain the salesperson's belief about the actual level of customer satisfaction as well as the expected material harm (or benefit) to themselves and the company from such an action. In a field setting such measures are rarely available. By contrast, in the laboratory such beliefs can be fixed by providing participants objective information about a true state and quantifying the payoff consequences from honest and deceitful communication. In addition, the laboratory environment allows individuals to be randomly assigned to different institutions, such as a public good setting with or without punishment (Chapters 1 and 2) or a work task with tournament or team incentives (Chapter 3). At the same time, the abstract environment of the laboratory has its limitations for tackling certain research questions. A principal objective of the work in Chapter 4 is to study the viability of the Pay-What-You-Want

pricing mechanism for purchasing ethical products. For this we needed participants to make actual purchase decisions and thus opted for a field experiment.

Chapter 1, entitled ‘Lying and Public Goods’ and joint work with Bernd Irlenbusch (University of Cologne) examines the behavioural implications of lying in the well-studied setting of public good provision, in which individual agents need to cooperate in order to achieve a socially optimal outcome. While cooperation is typically difficult to achieve, certain mechanisms such as costly peer-punishment are generally effective in mitigating the free-rider problem. In an experiment we evaluate the effectiveness of the punishment mechanism in a public good setting where individuals do not receive feedback about the contributions of others but have a possibility to communicate to one another what they have contributed. This setting gives rise to a number of constraints to full cooperation. First, it is possible that group members do not believe announcements of their fellow group members. From the perspective of maximizing contributions, this is problematic for subjects who contribute to the public good when they know that others are doing so as well. In addition, when peer punishment is introduced it is possible that punishment is assigned to high contributors whose reports are not believed or less punishment is assigned to low contributors who get away with an inflated announcement. We find evidence for both constraints in our experiment, which reduces overall contributions and earnings compared to the standard public good game.

Implicit in this work is that individuals make systematic mistakes when interpreting potentially dishonest messages. This is a necessary condition for deception to occur: the sender must believe that their message can influence the beliefs of the other party. Chapter 2, entitled ‘Fooling the Nice Guys’ and joint work with Bernd Irlenbusch (University of Cologne) investigates this in the same public good setting featured in Chapter 1. We find that a false consensus effect can partially explain how group members form beliefs based on the messages they receive. Using an independent proxy of contribution tendency, we find that subjects who are likely to contribute to the public good are more likely to believe messages that others are also contributing. While individuals with a tendency to free-ride show the opposite pattern, we cannot exclude the possibility that these individuals are simply well calibrated in their beliefs about actual contributions. Together, these first two chapters show that lying aversion explains behaviour in a symmetric public good setting and that own behavioural tendencies can partially explain how receivers interpret messages of others. Furthermore, the possibility for lying in a public setting constrains full cooperation even in the presence of an otherwise efficient solution mechanism.

Chapter 3 examines the reverse relationship by exploring the effects of mechanism design on lying behaviour. In a laboratory experiment subjects work under either a piece rate,

team incentives or a tournament scheme and then are presented with another task in which they can be dishonest for a monetary gain. Rather than testing for a direct effect on dishonesty, the results from this study are the first to provide support for the notion that monetary incentives can affect dishonesty in a subsequent unrelated task. In particular, working under the tournament incentive negatively affects honesty. In addition, when relative performance information is provided, this feedback appears to decrease honesty for workers who under- or outperform their work partner by a small amount. From a theoretical perspective these results are informative on what determines dishonest behaviour in individuals. In addition, they are instructive for mechanism designers who care about honesty.

The dissertation closes with Chapter 4, entitled ‘Are social investments rewarded?’ and joint work with Ayelet Gneezy (University of California, San Diego). It is slightly different from the previous chapters in that it focuses on ethical consumption, where individual consumers choose to purchase a product that directly or indirectly contributes to the welfare of a third party. The key questions in this paper are first whether the motivation of social preferences and self-identity concerns play a role in ethical consumption decisions and second, whether this would make a Pay-What-You-Want pricing mechanism more viable for ethical products. The Pay-What-You-Want pricing mechanism is suitable for studying this question because it allows people to determine their own price. As such we expect that if individuals have social preferences or self-identity concerns for ethical products, this should translate into higher payments. We test this in a field experiment by offering a regular and Fair Trade product to customers at a local Farmer’s Market. Customers are either presented the products separately or together. The results show that customers pay more for the Fair Trade product than the non-certified alternative when the two are offered together. However, this difference disappears when the products are offered separately. Specifically, payments for the regular product decrease when it is presented next to the Fair Trade alternative compared to when this product is offered on its own. Since there is no movement in payments for the Fair Trade, these results do not support that social preferences or self-identity concerns translate into higher payments.

Chapter 1

Lying in public good games with and without punishment

Joint work with Bernd Irlenbusch, University of Cologne

Abstract

We experimentally study a public good setting where accurate contribution feedback is not available and group members can send non-verifiable cheap talk messages about their contributions. As feedback, subjects receive only announced contributions or the announced or actual contribution with 50% probability. In this setting, we explore both information transmission and reception as well as the effectiveness of costly peer punishment. Overall, we find that cooperation breaks down in all announcement treatments except when actual contribution feedback is provided some of the time and punishment is available. We identify various constraints to full cooperation relative to the standard public good game. First, subjects make errors in adjusting their beliefs for the announcements of others and, on average, adjust their beliefs downward for a given announcement. Second, we find that significantly more punishment is assigned to high contributors compared to the standard public good game. Furthermore, punishment for low contributors appears to have a smaller disciplining effect. When actual contribution information is provided some of the time we find that these constraints are less severe compared to the setting where only announcements are available. However, when only announcements are displayed there is an overall decrease in punishment levels relative to the other treatments and it also fails to discipline low contributors. We do not find a mark-up in punishment for lying in any of the announcement treatments.

1.1 Introduction

A plethora of economic activities are characterised by public good structures where cooperation is essential for success. Examples include collaboration in teams, charity donations, and international endeavours to protect the environment. It therefore comes as no surprise that researchers have invested considerable effort to mitigate the free-riding problem in public good provision (Ledyard, 1995; Chaudhuri, 2011). Two particularly prominent insights emerged from this literature. First, many people are conditional cooperators meaning that their contributions to a public good increase when it is known that others contribute as well (Fischbacher et al., 2001; Keser and Van Winden, 2000). Second, bilateral punishment mechanisms can be used to discipline free-riders and these are effective to sustain cooperation even if punishment is costly for the punisher (Yamagishi, 1986; Ostrom et al., 1992; Gächter and Fehr, 2000; Fehr and Gächter, 2002). Both insights have been successfully used to design mechanisms that induce higher contributions (Ostrom, 1990; Frey and Meier, 2004; Gächter, 2007; List and Lucking-Reiley, 2002; Shang and Croson, 2009). Such mechanisms, however, crucially depend on the assumption that potential contributors have access to reliable information about the contributions of others. In many situations such an assumption seems unwarranted. Consider the case of fisheries management where accurate catch data is crucial in enforcing control systems such as total allowable catch and transferable quotas. While it is possible to track the vessel's movement and time at sea, it is difficult to record the exact catch size in an accurate and timely fashion (Beddington et al., 2007). While authorities rely on some form of monitoring, for example by letting observers perform random checks of the vessel's equipment, collecting fully accurate catch information is prohibitively costly. Unreliable information on contributions is also present in other settings. In teams, for example, group members work in spatial distances from each other such that individual effort levels are hardly mutually observable. Privacy considerations can also prevent the disclosure of reliable contribution information, such as in the case of charity donations.

Instead, the information that is often available is what others announce about their team efforts, fish catch, donations and so forth. For fisheries, it is common for individual fishermen to keep records of their catch in a manual or electronic logbook (Barkai et al., 2012). These numbers are consequently aggregated to determine quotas and forecasts, supplemented by other scientific measurements of the fish stock. The advantage of using logbooks is that information is immediate and collecting is relatively inexpensive. However, it is subject to misreporting. Fishermen can record a lower catch volume in

the books, which leads to a bias in official statistics¹ (Gagern et al., 2013; Pauly et al., 2013).

A public good setting where contribution feedback is not available, but communication is possible between group members generates important new questions for the public good literature. First, how honest are participants and to what degree do they trust the announcements of others? Conditioning own contributions on announcements that are not trusted seems problematic. Further, administering bilateral punishment is not straightforward. Would one refrain from punishing somebody who reports a high contribution? Or would one rather exert a particularly high punishment if one believes that the actual contribution was low and on top of that the announcement has been a lie? How do contributions develop over time when feedback is (partially) based on announcements?

To shed light on these questions we experimentally investigate the impact of participants' non-verifiable announcements about their own contributions on public good provision. First, we investigate to what extent group members lie about their contributions and how others perceive this information. Second, we examine possible inefficiencies that this creates in a public good setting with and without presence of costly peer punishment.

In our study we employ a standard repeated public good setting. The new feature is that participants make an announcement about their contribution after they decide about their actual contributions. They are free to announce whatever contribution they want irrespective of what they actually contributed. Payoffs are based on actual contributions and not on announcements. We employ a 2 x 2 experimental design. On one dimension we consider public good settings with and without punishment. To assess the effect of credibility of announcements, we vary the probability with which the announcements of the subjects are taken as feedback or whether true feedback is provided on the other dimension. Announcements are either taken as feedback with certainty (in treatments ANN and P-ANN) or with a probability of 0.5 (in treatments ACT/ANN and P-ACT/ANN). We also include a belief measure to evaluate to what extent announcements of others are believed and how subjects condition their contribution and punishment behaviour on these beliefs. For comparison we also include a treatment with a standard public goods game with punishment which entails only true feedback (treatment P-ACT).

¹There is some support that such misreporting is taking place. In 2010, the amount of Mediterranean Bluefin tuna reaching the market exceeded the reported catch amount by 40% (Gagern et al., 2013). Similarly, the Chinese fleet is estimated to have caught 4.6 million metric tons a year in distant waters between 2000 and 2011, of which less than 10% was reported to the UN Food and Agriculture Organization (Pauly et al., 2013).

We find that cooperation breaks down in all announcement treatments except when actual contribution feedback is provided some of the time and punishment is available (P-ACT/ANN). Here punishment holds contributions at intermediate levels, even though it is not efficient in terms of earnings. Driving these effects, we identify various constraints to full cooperation across the announcement treatments relative to the standard public good game. First, subjects make errors in adjusting their beliefs for the announcements of others and, on average, adjust their beliefs downward. Second, we find that significantly more punishment is assigned to high contributors compared to the standard public good game and that these contributors reduce their subsequent contributions. Furthermore, punishment for low contributors appears to have a smaller disciplining effect. When actual contribution information is provided some of the time we find that these inefficiencies are less severe compared to the setting where only announcements are available. However, when only announcements are displayed there is an overall decrease in punishment levels relative to the other treatments and it also fails to discipline low contributors. We do not find a mark-up in punishment for lying in any of the announcement treatments.

The rest of the paper is structured as follows. In the next section we summarize the relevant literature and derive hypotheses in section 3. Section 4 describes the experimental design. Section 5 presents our findings, followed by a discussion in section 6. Section 7 concludes.

1.2 Literature review

Several studies look at the effectiveness of public good provision and the punishment mechanism when the assumption of accurate contribution feedback is relaxed. [Ambrus and Greiner \(2012\)](#) evaluate a public good game with a binary strategy space of a full or zero contribution. In case subjects choose to contribute to the public good, there is a small probability that their contribution is displayed as zero to the other group members. In addition, subjects have the possibility to punish group members at a cost. They find that average earnings are lower in settings with noise and standard punishment technology. A stronger punishment technology, where each point invested in punishment reduces the target's earnings by 6 points, is more effective in maintaining high contributions, although average earnings do not improve beyond that of the no-punishment control group. The authors attribute this efficiency loss to continued use of the punishment mechanism in the treatments with noise. In the standard public good game, punishment is used in the initial rounds but then phases out, resulting in efficiency gains. [Grechenig et al. \(2010\)](#) find a similar result in a public good game

where subjects can contribute anything between 0 to 20 points. With some positive probability, the subject's actual contribution is replaced by a random number from the strategy space and given as feedback to the other group members. In other words, it is possible for a low contribution to be displayed as high, and vice versa. Costly peer punishment is effective in maintaining high contributions when actual contributions are displayed in 100% or 90% of the cases. Cooperation breaks down when accuracy drops to 50%. However, even under minimal noise (90% accuracy) average earnings are not higher than the treatments without punishment. There are several studies that manipulate contribution feedback but do not include a punishment mechanism. Work by [Nikiforakis \(2010\)](#) finds that subjects contribute less to the public good if they receive feedback about earnings rather than contributions. In the absence of any contribution feedback, [Neugebauer et al. \(2009\)](#) and [Sell and Wilson \(1991\)](#) find that contributions are stable over time compared to a control treatment where contribution feedback is provided. Finally, there is substantial literature on the role of communication in public good provision. Generally, communication improves public good provision ([Dawes et al., 1977](#); [Isaac and Walker, 1988](#); [Brosig et al., 2003a](#); [Bochet et al., 2006](#)) even when no contribution feedback is provided ([Wilson and Sell, 1997](#); [Cason and Khan, 1999](#)).

To our knowledge, two papers have thus far looked at lying in public good settings. The first is [Hoffmann et al. \(2013\)](#) who study the effect of inflated feedback on contributions. In the experiment, feedback about the group average contribution is exogenously inflated by 25%, or identical to one's own contribution if the individual is contributing above the group average. They find that inflated feedback is successful in raising contributions as long as high contributors remain unaware that they are contributing above the group average. The second paper, by [Serra-Garcia et al. \(2013\)](#), looks at the content of communication on lying and free-riding in a 2-player one-shot public good game. The experimental setting features an informed player who has private information about the MPCR to the public good and can communicate this to the uninformed player. They find that subjects lie less when the message describes future behaviour ('I contribute') compared to when they are describing a state ('the return is high').

Our work differs from and adds to these previous studies in three important ways. First, rather than introducing noisy feedback exogenously, any discrepancy between actual and displayed contribution information in our experiment is created by the subjects themselves. In other words, we look at endogenous feedback distortion, where accuracy in feedback depends on honesty. This makes it important to understand the degree to which subjects are honest, as well how they perceive the messages of other group members. From this perspective, the inclusion of our belief measure is an important addition to previous studies. Second, we evaluate this in a repeated public good setting in which

subjects do not receive accurate contribution feedback during the rounds of the experiment. In previous work on the role of communication, subjects typically communicate before making their contribution decisions. Subjects then receive accurate feedback on what their fellow group members actually decided before moving to the next round. Even though it is possible for subjects to make false promises in this context, any discrepancies are immediately revealed by the feedback mechanism. We focus on situations where such verification is not (immediately) possible. Finally, since we include a measure of beliefs, we can investigate how honest and dishonest messages affect subjects' perceptions about the contributions of others. This allows us to answer questions on conditional cooperation and motivations behind punishment behaviour when reliable contribution feedback is not available.

1.3 Hypotheses

To formulate our hypotheses we make several assumptions about the motivations of subjects in the public good game with respect to their contribution and lying behaviour. Note that these hypotheses are not meant to provide a definitive account of the underlying mechanisms. They simply serve to make plausible predictions about behaviour based on the canonical model of rational self-interested agents and well-supported behavioural alternatives. We entertain four constellations of motivations for subjects:

1. Only self-interested subjects and no cost of lying
2. Only self-interested subjects and moderate cost of lying
3. A proportion of conditional cooperators and no cost of lying
4. A proportion of conditional cooperators and moderate cost of lying

These four constellations speak first to the driver of contribution behaviour (self-interest or conditional cooperation) and second, to the motivation to misrepresent one's contribution (no or moderate costs of lying). We discuss each of these in turn.

The canonical model postulates that subjects are motivated exclusively by monetary self-interest. Since the marginal per capita rate of return to investment in the public good is lower than 1, it is individually rational for each subject to invest everything in the private account and contribute zero to the public good. In reconciling this assumption with experimental evidence on contributions in public good games, [Fischbacher et al. \(2001\)](#) and important follow-up work ([Frey and Meier, 2004](#); [Fischbacher and Gächter, 2010](#)) identify a proportion of subjects as conditional cooperators. Rather than being

driven by self-interest, these subjects are willing to contribute to the public good if other group members are also contributing. For these individuals, beliefs about what others are contributing are key in understanding contribution behaviour.

When it comes to misrepresenting one's contributions, we again start with the assumption of the canonical model that individuals do not experience any psychological disutility from communicating dishonest messages. The assumption that individuals have no costs of lying² has been challenged in a growing literature on lying aversion (Gneezy, 2005; Mazar et al., 2008; Sutter, 2009; Erat and Gneezy, 2012; Fischbacher and Föllmi-Heusi, 2013). For example, Gneezy et al. (2013a) and Gibson et al. (2013) identify different types of people according to their lying costs, i.e. those who are totally honest or dishonest, or those who vary their lying behaviour depending on the potential private rewards and harm caused to the other party. In formulating our hypotheses for instances (2) and (4), we follow this assumption that individuals are heterogeneous in their lying costs and that these costs, on average, are non-negligible.

1.3.1 Incentives for lying in the public good game

If subjects are motivated exclusively by monetary self-interest, they follow the dominant strategy of zero contributions to the public account. Their beliefs about what others are contributing is irrelevant for their own contribution decision. Since the subject's contribution decision is not dependent on beliefs about the contributions of others, it follows that communicating a number different from one's actual contribution does not yield any material benefit. Given that there are no incentives for lying, we expect contributions to be disclosed honestly whenever the utility function of subjects can be characterized exclusively by monetary self-interest. This prediction does not change when we introduce moderate lying costs in instance (2).

Hypothesis 1a. *If subjects are driven purely by self-interest, contributions to the public good are zero and there are no dishonest announcements irrespective of the lying costs of the subjects.*

This prediction changes when we assume that a proportion of subjects are conditional cooperators. Since the contribution decision of these subjects is based on their beliefs

²We use 'costs of lying' as a general term to refer to the psychological disutility experienced by telling a lie. We are not specific in whether these costs derive from an inherent aversion to telling lies (Gneezy, 2005; Vanberg, 2008) or through the experience of guilt (Battigalli et al., 2013).

about what others are investing in the public account, it can be beneficial to the individual subject to announce a higher contribution than what was actually contributed. If this inflated announcement translates into higher beliefs about actual group contributions, conditional cooperators can be expected to contribute more compared to a setting in which contribution feedback is accurate. Since the returns to investment in the public account are shared equally among the participants, this represents a monetary gain for the liar at the expense of the contributing group member. Given that this gain is only present when conditional cooperators are convinced that the group contributions are higher than they actually are, it follows that there are no incentives for subjects to underreport their actual contribution. Simply stated, subjects face a trade-off between reporting their actual contribution honestly or inflating it by communicating a higher number. Thus, the presence of conditional cooperators in the subject pool creates incentives for subjects to overstate their actual contributions.

In instance (3) where we assume no costs of lying, we expect subjects to overstate their actual contributions to the largest degree possible. For self-interested subjects, this would express itself as a contribution of zero to the public good coupled with a high announcement. We would expect higher contribution levels from conditional cooperators, but again coupled with inflated announcements. Since there are no incentives to underreport, it follows that this behaviour ‘contaminates’ higher announcements levels, since these can reflect both high actual contributions or an exaggerated report.

Hypothesis 1b. *If a proportion of subjects are conditional cooperators and the subjects experience no cost of lying, announcements will be strongly inflated relative to actual contributions.*

If subjects face non-negligible costs of lying (instance 4), we expect subjects to announce their contribution honestly or overstate by less compared to when lying costs are zero. This implies that high announcements are more credible than in instance (3), since there is now an increased likelihood that these announcements actually correspond to high contributions.

Hypothesis 1c. *If a proportion of subjects are conditional cooperators and the subjects experience a cost of lying, there will be a small or moderate inflation of announcements relative to actual contributions.*

The observant reader will notice that these predictions hinge on certain assumptions about how announcements are interpreted by the other group members. If we assume that receivers detect lies correctly and adjust their beliefs appropriately, lying cannot be successful in convincing conditional cooperators that contributions are higher than they actually are. Again, this removes the incentive for subjects to misrepresent their contributions. While the assumption of perfectly rational receivers is used in some theoretical work (Crawford, 2003; Kartik et al., 2007), a number of experimental papers show that individuals often make mistakes in detecting lies (Blume et al., 2001; Charness and Dufwenberg, 2006; Wang et al., 2010; Sheremeta and Shields, 2013) even though receivers' beliefs, on average, do respond to structural factors that affect the underlying deception rate (Belot et al., 2012; Sutter, 2009; Charness and Dufwenberg, 2010). For a detailed analysis on receivers' interpretation of cheap talk messages in a public good game with announcements, see Irlenbusch and Ter Meer (2013) or chapter 2 of this work. We follow the general behavioural assumption here in that the recipients of cheap talk messages do not accurately adjust for lying, but that subjects are attuned to the general incentive structure underlying lying behaviour. In our experiment, this implies that, on average, subjects should revise their beliefs downward rather than upward to account for the possibility that group members are overstating their contributions.

Hypothesis 2a. *If subjects are fully rational, there will be no discrepancy between announcements and subjects' beliefs about underlying actual contributions.*

Hypothesis 2b. *Subjects make errors when adjusting their beliefs and on average revise their beliefs downward for a given announcement.*

1.3.2 Treatment-specific hypotheses

Having set the stage regarding lying behaviour in the public good game, we now derive further hypotheses specific to our treatments.

The difference between the ACT/ANN and ANN treatments is that in ACT/ANN the subject's actual contribution is given as feedback to the other group members with probability 0.5, whereas this has a probability of 0 in ANN. In other words, in the ANN treatments only the subject's announced contribution is displayed as feedback. All of this is common knowledge to the subjects and clearly emphasized in the instructions and control questions (see appendix A.2). This weak form of monitoring in the experiment gives subjects in the ACT/ANN treatments more certainty that the information

they are receiving on the feedback screen is accurate, at least with 50% probability or higher if they believe group members are honest. This adds credibility to the reported contributions in the ACT/ANN treatments.

Hypothesis 3. *Reported contributions are more credible in the P-ACT/ANN and ACT/ANN treatments compared to the P-ANN and ANN treatments.*

This has implications for both contribution and punishment behaviour. Since displayed contributions are more likely to be credible in the ACT/ANN treatments it allows conditional cooperators to condition stronger on reported feedback.

Hypothesis 4. *Conditional on reported contributions, subjects in P-ACT/ANN and ACT/ANN contribute more to the public good than those in the P-ANN and ANN treatments.*

Hypothesis 5. *In the absence of punishment, contributions in ACT/ANN will be higher than in ANN.*

For our predictions on punishment, we start with the observation that, contrary to the canonical model of self-interested agents, a proportion of subjects are willing to exert costly punishment towards group members (Gächter and Fehr, 2000; Fehr and Gächter, 2002) and that, generally, low contributions are punished more frequently and severely than contributions closer to the social optimum (Herrmann et al., 2008). In a setting where contribution feedback is distorted, inferences about group members' contributions to the joint account are not as straightforward as in the standard public good game. Particularly for high announcements, it is possible that a discrepancy exists between actual contributions and beliefs, in which (i) a group member is believed to make a low contribution when this person's actual contribution is in fact high, or (ii) a group member is believed to make a high contribution when this person's actual contribution is in fact low. This implies that punishment is more likely to be misdirected in P-ACT/ANN and P-ANN due to erroneous beliefs compared to the standard public good game. This reduces the effectiveness of the punishment mechanism for two reasons. First, if there is a positive probability that the announcement is to some extent believed, the free-rider will receive less punishment than in a public good game where contribution feedback is accurate. This can reduce the disciplining effect of punishment for free-riders to increase their contributions (Fehr and Fischbacher, 2003). Second, high contributors that

are punished may react adversely by reducing their subsequent contributions (Herrmann et al., 2008). We have no a priori hypotheses on which of these mechanisms would underlie the reduced effectiveness of punishment in ACT/ANN and ANN compared to the standard public good game with punishment. However, our experimental data does allow us to evaluate the role of each of these explanations. Furthermore, if reported contributions are more credible in the ACT/ANN treatments, it follows that for subjects who are willing to punish, punishment is correctly targeted with a higher probability.

Hypothesis 6. *Punishment is less effective in raising contributions in P-ACT/ANN and P-ANN compared to the standard public good game, P-ACT.*

Hypothesis 7. *Punishment is less effective in raising contributions in P-ANN than P-ACT/ANN.*

In this section we have outlined several sources of inefficiency that are specific to the public good game with endogenous noise. First, the presence of conditional cooperators creates incentives for subjects to overstate their actual contributions. Depending on the lying costs of the subjects, we can expect contributions to be moderately or strongly inflated, leading to a contamination of high reported contributions. These can reflect an honest announcement or an exaggerated report. If subjects adjust their beliefs downward, we can expect conditional cooperators to contribute less than in the standard public good game for a given report. Two other possible inefficiencies originate from the punishment mechanism as described under hypothesis 6: if subjects are not well calibrated in their beliefs, free-riders can escape punishment when announcing high or high contributors receive punishment when their (honest) reports are not believed. Since each of these inefficiencies are absent in the standard public good game, we expect that the standard public good game with punishment results in higher overall contributions than both the P-ACT/ANN and P-ANN treatments. In addition, extending the argument made under hypothesis 5 and 7, we would also expect contributions to be higher in P-ACT/ANN compared to P-ANN.

Hypothesis 8. *Overall contributions will be higher in P-ACT than in P-ACT/ANN and P-ANN.*

Hypothesis 9. *Overall contributions are higher in P-ACT/ANN than in P-ANN.*

1.4 Experimental Design

In all experimental sessions, subjects played a four-person public good game with standard parameters (Fehr and Gächter, 2002). The game is repeated for 15 rounds and subjects stay in the same group throughout the experiment. At the start of each round subjects receive an endowment of 20 points, which they allocate either to themselves or to a shared account. Each point kept for oneself increases the subject's earnings by 1 point, whereas those allocated to the group account are multiplied by a factor of 1.6 and equally divided over the four group members.

We introduce communication through a post-hoc announcement mechanism, which is inserted immediately after the actual investment decision. Here, each subject makes a non-binding payoff-irrelevant announcement on how many points he or she has contributed to the group project on the previous screen. Subjects have the possibility to lie by reporting a lower or higher number than what they actually contributed. Thus, whether such a discrepancy between actual and announced contribution exists is entirely up to the individual subject. Both the actual investment decision and the announcement are made simultaneously by all group members. After the announcements have been made, subjects move to the feedback stage where they receive information about the individual contribution decisions of each of their fellow group members. Feedback is displayed anonymously and in random order to prevent subjects from tracking individual behaviour across periods.

Within this basic framework, we introduce two treatment variations. The first is the punishment mechanism, which is either present or absent. In the treatments with punishment, subjects have the possibility to assign punishment points in the feedback stage. Subjects are given 10 additional points per round that can be invested in punishment. Each point invested reduces the earnings of the targeted subject by three points. Any unused punishment points are added to the subject's individual earnings, thereby making punishment costly to administer. Each subject is subsequently informed about the sum of punishment points they received (if any) and the game is repeated until all fifteen rounds are finished. Subjects receive aggregate information on actual contributions and earnings only at the end of the experiment.

The payoff formula for each subject i is as follows:

$$\Pi_i = (20 - c_i) + (0.4 \sum_{k=1}^4 c_k) - (3 \sum_{k \neq i} p_k^i) + (10 - \sum_{k \neq i} p_i^k)$$

where c_i represents the contribution of subject i to the group project. p_j^i indicates how much punishment subject i receives from subject $j \neq i$ where $i, j \in \{1, \dots, 4\}$. Importantly, announcements are not payoff-relevant. Only the actual contributions and punishment of the subject and other group members enter the payoff function.

Our second treatment variation determines the information subjects receive in the feedback stage. In ACT, feedback on the contributions of the group members reflects their actual contribution decision in all instances. This is identical to the standard public good game, in which contribution feedback is always accurate. By contrast, in the ANN treatments, information displayed only reflects whatever was announced. ACT/ANN lies in between the two extremes. With 50% probability, the number displayed on the feedback screen reflects either the subject's actual or announced contribution. This is determined for each group member individually. Each displayed contribution on the feedback screen can reflect either the actual or announced contribution of the group member.

Given that the contribution information provided on the feedback screen is not necessarily accurate, we record what subjects believe about the actual contributions of the other group members. We elicit these beliefs in the feedback stage for each displayed contribution of the other group members, which provides us with three belief measures per subject per round. This belief elicitation is not incentivized, since past experimental work suggests it can affect contribution decisions (Gächter and Renner, 2010; Croson, 2000).

TABLE 1.1: Overview of the different treatments

	Feedback		
	Actual	Actual/Announced	Announced
Punishment	P-ACT (n = 56)	P-ACT/ANN (n = 52)	P-ANN (n = 56)
No Punishment		ACT/ANN (n = 56)	ANN (n = 56)

Number of participants in brackets. In P-ACT/ANN one of the groups of 4 participants could not be established because one of the registered subjects did not show up.

Table 1.1 summarizes our five treatments according to the variations of punishment and feedback. We label each treatment according to whether the punishment mechanism was present or absent (indicated by the letter 'P') and what information subjects are provided as feedback (actual contributions, announcements, or a mixture of both with equal probability). Thus, the label P-ACT/ANN refers to the treatment with punishment and actual or announced contributions as feedback.

The five treatments were conducted over 10 sessions (two per treatment) at the economics laboratory at the University of Cologne, Germany. We recruited a total of 276 undergraduate and graduate students to participate using the ORSEE online recruitment system (Greiner, 2004). This corresponds to 13 independent observations in the P-ACT/ANN treatment and 14 independent observations in each of the other treatments. The lower number in P-ACT/ANN was due to an insufficient number of participants arriving for the experiment. The mean age of participants is 23.3 years, with 52.7% female. The vast majority of participants were German nationals from a range of academic disciplines, including economics and business. No subject participated in any of the sessions more than once. Upon entering the lab, participants were seated at individually separated computers and given instructions. They had to successfully complete a set of control questions to ensure their understanding of the game before the experiment continued. Except for the instructions, the experiment was computerized and programmed using the z-Tree experimental software (Fischbacher, 2007). Each session lasted approximately 80 minutes and subjects were paid, on average, €12.52 at an exchange rate of 50 ECU to €1.

1.5 Results

1.5.1 Overall contributions and earnings

Figure 1.1 depicts contributions to the public good over the fifteen periods with the punishment treatments in the left panel and those without punishment on the right. It shows that contributions converge to the social optimum in the treatment with accurate feedback, P-ACT, but not in either announcement treatment, P-ACT/ANN and P-ANN. Contributions in P-ANN seem to fall over time to levels similar to that in the treatments without punishment. However, when accurate feedback is available some of the time, punishment seems to hold contributions at intermediate levels. Mann-Whitney U-tests (MWU)³ at the level of independent observation confirm that contributions in P-ACT are significantly higher than both P-ACT/ANN ($p < 0.01$) and P-ANN ($p < 0.01$). Yet, P-ACT/ANN does better than the treatments without punishment (both $p < 0.01$) and compared to P-ANN ($p = 0.029$) when we restrict our analysis to the final five periods of the game. Comparing contributions over the entire game results in weakly higher contributions in P-ACT/ANN compared to P-ANN ($p = 0.081$). We test for time trends non-parametrically using a binomial test on the Spearman rank correlation coefficient between contribution and period number for each independent observation. In the treatments without punishment and P-ANN, the rank correlation coefficient is

³Unless otherwise specified, all reported non-parametric tests are two-sided.

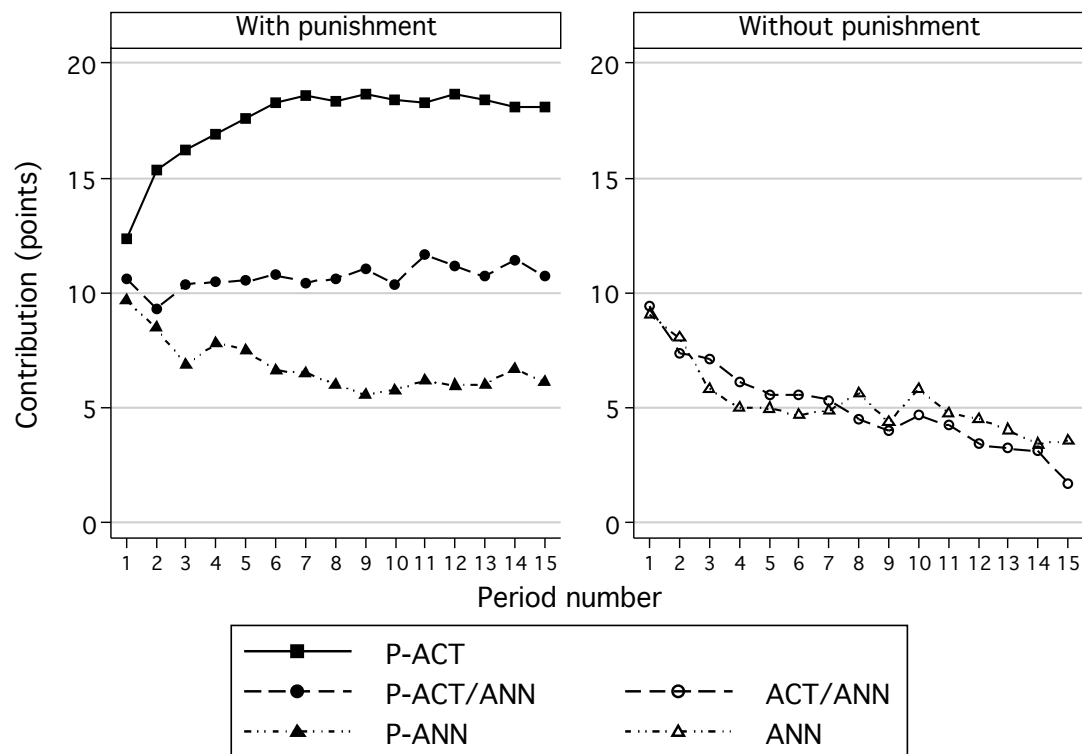


FIGURE 1.1: Contributions to the public good over time across treatments

negative significantly more often than chance, indicating that contributions fall over time ($p < 0.01$). While contributions rise in P-ACT ($p < 0.01$), no significant downward or upward trend was detected for P-ACT/ANN ($p = 0.58$). Table 1.2 provides detailed descriptive statistics on contributions and other variables of interest, such as average earnings. Despite moderate public good contributions in P-ACT/ANN, earnings in this treatment are significantly lower than in P-ACT but also compared to the no-punishment treatments (MWU, all $p < 0.01$).

Thus, these results support hypotheses 8 and 9 in that contributions are higher in P-ACT/ANN than in P-ANN, but that neither are as high compared to the standard public good game, P-ACT. We do not find that contributions are higher in ACT/ANN compared to the ANN treatment and thus fail to support hypothesis 5. We discuss these results in detail in section 1.6.

1.5.2 Lying and beliefs

A unique feature of our experimental design is that accurate contribution feedback is obscured and that subjects can send non-verifiable announcements about their contribution. As such, the degree of feedback distortion hinges on subjects' honesty in their

TABLE 1.2: General descriptive statistics

	Average contribution	Average lie	Average punishment	Average earnings (€)
<hr/>				
Actual Feedback				
<i>Punishment</i>	17.49 (4.55)		0.20 (0.86)	11.42
<hr/>				
Actual / Announced				
<i>Punishment</i>	10.68 (7.08)	4.00 (5.10)	0.53 (1.36)	9.03
<i>No Punishment</i>	5.03 (6.04)	5.83 (5.61)		9.91
<hr/>				
Announced feedback				
<i>Punishment</i>	6.79 (7.59)	10.35 (8.09)	0.14 (0.64)	9.71
<i>No Punishment</i>	5.24 (7.21)	10.54 (7.93)		9.94
<hr/>				

Standard deviations are show in brackets.

announcements as well as the beliefs about these announcements of others in the group. We evaluate these next.

Lying is prevalent in the experiment. On average, announcements are truthful⁴ less than a third of time. In line with previous work (Gneezy, 2005; Gibson et al., 2013), we find subjects that never lie ($\sim 10\%$), always lie ($\sim 21.8\%$) or show a mix between honest and dishonest announcements ($\sim 68.2\%$) across the treatments. The black line in figure 1.2 represents average reported contributions for each level of actual contribution, clustered in blocks of three. The actual underlying contribution is indicated by the solid gray reference line and beliefs about the underlying actual contribution are represented by the dashed black line. Average reports are significantly higher than actual contributions in all treatments (WSR, $p < 0.01$). When accurate contribution feedback is displayed some of the time, subjects overstate their contributions by an average of 4 and 5.83 points in the treatments with and without punishment respectively. When only announcements are displayed these averages are 10.35 and 10.54 points for the punishment and no-punishment treatments. This difference in average overstatements between the ACT/ANN and ANN treatments is significant (MWU, both $p < 0.01$)⁵.

⁴We label an announcement as truthful when it exactly corresponds with the subject's actual contribution in that period.

⁵This result remains significant at the 1% level when, instead of comparing absolute lies, we consider the discrepancy between announced and actual contribution as a percentage of how much the subject can

We find support for the hypothesis that subjects overstate their actual contributions, in line with what we would expect if a proportion of subjects are conditional cooperators and lying is not prohibitively costly for all subjects. However, we observe significantly higher overstatements in the ANN treatments compared to ACT/ANN. Thus, we find support for both hypothesis [1b](#) and [1c](#). We postpone our discussion of this result to section [1.6](#).

overstate. This addresses the concern that high contributors lie less because they have less possibility to overstate, since announcements are capped at 20 by design.

FIGURE 1.2: Beliefs and displayed contributions across treatments.

Note how the reference line representing actual contributions does not exactly form a 45-degree line. This is due to the overrepresentation of certain contribution levels, namely 5, 10 and 15, which skew the average slightly.

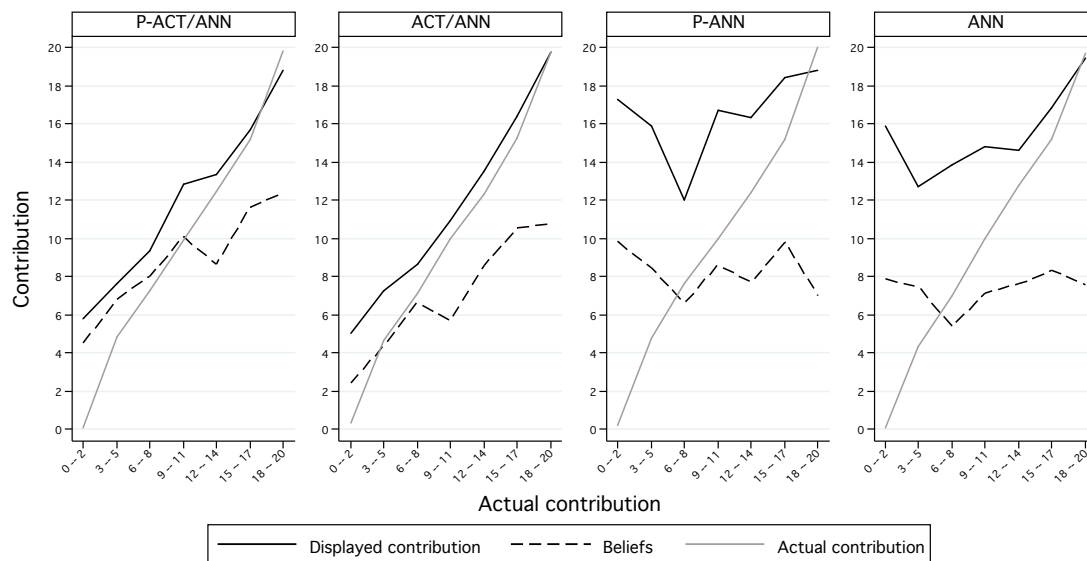
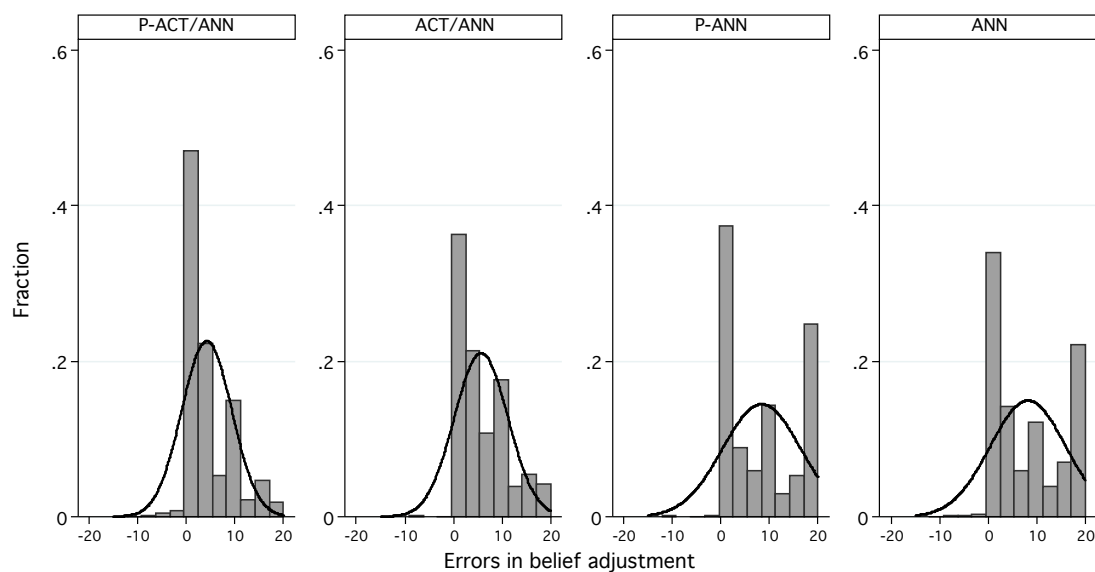


FIGURE 1.3: Errors in belief adjustment across treatments.



To evaluate perceptions of honesty, we compare subjects' beliefs about the actual contribution of each of the group members. Figure 1.3 displays the distribution of belief errors, i.e., the difference between the subject's belief and the actual contribution underlying the group member's announcement. The histograms show that while subjects are accurate in their beliefs approximately 35 - 45 percent of the time across treatments, the majority of belief errors are different from zero. This is significant for all treatments according to a signed rank test (WSR, $p < 0.01$). Furthermore, the figure shows that belief errors are on average positive, which indicates that subjects adjust their beliefs downward for a given announcement. Both these findings are in line with hypothesis 2b.

In terms of treatment differences, we hypothesized that the reported contributions in ACT/ANN would be more credible than in ANN (hypothesis 3). Since actual contribution feedback is sometimes displayed in the ACT/ANN treatments, subjects have more certainty that what they are observing as feedback is correct. The discrepancy between displayed contributions and what is believed is indeed smaller in treatments P-ACT/ANN and ACT/ANN (an average of 2.31 and 2.96 points, respectively) than in P-ANN and ANN (on average 8.40 and 8.20 points, respectively), which is significant in all pairwise comparisons (MWU, $p < 0.01$). The solid and dashed black lines in figure 1.2 show this difference graphically. For the P-ACT/ANN and ACT/ANN treatments, the difference between the displayed contributions and underlying beliefs is smaller than for the P-ANN and ANN treatments. It is important to note that in both treatments, subjects lower their beliefs considerably when receiving announcements from group members who are contributing at the social optimum (right hand size on the x-axis in figure 1.2). For the highest contributions, average beliefs level off at around 12 points in P-ACT/ANN, 11 points in ACT/ANN and between 7 and 8 points in the ANN treatments. In other words, someone who contributes 20 and announces this honestly is perceived by fellow group members to be contributing, on average, between 7 and 12 points depending on the treatment.

To evaluate whether this matters for contribution behaviour, we run a Tobit regression (see table 1.3 for results). We find support for hypothesis 4: in the ACT/ANN treatments, subjects condition their contribution more strongly on received announcements than in either of the ANN treatments. The coefficient for the average displayed reports in the previous period is significant with a positive sign, indicating that subsequent contributions increase for increases in average reports. The interaction terms between this variable and the treatment dummies are significant and negative for both the P-ANN and the ANN treatments, implying that reports are less important in the contribution decisions of the subjects in the treatments where only announcements are displayed.

TABLE 1.3: Tobit regression: the effect of reports on subject's contribution decision

<i>Dependent variable: Contribution</i>	
Period number	-0.271 *** [.075]
P-ANN	4.149 [4.266]
ACT/ANN	-8.795 * [4.525]
ANN	9.857 ** [4.431]
Av. displayed contribution (<i>t-1</i>)	1.449 *** [.282]
Av. displayed contribution (<i>t-1</i>) * P-ANN	-1.049 *** [.356]
Av. displayed contribution (<i>t-1</i>) * ACT/ANN	.491 [.507]
Av. displayed contribution (<i>t-1</i>) * ANN	-1.609 *** [.382]
Constant	-4.638 * [2.680]
N	3080
R^2	.061
Left-censored	1416
Right-censored	447

Robust standard errors are clustered at the group level and indicated in square brackets. The *, ** and *** indicate significant effects at the 10%, 5% and 1% level, respectively. Variable description: *Period number* : the period number; *P-ANN*, *ACT/ANN*; *ANN* : dummy for the respective treatment. Note that P-ACT/ANN is the baseline condition here; *Average displayed contribution (t-1)* : *The average displayed contribution the subject sees on the feedback screen in period t-1*; *Average displayed contribution (t-1) * P-ANN*; *ACT/ANN*; *ANN* : *The interaction term between the treatment and average displayed contribution in period t-1*;

1.5.3 The role of punishment

We now turn to the effects of the punishment mechanism. We hypothesized that erroneous beliefs in the ACT/ANN and ANN treatments could result in several inefficiencies relative to the standard public good game. First, compared to the P-ACT treatment, it is more likely that (i) a subject punishes a group member for a perceived low contribution when this person's actual contribution is in fact higher, and (ii) a subject punishes a group member less for a perceived high contribution when this person's actual contribution was in fact lower. Conditional on reactions to such punishment, this implies that the punishment mechanism to be less effective than in the standard public good game.

To evaluate this, we are interested in how punishment is assigned as well as how a subject's contribution responds to receiving punishment. To facilitate the presentation of our results, we restrict our attention to the extremes of the contribution spectrum: the low contributors, who contribute between 0 and 5 points to the public good; and the high contributors, who provide 15 to 20 points of their endowment. We confirm our findings with various regressions using the whole sample.

1.5.3.1 Punishment assigned

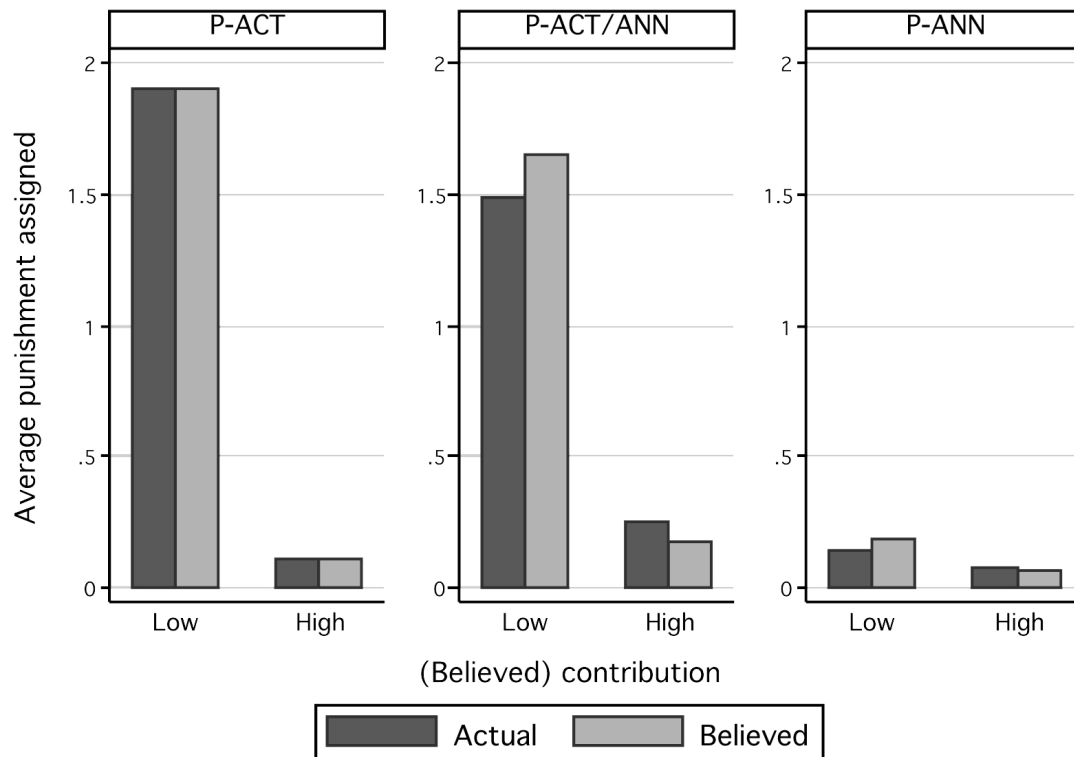
Figure 1.4 shows the difference between punishment assigned for actual contributions (dark bars) and for what the *believed* contribution of the target is (light bars). Since there is no discrepancy between actual and believed contributions in the standard public good game, these two bars are identical in P-ACT. In line with our hypothesis for the announcement treatments, subjects are assigning more punishment for contributions they believe are low compared to punishment assigned for contributions that are low in actuality. Reversely, subjects assign less punishment to contributions they perceive to be high compared to punishment for actual high contributions. In P-ACT/ANN, this difference between punishment assigned for actual and believed contributions is significant for high contributions (WSR, $p = 0.029$). In P-ANN these patterns are directionally true, but not significant at conventional levels.

These results refine insights from previous work on [Ambrus and Greiner \(2012\)](#) and [Grechenig et al. \(2010\)](#), who find more anti-social punishment in public good games where contribution feedback is noisy. However, for punishment to qualify as anti-social it is necessary that the punisher intends to punish such high contributions. While we corroborate their finding that anti-social punishment is higher in P-ACT/ANN than in P-ACT (MWU, $p = 0.015$), our results show that this difference falls away when we compare punishment according to the subject's beliefs (MWU, $p = 0.386$).

Further, it is important to note that the overall punishment pattern appears different in P-ANN than in the other two punishment treatments. Typically, low contributions are punished more frequently and severely than contributions closer to the social optimum ([Herrmann et al., 2008](#)). We find this pattern in P-ACT and P-ACT/ANN, where significantly more punishment is assigned to low compared to high contributors (WSR, P-ACT, $p = 0.01$; P-ACT/ANN, $p < 0.01$). However, in P-ANN, the difference between punishment for low and high contributions is minimal. Furthermore, average punishment assigned for low contributions in P-ANN is significantly below that assigned in the other two treatments (MWU, both $p < 0.01$).

FIGURE 1.4: Punishment assigned for (perceived) low and high contributions across treatments

The dark bars indicate the average punishment points assigned when the target's actual contribution lies between 0 and 5 for low contributors and between 15 to 20 points for contributions classified as high. The treatments where announcements are (sometimes) displayed, P-ACT/ANN and P-ANN, have an additional light bar that is constructed using the subject's beliefs about what the target is contributing. For low contributions, the light bar reflects how much punishment is assigned when the target's contribution is *believed* to lie between 0 and 5 points.



We confirm these insights from the non-parametric tests with several Tobit regressions, the results of which are presented in table 1.4. The variable *deviation* captures how much the target deviates from the social optimum of 20 points. In model 1 we use the target's actual deviation, whereas model 2 uses subjects' beliefs. Thus, the deviation variable in model 2 reflects how much the target is believed to be deviating from the social optimum. In model 1, deviation is strongly significant with a positive sign in the P-ACT treatment, indicating that the target receives more punishment the further removed their contribution is from the social optimum of 20 points. In line with the non-parametric results, the interaction term between the P-ANN treatment dummy and deviation is significant with a negative sign in both models, indicating that the target's deviation plays a much less prominent role in determining punishment when only announcements can be observed. For P-ACT/ANN, the interaction term is negative and weakly significant, suggesting that stronger deviations from the social optimum receive lower punishment compared to the standard public good game. This coefficient

becomes insignificant when we move to model 2, in line with the non-parametric result that subjects believe they are punishing low contributors more and high contributors less than they actually are. Importantly, we do not find that more average punishment is assigned in P-ACT/ANN compared to the standard public good game when controlling for deviation from the social optimum.

TABLE 1.4: Tobit regression: the effect of actual and perceived deviations from the social optimum on punishment assigned

<i>Dependent variable: Punishment assigned</i>		
	Model 1: Actual deviation	Model 2: Perceived deviation
Period number	-.062 [.041]	-.048 [.041]
Deviation	.391*** [.104]	.387*** [.102]
P-ACT/ANN	2.106 [1.285]	.877 [1.460]
P-ANN	.516 [1.114]	-.435 [1.034]
Deviation * P-ACT/ANN	-.211* [.113]	-.112 [.123]
Deviation * P-ANN	-.347*** [.109]	-.264** [.113]
Constant	-6.515*** [1.032]	-6.464*** [1.008]
N	7380	7380
R^2	.075	.091
Left-censored	6440	6440
Right-censored	23	23

Robust standard errors are clustered at the group level and indicated in square brackets. The *, ** and *** indicate significant effects at the 10%, 5% and 1% level, respectively. Variable description: *Period number* : the period number; *Deviation* : how much the target's contribution deviates from the social optimum of 20 points. Note how this variable is constructed using the subject's beliefs in model 2; *P-ACT/ANN*; *P-ANN* : dummy for the respective treatment; *Deviation * P-ACT/ANN*; *P-ANN* : interaction term between the treatment dummy and the target's deviation from the social optimum.

We also run several regressions to check for a mark-up in punishment for dishonesty. Using a Tobit regression (see appendix A.1 for results), we include variables to capture the discrepancy between the target's displayed contribution and how this is perceived by the subject (ie. the lie the target is perceived to be telling). In addition, we include a dummy with value 1 when the target is believed to be lying and 0 otherwise. Neither variable is significant in the P-ACT/ANN or P-ANN treatments.

1.5.3.2 Reactions to punishment

TABLE 1.5: Tobit regression: the effect of received punishment on contribution

<i>Dependent variable: Contribution</i>				
	Low contributors		High contributors	
	Model 1a	Model 1b	Model 2a	Model 2b
Period number	-.251*	-2.55*	.169	.162
	[.145]	[.130]	[.271]	[.272]
Pun. received ($t - 1$)	.649**	2.125***	-2.270***	-2.459***
	[.273]	[.427]	[.468]	[.538]
P-ACT/ANN	-4.748	1.976	-13.739***	-14.558***
	[4.011]	[2.369]	[4.411]	[4.919]
P-ANN	-7.322*	-1.180	-18.763***	-18.833***
	[4.267]	[2.471]	[4.755]	[4.857]
Pun. received ($t - 1$)		-1.858***		1.088
* P-ACT/ANN		[.451]		[1.627]
Pun. received ($t - 1$)		-2.721***		.134
* P-ANN		[.515]		[3.083]
Constant	2.546	-2.551	44.545***	44.648***
	[3.836]	[2.448]	[5.657]	[5.684]
N	581	581	1479	1479
R^2	.036	.053	.068	.068
Left-censored	392	392	30	30
Right-censored	0	0	1296	1296

Robust standard errors are clustered at the group level and indicated in square brackets. The *, ** and *** indicate significant effects at the 10%, 5% and 1% level, respectively. Variable description: *Period number* : the period number; *Punishment received ($t-1$)* : amount of punishment received by all group members in the previous period; *P-ACT/ANN* : dummy that takes the value 1 when the treatment is P-ACT/ANN and 0 otherwise; *P-ANN* : dummy that takes the value 1 when the treatment is P-ANN and 0 otherwise.

Table 1.5 presents the results of various Tobit regressions to evaluate the effect of punishment received across treatments. The regressions in model 1 are restricted to those classified as a low contributor in the preceding period (contributing 5 points or less), whereas model 2 focuses on high contributors who contributed 15 points or more in the previous round. We observe different effects of punishment across the treatments. For those classified as low contributors, punishment received in the last round has a positive effect on subsequent contribution when feedback is accurate (P-ACT). In P-ACT/ANN, where announcements are provided as feedback some of the time, this disciplining effect is lower, albeit still positive. When only announcements are displayed (P-ANN), the disciplining effect of punishment disappears and punishment in the previous round seems to have a negative effect on the subsequent contribution of low contributors. For

high contributions (model 2), the coefficient of punishment received in the last round is strongly significant with a negative sign. The interaction terms between this variable and the treatment dummies are not significant, indicating that high contributors who receive punishment reduce their subsequent contributions across all treatments. In other words, while the reaction to punishment of the high contributors is negative, it is not more negative in the treatments with announcements.

These results on punishment assigned and reactions to punishment largely support hypotheses 6 and 7 in that punishment is less effective in the treatments with announcements compared to the standard public good game and that punishment in P-ACT/ANN is more effective than in P-ANN. However, the reasons for the lack of effectiveness of punishment in P-ANN appears different from what we outlined under hypothesis 7. We discuss this and other results in the next section.

1.6 Discussion

In terms of overall contributions, we find that cooperation breaks down in all treatments except the standard public good game with punishment (P-ACT) and when actual contribution feedback is provided some of the time and punishment is available (P-ACT/ANN). In the latter treatment, punishment holds contributions at intermediate levels, although it is inefficient in terms of earnings. In line with our hypotheses, we find support for various constraints to full cooperation in the treatments with announcements compared to the standard public good game. First, subjects adjust their beliefs downward for a given reported announcement. While subjects are generally right to do so given the level of lying in the experiment, this also makes it more difficult for high contributors to signal their contribution to the others in the group. We indeed find that subjects systematically adjust their beliefs downward even for announcements from group members who are contributing close to the social optimum. For subjects classified as conditional cooperators, this discrepancy between actual contributions and beliefs implies that their contributions will be lower than in the standard public good game. When costly peer-punishment is introduced, we find that more punishment is assigned to high contributors in the P-ACT/ANN treatment than subjects believe they are assigning to high contributors. A regression on the reactions to punishment showed that high contributors react adversely to receiving such punishment. Further, there is a smaller disciplining effect of punishment of low contributors, who increase their subsequent contribution by less than in the standard public good game.

These inefficiencies were found to be less severe in the ACT/ANN treatments compared to when only announcements are displayed. First, we see fewer errors in the belief

adjustment of subjects in the ACT/ANN treatments compared to those in P-ANN and ANN. Furthermore, subjects condition their contributions more strongly on reports when actual contribution feedback is displayed some of the time. However, in the absence of punishment this increase in credibility does not raise contributions in ACT/ANN relative to the ANN treatment. A possible reason is that announcements themselves are less inflated in the ACT/ANN treatment compared to when only announcements are displayed.

For punishment we find support for hypothesis 7 that punishment is more effective in P-ACT/ANN than in P-ANN. However, it appears that the mechanisms behind this effect are slightly different between the two treatments. Rather than seeing more misdirected punishment in P-ANN compared to P-ACT/ANN, we find an overall decrease in punishment levels relative to the P-ACT and P-ACT/ANN treatments. In addition, punishment does not discipline low contributors. A possible reason for the lack of effectiveness of punishment in the P-ANN treatment is that subjects can hide entirely behind their announcements. Even if a low contributor is punished, she can adjust her announcement instead of her contribution. If subjects anticipate the limited role of punishment in raising actual contributions, it is possible that they decide not to assign it in the first place.

Finally, we find that a large amount of subjects overstate their actual contributions, but that overstatements are larger when only announcements are observed. This is somewhat surprising, since the form of monitoring employed in the treatment is very weak. Even though actual contributions are displayed some of the time, the displayed feedback cannot be tied to individual subjects in the experiment and we imposed no monetary penalties for lying. While our hypotheses were specific on the effect of credibility, it is also possible that this weak form of monitoring in the ACT/ANN treatments affected the cost of lying or guilt aversion sensitivity parameter of the subjects. This insight compliments work from the lying literature showing that the manner in which people communicate affects their tendency to lie (Charness and Dufwenberg, 2010; ?; Brosig et al., 2003b). A fruitful avenue for future research would be to analyze the effect of different forms of monitoring and communication vehicles on lying behaviour and subsequent public good contributions.

1.7 Conclusion

We study a public good setting in which accurate contribution feedback is not available, but group members can send non-verifiable cheap talk messages about their contributions. It extends past work in the public good literature that relaxes the key assumption

of accurate contribution feedback and in addition allows for communication between group members. By studying this setting in a controlled laboratory environment, we can explore both information transmission and reception as well as the effectiveness of costly peer punishment.

When actual contribution feedback is given some of the time, punishment appears to be moderately effective in terms of contributions but inefficient in terms of earnings. The constraints on full cooperation in public good games with lying are that subjects systematically adjust their beliefs downward for given reports, high contributors are more likely to receive punishment and there is a decreased disciplining effect of punishment on contributors. These results on punishment are conditional on actual contribution information being provided some of the time. When only announcements are observed, punishment does not discipline low contributors and is less severe than what is assigned in the standard public good game. These findings show that the established solution mechanism of costly peer punishment is less effective in a public good setting without accurate contribution feedback and communication.

Chapter 2

Fooling the Nice Guys: Explaining receiver credulity in a public good game with lying and punishment

Joint work with Bernd Irlenbusch, University of Cologne

Abstract

We demonstrate that receiver credulity can be understood through a false consensus effect: the likelihood with which individuals believe messages about the behaviour of others can be explained by their own behavioural tendencies in a comparable situation. In a laboratory experiment, subjects play a public good game with punishment in which feedback on actual contributions is obscured. Instead, subjects communicate what they have contributed through a post-hoc announcement mechanism. Using subjects' social value orientation as a proxy for their contribution tendency, we show that those high on the measure have inflated beliefs about the contribution of others. This, in turn, impacts their contribution and punishment decisions.

2.1 Introduction

Deception can be described as intentionally causing another person to believe what is false (Oxford English Dictionary, 2006). It thus involves two parties: the person doing the deceiving ('the sender') and the target of the deception ('the receiver'). Attempts at deception are largely successful because receivers tend to believe the message of the sender more often than they should. They are, in other words, overly credulous. Overcredulity appears a systematic and robust phenomenon across a wide range of settings, such as the sender-receiver game (Gneezy, 2005; Sutter, 2009; Wang et al., 2010; Erat, 2013; Besancenot et al., 2013), trust game (Charness and Dufwenberg, 2006) and prisoner's dilemma (Serra-Garcia et al., 2013). Furthermore, credulity persists even under repeated play (Blume et al., 2001) and role reversal (Sheremeta and Shields, 2013). Despite this evidence, the exact drivers of receiver credulity seem to be poorly understood.

In this paper we argue that receiver credulity (i.e. believing the messages of others) can, in part, be explained by the individual's own behavioural tendencies in a comparable situation. Under this so-called false consensus effect (Ross et al., 1977), individuals project their own behaviour, which they deem common and appropriate, onto the behaviour of others¹. The public good game offers an appropriate setting to evaluate this claim, since players decide on their own contribution as well as perceive the contribution decision of others. In addition, we can assess how subjects' beliefs about the messages of others influences subsequent decisions. Imagine an employee who needs to decide how many hours to invest in a group project. The false consensus effect suggests that someone who has a tendency to work hard herself is more likely to have the prior belief that others in the group will do likewise. If co-workers communicate to her that they are indeed putting in significant effort, we hypothesize that she is more likely to believe these messages compared to someone who is less inclined to work hard².

We study receiver credulity in a repeated public good game with lying and punishment. In the experiment, subjects do not receive accurate contribution feedback, but instead communicate their contribution to the others in the group through an announcement

¹For a review of the false consensus effect in social psychology, see Mullen et al. (1985). For applications in economic settings, see Madarász (2012).

²This does not imply that other considerations are unimportant. Past work shows that receivers' beliefs, on average, respond to structural factors that significantly affect the underlying deception rate. In the setting of a trust game, receivers correctly anticipate that promises made under free format communication have a stronger impact on behaviour than predetermined messages (Charness and Dufwenberg, 2010). As such, far fewer receivers act according to the sender's message when it has a pre-specified structure. Sutter (2009) compares the sender-receiver game at the individual and team level and finds that receivers are rightfully more skeptical of messages sent in a team environment. Finally, the work of Belot et al. (2012) finds that experimental subjects largely pick up on the appropriate cues when judging the trustworthiness of participants in a TV game show.

mechanism. These announcements are cheap talk and subjects can lie by announcing a lower or higher number than what they actually contributed. By eliciting subjects' beliefs about these announcements, we can assess the degree to which individuals are skeptical about the messages they receive and how this influences subsequent decisions. To obtain an independent proxy of an individual's contribution tendency, we measure subjects' Social Value Orientation (SVO) after the public good game. Higher scores on the measure reflect stronger other-regarding preferences, which, in turn, are correlated with higher contributions to the public good. This measure has been used in a wide range of public good experiments, most notably [Offerman et al. \(1996\)](#); [Sonnemans et al. \(1998\)](#); [van Dijk et al. \(2002\)](#), as well as other social dilemmas (see [Balliet et al. \(2009\)](#) and [Van Lange et al. \(2007\)](#) for reviews). According to the subject's SVO angle, we classify them as 'high' or 'low' types. If the false consensus effect predicts receiver credulity, we should observe that individuals who are likely to contribute to the public good ('high' types) will perceive such high contributions from others in the group. Reversely, individuals who are not likely to contribute to the public good will perceive low contributions from their fellow group members.

Our experimental evidence supports the false consensus effect. We find that individuals with a high tendency to contribute to the public good ('high types') believe announcements of others to be largely accurate, which results in inflated beliefs. These, in turn, impact their contribution and punishment decisions, resulting in significantly lower earnings compared to those low on the SVO measure. While low types adjust their beliefs more strongly than high types, they appear well calibrated in their beliefs about actual contributions to the public good. As such, we do not find conclusive evidence of a false consensus effect for low types.

The rest of the paper is structured as follows. Section 2 discusses the experimental design. Section 3 covers the analysis, focusing on belief formation and subsequent contribution and punishment decisions. Section 4 concludes.

2.2 Method

Subjects play a 4-player repeated public good game with punishment. The game consists of 15 periods and subjects stay in the same groups for the duration of the game (partner matching). There are two treatments: STANDARD and ANNOUNCE. We describe the STANDARD treatment first. At the start of each round every subject is endowed with 20 points, which can either be kept for oneself or allocated to a group project. Each point invested in the project is multiplied by 1.6 and split over all group members, irrespective of contribution. After the investment decision subjects enter a feedback

stage where they learn about the individual contributions of the others in the group. These are displayed in random order as to prevent subjects from tracking individual behaviour across periods. Furthermore, subjects also have the possibility to assign punishment, for which they receive 10 additional points per period. Each point invested in punishment reduces the earnings of the targeted subject by three points. Any points not used for punishment are added to the subject's individual earnings, thus making punishment costly to administer. Each subject is consequently informed about how many punishment points they received (if any) before starting the next period.

The payoff Π_i for each subject i in each period can be expressed as follows:

$$\Pi_i = (20 - c_i) + (0.4 \sum_{k=1}^4 c_k) - (3 \sum_{k \neq i} p_k^i) + (10 - \sum_{k \neq i} p_i^k)$$

where c_i represents the contribution of subject i to the group project and p_j^i indicates how much punishment subject i receives from subject $j \neq i$, $i, j \in \{1, \dots, 4\}$. Subscripts refer to the decision makers, whereas superscripts, when applicable, indicate to whom the action is directed.

After the public good game, a second part commences in which subjects complete an adapted version of [Liebrand \(1984\)](#) to measure their Social Value Orientation. In this separate task, subjects are presented with 32 binary allocation decisions where they divide points between themselves and a randomly selected other participant. Each of the 32 preferred allocations can be considered as a vector, where the sum describes an angle with the horizontal axis reflecting how much the individual cares about the payoffs of other person. After completing these two parts, subjects provide demographics and general comments through a questionnaire. They are then paid in private and dismissed.

In addition to the STANDARD treatment described above, we evaluate receiver credulity in the treatment ANNOUNCE. Immediately after the actual investment decision, each subject reports how many points they contributed to the project through an announcement. Subjects are free to report any number from the strategy space and thus have the possibility to lie by reporting a lower or higher number than what they actually contributed. Whether and to what degree such a discrepancy exists is entirely up to the individual subject. Importantly, these announcements are cheap talk: only the actual contribution is payoff-relevant for all players in the group. In the feedback stage, subjects only receive information about the *announced* contributions of the other group members. Since accurate feedback is not provided, we also elicit subjects' beliefs about

the actual contributions underlying the received announcements³. Subjects are only informed about aggregate actual contributions and personal earnings at the end of the experiment. All of the above is common knowledge to the subjects. In particular, it was made clear that the information received in the feedback stage reflects the announcements of the others in the group.

The two treatments were conducted over four sessions (two per treatment) with a total of 112 undergraduate and graduate students from the University of Cologne, Germany. This yields 14 independent observations per treatment. No subject participated in any of the sessions more than once. The mean age of participants was 23.3 years, with 52.7 percent female. The vast majority of participants were German nationals from a range of academic disciplines, including economics and business. Subjects were recruited using ORSEE (Greiner, 2004) and the experiment was programmed with the z-Tree software (Fischbacher, 2007). Upon entering the lab, subjects were seated at individual and visually separated computers. Before starting the experiment, written instructions were distributed and each subject had to complete a set of control questions to ensure understanding of the experimental procedure. At the end of the experiment the total sum of points was converted to Euros at an exchange rate of 50 points to €1. Each session lasted approximately 90 minutes and participants were paid, on average, €11 for the public good game and €3 for the SVO elicitation.

2.3 Results

2.3.1 SVO classification and general patterns

In ANNOUNCE, the SVO angle has a Spearman rank correlation coefficient of 0.21 with contribution decisions, which is strongly significant ($p < 0.01$)⁴. No subject was excluded according to the inconsistency requirement of Liebrand and McClintock (1988). To facilitate the presentation of our results, we classify subjects as either low or high types according to their SVO degree angle, taking the 25 percent subjects with the lowest and

³Given that we are also interested in subjects' contribution decisions, it was decided not to provide incentives for accurate beliefs. This decision is based on work showing that incentivized belief elicitation in repeated public good games can decrease (Croson, 2000) or increase (Gächter and Renner, 2010) contributions relative to a non-incentivized control treatment. In addition, Gächter and Renner (2010) find that the gain in accuracy from incentivized elicitation is small.

⁴We used the first-round contribution and the subject's SVO degree angle to ensure independence of observations. In STANDARD, the spearman correlation coefficient is 0.212 ($p < 0.01$). The mean SVO degree angle is 8.74 ($sd = 13.37$) and 14.76 ($sd = 18.29$) for the ANNOUNCE and STANDARD treatment, respectively.

highest degree angles respectively⁵. In both treatments, a large number of subjects have a slope of 0, indicating that they are completely individualistic. This makes it difficult to create a clear cutoff at 25 percent of the subjects with the lowest score. For this reason we include all subjects with a SVO degree angle of 0 and below as ‘low’ types. The high group in ANNOUNCE (STANDARD) thus comprises of 14 (14) subjects, whereas the low group consists of 27 (19). This corresponds to 14 and 8 independent observations in the ANNOUNCE treatment for low and high types, respectively. For the STANDARD treatment these numbers are 13 and 10. All reported non-parametric tests are two-tailed and respect the independence assumptions by using averages from a group of 4 players that interacted as one independent observation. For comparisons between types, we use only the independent observations from those groups in which both types are present.

The general descriptive statistics in table 2.1 reveal that the two types in ANNOUNCE differ along several dimensions. Wilcoxon signed-rank (WSR) tests confirm that high types have significantly higher contributions ($p < .012$) and earn less in the public good game ($p < .012$) compared to those classified as low types. Average announcements are not significantly different between types ($p = .528$). In the STANDARD treatment, differences in contributions between types are not significant.

Comparing announcements and actual contributions between types, it appears that high types tell smaller lies than low types. While this is significant (WSR, $p = .025$), it is possible that lying for the high types is limited by the experimental design, since announcements are capped at 20. As an alternative measure, we compare the difference between the subject’s announced and actual contribution as a percentage of how much the subject can overstate. For example, consider two individuals with an actual contribution of 5 and 15, respectively, who both overstate this contribution by three points. As a percentage, the first subject overstates by $3/(20 - 5) = 20\%$ compared to $3/(20 - 15) = 60\%$ for the second subject. Using this measure, high types overstate, on average, by 74.13% compared to 81.03% for the low types. This difference is not significant (MWU, $p = .345$). Similarly, using a measure comparing the percentage of honest announcements between types, excluding those subjects who contributed 20, we find that low types tell the truth 11.29% of the time, compared to 5.84% for high types. As such, we cannot exclude a ceiling effect in explaining the difference in lying between types.

⁵Our main results hold using an alternative classification according to the cooperative and individualistic types (Liebrand and McClintock, 1988). We also ran a robustness check using the SVO degree angle, rather than the type classification, as an independent variable in the belief formation regression. While this makes interpretation more difficult, our main results hold: announcements have a positive effect on beliefs for those with high SVO degree angles and a negative effect for subjects with degree angles of 0 or below. These results can be found in appendix sections B.2 and B.4.

TABLE 2.1: General descriptive statistics

	Average contribution	Average announced	Average adjustment	Average punishment	Average earnings (€)
Announce (N=56)					
<i>Overall</i>	6.79 (7.59)	17.14 (4.17)	8.40 (8.03)	0.14 (0.64)	9.71 (1.89)
<i>Low types</i> (n=27)	4.56 (6.62)	16.84 (4.40)	10.44 (8.05)	0.09 (0.50)	10.14 (1.80)
	<**	<	>**	<	>**
<i>High types</i> (n=14)	11.55 (7.36)	18.11 (7.21)	4.29 (3.73)	0.10 (0.45)	8.95 (1.63)
Standard (N=56)					
<i>Overall</i>	16.43 (6.08)			0.26 (1.16)	11.00 (1.76)
<i>Low types</i> (n=19)	15.85 (6.50)			0.29 (1.05)	10.88 (1.77)
	<			>	<
<i>High types</i> (n=14)	16.67 (6.22)			0.19 (0.95)	11.42 (0.93)

Standard deviations are shown in brackets. We use two-tailed Wilcoxon signed-rank tests for comparisons between types. The ** indicate significant effects at the 5% level.

2.3.2 The effect of announcements on beliefs

Using subjects' beliefs, we can estimate the degree to which subjects are skeptical about the messages they receive. For example, if a subject receives an announcement of 16 and reports 10 as her belief about the underlying contribution, she adjusts her belief, conditional on the message, by 6. In line with previous work, we find that subjects are too credulous on average. Actual contributions are overstated by an average of 10.35 points, while subjects adjust their beliefs by 8.40 points on average. However, belief adjustments differ significantly between types: low types adjust their beliefs downward by an average of 10.44 points, while high types adjust by 4.29 points. This difference is significant (WSR, $p = .025$)⁶.

⁶An important assumption underlying this conclusion is that each type is exposed to announcements that are similar in both their level and their underlying accuracy. For example, if announcements observed by high types are more truthful than those observed by low types, then high types rightfully adjust their beliefs by less. Similarly, if low types observe a higher absolute level of announcements, then their belief adjustment should be higher. Non-parametric tests show that observed announcements do not differ significantly between high and low types (WSR, $p = .528$). The difference in the rate with which observed announcements are truthful is weakly significant (WSR, $p = .079$). However, high

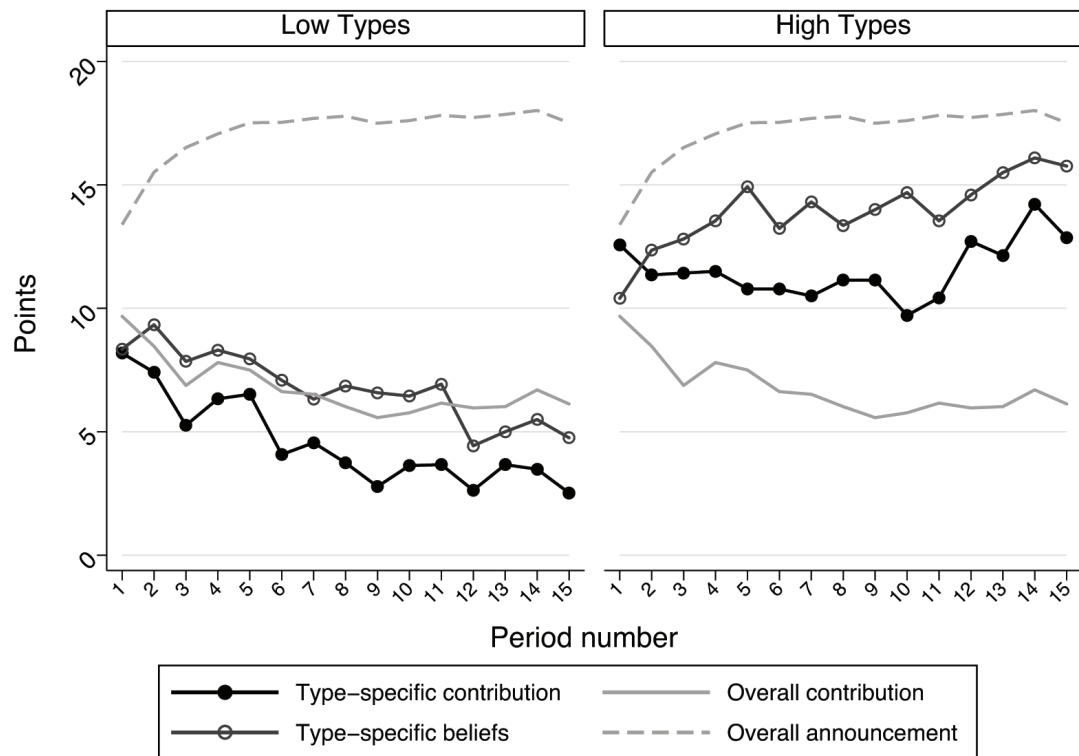


FIGURE 2.1: Average contributions, announcements and beliefs across periods in ANNOUNCE

Figure 2.1 depicts this difference in average credulity graphically. It maps both average beliefs and average contributions over the fifteen periods of the game for low and high types respectively. As a benchmark, the solid gray line represents the average actual contribution pattern of all subjects in ANNOUNCE, whereas the dashed line reflects overall average announcements. It is apparent from figure 2.1 that the beliefs of high types are not an accurate reflection of actual contributions. Indeed, beliefs for the high types are significantly above overall actual contributions (WSR, $p = .036$), which is in line with the prediction of the false consensus effect. However, we do not find an inverse relationship for the low types in that they are too skeptical about the actual contributions in their group. As the difference between actual contributions and the beliefs of low types is not significant (WSR, $p = .272$), it appears that they are, on average, well calibrated.

We support our analysis with several censored Tobit regressions examining the effect of announcements on belief formation. The results are presented in Table 2.2. Model 1 in Table 2.2 shows that average beliefs in the last period and subject type have a significant impact on beliefs. By contrast, the coefficient for average announcement is not

types observe announcements that are, on average, *less* truthful than those observed by low types. This strengthens the observation that high types should adjust more, rather than less.

TABLE 2.2: Tobit regressions - belief formation in ANNOUNCE

<i>Dependent variable: beliefs</i>		
	Model 1	Model 2
Period number	-.040 [.050]	-.047 [.050]
Av. beliefs ($t-1$)	1.185 *** [.090]	1.159 *** [.088]
Av. announcement	-.083 [.094]	-.253 ** [.111]
High type	3.941 *** [.873]	-6.555 *** [2.265]
High type * Av. announcement		.592 *** [.140]
Constant	-1.214 [1.743]	1.715 [1.816]
Controls	YES	YES
N	574	574
Pseudo R^2 (overall)	.222	.226
N left-censored	151	151
N right-censored	107	107

Robust standard errors are clustered at the group level and indicated in square brackets. The ** and *** indicate significant effects at the 5% and 1% level respectively. Dependent variable: *Average Beliefs*: the average of the three belief measures (one for each announcement of the other group members) in period t . Independent variables: *Av. Beliefs ($t-1$)*: lagged measure of average beliefs; *Av. Announcement*: average of the three announcements received by the subject on the feedback screen in period t ; *High type*: binary variable, (1 = High type; 0 = Low type); *High type * Av. Announcement*: the interaction between the subject's type and the average announcement received; *Controls*: include age, gender and field of study. Gender is a binary variable where 0 indicates male and 1 female; Field of study is a binary variable where 1 is assigned to those subjects studying economics or business and 0 otherwise. None of these controls is significant.

significant, indicating that, in general, announcements of others do not influence beliefs. Model 2 includes the interaction term between average announcement and the subject's type, which is strongly significant with a positive sign. Thus, announcements positively affect the beliefs of high types about underlying actual contributions. The coefficient for average announcement in model 2 is significant with a negative sign, indicating that low types decrease their beliefs in response to higher average announcements.

2.3.3 The effect on contributions and punishment

We run various Tobit regressions to assess the role of beliefs for the contribution decision, the results of which are included in appendix B.3. Not surprisingly, the contribution and beliefs from the previous period are a strong predictor of the contribution in the current period. In addition, high types also contribute significantly more than low types, which is significant at the 1 percent level. The coefficient of the interaction term between type and average beliefs in the previous period is negative, suggesting that beliefs inform the contribution decision to a lesser degree for high compared to low types. However, this is not significant ($p = .233$). These regression results indicate that the main difference between high and low types manifests itself at the level of belief formation. Announcements have a positive effect on beliefs for high types while having a negative effect for low types. However, when it comes to the actual contribution decision, high and low types act on their beliefs in a similar way.

In addition to the contribution decision, our experimental design allows for the analysis of punishment behaviour. Figure 2.2 displays the punishment reaction function for each type across treatments. The bars indicate the average punishment points assigned for each level of perceived contribution, indicated in blocks of three. Thus, the block 9-11 captures the punishment points assigned for beliefs about underlying actual contributions at 9, 10 or 11 for a given announcement. Since announcements are absent in the STANDARD treatment, the x-axis in the right-hand panel reflects actual contributions. The figure shows that high types administer more punishment when contributions are (perceived to be) low. In ANNOUNCE, high types assign an average of 0.38 points when the contribution is believed to lie between 0 and 2, compared to 0.07 points for perceived contributions between 18 and 20. Importantly, this punishment pattern is similar in the STANDARD treatment where credulity is not an issue⁷. The key insight is that in ANNOUNCE there are only few instances in which high types perceive actual contributions to be low. Out of all beliefs about actual contributions that high types report, only 10.5 percent fall in the range of 0-2, while 52.1 percent are located at the other extreme of 18-20. By contrast, low types believe only 9 percent of the actual contributions to be located in this latter range. Thus, despite the apparent willingness of high types to punish low contributions, their biased beliefs substantially reduce punishment.

⁷It should be noted that the level of average punishment assigned is substantially higher in the STANDARD treatment than in ANNOUNCE, particularly for low contributions (1.22 points and 1.39 points for contributions between 0-2 and 3-5, respectively). It is possible that uncertainty about whether punishment is actually justified depresses average punishment in the latter treatment.

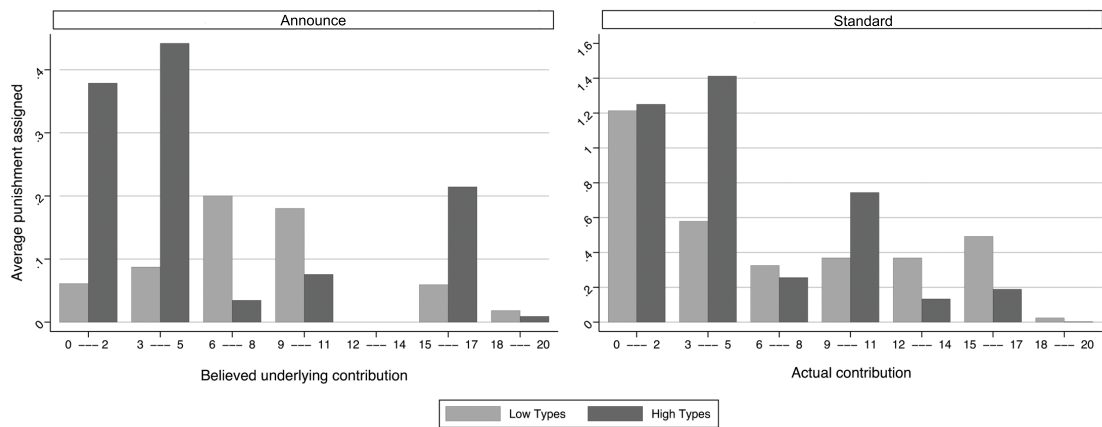


FIGURE 2.2: Punishment assigned by different types in the two treatments conditional on beliefs (ANNOUNCE) and actual contributions (STANDARD)

2.4 Conclusion

Our results show that in a symmetric setting of the public good game, an individual's own behavioural tendency is a useful lens to understand receiver credulity. In particular, we find that those high on Social Value Orientation believe announcements to be accurate to a larger extent than those low on the measure, resulting in inflated beliefs about actual contributions. This in turn influences their contribution and punishment decisions. The credulity of high types can thus be exploited by those subjects that contribute low and announce high. This is reflected in their respective earnings: low types earn significantly more than high types in ANNOUNCE, whereas no such difference exists in the STANDARD treatment. One feature of our design is that subjects do not receive any accurate information about actual contributions until the end of the experiment. It would be interesting to investigate whether the credulity of high types persists when accurate information becomes available⁸. For example, it might be possible for subjects to observe actual contributions at the aggregate level or receive accurate feedback with a certain probability. Future research can investigate whether high types only remain credulous for as long as they remain completely unaware of the true level of actual contributions.

⁸Work by [Gneezy et al. \(2013a\)](#) indicates that receivers who learn that they have been deceived become less credulous.

Chapter 3

The indirect effect of monetary incentives on deception

Abstract

This paper investigates whether working under competitive or cooperative incentives affects deception in a subsequent, unrelated task. I use a laboratory study with two stages. First, participants perform a real effort task under a piece rate, tournament or team incentive. Afterwards, they play a sender-receiver game in which the sender can gain financially at the expense of the receiver by sending a deceptive message. I find that senders who worked under the tournament incentive are less honest than those who worked under a piece rate. I find no increase in honesty for those who performed under team incentives relative to the piece rate. This only holds when participants are not informed about their relative performance during the work task. When such feedback is provided, I find that relative performance affects honesty across all incentive conditions. In particular, honesty decreases as relative performance differences become small.

3.1 Introduction

Pay-for-performance schemes, such as bonuses and tournament incentives, are an important means to induce effort of agents in the workplace. However, at the same time, there is an increasing concern that these incentive schemes motivate dishonest behaviour by linking monetary rewards to specific performance targets. Experimental evidence suggests that agents indeed respond to such incentives by lying about their work performance (Fischbacher and Föllmi-Heusi, 2013; Schweitzer et al., 2004; Cadsby et al., 2010; Conrads et al., 2014) and cheating on a task (Pascual-Ezama et al., 2013). This paper takes a broader perspective on the role of pay-for-performance incentives on dishonesty by focusing on a possible indirect effect. In particular, it considers whether the kind of work environment an agent is exposed to, as dictated by the incentive scheme, affects dishonest behaviour in a subsequent, unrelated task. The task is unrelated in the sense that any actions in the task have no bearing on the performance, and therefore pay level, of the prior work environment.

It has been argued that working under competitive incentives, such as tournament schemes, fosters an uncooperative mindset (Buser and Dreber, 2013) and a negative attitude towards others (Brandts et al., 2009). Other incentives, such as revenue-sharing schemes, have been found to foster social ties (van Dijk et al., 2002) and trust (Harbring, 2010). Under optimal mechanism design, the principal cares both about the direct and indirect effect of monetary incentives on behaviour. An indirect effect is particularly relevant in a work environment where a subject performs multiple tasks, but receives a pay-for-performance scheme on one of these activities (Holmstrom and Milgrom, 1991). Consider a salesperson who is incentivized to make sales, but also performs an auxiliary activity, such as writing a subjective review report on customer satisfaction that is not part of the pay-for-performance scheme. If monetary incentives in the main task influence dishonest behaviour in the auxiliary task, this can result in a loss of economic rents to the principal. In addition, the principal may care about fostering a general attitude of upholding ethical standards among her employees.

In a laboratory experiment, I examine the effect of cooperative and competitive incentives on subsequent dishonesty. Subjects are paired and perform a real-effort task under either a piece rate (baseline), a cooperative (team) or competitive (tournament) incentive. Further, as a robustness check to the effect of incentives, participants either receive or do not receive information about the work performance of their partner. Afterwards, the pair plays a sender-receiver game, where the sender is informed about a true state of the world which she then communicates to the receiver. By overstating the true state the sender secures a financial gain at the expense of the receiver, which I use as the measure of deception.

I find that compared to the piece rate condition, subjects exposed to the tournament incentive are less honest in the subsequent task. There is no increase in honesty for those working under team incentives compared to the piece rate. However, these incentive effects are not robust to relative performance feedback. When subjects are informed about the performance of their partner, honesty decreases as relative performance differences become small. This holds across all incentive conditions. These results suggest that even when no direct incentive for dishonesty is provided, the interaction in one's previous work environment can affect subsequent dishonest behaviour. In particular, tournament incentives as well as small relative performance differences can have a negative effect on subsequent honesty.

The rest of the paper is structured as follows. Section 2 discusses related literature and hypotheses. Section 3 outlines the experimental design and procedures, followed by the results in section 4. Section 5 concludes.

3.2 Literature review and hypotheses

Monetary incentives affect behaviour by changing the benefits and costs of a particular action. Yet, there is substantial evidence from behavioural economics that monetary incentives affect behaviour in more indirect ways, such as by altering social norms ([Gneezy and Rustichini, 2000a](#)), changing reputational concerns ([Ariely et al., 2009](#)), revealing unfavourable information about the principal ([Falk and Kosfeld, 2006](#)), reducing intrinsic motivation for the task ([Ryan and Deci, 2000](#)) or shifting the framework of the decision ([Tenbrunsel and Messick, 1999](#)). Through these mechanisms, studies have found behavioural responses contrary to what a canonical cost-benefit approach would predict when incentives are introduced in previously non-monetary contexts ([Gneezy and Rustichini, 2000b](#); [Heyman and Ariely, 2004](#)) or new monetary schemes come to replace others ([Burks et al., 2009](#); [Meier, 2007](#)).

A number of studies find a negative effect of competitive incentives, such as rank and tournament schemes, on subsequent cooperation. Work by [Harbring \(2010\)](#) shows that individuals who performed under a tournament incentive allocate less in a subsequent trust game compared to those who worked under a revenue-sharing scheme. [Brandts et al. \(2009\)](#) place subjects in a rivalrous task where two agents compete in order to be selected by the principal for a lucrative work task. They find that agents who competed hold a negative disposition towards one another as well as the principal. Finally, work by [Buser and Dreber \(2013\)](#) find that individuals working under a tournament incentive are less cooperative in a subsequent public good game compared to participants who performed under a piece-rate. While the exact mechanism driving these effects is

not immediately clear, [Buser and Dreber \(2013\)](#) find that their results hold even when workers play the subsequent public good game with someone they have not interacted with before. This results provides some evidence favouring the explanation that competitive work environments foster an uncooperative mindset that affects behaviour in subsequent, unrelated tasks.

Studies evaluating the effect of cooperative incentives, such as revenue-sharing schemes where employees share equally in the total output generated by the group, have largely found a positive effect on subsequent cooperation. [van Dijk et al. \(2002\)](#) show that individuals are more likely to share resources with others when they have previously interacted in a public good setting compared to a task where their payment depends only on individual effort. Work by [Pan and Houser \(2013\)](#) finds that individuals who were exposed to a work task requiring cooperation between group members showed more trusting behaviour in a subsequent trust game, irrespective of whether the trustee was a fellow group member or an outsider. However, the addition of a reward and punishment mechanism ([Falkinger et al., 2000](#)) or minimum binding contributions ([Reeson and Tisdell, 2008](#)) to public good settings reduce contributions in groups after these incentives are removed compared to a control group that played the public good game without these additional incentives.

Specific to dishonesty, [Gill et al. \(2013\)](#) present subjects with an opportunity to lie to obtain additional earnings after a work task with a lottery incentive. They find that subjects exposed to the lottery are less honest than those who received a fixed wage for their efforts. Also related is work by [Cappelen et al. \(2013b\)](#), who looked at the effect of a market prime on subsequent honesty. Subjects were asked to write about a recent experience where they bought or sold a good and then had the opportunity to behave dishonestly in a dice-rolling game. They find a slight, but non-significant increase in dishonesty for the market prime condition. My work differs from the design of [Gill et al. \(2013\)](#) and [Cappelen et al. \(2013b\)](#) in its focus on competitive and cooperative incentives. Furthermore, subjects are given actual monetary incentives rather than a priming task.

From the lying literature, it is not immediately straightforward that honesty is subject to such spillover effects. Previous work has shown that a majority of people experience some psychological disutility from being dishonest ([Fischbacher and Föllmi-Heusi, 2013](#)) and that this differs across individuals ([Gibson et al., 2013](#)). In addition, individuals are sensitive to the stake size in that they are more willing to lie when the monetary gain of doing so increases ([Gneezy, 2005](#); [Conrads et al., 2014](#)). In this literature, lying behaviour is understood by the individual's specific costs of lying as well as the particular incentives tied to the dishonest action. From this perspective, previous interaction

is irrelevant. I use this to formulate the null hypothesis:

Hypothesis 0. *Lying in the subsequent task does not differ across the incentive treatments.*

Alternatively, it is possible that the same mechanisms that affect subsequent cooperation translate to honesty as well. Following the abovementioned literature, if tournament incentives foster an attitude of uncooperativeness and recoil towards the work partner and decrease trust, we can expect that subsequent honesty is negatively affected after individuals have been exposed to such a work environment. Reversely, if team incentives foster social ties and trust, they may positively affect subsequent honesty.

Hypothesis 1. *Honesty in the subsequent task decreases for those who worked under a tournament incentive compared to a piece rate scheme.*

Hypothesis 2. *Honesty in the subsequent task increases for those who worked under a team incentive compared to a piece rate scheme.*

To check for the robustness of these incentive effects, I also include treatments in which individuals are informed about their own performance as well as that of their partner. Checking for the effect of relative performance was deemed important, because such information is often embedded in these monetary incentive schemes (Bandiera et al., 2007, 2013). At the same time, there is some evidence that relative performance feedback encourages comparisons between peers, which in turn can affect subsequent cooperation and effort provision (Larkin et al., 2012). Previous work suggests that those who outperform their partner may experience a sense of entitlement about their earnings (Gill et al., 2013) or are less concerned about the welfare of others (Cappelen et al., 2013a). Furthermore, work by Buser and Dreber (2013) finds that losing a competition as well as a lottery has negative effects for subsequent cooperation. This evidence informs the following null and alternative hypotheses on relative performance feedback.

Hypothesis 3a. *Feedback on partner's performance does not affect subsequent honesty in either incentive condition.*

Hypothesis 3b. *Feedback on partner's performance decreases subsequent honesty across all incentive conditions.*

3.3 Experimental design and procedures

The experiment consists of two stages: a real effort task and a sender-receiver game with deceptive messages. Both of these stages are played in pairs and subjects remain in the same pair throughout the entire experiment. Decisions in the experiment are anonymous to the other participants. Upon entering the lab, subjects are randomly assigned to a computer at an individually separated cabin. They are given instructions and a booklet with 15 pages of Latin text. During the first part of the experiment, the real effort task, participants are asked to find and identify specific letters in the text using directions on the computer screen. Each set of directions specifies the page, line, word and position where the letter is to be found. After entering the identified letter on the computer, directions for the next letter appear on the screen. Participants play five rounds of three minutes each, one of which is relevant for payment. Subjects are informed about their own performance at the end of each round, but only learn which of the five periods was selected for payment at the end of the experiment.

After the five rounds are finished, subjects move to stage two where they play a sender-receiver game. The sender receives information about the performance level, ranging from 10 to 25, of a randomly selected subject from another experiment who performed a similar work task. Upon learning about the actual performance level, the sender is asked to send a message to the receiver about this performance level. The receiver then chooses a number, ranging from 10 to 25, that determines payoffs for both players. Payoffs are such that if the receiver chooses the actual performance level, the payoff allocation is determined according to allocation X. If the number chosen by the receiver does not match the true state, payoffs for both sender and receiver are determined according to allocation Y. The receiver is informed about the basic structure of the game as well as that the message and action space ranges from 10 to 25, but does not know the actual performance level nor the exact details of the payoff allocation underlying X and Y.

After reading these general instructions for stage two, subjects are randomly assigned to the role of sender or receiver. Senders then receive private information about the actual performance level (12 in this experiment) and the exact payoff structure tied to the message space. Every sender receives the same performance level as private information and it is common knowledge that this performance level refers to that of a randomly selected participant in another experiment. Under payoff allocation X, implemented when the receiver chooses the actual performance level, both players receive 200 points. When the number chosen differs from the actual performance level, the exact payoffs under allocation Y depend on the message sent by the sender. Table 3.1 shows the message space with the exact payoffs for this allocation, with the true state 12 in italics. The table shows that if the sender sends the message that the performance level is 12, but

the receiver chooses a number different from 12, payoff allocation Y gives 200 points to both players for this particular message. If the sender decides to overstate the true state by sending, for example, message 17 and the receiver, again, chooses a number different from 12, then the payoffs under allocation Y are as specified under message 17, which gives 250 points to the sender and 150 to the receiver. Thus, even if a sender believes that the receiver will not follow her message, sending the honest message will still ensure a payoff of 200 points to both players. Prior to sending their message, senders answer several control questions to ensure their understanding of this payoff structure. The message is then transmitted to the receiver, who chooses a number between 10 and 25. Both players are then informed about their payoffs from both stages of the experiment, paid in private and dismissed.

TABLE 3.1: Payoff matrix Y, effective when the receiver does not choose the actual performance level

Message:	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Sender	200	200	200	210	220	230	240	250	260	270	280	290	300	300	300	300
Receiver	200	200	200	190	180	170	160	150	140	130	120	110	100	90	80	70

The second part of the experiment is a modified version of the setup of [Gneezy \(2005\)](#), where the sender receives private information and can reap some personal financial gain at the expense of the receiver by sending a message other than the true state. The main difference with the sender-receiver game in this paper is that the payoffs tied to the message space are such that the more the sender chooses to overstate the actual performance level, the more she gains at the expense of the receiver. Thus, this measure of dishonesty picks up not only whether senders are dishonest, but also the size of the lie they choose to tell. Besides an honest message, payoff allocation Y features both selfish black lies (message 13-22) and spiteful black lies (message 23-25). In case of the former, the sender's gain increases linearly by 10 currency points at the expense of the receiver for each unit increase of the message. However, from message 22 onwards, the sender's gain remains constant at 300 points, but the receiver continues to lose 10 points per unit increase of the message. These lies are spiteful, because they hurt the receiver without benefiting the sender ([Erat and Gneezy, 2012](#)). Given evidence that individuals in laboratory experiments are heterogeneous in their lying costs ([Gibson et al., 2013](#)), the design used in this paper allows for a more fine-grained measure to detect changes in lying behaviour beyond the binary outcome of honesty and dishonesty. In addition, the discrete message and action space, coupled with the payoff structure, makes the receiver's choice irrelevant from the sender's point of view. Even if the receiver decides not to follow the message of the sender, it is very unlikely that he will choose the actual

performance level. Moreover, since payoff allocation Y is implemented whenever the receiver chooses the wrong number, the sender's action effectively determines the payoffs for both parties. This addresses the potential concern that the sender chooses a different message because she believes the receiver anticipates her to be dishonest (Sutter, 2009).

I introduce two treatment variations in a 3x2 (incentive x feedback) between-subject design. The first varies the incentive scheme for the real effort task: a piece rate, a revenue-sharing scheme or a tournament incentive. Under the piece rate, both players in the pair receive individual payoffs of 30 points per correctly identified letter. In the team condition, subjects receive 15 points for each letter they or their partner identifies correctly. Subjects in the tournament compete for a prize of 1000 points that is awarded to the highest performer in that round; the loser receives 0 points. At the end of the second stage, one of the five rounds is randomly selected for payment. In the experiment every 100 points equals €1. The average performance of a pilot study was used to set the point allocation such that average earnings would not differ across incentive conditions. Indeed, according to two-tailed Mann-Whitney U-tests, neither average performance nor average earnings from the work task differ significantly across incentive treatments ($p > 0.10$).

For the second treatment variation, I run the three incentive conditions again, but now subjects receive information about the performance of their partner during the real effort task. Thus, in addition to knowing their own performance level, subjects are now also informed about the performance of their partner. The rest of the experiment is identical to the no-feedback treatment.

The sessions were conducted at the economics laboratory at the University of Cologne, Germany, in June and October of 2013. A total of 268 undergraduate and graduate level students from various disciplines participated in one of the six treatments. The median age of the participants was 23 and 66% were female. The recruitment procedure and experiment were entirely computerized using ORSEE (Greiner, 2004) and zTree (Fischbacher, 2007), respectively.

3.4 Results

The main variable of interest is the message sent by the sender in stage two of the experiment. I present the results of both the discrete (the message sent) and binary measure (whether the sender was honest or dishonest in her message, irrespective of how much the true performance level was overstated). The main results hold irrespective of

the measure. Table 3.2 provides an overview of the treatments and general descriptive statistics.

I analyze differences between treatments using non-parametric tests as well as probit and OLS regressions. I use Mann-Whitney U-tests (MWU) for comparing differences in the message sent and Fisher's Exact Tests (FET) as a conservative measure for comparing the proportion of honest senders. The latter treats honesty as a binary measure where the sender is considered honest if the message sent was 12 and dishonest otherwise. Unless otherwise noted, all non-parametric tests are two-sided.

TABLE 3.2: General descriptive statistics

	Independent observations	Av. performance work task	Honesty (%)	Av. message sent
No feedback				
<i>Piece rate</i>	21	10.75 (1.67)	47.62	16.00 (4.74)
<i>Team</i>	21	11.02 (1.63)	38.10	17.71 (5.03)
<i>Tournament</i>	21	11.30 (2.10)	14.29	20.57 (4.38)
Feedback				
<i>Piece rate</i>	23	10.54 (2.04)	21.74	19.50 (4.95)
<i>Team</i>	24	10.23 (1.73)	33.33	18.08 (5.52)
<i>Tournament</i>	24	10.23 (3.53)	33.33	18.38 (5.23)

Standard deviations are shown in brackets. In the piece rate condition with feedback one pair could not be formed because one of the registered subjects did not show up for the experiment.

3.4.1 Incentive effects

Figure 3.1 shows the distribution of messages across the incentive conditions without and with feedback, respectively. In general, the vast majority of messages across treatments are either truthful (message 12) or untruthful where the sender receives the largest financial gain without compromising efficiency (message 22). In all incentive treatments with feedback as well as the tournament treatment without feedback, there is a relatively large amount of senders (approximately 20%) who communicate that the performance level is 25, where the receiver loses money without a financial gain to the sender.

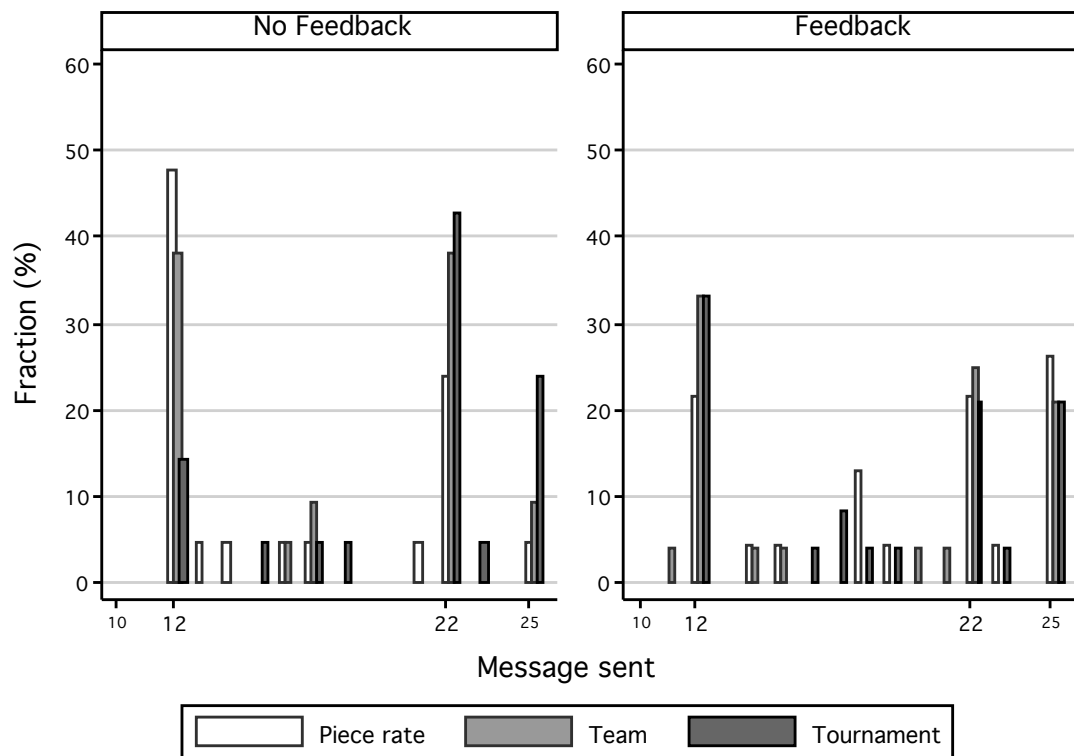


FIGURE 3.1: Message sent across the incentive treatments, without and with feedback

When no feedback is provided, senders in the tournament treatment are less likely to communicate the state honestly (14.3% vs. 47.6%, FET, $p = 0.04$) and send a more inflated message (20.6 vs. 16, MWU, $p < 0.01$) compared to those in the piece rate condition. The distribution of messages across these treatments is also significantly different according to a Kolmogorov-Smirnov test ($p = 0.03$). We find no increase in honesty under team incentives. Compared to the piece rate, senders who performed under the team incentive actually appear slightly less honest, although this is not significant (38.1% vs. 47.6%, FET, $p = 0.76$; 17.7 vs. 16, MWU, $p = 0.28$). These results only hold in the no-feedback treatment. When subjects are informed about the performance of their partner during the work task, I find no difference in honesty levels across incentive conditions (FET, $p > 0.10$, in all pairwise treatment comparisons). The role of relative performance feedback is discussed in the next section.

Table 3.3 presents the results of various probit and OLS regressions that support the results from the non-parametric tests. Specifically, the results show that senders who performed the work task under a tournament incentive are 30% less likely to be honest than those who worked under a piece rate. There is no significant effect of the incentive when subjects are informed about their partner's performance.

I thus find support for hypothesis 1 on the negative effective effect of the tournament incentive for subsequent honesty, but no evidence for hypothesis 2 on the increase in honesty following a work environment with team incentives.

TABLE 3.3: Probit and OLS regressions: the effect of incentives in the work task on subsequent honesty

<i>Dependent variable:</i>	Probit regression		OLS regression	
	<i>Honest message</i>		<i>Message sent</i>	
	No Feedback	Feedback	No Feedback	Feedback
Team	-.082 [.124]	.158 [.128]	1.703 [1.454]	-.1764 [1.567]
Tournament	-.302 ** [.125]	.120 [.129]	4.027 *** [1.473]	-1.159 [1.541]
Controls	YES	YES	YES	YES
N	63	71	63	71
(Pseudo) R ²	.124	.070	.199	.058
Log likelihood	-35.140	-40.101		

The probit regression reports average marginal effects. Standard errors are shown in square brackets. The ** and *** indicate significant effects at the 5% and 1% level, respectively. Variable description: *Team*: dummy for the team incentive treatment; *Tournament*: dummy for the tournament incentive treatment; *Controls*: gender and a dummy for when the subject majors in economics or business. Neither of these is significant.

3.4.2 Relative performance under feedback

From the descriptive statistics in table 3.2, it does not appear that relative performance information has an overall negative effect on dishonesty. While honesty is lower in the piece rate and team conditions compared to when no feedback is provided, these differences are not significant (FET, $p > 0.10$). However, it is possible that there are different effects on honesty depending on how the sender has performed relative to her partner. Table 3.4 presents the results of various probit regressions on sender's honesty. For conciseness, I present the results according to the binary measure of whether the sender was honest or dishonest in her message. The results hold when using the discrete measure (the message sent) as the dependent variable and can be found in appendix C.1.

Relative performance (models 1a and 2a) is measured by comparing the difference in performance level over the five rounds between the sender and the receiver. Larger numbers correspond to larger performance differences, where the sender outperforms or underperforms relative to the receiver by a wider margin. In the treatments where

TABLE 3.4: Probit regressions: the effect of average and relative performance in the work task on subsequent honesty

<i>Dependent variable: Honest message</i>				
	No Feedback		Feedback	
	Model 1a	Model 1b	Model 2a	Model 2b
Team	-.077 [.124]	-.064 [.122]	.213 [.118]	.200 [.118]
Tournament	-.298 ** [.125]	-.284 ** [.127]	.067 [.122]	.081 [.123]
Relative performance	.015 [.028]	.016 [.028]	.083 *** [.024]	.069 ** [.031]
Average performance		-.036 [.030]		-.020 [.026]
Controls	YES	YES	YES	YES
N	63	63	71	71
Pseudo R ²	.127	.145	.193	.200
Log likelihood	-34.990	-34.300	-34.783	-34.473

Average marginal effects reported. Standard errors are shown in square brackets. The ** and *** indicate significant effects at the 5% and 1% level, respectively. Variable description: *Team*: dummy for the team incentive treatment; *Tournament*: dummy for the tournament incentive treatment; *Relative performance*: the average performance difference between the sender and their partner over the five rounds in the work task; *Average performance*: the average performance of the sender over the five rounds in the work task; *Controls*: gender and a dummy for when the subject majors in economics or business. Neither of these is significant.

feedback is provided (model 2a), the variable is significant with a positive sign. This indicates that honesty increases as the relative performance difference between the sender and the receiver becomes large. This effect holds across the incentive conditions (see appendix C.1 for these regression results). When no feedback is provided (model 1a), the tournament incentive has a significant and negative effect on subsequent honesty. Relative performance is not significant. Taking these results together, relative performance affects subsequent honesty only when information is provided about the performance of the other player and this appears to dominate the effect of the incentive in the work task.

Models 1b and 2b look at the role of absolute performance in the no-feedback and feedback treatment. It is possible that differences in lying behaviour are driven by inherent individual differences in skill level. If this is the case, the variable average performance should be significant in both feedback treatments. The results in table 3.4 show that average performance is not significant in either treatment. Furthermore, when average performance is added to the model, the effects of the tournament incentive and

relative performance remain significant in the no-feedback and feedback treatments, respectively.

The regressions on average and relative performance can also exclude income effects as a driver of honesty. It is possible that subjects who perform well on the work task are more inclined to be honest, because they have already generated substantial earnings. Reversely, those who have earned little might be more inclined to lie to minimize inequality in earnings (Fehr and Schmidt, 1999). If this holds, we should expect subjects who earn little, particularly those in the tournament treatment, to lie more compared to subjects who generated higher earnings. This is not supported by the regressions in table 3.4, where subjects with higher average performance are not significantly more honest than those with lower average performance. Furthermore, regressing honesty on average earnings yields no significant coefficients for either the feedback or no-feedback treatments. These results can be found in appendix C.1.

These results indicate that while relative performance affects subsequent honesty levels, it does not decrease honesty per se (hypothesis 3b). Rather, senders who under- or outperform the receiver by a small amount appear less honest than when the relative performance difference is larger.

3.4.3 Efficiency

The tournament incentive does not appear to yield productivity gains in the work task. Average performance does not differ significantly between the incentive treatments (see table 3.2). In the absence of relative feedback information, subjects seem to perform slightly better. On average participants complete six more tasks in each incentive treatment compared to their feedback counterparts. These differences are not significant.

In the second stage, dishonest senders have the option to tell either a selfish black lie or a spiteful lie. The latter hurts efficiency, because the receiver loses money without an additional gain for the sender. Since spiteful lies are more common under the tournament incentive, average earnings are significantly lower compared to pairs in the piece rate treatment (MWU, $p = 0.04$). Again, this result only holds when no feedback is provided.

3.5 Conclusion

This paper demonstrates that the effect of monetary incentives on deception is not restricted to the specific task for which the incentive is designed. Even in a subsequent task, where the original incentive is no longer relevant, the decision to deceive is driven

by the incentive set in the work environment as well as one's relative performance in this environment.

I find that senders who worked under tournament incentives are less honest than those who worked under a piece rate. Due to the higher incidence of spiteful lies, efficiency in the second task is lowest for subjects that performed under the tournament incentive. I find no increase in honesty for subjects who worked under team incentives compared to the piece rate condition. This result is not robust to information about the worker's performance relative to their partner. When such feedback is provided, it affects honesty across the incentive conditions. In particular, honesty is lower when the relative performance difference between sender and receiver is small, compared to when the sender under- or outperforms the receiver by a larger amount. This effect of relative performance appears to override the individual treatment effect of the incentive scheme.

From the perspective of mechanism design, these results warrant caution in using tournament incentives as well as schemes that stress performance comparisons between peers. Such comparisons are a particular concern for tournament incentives, which typically incorporate a performance ranking among employees. However, information on a worker's relative performance is not necessarily absent from other schemes, such as a piece rate and team incentives.

The results on relative performance feedback suggest that subsequent honesty is affected depending on where the individual is in the relative performance distribution. It would be insightful to better understand whether mechanisms such as a sense of entitlement (Cappelen et al., 2013a) or general disutility from losing the competition (Buser and Dreber, 2013) become stronger when relative performance differences become small. In addition, it would be interesting

Since the experiment in this paper was not specifically designed to address underlying mechanisms, it would be insightful for future research to disentangle why tournament incentives and relative performance information affect subsequent honesty in this way. In particular, the results on relative performance feedback suggest that subsequent honesty is affected depending on where the individual is in the relative performance distribution. It would be insightful to better understand whether mechanisms such as a sense of entitlement (Cappelen et al., 2013a) or general disutility from losing the competition (Buser and Dreber, 2013) are stronger when relative performance differences are small. In addition, it would be interesting to examine whether subsequent dishonesty is affected by specific elements of the work interaction, such as the length of the work task and the possibility for helping (Drago and Garvey, 1998) and sabotage (Harbring and Irlenbusch, 2011).

Chapter 4

Are social investments rewarded? A Pay-What-You-Want field experiment with Fair Trade products

Joint work with Ayelet Gneezy, University of California, San Diego

Abstract

We investigate whether the addition of a social attribute (here: Fair Trade) affects the level of payments for products offered under Pay-What-You-Want pricing. In addition, we evaluate whether a supplier benefits from differentiating its product offering to include products that feature such a social attribute. In a field experiment we offer consumers a Fair Trade chocolate product and a non-Fair Trade equivalent. The two products are either offered separately or together. We find that when the products are offered separately, payments for the Fair Trade product are not significantly higher than that of the non-Fair Trade equivalent. When both products are offered, consumers are less likely to choose the non-Fair Trade product (20%) and pay significantly less compared to those choosing the Fair Trade alternative. By contrast, average payment levels for the Fair Trade product are not significantly different across conditions. Thus, the difference between the two treatments appears driven by a decrease in payments for the non-Fair Trade product in the joint condition. We find limited support for the argument that social preferences contribute to higher payments for the Fair Trade product.

4.1 Introduction

One of the most successful long-running applications of Pay-What-You-Want (PWYW) pricing is Humble Bundle, an online platform where customers can purchase a predetermined bundle of videogames by individual developers across various platforms. Besides deciding how much to pay, customers are also requested to indicate how their money should be allocated: to Humble Bundle, to the developers and to charity. After 3 years of operating, Humble Bundle has earned over \$50 million, of which 40% was directed to charity¹. However, not all applications of PWYW pricing are equally successful. In the same industry, individual developer Joost van Dongen launched his popular game Proun under Pay-What-You-Want. The game was downloaded 250.000 times, collecting a meagre \$10.000 in payments². This occurred despite the developer's best efforts to promote the PWYW pricing scheme and adding a soundtrack to the package as a bonus. In other applications, Panera Bread's restaurant in Portland reverted back to fixed prices after disappointing revenues, even though branches elsewhere have been more successful³.

The observation that PWYW succeeds in some settings, but fails to be profitable in others begs the question of what factors determine its effective application. Specifically, little is known about what drives first, the purchase decision and second, how much a customer will pay. One approach in the literature is that individuals with social preferences and self-identity concerns can be motivated to pay an amount higher than zero (Gneezy et al., 2010, 2012; Schmidt et al., 2012). In this paper, we examine whether such motives can explain why consumers would pay more for one product than for another. Specifically, does the addition of a social attribute to a product change purchase behaviours under a Pay-What-You-Want pricing scheme? When the purchase of the product affects the welfare of a third party, a consumer with self-image concerns may choose to offer a higher payment than for a product that lacks such a social attribute. Likewise, consumers may want to reciprocate social investments made by the supplier via a higher payment.

We address this question using a field experiment, where consumers at a Farmer's Market are offered a Fair Trade product and a non-certified equivalent under Pay-What-You-Want pricing. Besides the presence or absence of the Fair Trade label, the two products are identical in physical appearance and taste. We use no deception in this experiment. In a between-subject design, we offer the consumer the two products separately and analyze the likelihood to purchase as well as the amount paid. Second, we examine whether payments for the Fair Trade attribute can be leveraged by presenting the two

¹<http://www.rockpapershotgun.com/2013/08/23/interview-humble-bundle-on-humble-bundles>

²<http://www.joostdevblog.blogspot.nl/2011/10/proun-is-big-success-pay-what-you-want.html>

³<http://business.time.com/2012/02/27/at-paneras-pay-what-you-want-cafes-customers-usually-pay-full-price/>

products together. Since the two products are otherwise identical, the joint evaluation format emphasizes the Fair Trade attribute, or lack thereof.

We find that when the products are offered separately, payments for the Fair Trade product are not significantly higher than for the non-Fair Trade equivalent. When both products are offered, consumers are less likely to choose the non-Fair Trade product (20%), and those who do pay significantly less compared to those choosing the Fair Trade alternative. The difference between the two treatments appears driven by payments for the non-Fair Trade product. In the joint condition, these are lower compared to the average payment when the same product is presented on its own. By contrast, average payment levels for the Fair Trade product are not significantly different across conditions. Thus, our evidence does not support the argument that social preferences or identity concerns contribute to higher payments for the Fair Trade product. Rather, it appears that the presence of the Fair Trade product decreases valuations for the non-Fair Trade alternative relative to what customers pay for this product when it is offered on its own. As this is work in progress, we are currently running two additional treatments to exclude the possible confound of sorting in explaining the difference in payment in the joint condition. Preliminary results, albeit based on a low number of observations, suggest that the difference in payments remains when sorting is excluded.

The rest of the paper is structured as follows. The next section discusses how the mechanisms of social preferences and self-identity concerns affect payments and covers previous work on purchase behaviour for products with social attributes under Pay-What-You-Want and fixed pricing. Section 3 outlines our hypotheses. Section 4 discusses the experimental design and substantiates our use of the Fair Trade label. The results are presented in section 5, focusing on average amount paid and seller profits. Section 6 concludes with a discussion.

4.2 Literature review

4.2.1 Mechanisms

We briefly review the mechanisms of self-identity and social preferences and discuss their role in driving payments under Pay-What-You-Want pricing.

The notion of self-identity assumes that individuals care about the kind of person they consider themselves to be: their self-concept ([Bénabou and Tirole, 2011](#)). Generally, individuals tend to think of themselves according to certain favourable characteristics, such as honesty, fairness and generosity. Certain behaviour is in line with the individual's

self-concept, such as making a donation to charity, whereas other behaviour opposes it, such as failing to tip a waiter who has provided excellent service. Consequently, if the individual cares about the maintenance of her positive self-image, she is deterred from behaviours that violate her core values. The main difference with the concept of social image is that self-image concerns can deter norm violations even when the action occurs in private (Ariely et al., 2009).

Pay-What-You-Want incorporates self-image concerns by giving the consumer full responsibility over how much to pay for the product. Even though a payment of zero is possible, the consumer may find it difficult to reconcile such an action with her self-concept of being fair or generous. Thus, if a consumer feels that they are paying less than what is fair for the product, self-image concerns might incline them to either increase payment to a level that is closer to what they consider fair or forgo purchase altogether (Machado and Sinha, 2013).

Social preferences, most notably (reciprocal) altruism (Andreoni, 1990) and inequality aversion (Fehr and Schmidt, 1999), can also discourage low payments under Pay-What-You-Want. The consumer may suffer some psychological disutility from making a low payment when the supplier has incurred a cost in offering this product. Likewise, the Pay-What-You-Want mechanism presents an opportunity for individuals to show their altruism by making a generous offer. According to this view, we would expect reciprocal and altruistic consumers, as well as those experiencing advantageous inequality aversion, to make higher payments.

4.2.2 The Pay-What-You-Want literature

In line with the mechanism of self-identity, Gneezy et al. (2010) find that the addition of a charity component affects both average payment and purchase rate. In a large field experiment at a theme park, adding a charity component made fewer customers buy a souvenir photo of their ride in a roller coaster, but those that did paid significantly more (\$5.33 compared to \$0.92). In follow-up experiments, Gneezy et al. (2012) find a similar effect with souvenir photographs for a boat tour and with meal payments at a restaurant in Vienna. Further support comes from Gravert (2013), who finds that reminding customers in a charitable bookstore about their membership status increases average amounts paid during a special PWYW sale.

Schmidt et al. (2012) find support for social preferences in driving PWYW payments in a laboratory experiment. Participants who were assigned the role of buyer were willing to pay prices significantly above the supplier's production cost, which was common knowledge in the experiment. In addition, buyers also paid higher amounts to sellers who

had invested in product quality, even when such an investment decision was exogenously imposed. Finally, [Riener and Traxler \(2012\)](#) study payments over a duration of two years at a restaurant that runs exclusively on a PWYW pricing scheme. They find that above average hours of sunshine in the autumn season significantly increased the amount paid for the meal, even though this effect was negative in the summer season.

Other factors that have been investigated are anonymity and repeated interactions. [Gneezy et al. \(2012\)](#) find that amounts paid in a Viennese restaurant are higher when customers pay anonymously rather than engage with the waiter face-to-face. Studies at various PWYW restaurants find that payments are stable across repeated interactions ([Kim et al., 2010](#); [Gneezy et al., 2012](#); [Riener and Traxler, 2012](#)). In a laboratory study, [Schmidt et al. \(2012\)](#) also find that participants are willing to support the PWYW supplier over repeated interactions, although payments are less generous in the final period of the game.

4.2.3 Literature on WTP and ethical consumption

Related work using the paradigm of Willingness To Pay (WTP) provides further support for the importance of image concerns and social preferences in purchase behaviour. In an experimental study on eBay, [Elfenbein and McManus \(2010\)](#) find that bids are higher for a product where part of the payment is donated to charity compared to bids for an identical product that lacks such a charity component. In a laboratory study, [Frackenhohl and Pönitzsch \(2013\)](#) find that the addition of a charity donation increases the willingness to pay for a mug compared to when this charity donation is absent. In addition, the increase in willingness to pay for the mug with the donation exceeds the valuation of the charity donation when this is offered on its own. Work by [Friedrichsen and Engelmann \(2013\)](#) finds that participants were willing to pay more for a Fair Trade chocolate bar when they had to announce their reservation price to the other participants in the room compared to when they had to state their valuation privately.

Also related is the literature on ethical consumption. A number of empirical and field experimental studies report an increase in sales upon the introduction of a Fair Trade label on coffee ([Hainmueller et al., 2014](#); [Arnot et al., 2006](#)), an environmental label on apparel ([Hainmueller and Hiscox, 2014](#)), canned tuna marked as ‘dolphin-free’ ([Teisl et al., 2002](#)) and a label on cotton socks signaling good working conditions ([Prasad et al., 2004](#)). Related work in the laboratory finds that participants assigned the role of buyer are willing to pay more for a product to avoid imposing a negative externality on a third party ([Rode et al., 2008](#); [Danz et al., 2012](#); [Bartling and Weber, 2013](#)).

4.3 Hypotheses

The purpose of this paper is to examine whether the addition of a social attribute influences payment levels for a product. Thus, we build on the insight that self-identity concerns and social preferences can result in above-zero payments and examine to what extent these motivations affect the level of payment under PWYW pricing. In the experiment, we implement this by offering consumers a regular (non-Fair Trade) product and a Fair Trade certified equivalent.

We hypothesize that the addition of a social attribute, such as connecting a Fair Trade label or a charity donation to the purchase of the product, can leverage self-image concerns and social preferences. When a social attribute is involved, a low payment can signal, to the individual and to others, that the consumer does not value the cause that the attribute represents. Thus, from a self-identity perspective, frugality may be less desirable when the consumer's payment affects not only the seller, but also an external third party. From the perspective of social preferences, the addition of a social attribute can raise payments if the consumer chooses to reciprocate the supplier's social investment. Alternatively, consumers may derive some positive utility by contributing to the cause that the attribute represents and increase their payment accordingly.

Hypothesis 1. *In the separate condition, average payment for the Fair Trade product is higher than for the non-Fair Trade equivalent.*

We use the joint evaluation format as a means to stress the social attribute of the product. This insight draws on the literature of contextual inference (Kamenica, 2008). When options are presented separately, each of the product attributes enter the consumer's utility function with a certain weight. However, in the absence of a reference point some attributes are difficult to evaluate in isolation, which can lead to a poor translation of attribute importance to value estimation (List, 2002; Hsee et al., 1999). For example, in List (2002), a package of 13 baseball cards, of which 10 are in good condition and 3 in poor condition, was deemed less attractive than a package of 10 cards in good condition. Thus, when a comparative product is offered under joint evaluation, it creates an explicit reference point that helps individuals compare attributes and discern differences. This allows individuals to benchmark attributes that are difficult to evaluate in isolation (Hsee and Leclerc, 1998; González-Vallejo and Moran, 2001) as well as draw

attention to contrasting attributes (Bordalo et al., 2012; List, 2002)⁴. Coming back to the study by List (2002): when the two sets of baseball cards were presented together, consumers offered more for the package of 13 cards. In other words, individuals were more attentive to the attribute of quantity when the package of 10 baseball cards was presented next to a package containing 13 cards and adjusted their bids accordingly.

Two key differences between work in this literature and the present study is that our experiment features a product with a social attribute, as opposed to a self-interested feature, such as quantity or quality of the product. A possible critique is that a social attribute is more ambiguous as to its desirability. Some individuals might perceive the social attribute as negative, whereas others might be indifferent to the cause it represents. While this a valid critique, we believe the Fair Trade label is overall a desirable attribute, given that it enjoys wide consumer support and sales revenues are rapidly increasing in the United States and other key markets (Fair Trade Labeling Organization, 2012). A second important difference with previous work is that the social attribute is not explicitly featured on both products. While the Fair Trade chocolate cupcake is marked as ‘Fair Trade’, the alternative is simply marked as a ‘chocolate cupcake’. In other words, the latter is not labelled as ‘non-Fair Trade’, which makes the social attribute silent for the regular product when it is offered on its own. However, this only strengthens our argument that the joint presentation format will draw attention to the social attribute. Rather than the attribute being difficult to evaluate under separate evaluation, it is likely that the Fair Trade attribute is not taken into consideration at all when the regular product is presented on its own.

In the joint condition of the experiment we present the Fair Trade product alongside a non-Fair Trade equivalent. Since the two products are similar on all other observable dimensions such as physical appearance (shape, size and color) as well as taste, we expect the attribute of Fair Trade to be salient. We hypothesize that this salience can affect valuations in two ways. First, drawing attention to the Fair Trade attribute might raise

⁴Hsee et al. (1999) provide support for benchmarking under joint evaluation in the following experiment. Participants are given a hypothetical hiring decision, where they have to make a salary offer to a job candidate for the position of programmer. The first candidate, J, has a 3.0 GPA and has written 70 KY programs in the last two years. The other candidate, S, has a GPA of 4.9 and has written 10 KY programs in the same time span. While important, the attribute of programming experience is not informative in isolation, since it is difficult to evaluate whether writing 10 or 70 KY programs is a low or high number. Under separate evaluation, the authors find that participants offered a higher salary to candidate S, but made a higher offer to candidate J when the two candidates were evaluated side by side. Similar results were obtained by Hsee and Leclerc (1998) and González-Vallejo and Moran (2001) using other hypothetical scenarios. In support of attribute salience, Okada (2005) offers participants a \$50 grocery or dinner certificate as a reward for their participation in a study. Of those presented the two options separately, only 23.8% of subjects preferred the grocery certificate over the dinner option. However, under joint evaluation, 56% selected the grocery certificate. Okada argues that the attribute contrasting the two options, which she describes as the tradeoff between ‘utilitarian/hedonic’ properties, is made salient under joint evaluation. This in turn makes individuals opt for a product that is useful as opposed to just pleasurable.

the average amount paid for the Fair Trade product through either mechanism specified above. If this manipulation is successful, we should see an increase in amount paid for the Fair Trade product under joint evaluation compared to when this product is offered separately. Second, it is possible that the presence of the Fair Trade product conveys information to the consumer that the regular product is ‘non-Fair Trade’. Even though the regular product lacks a label, the attribute of ‘non-Fair Trade’ might not enter the utility function of most consumers when the product is presented on its own. Therefore, if the joint presentation makes clear that the regular product lacks the social attribute, we can expect payments for this product to decrease relative to the average amount paid in the separate condition.

Hypothesis 2. *In the joint condition, average payments for the Fair Trade product are higher than for the non-Fair Trade equivalent.*

Hypothesis 2a. *In the joint condition, average payments for the Fair Trade product increase relative to those in the separate condition.*

Hypothesis 2b. *In the joint condition, average payments for the non-Fair Trade product decrease relative to those in the separate condition.*

4.4 Experimental procedures

4.4.1 The Fair Trade label

Fair Trade certification represents a number of initiatives to alleviate poverty in developing countries by directing part of the proceeds to the farmers that grow the respective products. The label guarantees the farmers a minimum floor price for their output (or the market price if this is higher), ensures a safe work environment, freedom of association and the prohibition of child and forced labor. This is funded via a price premium (typically 20%) attached to products carrying the Fair Trade label ([Fair Trade Labeling Organization, 2012](#)). The collected premia are allocated via a democratic decision procedure to various social and business development projects in the community, including scholarships, leadership training and school building and renovation ([Fair Trade USA, 2014](#)). Different from a charity contribution, the Fair Trade label represents a market-based approach to alleviate poverty in developing countries. Individuals can voluntarily contribute to the cause through their consumption decisions by choosing to purchase the labelled product.

By sales volume, the main product categories that carry the Fair Trade label are flowers and plants, bananas, sugar and coffee (roasted, instant and raw cacao beans). Products are sold in over 125 countries worldwide with total sales revenues exceeding \$6.4 billion in 2012 (Fair Trade Labeling Organization, 2013). The largest markets are the United States, the United Kingdom and various countries in continental Europe such as Germany. In the United States, sales revenue has rapidly increased from \$289 million in 2004 to nearly \$1.4 billion in 2011 (Fair Trade Labeling Organization, 2005, 2012). However, Fair Trade consumption still accounts for less than 1% of total grocery market sales in the United States (Food Marketing Institute, 2014).

The Fair Trade label is particularly well suited to examine the role of social preferences and self-identity concerns in purchase behaviour. A useful feature is that the label refers exclusively to improving working conditions for farmers in developing countries. It does not impose quality standards on the product nor does it concern itself with the use of pesticides or treatment of livestock. Thus, in contrast to certified organic produce, there are no personal health or nutritional reasons to prefer Fair Trade over a regular alternative. Assuming that consumers understand the meaning of the Fair Trade label, the two products should be considered equivalent in terms of consumption value. There is some support for this assumption. A study by Lotz et al. (2013) does not find a difference in ex ante beliefs about taste for Fair Trade and non-Fair Trade chocolate among consumers in Germany.

4.4.2 General procedures

We run a field experiment at various Farmer's markets in the San Diego area between February and June 2014. A Farmer's Market is a weekly event where local vendors offer their products, typically produce, baked goods and crafts. There are a total of 53 Farmer's Markets organized every week across the San Diego area and are typically 4-5 hours in length. We operated a stand at 5 different markets (North Park, Pacific Beach, La Jolla, University Heights and Hillcrest) for a maximum of 4 occasions each. In total, we were at the Farmer's Market for 14 days generating 219 unique sales.

At the market we operate our own stand where we sell our product (chocolate cupcakes) exclusively under Pay-What-You-Want pricing. This avoids reputation issues with repeated customers as well as potential contagion from the sale of other products. The stand was framed by a large banner that read 'Delicious Cupcakes, Pay What You Want' but did not specify our treatment conditions.

Our product offering consists of two products: a regular chocolate cupcake and a Fair-Trade certified equivalent. Both products are identical in physical appearance (see

appendix D.1 for a graphical representation) and were presented to the customer on the stand display with signs indicating the respective product. We denoted the regular chocolate cupcake as ‘Chocolate Cupcake’ and the Fair Trade equivalent as a ‘Fair Trade Chocolate Cupcake’. Besides several display items, the cupcakes were individually boxed and labelled with a sticker indicating which of the two products was purchased (Figure 4.1 and 4.2). The Fair Trade sticker featured the official Fair Trade logo used in the United States. The sticker for the regular product featured an image of a cupcake. We made the signage and stickers as identical as possible in terms of style, but different with respect to content. Customers were limited to one cupcake per person.

The experiment uses no deception whatsoever. The Fair Trade cupcakes we offered were indeed distinct from the regular cupcakes in that they are made using Fair Trade certified cacao powder. This was arranged via a special order to a small local bakery in the San Diego area. We paid \$1.75 and \$2.10 for each regular and Fair Trade chocolate cupcake, respectively.



FIGURE 4.1: Sticker for the regular (non-Fair Trade) product



FIGURE 4.2: Sticker for the Fair Trade certified product

For each purchase, we record the amount paid by the customer. We also document the observable demographics of gender, age and ethnicity to the best of our abilities. In addition, we recorded the time of purchase as well as whether the customer was alone or accompanied by a group and whether this group contained any young children. Finally, we also counted the number of passerbys every thirty minutes to calculate the overall purchase rate.

4.4.3 Treatments and randomization

Once customers approached the stand they were exposed to our different treatment conditions: the regular cupcake is presented separately (SEP REG), the Fair Trade cupcake is presented separately (SEP FT) or a joint condition where the regular and Fair Trade cupcake are offered together (JOINT). We randomized at the market level, changing

conditions every 7 sales⁵, which implied changing the display with the appropriate signs and a slight change in the script to inform the customer what we were offering. Each market was exposed to all treatment conditions at least once. In advance, we created a random series of the treatment conditions for each market (see appendix D.2 for details) that dictated the order in which conditions would change. In case the Farmer's Market ended before we were able to conclude a treatment, we would continue with this condition upon our next visit at that particular market.

4.5 Results

4.5.1 Checking randomization

We collected 219 unique sales over 5 markets. The proportion of conditions in each market are not significantly different according to a Chi-square test, $\chi^2(6, N = 212) = 8.82, p = 0.18$. However, this result is conditional on excluding the data from the University Heights market ($n = 7$). Due to practical limitations, we were only able to operate a stand at University Heights for 1 day. It turned out that this market as a whole was doing poorly and was terminated the following week. Since the number of sales at this market were low we were only able to effectively run the SEP FT and JOINT conditions. In the following analyses we still include the sales data from University Heights. However, all the results we report are robust when the observations from this market are excluded.

The conditions also appear balanced in terms of demographics of the customer, which are displayed in appendix D.3. There are no significant differences in observed gender, ethnicity, age and group size between the separate regular, separate Fair Trade and joint (overall) treatments. However, within the joint treatment the proportion of women choosing the regular product in the joint condition is significantly lower than those choosing the Fair Trade equivalent (Fisher's Exact Test, $p = 0.029$).

4.5.2 Profile of the customer

Of those that could be classified according to observable demographics, consumers purchasing our product are predominately female (62.9%) with a mean age in the high 30s. The vast majority of purchasers are Caucasian (79.2%), with Asian and African

⁵During our first visit at each market, we randomized every 5 sales to ensure that we would be able to run all conditions on the first day. This was to address a possible 'first day' effect, where consumers are more likely to approach the stand because it is new at the market. During each consecutive visit, we randomized every 7 sales.

Americans being the next two largest ethnic categories (7% and 3%, respectively). Approximately 60% of customers purchase the product on their own, whereas slightly more than 32% come in pairs. The remaining 8% are in groups larger than two people. Approximately 11% of customers were accompanied by one or more young children.

4.5.3 Amount paid

Figure 4.3 presents the average amount paid by condition. The joint condition is broken up by the type of product (regular or Fair Trade) chosen by the customer. A total of 14 customers, approximately 17% of consumers in the joint condition, chose the regular product. Unless otherwise noted, all reported statistical tests are two-sided.

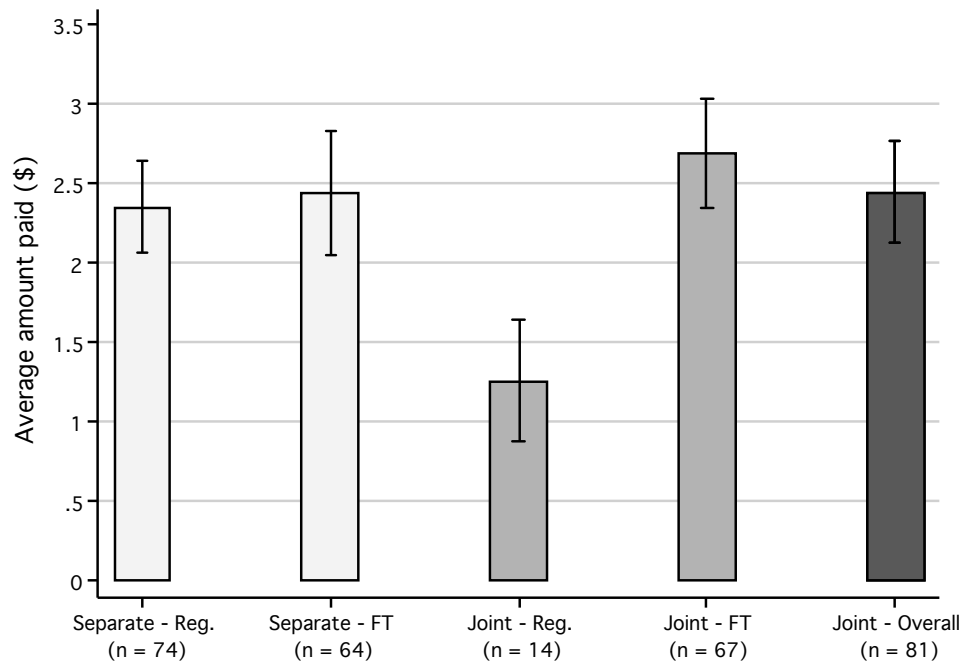


FIGURE 4.3: Average amount paid by condition

When the two products are presented separately, customers pay an average of \$2.35 for the regular chocolate cupcake and \$2.44 for the Fair Trade equivalent. These amounts are not different according to a standard t-test ($M_{SEP-REG} = 2.35$ vs. $M_{SEP-FT} = 2.44$, $t(136) = -0.351$, $p = .726$) and Mann-Whitney U-test ($p = .867$). Moreover, the distribution of payments between these two treatments is not significant according to a Kolmogorov-Smirnov test ($p = .988$). These results indicate that customers do not pay more for a product with a Fair Trade label compared to a non-certified equivalent when the two are presented separately. Thus, we do not find support for hypothesis 1.

In the joint condition, the relative difference between the two products becomes significant. Customers who choose the regular product pay an average of \$1.26 for the regular chocolate cupcake and \$2.70 for the Fair Trade alternative. These differences are strongly significant according to a t-test ($M_{JOINT-REG} = 1.26$ vs. $M_{JOINT-FT} = 2.70$, $t(79) = -3.705$, $p < .001$) and Mann-Whitney U-test ($p < .001$). Comparing payments for each of the products in the joint condition to those in the separate condition, we find that the difference appears to be driven by regular product. For the Fair Trade product, payments in the joint condition (\$2.70) rise slightly compared to the separate presentation (\$2.44). However, this difference in the amount paid is not significant (MWU, $p = .246$). Payments for the regular product in the joint condition are significantly below those in the separate condition ($M_{SEP-REG} = 2.35$ vs. $M_{JOINT-REG} = 1.26$, $t(86) = 3.205$, $p = .002$) and Mann-Whitney U-test ($p < .001$). Furthermore, under joint evaluation, nearly 80% of customers who choose the regular cupcake pay \$1. In the separate condition only 21% of customers pay this amount for the same product. These distributions are significantly different according to a Kolmogorov Smirnov test ($p < .001$).

TABLE 4.1: OLS regression: Drivers of the amount paid

<i>Dependent variable: Amount paid</i>				
	Model 1	Model 2	Model 3	Model 4
Separate - Fair Trade	.084 [.243]	.0783 [.257]	.230 [.321]	.396 [.438]
Joint - Regular	-1.093 *** [.224]	-1.075 *** [.224]	-1.105 *** [.328]	-1.274 *** [.394]
Joint - Fair Trade	.343 [.226]	.348 [.230]	.441 [.269]	.337 [.342]
Pacific Beach		.165 [.252]	-.359 [.388]	-.721 [.731]
La Jolla		.081 [.239]	-.381 [.356]	-.475 [.460]
University Heights		.333 [.651]	-.210 [.754]	-.030 [.928]
Hillcrest		.244 [.323]	-.306 [.408]	-.155 [.643]
Market session			-.045 [.137]	.044 [.164]
Gender			-.195 [.252]	-.118 [.294]
Age category			.662 ** [.310]	.598 [.388]
Age category ²			-.099 ** [.039]	-.088 * [.051]
Ethnic majority				.068 [.337]
Group size				-.032 [.229]
Constant		2.226 *** [.214]	2.042 *** [.626]	2.628 *** [.837]
RA controls	NO	NO	NO	YES
N	219	219	177	141
R ²	.057	.061	.105	.139

All models have robust standard errors for independent purchases and are shown in square brackets. The ***, ** and * indicate significant effects at the 1%, 5% and 10% level, respectively. Models 3 and 4 exclude observations where the customer could not be classified according to one or all of the demographic variables included in the model. Variable description: *Separate - Fair Trade / Joint - Regular / Joint - Fair Trade*: dummy for the respective treatment; *Pacific Beach / La Jolla / University Heights / Hillcrest*: dummy for the respective market; *Market session*: the occasion number that our stand operated at the market; *Gender*: the gender of the customer; *Age category / Age category²*: an ordinal variable of age at 10 year intervals. Higher numbers indicate higher age; *Ethnic majority*: dummy with a value of 1 if the customer is Caucasian and 0 otherwise; *Group size*: the number of individuals in the group; *RA controls*: Includes dummies for the research assistant(s) operating the stand. For the coefficients of these controls, see appendix D.5.

We complement this non-parametric analysis with various OLS regressions, the results of which are shown in table 4.1. The dummy for the joint condition with the regular

product choice is strongly significant and negative compared to when the regular product is offered on its own. Specifically, moving from the separate regular to the joint regular condition lowers payments by \$1.09 - \$1.27, depending on the model. This is a decrease of approximately 50% compared to what is paid for this product in the separate condition. Differences between the separate regular condition and the other treatments are not significant. Models 2 - 4 include various control variables, such as dummies for each of the markets, the number of occasions our stand was present at the market, the research assistant as well as various observable demographics of the purchaser. None of these controls except age is significant. In particular, we find an inverted-u relationship between age and amount paid, where young and very old customers pay less than those of middle age. However, this effect is not entirely robust in model 4 which includes all the controls.

These results support hypothesis 2: in the joint condition, consumers pay more for the Fair Trade product compared to the non-Fair Trade equivalent. Conditional on product choice, this difference appears to be driven by payments for the non-Fair Trade product decreasing relative to those in the separate condition, which supports hypothesis 2b. Average payments for the Fair Trade product do not appear to increase in the joint condition compared to amounts paid for this product when it is presented on its own.

4.5.4 Purchase rates

Another indicator of purchase behaviour is the purchase rate. We take the count of the passerbys for each market to calculate overall traffic over each thirty minute interval. Using the recorded times of purchase, we can calculate the amount of purchases per condition for every half hour we were present at the market. Since both traffic and amount sold can differ substantially by market session, we average the total number of purchases and traffic numbers by condition for each market session.

Purchase rates for the separate conditions are 2.42% for the regular product and 3.74% for the Fair Trade alternative. This difference is weakly significant (MWU, $p = .08$). At a rate of 1.46%, the amount of customers purchasing in the joint condition is substantially lower. This is significant compared to both separate conditions (MWU, both $p < .01$). An OLS regression with controls, the results of which are included in appendix D.5, confirm these results. In the regression, the dummies for the La Jolla and Hillcrest markets are significant and negative, indicating that purchase rates in these markets, irrespective of the treatment, are lower compared to the benchmark of the North Park market. A possible explanation is that there is more competition at the La Jolla and Hillcrest markets. Since these are among the largest markets in San Diego, there are more

vendors offering similar kinds of products. In addition, stand space at these markets is in high demand, which could imply that only very successful vendors are stationed there.

4.5.5 Profits

TABLE 4.2: Marginal and estimated average profit (per 10.000 traffic) per condition

	Separate Regular	Separate Fair Trade	Joint Regular	Joint Fair Trade	Joint Overall
Marginal revenue	\$2.35	\$2.44	\$1.26	\$2.70	\$2.45
Marginal costs	\$1.75	\$2.10	\$1.75	\$2.10	\$2.04
Marginal profit	\$0.60	\$0.34	-\$0.49	\$0.60	\$0.41
Average revenue	\$570.20	\$911.58	\$31.77	\$325.22	\$274.50
Average costs	\$423.77	\$785.11	\$44.07	\$253.16	\$217.02
Average profit	\$146.44	\$126.47	-\$12.30	\$72.06	\$57.48

Finally, to determine whether offering the Fair Trade product is a sensible strategy, we evaluate the profitability of each of the different product offerings. The top rows of table 4.2 display marginal profits for the regular and Fair Trade product across the separate and joint conditions. Even though marginal revenue for the two products is similar under separate evaluation, the Fair Trade product comes with higher costs. We paid \$2.10 for each Fair Trade cupcake and \$1.75 for the non-Fair Trade alternative. Thus, under separate evaluation, marginal profits of \$0.60 for the non-Fair Trade product are markedly higher than the \$0.34 profit for each Fair Trade cupcake. When the two products are offered together, marginal profits for the Fair Trade product increase to \$0.60. For the regular product, marginal profits are negative at \$0.49 per cupcake. Thus, in terms of marginal profits the Fair Trade cupcake does better when the product is presented alongside a non-certified equivalent. However, marginal profit is not higher compared to that of the regular cupcake when this is presented on its own.

One caveat to this analysis is that the purchase rates for the joint condition are significantly lower compared to both separate conditions. Taking into account these purchase rates, we can estimate average overall profits for every 10.000 passerbys of traffic (bottom rows of table 4.2). Overall, estimated profits are highest for the non-certified product when it is offered on its own, generating \$146.44 compared to \$126.47 for the Fair Trade cupcake. While marginal profits for the Fair Trade product under joint evaluation are higher than in the separate condition, the low purchase rate depresses overall profits to slightly over \$72. Based on these estimates, the supplier would maximize profits by

offering the non-certified product on its own. However, conditional on differentiating, offering the Fair Trade product generates higher profits.

These conclusions are of course conditional on the marginal cost difference for each of the products. It is possible that overall profits for the Fair Trade product in the separate condition can do better than the non-certified equivalent if the seller is able to acquire the Fair Trade ingredients at a lower cost. In our study we paid a mark-up of 20% for the Fair Trade product. If this would drop to 16.6% (a \$0.06 reduction) the estimated profit for the Fair Trade product per 10.000 passerbys would match that of the regular product when this is offered on its own.

4.6 Discussion and conclusion

In a field experiment we offer consumers a Fair Trade and non-Fair Trade product under a Pay-What-You-Want pricing scheme. Under joint evaluation, the consumers who choose the non-Fair Trade product pay significantly less than those choosing the Fair Trade equivalent. However, we find no such difference when the two products are offered separately. This evidence does not support that social preferences or self-identity concerns contribute to higher payments for the Fair Trade product. Under social preferences, we would expect customers to pay more for the Fair Trade product if they desire to reciprocate the supplier's social investment or if the customer derives some positive utility from contributing to a better existence for coffee bean farmers. Individuals with self-identity concerns would choose to pay more for the Fair Trade product if they maintain themselves as individuals who are fair and care for the well-being of a third party. However, customers do not pay more for the Fair Trade product when this is offered on its own. Furthermore, under joint evaluation, there is no increase in payments for the Fair Trade product in the joint condition compared to what is paid when the product is presented separately. Our interpretation of this result is that the presence of the Fair Trade product in the joint evaluation format depresses valuations for the non-certified option. Whereas under separate evaluation the consumer purchases a 'chocolate cupcake', in the joint condition this product is now explicitly 'non-Fair Trade'. Thus, when the two products are offered together, it becomes clear that the regular product lacks the social attribute, which depresses valuations. This results in a significant difference in payments for the two products and an increase in marginal profits for the Fair Trade product. However, overall profits are highest when the non-Fair Trade product is offered on its own.

These results warrant caution in attributing the success of the Fair Trade label to social preferences and/or self-identity concerns. This explanation is in line with findings

from several studies in the ethical consumption literature. Work by [Teisl et al. \(2002\)](#) finds that the increased demand for dolphin-free tuna in supermarkets came largely from consumers substituting away from non-labelled alternatives. In a controlled field experiment, [Hainmueller et al. \(2014\)](#) find a similar substitution effect for Fair Trade coffee. While the introduction of Fair Trade labelled coffee increased demand by 10%, this was offset by a reduction of 9% in demand for non-labelled coffee. Our results suggest that such a substitution effect might be driven by the joint evaluation format, where the introduction of a Fair Trade product decreases valuations for the non-certified alternative.

This conclusion is subject to an important caveat, namely that sorting poses a viable alternative explanation for our results. Contrary to the separate treatments, the joint evaluation format presents consumers with a choice between the Fair Trade and non-Fair Trade product. It is possible that payments for the non-Fair Trade product are lower in this condition, simply because individuals with a low valuation sort into buying this product, whereas consumers with a high valuation opt for the Fair Trade alternative. This would imply that valuations for the two products are actually consistent with what we obtained in the separate treatments, but that the possibility to sort into different products results in different averages in the joint condition. However, it is unclear a priori on what grounds consumers with high valuations are expected to sort into buying either product. For example, it is not necessarily obvious that consumers who value the Fair Trade attribute necessarily have a higher valuation of the overall product. Indeed, our data shows that a substantial portion of consumers (17.9%) who choose the Fair Trade product pay \$1 or less, implying that not all consumers who pay little self-select into the non-Fair Trade product. Still, it is important to address this possible confound to support our argument that valuations for the non-Fair Trade product decrease in the joint condition. We do this by running two additional treatments, which are scheduled to finish by November 2014. These treatments are identical to the two separate conditions, where consumers are presented either the Fair-Trade chocolate cupcake or the non-Fair Trade alternative. However, when we introduce customers to our product offering, we include the statement that “Usually we also have (Fair Trade) chocolate cupcakes, but these are unavailable today”. We thus make consumers aware that an alternative exists, but do not actually give the consumer the option to choose this product. If we still see a decrease in payments for the non-Fair Trade product in this condition, we can exclude sorting as a possible confound⁶.

⁶Preliminary results, based on 7 unique sales per condition, suggest an effect in this direction. When the non-Fair Trade product is offered on its own but consumers are made aware that a Fair Trade alternative is usually available, consumers pay an average of \$1.62 ($sd = 1.06$). For the Fair Trade treatment average payments are \$3.07 ($sd = 1.88$).

Appendix A

Appendix Chapter 1: Lying and Public Goods

A.1 Additional regression results

TABLE A.1: Tobit regression: the effect of beliefs on the contribution decision across treatments

<i>Dependent variable: Contribution</i>		
	Model 1	Model 2
Period number	-.180 *** [.053]	-.176 *** [.049]
Average beliefs ($t - 1$)	1.479 *** [.114]	1.914 *** [.334]
P-ACT/ANN	-4.312 ** [1.689]	3.062 [5.179]
P-ANN	-9.881 *** [1.662]	-1.989 [5.277]
NoP-ACT/ANN	-7.132 *** [1.986]	-2.617 [5.246]
NoP-ANN	-11.441 *** [2.181]	-1.759 [5.572]
P-ACT/ANN *Av. beliefs ($t - 1$)		-.469 [.378]
P-ANN *Av. beliefs ($t - 1$)		-.520 [.360]
NoP-ACT/ANN *Av. beliefs ($t - 1$)		.002 [.383]
NoP-ANN * Av. beliefs ($t - 1$)		-.734 * [.435]
Constant	2.041 [2.041]	-5.030 [4.836]
N	3864	3864
R^2	.196	.200
Left-censored	1225	1225
Right-censored	1031	1031

Robust standard errors are clustered at the group level and indicated in square brackets. The *, ** and *** indicate significant effects at the 10%, 5% and 1% level, respectively. Variable description: *Period number* : the period number; *Average beliefs (t-1)* : the sum of beliefs about the actual contribution of each of the group members in the previous period; *P-ACT/ANN* : dummy that takes the value 1 when the treatment is P-ACT/ANN and 0 otherwise; *P-ANN* : dummy that takes the value 1 when the treatment is P-ANN and 0 otherwise; *NoP-ACT/ANN* : dummy that takes the value 1 when the treatment is NoP-ACT/ANN and 0 otherwise; *NoP-ANN* : dummy that takes the value 1 when the treatment is NoP-ANN and 0 otherwise.

TABLE A.2: Tobit regression: the role of lies on punishment assigned

<i>Dependent variable: Punishment assigned</i>		
	P-ACT/ANN	P-ANN
Period number	-.038 [.052]	-.138 ** [.066]
Deviation	.296 *** [.086]	.322 *** [.112]
Perceived liar (dummy)	.665 [.886]	.317 [1.230]
Size of the lie	-.130 [.080]	-.223 [.140]
Constant	-6.256 *** [1.481]	-6.827 *** [1.392]
N	2298	2512
R^2	.062	.060
Left-censored	1779	2331
Right-censored	17	1

Robust standard errors are clustered at the group level and indicated in square brackets. The *, ** and *** indicate significant effects at the 10%, 5% and 1% level, respectively. Variable description: *Deviation* : how much the target's contribution deviates from the social optimum of 20 points; *Perceived liar* : a dummy with a value of 1 when the target's contribution is believed not to coincide with the made announcement; and 0 when the target is believed to be honest; *Size of the lie* : the discrepancy between the target's displayed contribution and the subject's beliefs about the actual contribution of the target.

A.2 Instructions public good game (P-ACT/ANN treatment)

The instructions below are for the P-ACT/ANN treatment (announcements are displayed 50% probability and participants can administer costly peer punishment) and are translated from the original German. Instructions for the other treatments (in English and German) are available upon request.

General instructions

You are now participating in a scientific experiment. In this experiment you can earn money depending on your own decisions and those of other participants. How you can earn money will be made clear to you in the following instructions. Please read these carefully.

During the experiment communicating with other participants is not permitted. Not following these rules results in the termination of the experiment and all payments. When you have questions, please raise your hand out of the cabin. A member of the student team will come to you and answer your question in private.

During the experiment your earnings are calculated in points. The total number of points you earn will be converted to Euro at the following exchange rate:

$$50 \text{ points} = \text{€} 1$$

The €2.50, which you received for showing up to this experiment, are converted into points. This means you start the experiment with 125 points.

The converted amount in Euros are paid to you in cash at the end of the experiment. The payment will happen anonymously, meaning that no participant will know how much other participants were paid. All decisions during the experiment are also made anonymously, meaning that no participant will find out the identity of those behind the decisions made.

The Experiment

The experiment lasts a total of 15 periods. In each of these 15 periods, you are in a group with three other participants. The group composition stays the same during the whole duration of the experiment. You thus play with the same three participants in a group during all 15 periods. Every period is divided into two phases.

Contribution Phase

At the start of each period, each participant receives 20 points. We refer to these points as ‘endowment’. In the first phase you need to decide how many of these 20 points you want to contribute to a group project and how many you would like to keep for yourself. Every point that you keep for yourself increases your earnings by 1 point. Every point that you contribute to the project increases your own earnings by $(0.4 * 1 =)$ 0.4 points and also raises the earnings of each of your group members by 0.4 points. Likewise, every point that other group members contribute to the group project increases your earnings by 0.4 points. Imagine that all group members together contributed 60 points. In this case each group member receives $(0.4 * 60 =)$ 24 points from the project. If the sum of all contributions is 9 points, then each group member receives $(0.4 * 9 =)$ 3.6 points from the project. Earnings are determined in the same way for every group member. This means that every group member receives the same share from the group project.

Your earnings in this phase can be calculated using the following formula:

$$\text{Earnings} = (\text{Endowment} - \text{Your contribution}) + (0,4 * \text{Sum of all contributions})$$

Imagine that every group member contributed 10 points to the project. This means that you keep $(20 - 10 =)$ 10 points for yourself. The sum of all contributions in the group is $(10 + 10 + 10 + 10 =)$ 40 points. As such you earn $(0.4 * 40 =)$ 16 points from the project. Your total earnings in this period are $(10 + 16 =)$ 26 points.

After each period you receive information about the contributions of the other group members. However, in this experiment you cannot observe the contribution decision of your fellow group members. Likewise, other group members cannot observe your contribution decision. For this reason, after each group member has made their contribution decision, you can announce your contribution to the others in your group. The amount that you decide to announce is at your discretion.

After all group members have made their decisions, you receive information about the contributions of your fellow group members in the previous period. The computer randomly selects either your actual contribution or your announcement is displayed on the feedback screen. With **50% probability** your actual contribution will be displayed and with **50% probability** your announced contribution will be displayed. The displayed contribution is determined in this way for each group member. As such it is possible in

any given round that, for example, the announced contribution of the first group member and the actual contribution of the second and third group members is displayed. For this reason, we also ask you for your beliefs about the actual contribution of each group member.

The table below shows a screenshot of the feedback screen in a given period:

	Contribution	Estimated contribution (actual/announced)	Estimated actual contribution	Reduction points
You	—			
Group member	—	<input type="radio"/> Actual <input type="radio"/> Announced	<input type="text"/>	<input type="text"/>
Group member	—	<input type="radio"/> Actual <input type="radio"/> Announced	<input type="text"/>	<input type="text"/>
Group member	—	<input type="radio"/> Actual <input type="radio"/> Announced	<input type="text"/>	<input type="text"/>

Please note that the order in which group members are displayed is reshuffled each round.

To summarize, in the contribution phase you make two decisions about your contribution: the first decision about how many points you contribute to the project. And second, a decision about how many points you announce to your fellow group members about your contribution. All group members make their decisions simultaneously. This means that no one is informed about the decisions of the others before making his or her own decision. On the feedback screen you will be informed about the actual or announced contributions of your fellow group members and will be asked to state your beliefs about their actual contribution. Please note: even though your announced contribution can be displayed as feedback, only the actual contribution decision influences your earnings.

Reduction Phase

In the last column of the above table on the feedback screen you have the possibility to assign reduction points to your fellow group members. This will be explained below.

In the reduction phase you receive 10 additional points. Every group member now has to decide, whether to reduce the earnings of the others by assigning reduction points or to leave earnings unchanged. Your fellow group members can thus have the possibility to reduce your earnings if they want. In this phase, all decisions are made simultaneously. Every reduction point that you assign to a group member reduces the earnings of this participant by **3 points**. Please note that – even in the case of a loss - you will receive at least €2.50 for your participation at the end of the experiment.

When you do not want to change the earnings of your fellow group members, then you do not assign any reduction points. The points that you do not assign to reductions are added to your personal earnings. For example, when you assign two reduction points, a total of $(10 - 2 =) 8$ points are added to your personal earnings. In other words, assigning reduction points to your fellow group members is costly to you.

Imagine that you assign 3 reduction points to group member 1 (this reduces the earnings of group member 1 by 9 points) and 0 reduction points to group member 2 and 3 (this does not change the earnings of group member 2 and 3). After all group members have made their decisions, you learn that the other group members assigned you a total of 2 reduction points. This means, that your personal earnings are reduced by $(2 * 3 =) 6$ points, while the $(10 - 3 =) 7$ reduction points you did not assign are added to your personal earnings.

When you assign reduction points, you need to indicate to which group member you assign these. Because the announced contributions of the group members are anonymous and displayed in random order, you can indicate the number of points you want to assign in the corresponding row on the feedback screen. Given that you have a total of 10 reduction points, the maximum you can assign is 10 points.

Please note that you do not learn the individual reduction decisions of your fellow group members. This means that you receive information about how many reduction points you received in total, but not how many points each group member separately assigned to you. Further, you only learn about how many reduction points you received and not how many points other group members received. After you receive this information, the next period begins.

Earnings formula and example

The proceedings in each period are as follows:

+20 (Endowment)		+10 (Reduction-endowment)
Contribution	+ Announced/ Actual contribution	+ Information about contributions and the possibility to assign reduction points
		- Information about your assigned and received reduction points
Contribution Phase		Reduction Phase

Your earnings in each period can be calculated using the following formula. When you have questions about this, please notify us.

$$\text{Earnings} = (\text{Endowment} - \text{Your contribution}) + (0,4 * \text{Sum of all contributions}) + (10 - \text{Reduction points assigned by you}) - (3 * \text{Total reduction points assigned to you})$$

This formula shows that your earnings consist of four parts:

1. The points that you decide to keep for yourself: (Endowment – Your contribution)
2. The points from the project, which is 40% of the sum of all contributions.
3. The points that you do not assign as reduction points: (10 – Reduction points assigned by you).
4. The reduction points assigned to you multiplied by a factor 3.

Example (the numbers in this example were determined randomly)

Imagine that you and every other group member contributed 5 points to the project. This means that you keep $(20 - 5 =)$ 15 points for yourself. The sum of all contributions in the group is $(5 + 5 + 5 + 5 =)$ 20 points. Therefore you receive $(0.4 * 20 =)$ 8 points from the project. In the reduction phase you decide to assign 1 reduction point to another group member, which reduces the earnings of this participant by $(1 * 3 =)$ 3 points. From the 10 reduction points that you could assign, you have $(10 - 1 =)$ 9 points left over. These are added to your personal earnings. You receive 2 reduction points from the other group members, which reduces your earnings by $(2*3 =)$ 6 points.

Your earnings in this period are:

15	+	8	+	9	-	6	=	26 points
(Endowment - Your contribu- tion)		(0.4*Sum of all contributions)		(10 - Reduction points assigned by you)		(3*total reduc- tion points as- signed to you)		

Important: Even though your announced contribution is shown on the feedback screen, only your actual contribution decision influences your payoffs.

When you have read and understood these instructions, please complete the practice questions on your screen. These are meant to familiarize you with the decision procedures. When all participants have answered all practice questions correctly, the experiment begins.

Thank you for participating.

Appendix B

Appendix Chapter 2: Fooling the Nice Guys

B.1 The 32 allocation decision tasks of the ring measure (Liebrand, 1984)

	Option A		Option B	
	You	Other	You	Other
1	0	+500	+304	+397
2	+304	+397	+354	+354
3	+354	+354	+397	+304
4	+397	+304	+433	+250
5	+433	+250	+462	+191
6	+462	+191	+483	+129
7	+483	+129	+496	+65
8	+496	+65	+500	0
9	+500	0	+496	-65
10	+496	-65	+483	-129
11	+483	-129	+462	-191
12	+462	-191	+433	-250
13	+433	-250	+397	-304
14	+397	-304	+354	-354
15	+354	-354	+304	-397
16	+304	-397	0	-500
17	0	-500	-304	-397
18	-304	-397	-354	-354
19	-354	-354	-397	-304
20	-397	-304	-433	-250
21	-433	-250	-462	-191
22	-462	-191	-483	-129
23	-483	-129	-496	-65
24	-496	-65	-500	0
25	-500	0	-496	+65
26	-496	+65	-483	+129
27	-483	+129	-462	+191
28	-462	+191	-433	+250
29	-433	+250	-397	+304
30	-397	+304	-354	+354
31	-354	+354	-304	+397
32	-304	+397	0	+500

TABLE B.1: The 32 allocation decision tasks comprising the ring measure (Liebrand, 1984)

B.2 SVO angles and corresponding classifications based on our 25% and [Liebrand and McClintock \(1988\)](#)

SVO degree	Frequency	Classification	
		L& McC	25%
-3.73	1	Individualistic	Low
-1.33	1	Individualistic	Low
-0.81	1	Individualistic	Low
0.00	24	Individualistic	Low
0.36	1	Individualistic	NA
2.32	1	Individualistic	NA
3.73	6	Individualistic	NA
7.43	1	Individualistic	NA
7.48	1	Individualistic	NA
10.66	1	Individualistic	NA
11.18	1	Individualistic	NA
11.35	1	Individualistic	NA
14.54	1	Individualistic	NA
14.58	1	Individualistic	NA
14.96	1	Individualistic	High
18.55	1	Individualistic	High
18.72	1	Individualistic	High
22.46	1	Individualistic	High
22.48	1	Individualistic	High
22.51	1	Cooperative	High
23.44	1	Cooperative	High
24.07	1	Cooperative	High
26.23	1	Cooperative	High
26.28	1	Cooperative	High
37.51	1	Cooperative	High
55.33	1	Cooperative	High
57.96	1	Cooperative	High

N=56. Mean=8.74. Classification: individualistic (n=48), cooperative (n=8), low (n=27), high (n=14). An individual is classified as individualistic if her SVO angle lies between -22.5 and 22.5 degrees. Subjects in the range of 22.5 and 65.7 degrees are classified as cooperative. Those with degrees lower than -22.5 are classified as competitive, while those higher than 65.7 degrees are altruists. The table below includes the SVO degree angles of all subjects as well as the corresponding classification according to [Liebrand and McClintock \(1988\)](#) and our 25% categorization (25%).

TABLE B.2: SVO angles of all experimental subjects and corresponding classifications based on our 25% and [Liebrand and McClintock \(1988\)](#)

B.3 Tobit regression: Drivers of the contribution decision

As stated in the main text, the contribution and average beliefs in the previous period are a significant predictor of the contribution in the current period. Subjects also contribute more when they are a high type compared to a low type. Importantly, the interaction term between type and the lagged measure of average beliefs is not significant. This indicates that types do not differ in the degree to which beliefs inform their contribution decision.

<i>Dependent variable: Contribution</i>		
	Model 1	Model 2
Period number	-.091 [.122]	-.066 [.129]
Contribution ($t-1$)	1.002 *** [.185]	1.002 *** [.187]
Av. beliefs ($t-1$)	.527 *** [.103]	.624 *** [.145]
High type	4.201 *** [.999]	6.855 *** [2.673]
High type * Av. beliefs ($t-1$)		-.239 [.200]
Constant	-8.894 *** [2.850]	-9.884 *** [3.287]
Controls	YES	YES
N	574	574
Pseudo R^2 (overall)	.198	.199
N left-censored	243	243
N right-censored	96	96

Robust standard errors are clustered at the group level and indicated in square brackets. The *** indicate significant effects at the 1% level. Dependent variable: *Contribution*: the subject's contribution to the public good. Independent variables: *Period number*: the period number; *Contribution ($t-1$)*: lagged measure of contribution; *Av. Beliefs ($t-1$)*: lagged measure of average beliefs. *High type*: binary variable, where 1 indicates that the subject is classified as a high type and 0 when she is a low type; *High type * Av. Beliefs ($t-1$)*: captures the interaction between the subject's type and lagged average beliefs; *Controls*: include age, gender and field of study. Gender is a binary variable where 0 indicates male and 1 female; Field of study is a binary variable where 1 is assigned to those subjects studying economics or business and 0 otherwise. None of these is significant.

TABLE B.3: Tobit regression: Drivers of the contribution decision

B.4 Robustness checks

B.4.1 Analysis according to classification of cooperative and individualistic types

We can repeat our analysis using an alternative classification according to cooperative and individualistic types (Liebrand and McClintock, 1988). Even though the differences between types become more pronounced than those obtained through the 25%-classification, we fail to reach strong significant results because of a low number of independent observations. The 8 subjects classified as cooperative types are spread over 6 different groups, which gives us 6 independent observations for the non-parametric tests. Compared to those classified as individualistic, cooperative types contribute more and discount less, and these differences are weakly significant (WSR, $p = 0.074$ for both). In a regression on belief formation using this classification, the interaction term between type and announcement remains positive and strongly significant ($p = 0.01$).

	Average contribution	Average belief	Average adjustment	Average punishment
Overall (N=56)	6.79 (7.59)	8.74 (7.51)	8.40 (8.09)	0.14 (0.64)
<i>Individualistic</i> (n=48)	6.00 (7.36)	7.74 (7.24)	9.16 (8.14)	0.16 (0.68)
<i>Cooperative</i> (n=8)	11.50 (7.24)	14.71 (6.18)	3.85 (5.45)	0.03 (0.20)

Standard deviations are show in brackets.

TABLE B.4: General descriptive statistics for alternative classification

B.4.2 SVO degree angle as an independent variable in the belief formation regression

Our main results hold when we replace the type classification in the belief formation regression with the SVO degree angle. In Model 2, average announcement is significant and has a negative coefficient. Combining this with the interaction term, announcements have a negative impact on beliefs for those with a SVO degree angle of 0 (ie. completely individualistic). The impact of average announcement on beliefs becomes positive for subjects with a SVO degree angle of approximately 10 and higher.

<i>Dependent variable: Average beliefs</i>		
	Model 1	Model 2
Period number	-.023 [.045]	-.028 [.046]
Av. beliefs ($t-1$)	1.224 *** [.107]	1.196 *** [.114]
Av. announcement	-.036 [.079]	-.201 ** [.099]
SVO degree angle	.118 *** [.037]	-2.71 ** [.126]
SVO degree angle * Av. announcement		.022 *** [.008]
Constant	-2.717 ** [1.285]	.424 [1.749]
N	784	784
Pseudo R^2 (overall)	.227	.230
N left-censored	209	209
N right-censored	136	136

Robust standard errors are clustered at the group level and indicated in square brackets. The ** and *** indicate significant effects at the 5% and 1% level, respectively. Dependent variable: *Average beliefs*: the average of the three belief measures (one for each announcement of the other group members) in period t . Independent variables: *Period number*: the period number; *Av. Beliefs ($t-1$)*: lagged measure of average beliefs; *Av. announcement*: average of the three announcements received by the subject on the feedback screen in period t ; *SVO degree angle*: the subject's SVO degree angle; *SVO degree angle * Av. Announcement*: the interaction between the subject's SVO degree angle and the average announcement received.

TABLE B.5: Tobit regression: the effect of the SVO degree angle on belief formation

B.5 Instructions public good game (ANNOUNCE treatment)

The instructions below are for the ANNOUNCE treatment and are translated from the original German. Instructions for the other treatments (in English and German) are available upon request.

General instructions

You are now participating in a scientific experiment. In this experiment you can earn money depending on your own decisions and those of other participants. How you can earn money will be made clear to you in the following instructions. Please read these carefully.

During the experiment communicating with other participants is not permitted. Not following these rules results in the termination of the experiment and all payments. When you have questions, please raise your hand out of the cabin. A member of the student team will come to you and answer your question in private.

During the experiment your earnings are calculated in points. The total number of points you earn will be converted to Euro at the following exchange rate:

$$50 \text{ points} = \text{€}1$$

The €2.50, which you received for showing up to this experiment, are converted into points. This means you start the experiment with 125 points.

The converted amount in Euros are paid to you in cash at the end of the experiment. The payment will happen anonymously, meaning that no participant will know how much other participants were paid. All decisions during the experiment are also made anonymously, meaning that no participant will find out the identity of those behind the decisions made.

The Experiment

The experiment lasts a total of 15 periods. In each of these 15 periods, you are in a group with three other participants. The group composition stays the same during the whole duration of the experiment. You thus play with the same three participants in a group during all 15 periods. Every period is divided into two phases.

Contribution Phase

At the start of each period, each participant receives 20 points. We refer to these points as ‘endowment’. In the first phase you need to decide how many of these 20 points you want to contribute to a group project and how many you would like to keep for yourself. Every point that you keep for yourself increases your earnings by 1 point. Every point that you contribute to the project increases your own earnings by $(0.4 * 1 =)$ 0.4 points and also raises the earnings of each of your group members by 0.4 points. Likewise, every point that other group members contribute to the group project increases your earnings by 0.4 points. Imagine that all group members together contributed 60 points. In this case each group member receives $(0.4 * 60 =)$ 24 points from the project. If the sum of all contributions is 9 points, then each group member receives $(0.4 * 9 =)$ 3.6 points from the project. Earnings are determined in the same way for every group member. This means that every group member receives the same share from the group project.

Your earnings in this phase can be calculated using the following formula:

$$\text{Earnings} = (\text{Endowment} - \text{Your contribution}) + (0,4 * \text{Sum of all contributions})$$




Imagine that every group member contributed 10 points to the project. This means that you keep $(20 - 10 =)$ 10 points for yourself. The sum of all contributions in the group is $(10 + 10 + 10 + 10 =)$ 40 points. As such you earn $(0.4 * 40 =)$ 16 points from the project. Your total earnings in this period are $(10 + 16 =)$ 26 points.

After each period you receive information about the contributions of the other group members. However, in this experiment you cannot observe the contribution decision of your fellow group members. Likewise, other group members cannot observe your contribution decision. For this reason, after each group member has made their contribution decision, you can announce your contribution to the others in your group. The amount that you decide to announce is at your discretion.

After all group members have made their decisions, you receive information about the announced contributions of your fellow group members in the previous period. In addition, we ask for your beliefs about the underlying actual contribution of each of your group members.

The table below shows a screenshot of the feedback screen in a given period:

Please note that the order in which group members are displayed is reshuffled each round.

	Announced contribution	Estimated actual contribution
Group member	—	
Group member	—	
Group member	—	

To summarize, in the contribution phase you make two decisions about your contribution: the first decision about how many points you contribute to the project. And second, a decision about how many points you announce to your fellow group members about your contribution. All group members make their decisions simultaneously. This means that no one is informed about the decisions of the others before making his or her own decision. Please note: even though your announced contribution will be displayed as feedback, only the actual contribution decision influences your earnings.

Reduction Phase

In the last column of the above table on the feedback screen you have the possibility to assign reduction points to your fellow group members. This will be explained below.

In the reduction phase you receive 10 additional points. Every group member now has to decide, whether to reduce the earnings of the others by assigning reduction points or to leave earnings unchanged. Your fellow group members can thus have the possibility to reduce your earnings if they want. In this phase, all decisions are made simultaneously. Every reduction point that you assign to a group member reduces the earnings of this participant by **3 points**. Please note that – even in the case of a loss - you will receive at least €2.50 for your participation at the end of the experiment.

When you do not want to change the earnings of your fellow group members, then you do not assign any reduction points. The points that you do not assign to reductions are added to your personal earnings. For example, when you assign two reduction points, a total of $(10 - 2 =) 8$ points are added to your personal earnings. In other words, assigning reduction points to your fellow group members is costly to you.

Imagine that you assign 3 reduction points to group member 1 (this reduces the earnings of group member 1 by 9 points) and 0 reduction points to group member 2 and 3 (this does not change the earnings of group member 2 and 3). After all group members have

made their decisions, you learn that the other group members assigned you a total of 2 reduction points. This means, that your personal earnings are reduced by $(2 * 3 =)$ 6 points, while the $(10 - 3 =)$ 7 reduction points you did not assign are added to your personal earnings.

When you assign reduction points, you need to indicate to which group member you assign these. Because the announced contributions of the group members are anonymous and displayed in random order, you can indicate the number of points you want to assign in the corresponding row on the feedback screen. Given that you have a total of 10 reduction points, the maximum you can assign is 10 points.

Please note that you do not learn the individual reduction decisions of your fellow group members. This means that you receive information about how many reduction points you received in total, but not how many points each group member separately assigned to you. Further, you only learn about how many reduction points you received and not how many points other group members received. After you receive this information, the next period begins.

Earnings formula and example

The proceedings in each period are as follows:

+20 (Endowment)	+10 (Reduction-endowment)
Contribution + Announced contribution	+ Information about contributions and the possibility to assign reduction points - Information about your assigned and received reduction points
Contribution Phase	Reduction Phase

Your earnings in each period can be calculated using the following formula. When you have questions about this, please notify us.

$$\text{Earnings} = (\text{Endowment} - \text{Your contribution}) + (0,4 * \text{Sum of all contributions}) + (10 - \text{Reduction points assigned by you}) - (3 * \text{Total reduction points assigned to you})$$

This formula shows that your earnings consist of four parts:

1. The points that you decide to keep for yourself: (Endowment – Your contribution)
2. The points from the project, which is 40% of the sum of all contributions.
3. The points that you do not assign as reduction points: (10 – Reduction points assigned by you).
4. The reduction points assigned to you multiplied by a factor 3.

Example (the numbers in this example were determined randomly)

Imagine that you and every other group member contributed 5 points to the project. This means that you keep $(20 - 5 =)$ 15 points for yourself. The sum of all contributions in the group is $(5 + 5 + 5 + 5 =)$ 20 points. Therefore you receive $(0.4 * 20 =)$ 8 points from the project. In the reduction phase you decide to assign 1 reduction point to another group member, which reduces the earnings of this participant by $(1 * 3 =)$ 3 points. From the 10 reduction points that you could assign, you have $(10 - 1 =)$ 9 points left over. These are added to your personal earnings. You receive 2 reduction points from the other group members, which reduces your earnings by $(2*3 =)$ 6 points.

Your earnings in this period are:

$$\begin{array}{rcccccc}
 15 & + & 8 & + & 9 & - & 6 & = & 26 \text{ points} \\
 \text{(Endowment -} & & \text{(0.4*Sum of all} & & \text{(10 - Reduction} & & \text{(3*total reduc-} & & \\
 \text{Your contribu-} & & \text{contributions)} & & \text{points assigned} & & \text{tion points as-} & & \\
 \text{tion)} & & & & \text{by you)} & & \text{signed to you)} & &
 \end{array}$$

Important: Even though your announced contribution is shown on the feedback screen, only your actual contribution decision influences your payoffs.

When you have read and understood these instructions, please complete the practice questions on your screen. These are meant to familiarize you with the decision procedures. When all participants have answered all practice questions correctly, the experiment begins.

Thank you for participating.

B.6 Instructions ring measure, translated from the original German

In the upcoming task we ask you to make several allocation decisions. For this you are paired with another randomly selected participant. Through the allocation decisions you and this participant can earn points. In each allocation you must repeatedly choose

between two allocations X and Y (for example, allocation X: 10 points for you and 12 points for the other or allocation Y: 8 points for you and 20 points for the other). The points that you allocate to yourself are converted to Euros at an exchange rate of 500 points = €1, and paid to you at the end of the experiment. As a randomly selected participant is connected to you for this task, likewise are you paired with a randomly selected participant. Through the allocation decisions of this participant, he or she allocates points to you. This means that the participant to whom you allocate points is a different person from the one who allocates points to you. The points that this participant allocates to you are added to your earnings and are also paid to you at the end of the experiment at an exchange rate of 500 points = €1.

Appendix C

Appendix Chapter 3: The indirect effect of monetary incentives on deception

C.1 Additional regression results

TABLE C.1: OLS regressions: the effect of average and relative performance in the work task on the message sent

<i>Dependent variable: Message sent</i>				
	No Feedback		Feedback	
	Model 1a	Model 1b	Model 2a	Model 2b
Team	1.672 [1.471]	1.616 [1.459]	-2.098 [1.514]	-1.583 [1.533]
Tournament	4.009 *** [1.487]	3.870 ** [1.485]	-.422 [1.512]	-1.000 [1.507]
Relative performance	-.082 [.320]		-.788 ** [.318]	
Average performance		.306 [.336]		.516 ** [.252]
Constant	18.392 *** [1.909]	14.822 *** [4.118]	21.016 *** [1.698]	13.891 *** [2.813]
Controls	YES	YES	YES	YES
N	63	63	71	71
R ²	.200	.210	.141	.115

Standard errors are shown in square brackets. The ** and *** indicate significant effects at the 5% and 1% level, respectively. Variable description: *Team*: dummy for the team incentive treatment; *Tournament*: dummy for the tournament incentive treatment; *Relative performance*: the average performance difference between the sender and receiver over the five rounds of the work task; *Average performance*: the average performance of the sender over the five rounds of the work task; *Controls*: gender and a dummy for when the subject majors in economics or business. Neither of these is significant.

TABLE C.2: Probit regression: the effect of relative performance on honesty across incentive conditions

<i>Dependent variable: Honest message</i>		
	Feedback	
	Model 1a	Model 1b
Team	.213 [.118]	.400 [.210]
Tournament	.067 [.122]	.204 [.236]
Relative performance	.083 *** [.024]	.137 ** [.067]
Team * Rel. performance		-.106 [.091]
Tourn. * Rel. performance		-.059 [.081]
Controls	YES	YES
N	71	71
Pseudo R ²	.193	.210
Log likelihood	-34.783	-34.043

Average marginal effects reported. Standard errors are shown in square brackets. The ** and *** indicate significant effects at the 5% and 1% level, respectively. Variable description: *Team*: dummy for the team incentive treatment; *Tournament*: dummy for the tournament incentive treatment; *Relative performance*: the average performance difference between the sender and receiver over the five rounds of the work task; *Team/Tourn. * Rel. performance*: the interaction term between the treatment dummy and relative performance; *Controls*: gender and a dummy for when the subject majors in economics or business. Neither of these is significant.

TABLE C.3: Probit regression: the effect of average earnings on honesty across incentive conditions

<i>Dependent variable: Honest message</i>		
	No Feedback	Feedback
Team	-.081 [.126]	.179 [.147]
Tournament	-.264 ** [.118]	.174 [.151]
Average earnings	-.001 [.000]	-.000 [.000]
Controls	YES	YES
N	63	71
Pseudo R ²	.158	.085
Log likelihood	-33.761	-39.435

Average marginal effects reported. Standard errors are shown in square brackets. The ** indicate significant effects at the 5% level. Variable description: *Team*: dummy for the team incentive treatment; *Tournament*: dummy for the tournament incentive treatment; *Average earnings*: the average earnings of the sender in the five rounds of the work task; *Controls*: gender and a dummy for when the subject majors in economics or business. Neither of these is significant.

C.2 Experimental instructions

The instructions below are for the tournament treatment with feedback and were translated from the original German. In the experiment, the instructions for part 2 were only handed out to participants once part 1 was concluded. Instructions for the other treatments (in English and German) as well as the booklet with Latin text are available upon request.

C.2.1 Instructions

Please read these instructions carefully. You can earn money depending on the decisions made by you and the other participants in this experiment. When you have questions, please raise your hand and a research assistant will answer your question privately. Throughout the entire experiment you make decisions anonymously: no participant knows the identity of the other the participants.

In this experiment you earn points. When the experiment is over you will receive €1 for every 100 points earned. This payment and the additional €2.5 that you receive for arriving to the experiment on time are paid immediately and anonymously after the experiment ends.

The experiment consists of two parts.

C.2.2 Part 1

Pairs and periods

- You are randomly matched with another participant in this experiment. This will not change in the course of the experiment.
- This part consists of five rounds of three minutes each.

Task

- In each round you are given a series of tasks with directions to find a letter in the booklet in front of you. The directions indicate the page, line number, word and

position where the letter can be found. All letters of the alphabet are possible, but spaces and punctuation marks, such as commas and full stops, should not be counted. Note that the line number is indicated at the left margin of the page.

- Type the found letter in the text box on the screen and press 'OK' to move to the next task with directions for a new letter.

Information

- A task is correctly solved when the right letter is identified. However, you are only informed about the number of correctly solved tasks at the end of each period. This means that while the round is in progress, you are not informed whether the letter you entered is correct or not.
- A timer in the top right corner of the screen counts down from three minutes. You can continue working for as long as you have time remaining. When the three minutes are over, you move to the feedback screen. Here you learn how many tasks you solved correctly as well as how many were solved by the participant matched to you.

Payment

- Payment depends on your performance as well as the performance of the other participant.
- In each round, the participant who solved the most tasks correctly receives 1000 points. The other participant receives 0 points. Note that a task is only correctly solved when the correct letter has been identified.
- After all five rounds are finished, one round will be randomly selected for payment. However, you will only be informed which round was selected when the experiment is concluded.
- Example: Assume that round 4 is randomly selected for payment. In this round, you correctly solved 5 tasks and the other participant solved 3. Since your performance is higher, you earn 1000 points and the other participant earns 0 points.

Practice round

- You start with a practice round to familiarize yourself with the task of part 1. Any earned points in this round do not count towards your final payoffs.

- The practice round lasts 3 minutes. If you wish, you can finish the round early by clicking the ‘SKIP’ button. However, you will still need to wait for all other participants to finish their practice round before the experiment continues. This ‘SKIP’ option is not available during the actual five rounds of the experiment.

C.2.3 Part 2

Pairs

- In part 2 you are matched with another participant. This is the same person you were matched with in part 1.
- There are two roles: player A and player B. At the start of part 2 one participant in the pair is randomly assigned the role of player A. The other participant is assigned the role of player B.

Decision

- In a previous session, participants completed a task similar to part 1 of this experiment. In this previous session, performance levels ranged from 10 to 25 completed tasks. We randomly selected a performance level from one participant in a randomly selected round from this experiment.
- Player A will be informed about this performance level. Player B will not know it.
- Player A is asked to send a message to player B. Player A can freely choose from the following messages:

Message: “The performance level was 10”	Message: “The performance level was 18”
Message: “The performance level was 11”	Message: “The performance level was 19”
Message: “The performance level was 12”	Message: “The performance level was 20”
Message: “The performance level was 13”	Message: “The performance level was 21”
Message: “The performance level was 14”	Message: “The performance level was 22”
Message: “The performance level was 15”	Message: “The performance level was 23”
Message: “The performance level was 16”	Message: “The performance level was 24”
Message: “The performance level was 17”	Message: “The performance level was 25”

- Player B receives the message sent by player A and then chooses a number between 10 and 25.

Payment

- There are two different payment options, option X and Y. Each option allocates a certain number of points to player A and a certain number of points to player B. In addition, the message that player A chooses to send determines the point allocation of payment option Y.
- The decision of player B determines which payment option is implemented.
 - If player B chooses a number that contains the actual performance level, player A and B are paid according to option X.
 - If player B chooses a number that differs from the actual performance level, player A and B are paid according to option Y.

Information

- Player A is informed about the actual performance level from the randomly selected participant from the previous experiment. Player B is not informed about this.
- Player A is informed about the particular point values of payment options X and Y. Player B does not receive information about these values.
- Player A and B are informed about their earnings from part 2 at the end of the experiment.

C.2.4 Private instructions for player A in part 2

Important: Additional instructions player A!

In a previous session, participants completed a task similar to part 1 of this experiment. In this previous session, performance levels ranged from 10 to 25 correctly solved tasks. We randomly selected a performance level from one participant in a randomly selected round from this participant. The actual performance of the participant from the previous session in the experiment is '12'. The other participant does not know that the performance level is 12.

We now ask you to send a message to player B. Depending on the number chosen by the receiver, payoff allocation X or Y is implemented, which allocates a specific number of points to you and the other participant. Only you are informed about the point values of this payment option. Player B is not informed about this. The message that you choose will be sent to player B. Player B then chooses a number and this determines the payoffs for both you and the other participant.

If player B chooses a number that contains the actual performance level, then you and player B are paid according to option X. This means that you earn 200 points and player B also earns 200 points.

If player B chooses a number than differs from the actual performance level, then you and player B are paid according to option Y. The specific point values for Y are as follows:

- If you send message “*the performance level was 10*”, and player B chooses a number different from 12, then you receive **200 points** and player B receives **200 points**.
- If you send message “*the performance level was 11*”, and player B chooses a number different from 12, then you receive **200 points** and player B receives **200 points**.
- If you send message “*the performance level was 12*”, and player B chooses a number different from 12, then you receive **200 points** and player B receives **200 points**.
- If you send message “*the performance level was 13*”, and player B chooses a number different from 12, then you receive **210 points** and player B receives **190 points**.
- If you send message “*the performance level was 14*”, and player B chooses a number different from 12, then you receive **220 points** and player B receives **180 points**.
- If you send message “*the performance level was 15*”, and player B chooses a number different from 12, then you receive **230 points** and player B receives **170 points**.
- If you send message “*the performance level was 16*”, and player B chooses a number different from 12, then you receive **240 points** and player B receives **160 points**.
- If you send message “*the performance level was 17*”, and player B chooses a number different from 12, then you receive **250 points** and player B receives **150 points**.
- If you send message “*the performance level was 18*”, and player B chooses a number different from 12, then you receive **260 points** and player B receives **140 points**.

- If you send message “*the performance level was 19*”, and player B chooses a number different from 12, then you receive **270 points** and player B receives **130 points**.
- If you send message “*the performance level was 20*”, and player B chooses a number different from 12, then you receive **280 points** and player B receives **120 points**.
- If you send message “*the performance level was 21*”, and player B chooses a number different from 12, then you receive **290 points** and player B receives **110 points**.
- If you send message “*the performance level was 22*”, and player B chooses a number different from 12, then you receive **300 points** and player B receives **100 points**.
- If you send message “*the performance level was 23*”, and player B chooses a number different from 12, then you receive **300 points** and player B receives **90 points**.
- If you send message “*the performance level was 24*”, and player B chooses a number different from 12, then you receive **300 points** and player B receives **80 points**.
- If you send message “*the performance level was 25*”, and player B chooses a number different from 12, then you receive **300 points** and player B receives **70 points**.

Please answer the following questions to ensure your understanding of the above information.

Question 1: If you send the message “The performance level was 12” and player B chooses number 12, then the payoffs are:

- (a) You: 0 points; Player B: 0 points
- (b) You: 200 points; Player B: 200 points
- (c) You: 230 points; Player B: 170 points
- (d) You: 170 points; Player B: 230 points

Question 2: If you send the message “The performance level was 15” and player B chooses number 15, then the payoffs are:

- (a) You: 0 points; Player B: 0 points
- (b) You: 200 points; Player B: 200 points
- (c) You: 230 points; Player B: 170 points
- (d) You: 170 points; Player B: 230 points

Appendix D

Appendix Chapter 4: Are social investments rewarded?

D.1 Photos of stand materials



FIGURE D.1: Stand display, separate Fair Trade condition



FIGURE D.2: Stand display, separate regular condition



FIGURE D.3: Detail of product and sign, separate regular condition



FIGURE D.4: The stand from a distance with banner and research assistant

D.2 Randomization

TABLE D.1: Overview of dates and randomization for each market

Market	Dates	Randomization
North Park	13/02, Thursday	R → FT → J → J → R
	20/02, Thursday	
Pacific Beach	18/02, Tuesday	R → FT → J → J → R →
	25/02, Tuesday	FT → R → FT
	04/03, Tuesday	
La Jolla	23/02, Sunday	R → J → FT → J → R →
	16/03, Sunday	FT → FT → J
	30/03, Sunday	
	27/04, Sunday	
University Heights	19/04, Saturday	FT → J
Hillcrest	04/05, Sunday	J → FT → R → J →
	11/04, Sunday	R → FT → R → J → FTn →
	25/05, Sunday	Rn → J
	01/06, Sunday	

Abbreviations for conditions are as follows. *R*: separate regular; *FT*: separate Fair Trade; *J*: joint; *FTn*: joint, but regular unavailable (new treatment); *Rn*: joint, but Fair Trade unavailable (new treatment).

Table D.1 displays the order of treatments we ran at each market. The first two columns list the respective market and the dates we operated a stand there. In advance, we generated a random series of the treatments for each market. In Pacific Beach, for example, this series turned out to be ‘Separate regular’, followed by ‘Separate Fair Trade’, then ‘Joint’, then ‘Joint’ again, then ‘Separate regular’, then ‘Separate Fair Trade’ and so forth. We changed condition every 5 sales on the first day at the market, and then every 7 sales on subsequent visits to that market.

Table D.2 lists various observable demographics of the purchaser across the different conditions and product choices. Note that the number of observations is lower than our full sample. Practical limitations, such as one customer interaction quickly followed by another, inhibited us from recording all observable demographics. However, this affected all treatments equally.

We use non-parametric tests to evaluate whether customers across the treatment conditions differ in terms of the recorded demographics. For age and group size a Kolmogorov-Smirnov tests shows no significant differences between the separate regular, separate Fair Trade and joint conditions at the 10% level. For gender, we find no significant differences

using a test of equality of proportions, again at the 10% level or stronger. For ethnicity, the consumers purchasing Fair Trade in the separate condition are more likely to be Caucasian than not (91.1%) compared to consumers in the separate regular condition (72.5%) (equality of proportions test, $p = 0.02$). Finally, within the joint treatment, a larger proportion of those purchasing Fair Trade are female than those choosing the regular product (equality of proportions test, $p < 0.01$).

TABLE D.2: Demographics of purchaser by condition

	Separate		Joint		Overall $n = 74$
	Regular $n = 60$	Fair Trade $n = 49$	Regular $n = 12$	Fair Trade $n = 52$	
Sex (% female)	56.7%	67.3%	25.0%	67.9%	60.3%
Ethnicity (% Caucasian)	72.5%	91.1%	66.7%	82.7%	79.7%
Av. age category	3.43 (1.48)	4.20 (1.57)	3.78 (1.47)	3.88 (1.76)	3.85 (1.42)
Av. group size	1.53 (0.81)	1.53 (0.63)	1.42 (0.79)	1.27 (0.45)	1.30 (0.52)

Standard deviations are shown in brackets

D.3 Details of the markets

TABLE D.3: General descriptive statistics by market

	Sex (% female)	Av. age category	Ethnicity (% Cauc.)	Av. amount paid	Av. traffic
North Park ($n = 46$)	61.1	20-30	66.7	\$2.29 (1.26)	755
Pacific Beach ($n = 61$)	56.8	40-50	82.1	\$2.41 (1.30)	903
La Jolla ($n = 61$)	64.8	40-50	89.8	\$2.36 (1.27)	1452
Univ. Heights ($n = 7$)	42.9	40-50	85.7	\$2.71 (1.70)	NA
Hillcrest ($n = 44$)	69.0	30-40	77.6	\$2.51 (1.79)	1367
Overall ($N = 219$)	62.3	30-40	80.2	\$2.41	1166

Standard deviations are shown in brackets

TABLE D.4: Purchase rates and total traffic by market

	North Park	Pacific Beach	La Jolla	Hillcrest	Overall
Separate regular	2.53%	3.71%	1.23%	1.53%	2.25%
Separate Fair Trade	6.25%	3.51%	2.07%	3.08%	3.73%
Joint	3.34%	2.37%	0.88%	0.92%	1.88%
Total traffic	1509	2710	4356	5468	14043

D.4 Script

The script was provided to the research assistants at the Farmer's Market to ensure consistency in their interaction with the customer. It was emphasized that the wording to explain the product offering, the Pay-What-You-Want pricing mechanism and the meaning of the Fair Trade label had to be strictly adhered to. However, they were allowed to make additional statements if they felt this would help their interaction with the customer. Research assistants were only sparsely informed about the purpose of the study and what we were doing with the proceeds. This made their answer of "I don't know" to certain questions, such as cost structure, legitimate.

D.4.1 Main interaction

- Hello. Are you interested in a chocolate cupcake?
- We have a chocolate cupcake. (SEP REG)
- We have a Fair Trade certified chocolate cupcake. (SEP FT)
- We have a chocolate cupcake and a Fair Trade certified chocolate cupcake. (JOINT)
- We have a chocolate cupcake. Usually we also sell Fair Trade certified ones, but these are not available today. (Rn)
- We have a Fair Trade certified chocolate cupcake. Usually we also sell regular chocolate ones, but these are not available today. (FTn)
- Today we have a Pay-What-You-Want offer. It means you can take a cupcake and choose how much to pay for it.
- Thank you for stopping by. Have a nice day.

D.4.2 Suggested answers to questions from customers

- *What does Pay-What-You-Want mean?* It means you can choose how much to pay for a cupcake. This means you get to choose the price and that any price you choose to pay is acceptable.
- *Does this mean I can pay anything I want?* Yes, you are in control, so you can choose how much you want to pay.
- *Can I get the product for free?* If you decide to pay \$0, then yes, you can get it for free.

- *Is \$1/\$2/\$3/\$4/\$5 ok?* However much you want to pay is ok.
- *How much do other people pay? / What is the average that people have paid so far?* I'm not sure, but you can choose to pay any price that you want.
- *Why are you doing this?* We are interested to see how much people will pay for a cupcake.
- *Is this tied to a marketing promotion? What strings are attached?* No, there are no strings attached.
- *Where does the money you make go?* The money is used to cover the costs of the cupcakes. That's all I've been told.
- *Why are you offering cupcakes?* Because these cupcakes are delicious and we'd like to share them with you.
- *Do you make these cupcakes yourself?* No, they are made by Sugar and Scribe Bakery. They are located in Pacific Beach.
- *Do you work at Sugar and Scribe Bakery?* I do not. I'm just here on the Farmer's Market.
- *Why do you offer two kinds of cupcakes? (JOINT)* So that you can choose the one you prefer.
- *Can I buy more than one cupcake?* No, we have a limit of one cupcake per person, sorry.
- *Is this your own business? / Is this a commercial business?* The money we make goes to covering the costs of the cupcake. Our goal is to sell cupcakes under Pay-What-You-Want at various Farmer's Markets across San Diego.
- *What kind of cupcake is it?* It's a chocolate cupcake with golden sprinkles. They come in a box, so it is easy for you to take home.
- *Do you offer any other flavours/sizes of cupcakes?* No, at the moment we only offer these cupcakes.
- *Are the cupcakes organic / suitable for vegetarians / suitable for vegans?* No, they are made with non-certified organic ingredients / Yes, they are suitable for vegetarians / No, they contain egg.
- *What does Fair Trade mean?* Fair Trade is about decent working conditions, local sustainability and fair terms of trade for farmers and workers in the developing

world who produce the chocolate. The organization Fair Trade International certifies products that meet the Fair Trade standards and these products receive the Fair Trade mark. It's an indication to the consumer that the product meets certain social, economic and environmental requirements [Can hand out an information sheet about the Fair Trade label].

- *What is Fair Trade about the cupcake?* The chocolate used to make the cupcake is Fair Trade certified.
- *Can I try a sample?* Unfortunately we do not offer samples at this time, sorry.
- *How much is the cupcake worth?* I don't know.
- *Why are you selling these cupcakes under Pay-What-You-Want?* It's a good product. We simply decided to allow people to decide how much they want to pay for it.
- *Are you also at other Farmer's Markets?* Yes, we are in Pacific Beach on Tuesday, North Park on Thursday and La Jolla and Hillcrest on Sunday, for at least two weeks.
- *Why are the (Fair Trade) chocolate ones not available?* I don't really know, I just know we don't have them today, sorry.

D.5 Additional regression results

TABLE D.5: OLS regression: Drivers of the amount paid

<i>Dependent variable: Purchase rate (%)</i>			
	Model 1	Model 2a	Model 2b
Separate - Regular	-1.156 ** [.545]	-1.156 [.927]	-1.117 [.903]
Joint	-1.748 *** [.546]	-1.748 ** [.688]	-1.694 ** [.731]
Pacific Beach	-.461 [.552]	-.461 [.666]	-.588 [.669]
La Jolla	-2.280 *** [.571]	-2.280 ** [.750]	-2.329 ** [.774]
Hillcrest	-2.057 *** [.588]	-2.057 ** [.706]	-1.923 ** [.744]
Average temperature			-.103 *** [.030]
Before/After lunch			-.150 [.147]
Constant	4.731 *** [.620]	4.731 *** [1.108]	6.691 *** [.940]
N	90	90	90
Adjusted R ²	.325	.370	.370

Standard errors are shown in square brackets. Models 2a and 2b use robust standard errors for time blocks. The ***, ** and * indicate significant effects at the 1%, 5% and 10% level, respectively. Variable description: *Separate regular / Joint*: dummy for the respective treatment; *Pacific Beach / La Jolla / Hillcrest*: dummy for the respective market; *Average temperature*: Average daytime temperature; *Before/After lunch*: Dummy with value 1 for purchase rates recorded after lunch, 0 otherwise.

TABLE D.6: OLS regression: Drivers of the amount paid

<i>Dependent variable: Amount paid</i>				
	Model 1	Model 2	Model 3	Model 4
Separate - Fair Trade	.084 [.243]	.0783 [.257]	.230 [.321]	.396 [.438]
Joint - Regular	-1.093 *** [.224]	-1.075 *** [.224]	-1.105 *** [.328]	-1.274 *** [.394]
Joint - Fair Trade	.343 [.226]	.348 [.230]	.441 [.269]	.337 [.342]
Pacific Beach		.165 [.252]	-.359 [.388]	-.721 [.731]
La Jolla		.081 [.239]	-.381 [.356]	-.475 [.460]
University Heights		.333 [.651]	-.210 [.754]	-.030 [.928]
Hillcrest		.244 [.323]	-.306 [.408]	-.155 [.643]
Market session			-.045 [.137]	.044 [.164]
Gender			-.195 [.252]	-.118 [.294]
Age category			.662 ** [.310]	.598 [.388]
Age category ²			-.099 ** [.039]	-.088 * [.051]
Ethnic majority				.068 [.337]
Group size				-.032 [.229]
Research assistant 2				.060 [.923]
Research assistant 3				-.437 [.681]
Research assistant 4				-.013 [1.100]
Research assistant 5				-.778 [.850]
Constant		2.226 *** [.214]	2.042 *** [.626]	2.628 *** [.837]
N	219	219	177	141
R ²	.057	.061	.105	.139

All models have robust standard errors for independent purchases and are shown in square brackets. The ***, ** and * indicate significant effects at the 1%, 5% and 10% level, respectively. Models 3 and 4 exclude observations where the customer could not be classified according to one or all of the demographic variables included in the model. Variable description: *Separate - Fair Trade / Joint - Regular / Joint - Fair Trade*: dummy for the respective treatment; *Pacific Beach / La Jolla / University Heights / Hillcrest*: dummy for the respective market; *Market session*: the occasion number that our stand operated at the market; *Gender*: the gender of the customer; *Age category / Age category²*: an ordinal variable of age at 10 year intervals. Higher numbers indicate higher age; *Ethnic majority*: dummy with a value of 1 if the customer is Caucasian and 0 otherwise; *Research assistant 2-5*: dummy for the research assistant(s) operating the stand.

Bibliography

- Abeler J., Becker A., Falk A., 2012. Truth-telling: A representative assessment. Technical report, Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit.
- ACFE, 2012. Report to the nations on occupational fraud and abuse: 2012 global fraud study. Technical report, Association of Certified Fraud Examiners.
- Ambrus A., Greiner B., 2012. Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review* 102, 3317–32.
- Andreoni J., 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. *The Economic Journal* 464–477.
- Andreoni J., Miller J., 2002. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica* 70, 737–753.
- Ariely D., Bracha A., Meier S., 2009. Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *The American Economic Review* 544–555.
- Arnot C., Boxall P.C., Cash S.B., 2006. Do ethical consumers care about price? a revealed preference analysis of fair trade coffee purchases. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 54, 555–565.
- Balliet D., Parks C., Joireman J., 2009. Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations* 12, 533–547.
- Bandiera O., Barankay I., Rasul I., 2007. Incentives for managers and inequality among workers: evidence from a firm-level experiment. *The Quarterly Journal of Economics* 122, 729–773.
- Bandiera O., Barankay I., Rasul I., 2013. Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association* 11, 1079–1114.
- Barkai A., Meredith G., Felaar F., Dantie Z., de Buys D., 2012. The advent of electronic logbook technology-reducing cost and risk to both marine resources and the fishing

- industry. In: Proceedings of World Academy of Science, Engineering and Technology, number 67. World Academy of Science, Engineering and Technology.
- Bartling B., Weber R.A., 2013. Do markets erode social responsibility? Technical report, CESifo Working Paper.
- Battigalli P., Charness G., Dufwenberg M., 2013. Deception: The role of guilt. *Journal of Economic Behavior & Organization* 93, 227–232.
- Becker G.S., 1968. Crime and punishment: An economic approach. *Journal of Political Economy* 76, 169–217.
- Beddington J.R., Agnew D.J., Clark C.W., 2007. Current problems in the management of marine fisheries. *Science* 316, 1713–1716.
- Belot M., Bhaskar V., Van De Ven J., 2012. Can observers predict trustworthiness? *Review of Economics and Statistics* 94, 246–259.
- Belot M., Schröder M., 2013. Sloppy work, lies and theft: A novel experimental design to study counterproductive behaviour. *Journal of Economic Behavior & Organization* 93, 233–238.
- Bénabou R., Tirole J., 2011. Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics* 126, 805–855.
- Besancenot D., Dubart D., Vranceanu R., 2013. The value of lies in an ultimatum game with imperfect information. *Journal of Economic Behavior & Organization* 93, 239–247.
- Blume A., DeJong D.V., Kim Y.G., Sprinkle G.B., 2001. Evolution of communication with partial common interest. *Games and Economic Behavior* 37, 79–120.
- Bochet O., Page T., Putterman L., 2006. Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization* 60, 11–26.
- Bordalo P., Gennaioli N., Shleifer A., 2012. Salience in experimental tests of the endowment effect. Technical report, National Bureau of Economic Research.
- Brandts J., Riedl A., van Winden F., 2009. Competitive rivalry, social disposition, and subjective well-being: An experiment. *Journal of Public Economics* 93, 1158–1167.
- Brosig J., Weimann J., Ockenfels A., 2003a. The effect of communication media on cooperation. *German Economic Review* 4, 217–241.
- Brosig J., Weimann J., Ockenfels A., 2003b. The effect of communication media on cooperation. *German Economic Review* 4, 217–241.

- Burks S., Carpenter J., Goette L., 2009. Performance pay and worker cooperation: Evidence from an artefactual field experiment. *Journal of Economic Behavior & Organization* 70, 458–469.
- Buser T., Dreber A., 2013. The flipside of comparative payment schemes. Technical report, Tinbergen Institute.
- Cadsby C.B., Song F., Tapon F., 2010. Are you paying your employees to cheat? an experimental investigation. *The BE Journal of Economic Analysis & Policy* 10.
- Cappelen A.W., Nielsen U.H., Sørensen E.Ø., Tungodden B., Tyran J.R., 2013a. Give and take in dictator games. *Economics Letters* 118, 280 – 283.
- Cappelen A.W., Sørensen E.Ø., Tungodden B., 2013b. When do we lie? *Journal of Economic Behavior & Organization* 93, 258–265.
- Cason T.N., Khan F.U., 1999. A laboratory study of voluntary public goods provision with imperfect monitoring and communication. *Journal of development Economics* 58, 533–552.
- Charness G., Dufwenberg M., 2006. Promises and partnership. *Econometrica* 74, 1579–1601.
- Charness G., Dufwenberg M., 2010. Bare promises: An experiment. *Economics Letters* 107, 281–283.
- Charness G., Rabin M., 2002. Understanding social preferences with simple tests. *Quarterly journal of Economics* 817–869.
- Chaudhuri A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics* 14, 47–83.
- Conrads J., Irlenbusch B., Rilke R.M., Schielke A., Walkowitz G., 2014. Honesty in tournaments. *Economics Letters* 123, 90–93.
- Crawford V.P., 2003. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review* 133–149.
- Croson R.T., 2000. Thinking like a game theorist: factors affecting the frequency of equilibrium play. *Journal of economic behavior & organization* 41, 299–314.
- Danz D., Engelmann D., Kübler D., 2012. Do legal standards affect ethical concerns of consumers? an experiment on minimum wages. University of Mannheim, Department of Economics, Working Paper 12 3.

- Dawes R.M., McTavish J., Shaklee H., 1977. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of personality and social psychology* 35, 1.
- Drago R., Garvey G.T., 1998. Incentives for helping on the job: Theory and evidence. *Journal of Labor Economics* 16, 1–25.
- Elfenbein D.W., McManus B., 2010. A greater price for a greater good? evidence that consumers pay more for charity-linked products. *American Economic Journal: Economic Policy* 2, 28–60.
- Erat S., 2013. Avoiding lying: the case of delegated deception. *Journal of Economic Behavior & Organization* 93, 273–278.
- Erat S., Gneezy U., 2012. White lies. *Management Science* 58, 723–733.
- Fair Trade Labeling Organization, 2005. Annual report 2004-2005. Technical report. URL http://www.fairtrade.net/fileadmin/user_upload/content/FL0_AR_2004_05.pdf.
- Fair Trade Labeling Organization, 2012. Annual report 2011-2012. Technical report. URL http://www.fairtrade.net/fileadmin/user_upload/content/2009/resources/2011-12_AnnualReport_web_version_small_FairtradeInternational.pdf.
- Fair Trade Labeling Organization, 2013. Annual report 2012-2013. Technical report. URL http://www.fairtrade.net/fileadmin/user_upload/content/2009/resources/2012-13_AnnualReport_FairtradeIntl_web.pdf.
- Fair Trade USA, 2014. What is fair trade? <http://fairtradeusa.org/what-is-fair-trade/faq>. Accessed: 2014-07-02.
- Falk A., Kosfeld M., 2006. The hidden costs of control. *The American economic review* 1611–1630.
- Falkinger J., Fehr E., Gächter S., Winter-Ebmer R., 2000. A simple mechanism for the efficient provision of public goods: Experimental evidence. *The American Economic Review* 90, pp. 247–264.
- Fehr E., Fischbacher U., 2003. The nature of human altruism. *Nature* 425, 785–791.
- Fehr E., Gächter S., 2002. Altruistic punishment in humans. *Nature* 415, 137–140.
- Fehr E., Schmidt K.M., 1999. A theory of fairness, competition, and cooperation. *The quarterly journal of economics* 114, 817–868.

- Fischbacher U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171–178.
- Fischbacher U., Föllmi-Heusi F., 2013. Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association* 11, 525–547.
- Fischbacher U., Gächter S., 2010. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review* 541–556.
- Fischbacher U., Gächter S., Fehr E., 2001. Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters* 71, 397–404.
- Food Marketing Institute, 2014. Supermarket facts. <http://www.fmi.org/research-resources/supermarket-facts>. Accessed: 2014-07-02.
- Frackenpohl G., Pönitzsch G., 2013. Bundling public with private goods. Technical report, Bonn Econ Discussion Papers.
- Frey B.S., Meier S., 2004. Social comparisons and pro-social behavior: Testing” conditional cooperation” in a field experiment. *American Economic Review* 1717–1722.
- Friedrichsen J., Engelmann D., 2013. Who cares for social image? interactions between intrinsic motivation and social image concerns. Technical report, CESifo Working Paper.
- Gächter S., 2007. Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. na.
- Gächter S., Fehr E., 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980–994.
- Gächter S., Renner E., 2010. The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics* 13, 364–377.
- Gagern A., van den Bergh J., Sumaila U.R., 2013. Trade-based estimation of bluefin tuna catches in the eastern atlantic and mediterranean, 2005–2011. *PloS one* 8, e69959.
- Gibson R., Tanner C., Wagner A.F., 2013. Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review* 103, 532–48.
- Gill D., Prowse V., Vlassopoulos M., 2013. Cheating in the workplace: An experimental study of the impact of bonuses and productivity. *Journal of Economic Behavior & Organization* 96, 120–134.
- Gneezy A., Gneezy U., Nelson L.D., Brown A., 2010. Shared social responsibility: A field experiment in pay-what-you-want pricing and charitable giving. *Science* 329, 325–327.

- Gneezy A., Gneezy U., Riener G., Nelson L.D., 2012. Pay-what-you-want, identity, and self-signaling in markets. *Proceedings of the National Academy of Sciences* 109, 7236–7240.
- Gneezy U., 2005. Deception: The role of consequences. *American Economic Review* 95, 384–394.
- Gneezy U., Rockenbach B., Serra-Garcia M., 2013a. Measuring lying aversion. *Journal of Economic Behavior & Organization* 93, 293–300.
- Gneezy U., Rustichini A., 2000a. A fine is a price. *Journal of Legal Studies* 29, 1.
- Gneezy U., Rustichini A., 2000b. Pay enough or don't pay at all. *Quarterly journal of economics* 791–810.
- Gneezy U., Saccardo S., Van Veldhuizen R., 2013b. Bribery: Greed versus reciprocity .
- González-Vallejo C., Moran E., 2001. The evaluability hypothesis revisited: Joint and separate evaluation preference reversal as a function of attribute importance. *Organizational Behavior and Human Decision Processes* 86, 216–233.
- Gravert C., 2013. Pride and patronage - the effect of identity on pay-what-you-want prices at a charitable bookstore. Technical report, Aarhus University Economics Working Papers.
- Grechenig K., Nicklisch A., Thöni C., 2010. Punishment despite reasonable doubt—a public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies* 7, 847–867.
- Greiner B., 2004. The online recruitment system orsee 2.0 - a guide for the organization of experiments in economics. Working Paper Series in Economics 10, University of Cologne, Department of Economics.
- Hainmueller J., Hiscox M., 2014. Buying green? field experimental tests of consumer support for environmentalism. Technical report, MIT Political Science Department Research Paper No. 2012-14.
- Hainmueller J., Hiscox M., Sequeira S., 2014. Consumer demand for the fair trade label: Evidence from a field experiment. *Review of Economics and Statistics* Forthcoming.
- Harbring C., 2010. On the effect of incentive schemes on trust and trustworthiness. *Journal of Institutional and Theoretical Economics (JITE)* 166, 690–714.
- Harbring C., Irlenbusch B., 2011. Sabotage in tournaments: Evidence from a laboratory experiment. *Management Science* 57, 611–627.

- Herrmann B., Thöni C., Gächter S., 2008. Antisocial punishment across societies. *Science* 319, 1362–1367.
- Heyman J., Ariely D., 2004. Effort for payment a tale of two markets. *Psychological science* 15, 787–793.
- Hoffmann M., Lauer T., Rockenbach B., 2013. The royal lie. *Journal of Economic Behavior & Organization* 93, 305–313.
- Holmstrom B., Milgrom P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization* 24–52.
- Hsee C.K., Leclerc F., 1998. Will products look more attractive when presented separately or together? *Journal of Consumer Research* 25, 175–186.
- Hsee C.K., Loewenstein G.F., Blount S., Bazerman M.H., 1999. Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin* 125, 576.
- Irlenbusch B., Ter Meer J., 2013. Fooling the nice guys: Explaining receiver credulity in a public good game with lying and punishment. *Journal of Economic Behavior & Organization* 93, 321–327.
- Isaac R.M., Walker J.M., 1988. Communication and free-riding behavior: The voluntary contribution mechanism. *Economic inquiry* 26, 585–608.
- Kamenica E., 2008. Contextual inference in markets: On the informational content of product lines. *The American Economic Review* 98, 2127–2149.
- Kartik N., Ottaviani M., Squintani F., 2007. Credulity, lies, and costly talk. *Journal of Economic theory* 134, 93–116.
- Keser C., Van Winden F., 2000. Conditional cooperation and voluntary contributions to public goods. *The Scandinavian Journal of Economics* 102, 23–39.
- Kim J.Y., Natter M., Spann M., 2010. Kish—where customers pay as they wish. *Review of Marketing Science* 8, 3.
- Larkin I., Pierce L., Gino F., 2012. The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal* 33, 1194–1214.
- Ledyard J., 1995. 0. 1995. public goods: A survey of experimental research. *Handbook of Experimental Economics*, Princeton University Press, Princeton 111–194.

- Liebrand W.B., 1984. The effect of social motives, communication and group size on behaviour in an n-person multi-stage mixed-motive game. *European Journal of Social Psychology* 14, 239–264.
- Liebrand W.B., McClintock C.G., 1988. The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *European journal of personality* 2, 217–230.
- List J.A., 2002. Preference reversals of a different kind: The “more is less” phenomenon. *American Economic Review* 1636–1643.
- List J.A., Lucking-Reiley D., 2002. The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. *Journal of Political Economy* 110, 215–233.
- Lotz S., Christandl F., Fetchenhauer D., 2013. What is fair is good: Evidence of consumers’ taste for fairness. *Food Quality and Preference* 30, 139–144.
- Machado F., Sinha R.K., 2013. The viability of pay what you want pricing. Technical report.
- Madarász K., 2012. Information projection: Model and applications. *The review of economic studies* 79, 961–985.
- Mazar N., Amir O., Ariely D., 2008. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research* 45, 633–644.
- Meier S., 2007. Do subsidies increase charitable giving in the long run? matching donations in a field experiment. *Journal of the European Economic Association* 5, 1203–1222.
- Mullen B., Atkins J.L., Champion D.S., Edwards C., Hardy D., Story J.E., Vanderklok M., 1985. The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology* 21, 262–283.
- Neugebauer T., Perote J., Schmidt U., Loos M., 2009. Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *Journal of Economic Psychology* 30, 52–60.
- Nikiforakis N., 2010. Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior* 68, 689–702.
- Offerman T., Sonnemans J., Schram A., 1996. Value orientations, expectations and voluntary contributions in public goods. *Economic Journal* 106, 817–45.

- Okada E.M., 2005. Justification effects on consumer choice of hedonic and utilitarian goods. *Journal of Marketing Research* 42, 43–53.
- Ostrom E., 1990. *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.
- Ostrom E., Walker J., Gardner R., 1992. Covenants with and without a sword: Self-governance is possible. *American Political Science Review* 86, 404–417.
- Oxford English Dictionary, 2006. *Oxford English Dictionary*.
- Pan X.S., Houser D., 2013. Cooperation during cultural group formation promotes trust towards members of out-groups. *Proceedings of the Royal Society B: Biological Sciences* 280.
- Pascual-Ezama D., Prelec D., Dunfield D., 2013. Motivation, money, prestige and cheats. *Journal of Economic Behavior & Organization* 93, 367–373.
- Pauly D., Belhabib D., Blomeyer R., Cheung W.W.W.L., Cisneros-Montemayor A.M., Copeland D., Harper S., Lam V.W.Y., Mai Y., Le Manach F., Åsterblom H., Mok K.M., van der Meer L., Sanz A., Shon S., Sumaila U.R., Swartz W., Watson R., Zhai Y., Zeller D., 2013. China's distant-water fisheries in the 21st century. *Fish and Fisheries* 1467–2979.
- Prasad M., Kimeldorf H., Meyer R., Robinson I., 2004. Consumers of the world unite a market-based response to sweatshops. *Labor Studies Journal* 29, 57–79.
- Reeson A.F., Tisdell J.G., 2008. Institutions, motivations and public goods: An experimental test of motivational crowding. *Journal of Economic Behavior & Organization* 68, 273–281.
- Riener G., Traxler C., 2012. Norms, moods, and free lunch: Longitudinal evidence on payments from a pay-what-you-want restaurant. *The Journal of Socio-Economics* 41, 476–483.
- Rode J., Hogarth R.M., Le Menestrel M., 2008. Ethical differentiation and market behavior: An experimental approach. *Journal of Economic Behavior & Organization* 66, 265–280.
- Ross L., Greene D., House P., 1977. The 'false consensus effect': An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology* 13, 279–301.
- Ryan R.M., Deci E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 68.

- Schmidt K.M., Spann M., Zeithammer R., 2012. Pay what you want as a marketing strategy in monopolistic and competitive markets .
- Schweitzer M.E., Ordóñez L., Douma B., 2004. Goal setting as a motivator of unethical behavior. *Academy of Management Journal* 47, 422–432.
- Sell J., Wilson R.K., 1991. Levels of information and contributions to public goods. *Social Forces* 70, 107–124.
- Serra-Garcia M., Van Damme E., Potters J., 2013. Lying about what you know or about what you do? *Journal of the European Economic Association* 11, 1204–1229.
- Shang J., Croson R., 2009. A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal* 119, 1422–1439.
- Sheremeta R.M., Shields T.W., 2013. Do liars believe? beliefs and other-regarding preferences in sender–receiver games. *Journal of Economic Behavior & Organization* 94, 268–277.
- Sonnemans J., Schram A., Offerman T., 1998. Public good provision and public bad prevention: The effect of framing. *Journal of Economic Behavior & Organization* 34, 143–161.
- Sutter M., 2009. Deception through telling the truth?! experimental evidence from individuals and teams*. *The Economic Journal* 119, 47–60.
- Teisl M.F., Roe B., Hicks R.L., 2002. Can eco-labels tune a market? evidence from dolphin-safe labeling. *Journal of Environmental Economics and Management* 43, 339–359.
- Tenbrunsel A.E., Messick D.M., 1999. Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly* 44, 684–707.
- van Dijk F., Sonnemans J., van Winden F., 2002. Social ties in a public good experiment. *Journal of Public Economics* 85, 275–299.
- Van Lange P.A., De Cremer D., Van Dijk E., Van Vugt M., 2007. Self-interest and beyond. *Social psychology: Handbook of basic principles* 540–61.
- Vanberg C., 2008. Why do people keep their promises? an experimental test of two explanations¹. *Econometrica* 76, 1467–1480.
- Wang J.T.y., Spezio M., Camerer C.F., 2010. Pinocchio’s pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *The American Economic Review* 100, 984–1007.

Wilson R.K., Sell J., 1997. "liar, liar..." cheap talk and reputation in repeated public goods settings. *Journal of Conflict Resolution* 41, 695–717.

Yamagishi T., 1986. The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology* 51, 110–116.