

Studies and Experiments on Information  
Asymmetries and the Intergenerational Free  
Rider Problem in Climate Change

Inaugural-Dissertation  
zur Erlangung des akademischen Grades eines Doktors  
der Wirtschafts- und Sozialwissenschaften  
der Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von  
Aurel Stenzel  
aus Leutkirch im Allgäu

Kiel, 2024

Gedruckt mit Genehmigung der  
Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

Erstbegutachtung:	Prof. Dr. Till Requate
Zweitbegutachtung:	Prof. Dr. Jana Friedrichsen
Tag der Abgabe der Arbeit:	08.10.2024
Tag der mündlichen Prüfung:	13.02.2025

## Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Till Requate and Israel Waichman, for their invaluable guidance and support. Re-entering the academic world after a decade of full-time employment presented a number of challenges, and their expertise and patience proved invaluable in helping me navigate this transition. Israel, a special thanks to you for working closely with me in the last couple of years. From countless discussions - in many different coffee places across Berlin and Munich or in front of our screens - to the hundreds (if not thousands) of WhatsApp messages exchanged, your constant support, encouragement, and practical advice were essential in pushing this thesis forward.

Dear Fritzi, this is the third time I've had the opportunity to mention you in an academic thesis of mine - let's see if it's my last. Your constant love, patience, and support through all the ups and downs of this journey (and the ones before) have meant the world to me. Without you, none of this would have been possible. To our Effi, who has no idea yet what a PhD is but still gave me the energy and love to finish this project. To Isabella, I cannot wait for my *Abenteuertag* to come. Last but not least, I thank my parents and family for the foundation they laid for everything I do today.

# Contents

<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Supply Chain Data Sharing for Scope 3 Emissions</b>	<b>6</b>
2.1 Introduction . . . . .	7
2.2 Scope 3 emissions: Introduction and current estimation . . . . .	9
2.3 The potential benefits of accurately calculating and reporting Scope 3 emissions . . . . .	11
2.4 Obstacles to primary data sharing . . . . .	12
2.4.1 Legal and regulatory challenges . . . . .	13
2.4.2 Interoperability . . . . .	14
2.4.3 Data privacy . . . . .	15
2.5 Overcoming the obstacles to primary data sharing . . . . .	15
2.5.1 Overcoming legal and regulatory challenges . . . . .	15
2.5.2 Overcoming missing interoperability . . . . .	16
2.5.3 Overcoming data privacy concerns . . . . .	18
2.5.4 Summary . . . . .	19
2.6 Concluding remarks . . . . .	20
Bibliography . . . . .	21
<b>3 The Efficacy of (Self-)Verification Instruments in Risky In-     vestments: An Experiment</b>	<b>29</b>
3.1 Introduction . . . . .	30
3.2 Related literature . . . . .	33
3.2.1 Cheap talk communication . . . . .	33
3.2.2 Lying and deception in sender-receiver games . . . . .	34
3.2.3 Verification by receivers and self-verification by senders .	36
3.3 Design, predictions, and experimental procedure . . . . .	38



3.3.1	Experimental design and workhorse game . . . . .	38
3.3.2	Predictions . . . . .	41
3.3.3	Experimental procedure . . . . .	43
3.4	Results . . . . .	44
3.4.1	Investment per firm's probability of success . . . . .	44
3.4.2	Truth-telling and strategic behavior . . . . .	47
3.4.3	Investment per free message . . . . .	48
3.4.4	Verification use . . . . .	50
3.4.5	Investor Efficiency . . . . .	53
3.5	Revisiting our predictions: Alternative explanations for the ob- served investment behavior . . . . .	55
3.6	Conclusions and discussion . . . . .	57
	Bibliography . . . . .	58
	Appendix . . . . .	64
3.A	Theoretical predictions . . . . .	64
3.B	Risk aversion . . . . .	72
3.B.1	Reflecting risk aversion . . . . .	72
3.B.2	Estimating the risk aversion of our participants . . . . .	73
3.C	Alternative models . . . . .	75
3.C.1	Inequity aversion . . . . .	75
3.C.2	Level- $k$ reasoning . . . . .	76
3.C.3	Intrinsic costs of lying and belief updating . . . . .	77
3.D	Further analysis . . . . .	78
3.D.1	Investment per information available in Costly Verification (T2) . . . . .	79
3.D.2	Deceptive behavior by firms . . . . .	80
3.D.3	Verification in Costly Verification (T2) . . . . .	81
3.E	Robustness . . . . .	81
3.F	Translated instructions (computer screens) . . . . .	89
3.F.1	Baseline . . . . .	90
3.F.2	Costly Verification (T2) . . . . .	96
3.F.3	Costly Self-Verification (T3) . . . . .	98
<b>4</b>	<b>Verification and Blockchain-Based Self-Verification Instru- ments for Risky Investment Decisions</b>	<b>102</b>
4.1	Introduction . . . . .	103
4.2	Related literature . . . . .	105

4.3	Design, predictions, and experimental procedure . . . . .	106
4.3.1	Experimental design . . . . .	106
4.3.2	Predictions . . . . .	108
4.3.3	Experimental procedure . . . . .	108
4.4	Results . . . . .	109
4.4.1	Effectiveness in conditioning investment on true probability of success . . . . .	109
4.4.2	High effectiveness in Costly Verification even without instrument use . . . . .	112
4.4.3	Measuring Investor and Societal Efficiency . . . . .	114
4.5	Conclusions . . . . .	116
	Bibliography . . . . .	118
	Appendix . . . . .	122
4.A	Further analysis . . . . .	122
4.A.1	Investment per true probability of success . . . . .	122
4.A.2	Investment per true probability of success with and without (self-)verification . . . . .	125
4.A.3	Percentage cheating per true probability of success . . . . .	129
4.A.4	Use of (self-)verification per true probability of success . . . . .	133
4.A.5	Investors' beliefs per cheap talk message . . . . .	135
4.A.6	Examples for Investor Efficiency and Societal Efficiency . . . . .	139
4.A.7	Mann-Whitney U-tests . . . . .	140
4.B	Translated instructions (computer screens) . . . . .	141
4.B.1	Baseline . . . . .	141
4.B.2	Costly Noisy Verification (T2) . . . . .	145
4.B.3	Self-Verification (T5) . . . . .	148
<b>5</b>	<b>Punishment with Communication Sustains Altruistic Cooperation Across Generations</b>	<b>151</b>
5.1	Introduction . . . . .	152
5.2	Results . . . . .	156
5.2.1	Intergenerational cooperation . . . . .	156
5.2.2	Individual contributions . . . . .	157
5.2.3	Determinants of received punishment . . . . .	159
5.2.4	Relative benefits of different institutions . . . . .	159
5.3	Discussion . . . . .	159
5.4	Methods . . . . .	161

5.4.1 Design . . . . .	161
5.4.2 Procedure . . . . .	162
Bibliography . . . . .	162
Appendix . . . . .	166
5.A Experimental procedure and treatments . . . . .	166
5.A.1 Terminology and experimental design . . . . .	166
5.A.2 Procedure and treatments . . . . .	167
5.B Theoretical framework . . . . .	169
5.B.1 Introduction . . . . .	169
5.B.2 Theoretical framework . . . . .	171
5.B.3 Predictions . . . . .	176
5.C Pairwise tests . . . . .	177
5.D Regressions and their tests . . . . .	179
5.E Individual contributions in all periods . . . . .	184
5.F Robustness test regarding the lost observation . . . . .	184
5.G Bootstrap simulations . . . . .	185
5.H Analysis of communication . . . . .	188
5.H.1 Results of the communication analysis . . . . .	188
5.H.2 Annotation guidelines . . . . .	189
5.I Translated instructions (computer screens) . . . . .	193
<b>6 Conclusions</b>	<b>200</b>
<b>Bibliography</b>	<b>205</b>
<b>Erklärungen</b>	<b>209</b>

## List of figures

2.1	Scope 1, 2, and 3 emissions . . . . .	10
3.1	The Probabilistic Investment Game (repeated design) . . . .	39
3.2	Average investments in firms with a high and low probability of success . . . . .	45
3.3	Deception by firms (with a low probability of success) . . . .	48
3.4	Average investment per free and self-verified message . . . .	49
3.5	(Self-)verification of messages . . . . .	51
3.6	Average Investor Efficiency . . . . .	54
3.B.1	Investor’s self-indicated risk aversion (box plot) . . . . .	74
3.B.2	Distribution of risk aversion across treatments . . . . .	75
3.B.3	Average investments after (self-)verification of firms with a high probability of success . . . . .	75
3.D.1	Investment per free and verified messages in Costly Verifica- tion (T2) . . . . .	79
3.D.2	Share of deception over time . . . . .	80
4.1	The Probabilistic Investment Game (one-shot design) . . . .	105
4.2	Scatter plot and linear prediction: Investments per true prob- ability of success . . . . .	110
4.3	Investment per true probability of success with and without using (self-)verification . . . . .	111
4.4	Average percentage cheating per true probability of success range . . . . .	113
4.5	Investor and Societal Efficiency per (self-)verification instru- ment use in the different treatments (box plot) . . . . .	116
4.A.1	Messages per true probability of success (scatter plot) . . . .	129
4.A.2	“Percentage cheating” per true probability of success (scatter plot) . . . . .	130
4.A.3	“Percentage cheat” and “Log(percentage cheat)” as depen- dent variables . . . . .	131

4.A.4	(Self-)verification use per true probability of success (scatter plot and linear prediction) . . . . .	133
4.A.5	Investor's guess and the cheap talk message received (scatter plot and linear prediction) . . . . .	136
5.1	Repeated Intergenerational Goods Game (IGG) . . . . .	153
5.2	Sustained generations across treatments . . . . .	156
5.3	Frequency of individual contributions in the first periods of all generations . . . . .	157
5.A.1	Age distribution of participants . . . . .	169
5.B.1	Possible group compositions and example likelihoods . . . . .	172
5.B.2	Group composition and possible outcomes . . . . .	173
5.B.3	Numerical example for the likelihood of a sustained generation in COMMS and PUN&COMMS . . . . .	174
5.B.4	Impact of change in the distribution of types on the likelihood of a sustained generation in COMMS and PUN&COMMS . . . . .	175
5.E.1	Frequency of individual contributions in all periods of all generations . . . . .	184
5.F.1	Sustained generations across treatments considering the lost observation . . . . .	185
5.G.1	Simulation result for sustained generations across treatments . . . . .	187
5.H.1	Communication analysis . . . . .	189
5.H.2	Example of one generation's chat . . . . .	191
5.H.3	Questions . . . . .	192
5.H.4	Example for Questions 1 and 2 . . . . .	192
5.H.5	Example for Questions 3, 3a, and 3b . . . . .	192

## List of tables

2.1	Obstacles to data sharing and current approaches . . . . .	19
3.1	Treatments . . . . .	40
3.2	Investment per probability of success (random-effects estimations) . . . . .	46
3.3	Investment per free message . . . . .	50
3.4	Firms' self-verification use . . . . .	52
3.5	Investor Efficiency . . . . .	55
3.D.1	Deception per firm and probability of success . . . . .	81
3.D.2	Verification use in Costly Verification (T2) . . . . .	82
3.E.1	Investment per probability of success . . . . .	83
3.E.2	Investment per free communication treatment . . . . .	84
3.E.3	Firms' self-verification use . . . . .	85
3.E.4	Investor Efficiency . . . . .	87
3.E.5	Deception per firm and probability of success . . . . .	88
3.E.6	Verification use in Costly Verification (T2) . . . . .	89
4.1	Treatments . . . . .	107
4.2	Investment per true probability of success . . . . .	110
4.3	Investment per true probability of success and instrument use	112
4.4	"Log(percentage cheating)" per true probability of success .	114
4.5	Investor's guess and the cheap talk message received . . . .	115
4.A.1	Investment per true success probability, Model 1 and 2 . . .	123
4.A.2	Wald tests whether true probability of success affects investment in each of the treatments . . . . .	124
4.A.3	Investment per true probability of success with and without (self-)verification, Model 1 and 2 . . . . .	126
4.A.4	Wald tests whether true probability of success affects investment in each of the treatments with and without (self-)verification, Model 1 . . . . .	127

4.A.5	Wald tests whether true probability of success affects investment in each of the treatments with and without (self-)verification, Model 2 . . . . .	128
4.A.6	“Percentage cheating” per true probability of success . . . .	131
4.A.7	Wald tests whether true probability of success affects percentage cheating in each of the treatments . . . . .	132
4.A.8	(Self-)verification use depending on true probability of success (linear probability model) . . . . .	134
4.A.9	Wald tests whether true probability of success affects (self-)verification . . . . .	134
4.A.10	Investor’s guess and the cheap talk message received . . . .	137
4.A.11	Wald tests whether the cheap talk message affects the investor’s guess of true probability of success, Model 1 and 2 .	138
4.A.12	Wald tests (whether the cheap talk message affects the investor’s guess of true probability of success, Model 3 . . . .	139
4.A.13	Mann-Whitney U-test for stepwise difference in Investor Efficiency . . . . .	140
4.A.14	Mann-Whitney U-test for stepwise difference in Societal Efficiency . . . . .	140
5.A.1	Treatments . . . . .	167
5.C.1	Occurrence of sustained generations in the different treatments	177
5.C.2	Success rate in sustaining generations across all possible generations . . . . .	178
5.C.3	Total payoffs per cohort . . . . .	178
5.D.1	Linear regressions: Individual contributions . . . . .	179
5.D.2	Linear probability model regressions: Individual contributions of 5 ECU (focal point cooperation) . . . . .	180
5.D.3	Linear probability model regressions: Individual contributions of 0 ECU . . . . .	181
5.D.4	Linear probability model regressions: Punishment received .	181
5.D.5	Pairwise tests of Table 5.D.4 . . . . .	182
5.D.6	Random-effects panel data linear probability model regressions: Punishment received . . . . .	183
5.D.7	Pairwise tests of Table 5.D.6 . . . . .	183
5.F.1	Descriptive statistics for sustained generations . . . . .	185

5.F.2	Occurrence of sustained generations in the different treatments, based on Table 5.C.1 . . . . .	186
5.F.3	Success rate in sustaining generations across all possible generations, based on Table 5.C.2 . . . . .	186



# Chapter 1

## Introduction

Climate change: The ultimate  
challenge for economics

---

*William Nordhaus*

Climate change, caused by greenhouse gases (GHG) released by human activities, poses a significant challenge to humanity (IPCC, 2022). The accumulation of GHG in the atmosphere might cause irreversible impacts on the Earth's climate (Lenton et al., 2008; IPCC, 2022). Despite widespread recognition and communication to the public, current global efforts to mitigate are insufficient and likely to result in disastrous consequences, such as rising sea levels (e.g., Tebaldi et al., 2021), droughts (Naumann et al., 2021), and intense rainfall (Reed et al., 2022). The timeframe for effective climate action is rapidly closing (Tosun, 2022; IPCC, 2022). However, several market failures hinder the adequate mitigation of GHG emissions. This dissertation addresses two of these market failures: information asymmetries and the intergenerational free rider problem.

The first part of the thesis focuses on information asymmetries regarding GHG emissions present in the relationship between investors and firms. Financial investors are demanding compensation for carbon risks due to increasing regulatory pressure to transition to a low-carbon economy. Numerous studies have examined whether investors and financial markets incorporate carbon emission risk into their investment decisions (for a review, see Wang et al., 2022). These studies demonstrate that carbon emission risk increases the cost of equity and debt for firms (e.g., Jung et al., 2018; Bernardini et al., 2021; Bolton & Kacperczyk, 2021; Trinks et al., 2022). However, assessing a firm's exposure to these risks is challenging due to insufficient data and information, particularly when including indirect emissions created throughout a

firm’s value chain (known as “Scope 3”). These emissions often constitute the majority of a firm’s carbon footprint and can be difficult to quantify (WBCSD, 2021). The existence of information asymmetries regarding GHG emissions can result in suboptimal resource allocation, which could ultimately lead to overpollution. The second part of the thesis discusses the intergenerational free rider problem present in climate change. Nordhaus (2019), like others, describes GHG-induced climate change as a global public good (e.g., Barrett, 2007; Chaudhuri, 2011; Buchholz & Sandler, 2021): Mitigation efforts are expensive for individual nations, while the benefits are shared globally (e.g. Battersby, 2017; Nordhaus, 2019; Buchholz & Sandler, 2021; Boon-Falleur et al., 2022). The significant differences in wealth, potential losses from climate change, and past emissions among countries can further decrease their willingness to cooperate (e.g., Lange et al., 2010; Milinski et al., 2011; Tavoni et al., 2011; Burton-Chellew et al., 2013; Brown & Kroll, 2017; Waichman et al., 2021). Further, addressing the intergenerational challenge of climate change requires a unique form of cooperation, altruistic cooperation: Present generations must be willing to bear the full costs while future generations reap the benefits (e.g., Jacquet et al., 2013; Hauser et al., 2014; Lohse & Waichman, 2020). The issue of intergenerational free riding hinders the effectiveness of global efforts to mitigate climate change.

Behavioral research plays a crucial role in identifying solutions to market failures and can improve the effectiveness of economic tools and policy interventions. Behavioral research investigates individual motives and the reasons behind failed coordination in randomized controlled trials. Applying these results to climate change helps researchers understand the underlying reasons for insufficient action. My dissertation comprises four research papers, three of which are experimental studies. The first paper focuses on information asymmetries related to GHG emissions in supply chains and discusses how these asymmetries impede effective climate action. The two subsequent experimental studies investigate behavioral aspects to determine whether new emerging (self-)verification technologies can overcome such information asymmetries. The last study experimentally investigates institutions designed to tackle the intergenerational free rider problem. We assess the efficacy of punishment with and without communication in promoting altruistic cooperation, equivalent to mitigating climate change resulting from greenhouse gases emitted by human activities.

In Chapter 2, we provide an illustration of information asymmetries in

relation to value chain emissions of firms (Stenzel & Waichman, 2023). At the moment, companies do not have sufficient access to the data they need to calculate their emissions footprint along their supply chain (Scope 3). They currently rely on industry averages and approximations, resulting in inadequate motivation for firms to implement cleaner-than-average technologies. In contrast, primary data sharing throughout the supply chain would give firms accurate measurements of Scope 3 emissions. We examine the advantages of sharing primary data regarding emissions for increased transparency, e.g., the reduction of uncertainties regarding the carbon emission risk of investment decisions. Markets penalize firms for their carbon emissions, but a further penalty is imposed on firms that do not disclose emissions information (e.g., Matsumura et al., 2014; Bui et al., 2020; Mahmoudian et al., 2023). To avoid this, companies require access to data and information provided by their suppliers, which lie beyond their control. We outline the three primary hurdles to such data sharing, namely legal and regulatory challenges, a lack of technical interoperability, and data privacy concerns. Finally, we provide an overview of existing initiatives and the strategies that they are pursuing.

In Chapter 3 and 4, we use a novel experiment designed by Till Requate, Israel Waichman, and me. We aimed to empirically explore whether an emerging technology in the form of (self-)verification instruments enables investors to overcome the information asymmetry in risky investment decisions. In the “Probabilistic Investment Game,” each investor is endowed with a sum of money that they can invest in a firm. Each firm is assigned a success probability drawn from a uniform distribution. The range of success probabilities is common knowledge to all firms and investors and specified in a way that not investing is the investor’s a priori best choice. The baseline game is composed of two stages. First, the firm learns its probability of success and sends a free message to the investor, either revealing the actual or any success probability of the firm’s choice (“cheap talk”). Second, the investor receives the free message and determines the share of their endowment to invest in the firm. Ultimately, the success of the investment is reliant on a lottery based on the firm’s actual probability of success. If successful, both the investor and the firm receive a profit equal to the investment amount. Otherwise, the investment is lost.

We investigate the impact of investor-paid verification instruments and firm-paid self-verification instruments, comparing them against a trust-based baseline. Verification treatments involve investors deciding whether to verify

a firm’s probability of success at a cost after receiving the free message. In the self-verification treatments, the firms choose whether to self-verify their likelihood of success to investors by paying a cost or by sending a free message that may or may not disclose the true probability of success. In accordance with technological advancements, we vary the costs of self-verification among treatments. The expensive (self-)verification technology simulates the status quo, wherein investors perform expensive due diligence or firms engage in costly self-verification, such as receiving certification through costly intermediaries. On the other hand, the cheap self-verification method models the utilization of emerging data sharing technologies.

First, we conducted a one-shot variation of the Probabilistic Investment Game where firms and investors made only one decision. Our assumption was that verification and self-verification tools are more important in one-shot than in repeated interactions if a firm’s deceitful actions lack repercussions or externalities in the future. However, we refined the design following the analysis of the initial findings. We introduced repetition to allow for learning and experience but not for reputation. The game was repeated for 15 rounds with fixed roles and random matching (i.e., firms and investors remain in their roles throughout the experiment but do not stay with the same investor and firm).

In Chapter 5, we examine the intergenerational free rider problem. Although humans possess a unique ability to cooperate with strangers, this trait is thought to have evolved to promote mutually beneficial cooperation within small groups through principles of reciprocity and reputation. However, addressing the free rider challenge for future generations requires a unique form of cooperation. Current generations must be willing to bear the full burden of cooperation while future individuals reap the benefits. The extent of altruistic cooperation and the institutions required to sustain it are still uncertain and are explored in Chapter 5. We evaluate whether decentralized institutions, in the form of costly punishment with and without communication, can encourage altruistic cooperation using a repeated version of the “Intergenerational Goods Game” (IGG, see Hauser et al., 2014; Lohse & Waichman, 2020). The repeated IGG proceeds as follows: Individuals are divided into four generations, each represented by a group of three members. In the first generation, every member is given an endowment. Over the course of five periods, each member makes five decisions on how much to contribute to a group account. If the members have contributed half of their total available endowment at

the end of the five periods, the experiment will proceed to the next generation. Otherwise, the experiment ends, and future generations do not receive an additional endowment. Generation 2 faces the same challenge. If the game proceeds to Generation 3 and their contributions reach the threshold, each member of Generation 4 will receive an additional payoff, and the game will conclude without any further decisions being made. In addition to this Baseline treatment, we introduced three additional treatments. In the punishment treatment, members can reduce their peers' payoffs at a cost after each of the five contribution decisions. In the communication treatment, members of the current generation are permitted to engage in free communication with each other before making a contribution decision. Finally, a treatment combining punishment and communication was tested.

In sum, this thesis discusses information asymmetries (e.g., regarding GHG emissions) and the intergenerational free rider problem, both of which hinder effective climate action. Chapter 2 deals with data sharing among firms along a supply chain for the calculation of Scope 3 emissions, giving a specific example of information asymmetries in practice and shall be cited as

Stenzel, A., & Waichman, I. (2023). Supply-chain data sharing for Scope 3 emissions. *npj Climate Action*, 2, 7.

Chapter 3 uses the repeated version of the Probabilistic Investment Game to experimentally investigate whether technology, in the form of (self-)verification instruments, can overcome information asymmetries. It shall be cited as

Stenzel, A., Requate, T., & Waichman, I. (2024). The efficacy of (self-) verification instruments in risky investments: An experiment.

Chapter 4 investigates the same research question using the one-shot design. The chapter shall be cited as

Stenzel, A. (2022). Verification and Blockchain-based self-verification instruments for risky investment decisions: An experiment.

Chapter 5 examines decentralized institutions that have the potential to solve the intergenerational free rider challenge. The following citation should be used:

Stenzel, A., Lohse, J., Requate, T., & Waichman, I. (2024). Punishment with communication sustains altruistic cooperation across generations.

Finally, Chapter 6 concludes and puts the different results in perspective.

## Chapter 2

### Supply Chain Data Sharing for Scope 3 Emissions

#### Bibliographic Information

This chapter is reproduced with permission from Springer Nature and has been published in a modified version as: Stenzel, A., & Waichman, I. (2023). Supply chain data sharing for Scope 3 emissions. *npj Climate Action*, 2(1), 7.

#### Acknowledgements

We are grateful to Nicolas Duvoisin, Yasha Tarani, and Francesco Vogelezang for their invaluable comments and discussion.

#### Competing Interest

The authors declare no competing interests. Aurel Stenzel is co-founding member and co-chairman of the SINE Foundation. SINE is a non-profit think tank combining academics and technology entrepreneurs to promote data sharing for the common good. SINE is a technology partner in the Partnership for Carbon Transparency. Israel Waichman is a professor of Economics at Bard College Berlin and a co-founding member of the SINE Foundation.

#### Copyright Notice

Clarification of the copyright adjusted according to the guidelines of the publisher.

## 2.1 Introduction

Reducing greenhouse gas (GHG) emissions is among the momentous and most urgent challenges humanity is facing (IPCC, 2022; Tosun, 2022). The problem is that mitigating climate change is essentially a global social dilemma, where mitigation efforts are very costly for individual countries (or firms), but the benefits are shared globally and would be realized in the future (Barrett, 2007; Hauser et al., 2014; Battersby, 2017; Nordhaus, 2019; Lohse & Waichman, 2020; Buchholz & Sandler, 2021; Boon-Falleur et al., 2022). An additional barrier for successful mitigation stems from the considerable heterogeneities between countries regarding wealth, projected loss from climate change, historical emissions, etc. (Mendelsohn et al., 2006; Heyward, 2007; Lange et al., 2010; Burton-Chellew et al., 2013; Brown & Kroll, 2017; Bathiany et al., 2018; Waichman et al., 2021; IPCC, 2022).

We are in the midst of a digital revolution in production (Bojanova, 2014). With the Internet of Things and an increasing number of connected devices, raw data are often created as by-products of firms' daily operations at very low costs; for example, logistic data from a company (such as volumes of truckloads, GPS truck positions, truck routes, etc.). Following Varian's (2018) definition, data are raw alphanumeric values that need to be organized and contextualized to be turned into information. Knowledge is the conclusions and insights learned from that information. Data-based knowledge allows firms to gain better control over their production and supply chains and the opportunity to reduce their costs and emissions. While firms are hesitant to share confidential information and knowledge with other firms, they are more willing to share encrypted data (Jones & Tonetti, 2020).

This paper deals with data sharing among firms along a supply chain for the calculation of product-level Scope 3 emissions; e.g., data on all material and energy inputs, purchased product components, transportation, and their direct emissions. It answers the following questions: What are the benefits of primary data sharing for Scope 3 emissions? What are the barriers to successful emission data sharing along the supply chain? And what are the current developments in overcoming these obstacles? Before answering these questions, we provide some background on Scope 3 emissions and their importance.

Effective data sharing became possible due to recent technological developments (e.g., digitalization in production and processes, increase in computer

power, new encryption technologies, etc.). As mitigating climate change is a social dilemma, data sharing can increase the transparency of the situation and the accountability of the actors (Li et al., 2019), and can even transfer the social dilemma to a coordination game (Barrett & Dannenberg, 2012). Thus, the overarching motivation for sharing primary data for Scope 3 emissions is to optimally mitigate GHG emissions. In particular, by sharing company data with other firms, data are contextualized and turned into information. Latest innovations in cryptography allow the computation or analysis of shared data without requiring the parties to reveal their private inputs. The encryption reduces the risk for data owners to reveal sensitive data or lose control over it, and thus increases the willingness of firms to engage in data sharing (Jones & Tonetti, 2020). With information derived from their data sharing activities, firms can conclude (i.e., gain knowledge) about parts of the supply chain (e.g., emission reduction potentials).

We focus on Scope 3 emissions along the upstream supply chain (“cradle-to-gate”). In this regard, an important distinction is between “cradle-to-gate” and “cradle-to-grave.” The former includes all upstream emissions in the production process (i.e., along the supply chain up to the reporting firm), whereas the latter also includes downstream emissions from the consumption of the good. Currently, much of Scope 3 data estimates are calculated based on industry averages, financial information, and approximations (Li et al., 2019; WBCSD, 2021). Such approximations sometimes yield close estimates to the actual emissions of a product, but other times the derived values are much less accurate. For instance, Downie and Stubbs (2012) find that when conversion information is not available, there are wide discrepancies in reported emissions for the same activities among Australian firms. In this respect, corporate environmental performance (CEP) rating schemes do not yet provide a sufficiently good measure for emissions, and current CEP rating databases lack consistency (Rekker et al., 2021). The problem with using industry averages is that they may lead to insufficient incentives for firms to adopt cleaner-than-average technologies. By contrast, primary emission data that are shared along the supply chain can provide exact emissions at the product- and firm-level. Thus, it allows firms to tackle emissions reduction at the most effective point on the supply chain and also provides incentives for the adoption of cleaner technologies beyond the industry average.

The structure of the paper is as follows: In the next section, we introduce the notion of Scope 3 emissions. Section 3 reviews the main obstacles to

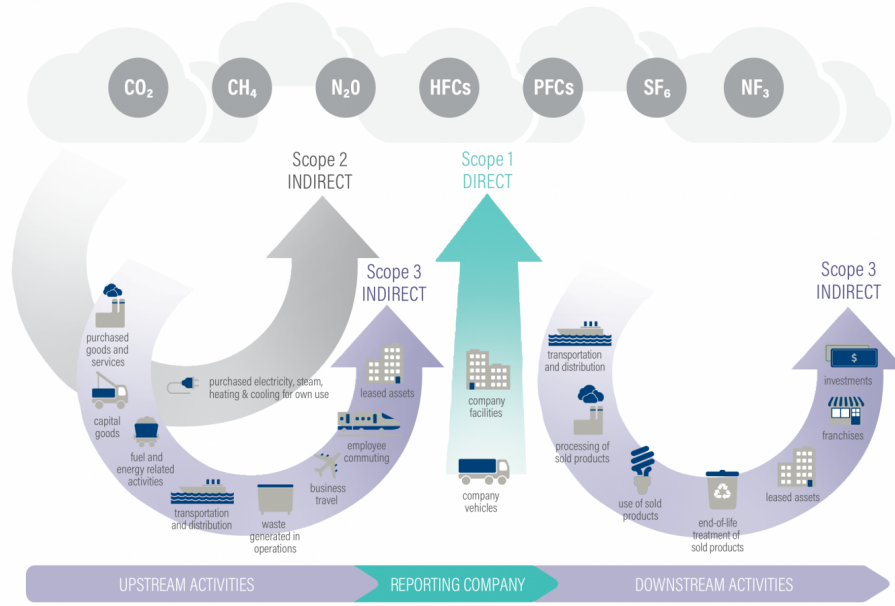


sharing primary data for Scope 3 emission reporting. Section 4 explains what is required to overcome the main data sharing obstacles and discusses the current approaches in overcoming these obstacles. Finally, Section 5 concludes.

## **2.2 Scope 3 emissions: Introduction and current estimation**

Emissions generated in the production process can be classified as direct or indirect emissions. Direct emissions are those created by the reporting entity itself, whereas indirect emissions are embodied in products as raw materials that are transformed along the supply chain (Scott et al., 2018). In this respect, the GHG Corporate Protocol provides a standardized definition for direct and indirect emissions in the production process (WRI/WBCSD, 2004). It classifies emissions as follows: Scope 1 emissions are direct emissions from owned or controlled sources, while Scope 2 and 3 emissions are indirect emissions. Scope 2 are emissions from the generation of acquired and consumed electricity, steam, heat, or cooling (WRI/WBCSD, 2015). Scope 3 emissions include all other emissions along the supply chain, both upstream emissions related to the purchased goods and services and downstream emissions related to sold goods and services (WRI/WBCSD, 2011, p. 31). In this paper, we focus on upstream Scope 3 emissions consisting of emissions from purchased goods and services, capital goods, transportation of goods and services, waste generated in operations, business travel, employees commuting, and leased assets, see Figure 2.1.

Under the corporate standard, firms are required to quantify and report Scope 1 and 2 emissions. However, to date, the reporting of Scope 3 emissions is recommended, but optional (Mytton, 2020; WBCSD, 2021). This is especially problematic because according to Matthews et al. (2008), Scope 3 emissions represent, on average, 84% of a firm’s total carbon emission footprint (whereas the remaining 16% are Scope 1 and 2). Moreover, in nearly two-thirds of all economic sectors, Scope 3 emissions represent more than 75% of their carbon emission footprint (Matthews et al., 2008). Similarly, Huang et al. (2009) find that Scope 3 emissions are responsible for 70–80% of firms’ total analyzed footprint for most manufacturing industries. Moreover, Hertwich and Wood (2018) find that while direct (Scope 1) emissions increased by 47% between 1995 and 2015, indirect emissions grew considerably more: Scope 2 emissions increased by 78% and Scope 3 emissions by 84%.



Source: (WRI/WBCSD, 2011, p. 5)

**Figure 2.1:** Scope 1, 2, and 3 emissions

To assess its Scope 3 emissions, a firm must map its total value chain (i.e., cradle-to-gate) and identify all emission sources. Often, a firm does not have access to the emission data from other firms along the supply chain, especially when firms are further away in the supply chain (Jaeger et al., 2022). The data used to calculate Scope 3 emissions should ideally be primary data. Primary data are data on actual emissions that are directly collected by firms (e.g., meter readings, purchase records, utility bills, engineering models, direct monitoring, etc.). With digital production processes, such data are produced as by-products or at very low costs. When we refer to data sharing, we refer to sharing of such primary data. When primary data are not available or shared along the supply chain, it becomes necessary to use secondary data to calculate Scope 3 emissions. Secondary data are not directly collected by the firms and are, e.g., industry average data, financial data, proxy data, and other generic data (WRI/WBCSD, 2013, p. 21). Such data could be used with so-called “environmentally extended input output (EEIO) models” (Matthews et al., 2008; Huang et al., 2009; Wiedmann, 2009; Kitzes, 2013; Hertwich & Wood, 2018). EEIO models attribute emissions to final demands, providing a method of accounting for the embodied emissions in different sectors and regions for different products.

## 2.3 The potential benefits of accurately calculating and reporting Scope 3 emissions

To overcome climate change, governments, firms, and consumers need to reduce GHG emissions. Reporting Scope 3 emissions increases transparency as it reveals the emissions in production processes, i.e., it allows firms to identify emissions as well as the most promising intervention points. Moreover, reporting leads to (at least some) accountability. Increasing accountability could increase cooperation through the actions of “conditional cooperators” (Keser & Van Winden, 2000; Fischbacher et al., 2001; Rustagi et al., 2010). Additionally, reporting opens the possibility for various types of sanctioning that have been shown to overcome social dilemmas (Fehr & Gächter, 2002; Tyran & Feld, 2006) and could facilitate coordination between firms along the supply chain. However, when firms report their Scope 3 emissions, the overall emissions along the supply chain may become a liability for the firms, as they may be forced to reduce those emissions (e.g., by adopting low-pollution technologies, switching to “cleaner” suppliers, and improving material efficiency (Liu et al., 2013; Scott et al., 2018)). Yet, for the individual firms, emission reporting helps in identifying potential cost savings as well as potential and future climate change-related risk factors, and in seeking new business opportunities (Kauffmann & Less, 2012).

Scope 3 reporting based on primary data (i.e., representing the actual production technologies of the individual firms instead of industry averages) is also useful for investors seeking to assess a firm’s strengths and reduce uncertainties. In fact, Matsumura et al. (2014) find “that the markets penalize all firms for their carbon emissions, but a further penalty is imposed on firms that do not disclose emissions information.” Moreover, if consumers are aware of the embodied emissions of competing products, they may consider this information in their purchase decisions (Korkala et al., 2014; Jürkenbeck et al., 2021). This would put pressure on firms to reduce emissions along the supply chain that could result in switching to low-emission suppliers, provide incentives for adopting low-pollution technologies, increasing material efficiency, etc. However, if Scope 3 emissions are reported based on secondary data of industry averages, then firms do not have an incentive to improve their processes beyond the industry average.

The current situation, where only Scope 1 and 2 emissions must be reported, may lead to a re-allocation of emissions to industries with stronger

lobbies (i.e., with a lower chance to be regulated), or to “carbon leakage” to regions with less strict environmental standards. Since GHG emissions matter globally, this may impair the fight against climate change. In this context, Mytton (2020) reports that when firms deploy information technology to the cloud, they switch their emissions from Scope 1 and 2 to Scope 3 emissions. Given the current reporting policy, firms may avoid reporting their Scope 3 emissions and thus brand themselves as “cleaner” (as they have now lower Scope 1 and 2 emissions), although they did not reduce emissions along the supply chain. Another point is that for emission trading institutions (e.g., EU ETS) to work efficiently, firms need to be aware of their actual emissions. Precise reporting (which can only be attained by sharing primary emission data along the supply chain) will also increase the efficiency of the emission permit trading systems.

Therefore, the overall consequences of not reporting Scope 3 emissions based on primary data (i.e., using secondary data that may or may not be close to the true values) may lead to a distorted climate policy (Hertwich & Wood, 2018), in particular to a less-than-optimal emission mitigation. Finally, firms that share data with each other may establish partnerships and also generate new business models, more efficient supply chains, and revenue from the monetization of their data (European Commission, 2018).

## 2.4 Obstacles to primary data sharing

Today’s value chains are complex networks that include multiple tiers of different suppliers across the globe (Hearnshaw & Wilson, 2013; Patchell, 2018). To (precisely) calculate Scope 3 emissions based on primary data, firms need access to data that are not under their control; all suppliers (in all tiers) need to measure and share relevant data with each other. However, some firms lack the in-house knowledge and personnel to prepare the required data for Scope 3 in a sharable way and would need to make significant up-front investments (OECD, 2019). This is especially problematic for small and medium-sized firms with smaller operations, less personnel, and smaller budgets. And since sharing such data could also lead to a competitive disadvantage, as it might reveal confidential information on the production processes (e.g., costs, technology), firms are often hesitant to make the required investments (Kraemer et al., 2021).

In our analysis, we do not reflect on questions of data measurement. In

fact, FERF (2014) has estimated that the internal costs of accounting constitute between 1 (for large firms) and 3 percent (for smaller firms) of the revenues (Patchell, 2018, p. 949). Instead, we focus on the obstacles to sharing available emission data and review the following issues: lack of legal clarity and regulatory concerns, lack of data and action interoperability (allowing two or more IT systems to exchange data which is understood and can be re-used by the other system), and high risk in sharing sensitive data (OECD, 2019; Martens et al., 2020; WBCSD, 2021; European Commission, 2021, 2018).

#### **2.4.1 Legal and regulatory challenges**

The rapid growth of global data volumes raises multiple legal and regulatory questions, for example, with regard to data ownership, access to, and re-use of data (Drexler, 2017; Wiebe, 2017; Hummel et al., 2021; Asswad & Gómez, 2021). For Scope 3 emissions, currently, firms do not have access to all the data they need and face (contractual) limitations when they re-share received data as part of their own emission footprint to the next tier in the supply chain. In addition, individual countries have started to protect data for local value creation by implementing an increasing amount of data localization measures (Wu, 2021); i.e., measures that limit data flows across borders to not lose control over the data. An example for such a data localization measure is article 27 of the Data Act by the European Commission: It requires cloud computing providers to prevent the international transfer of non-personal data where such transfer might create a conflict with Union law.<sup>1</sup> A review of different regulations and their impact on non-personal data sharing in general is offered by SCDS (2020) and Leistner and Antoine (2022). For example, data sharing between firms may infringe anti-competition law, e.g., by fixing prices (OECD, 2019). In particular, Scope 3 emissions data could include competition-relevant information such as production capacity, truck routes, etc. (WBCSD, 2021). Contrary to information and knowledge, data and their specific use regarding Scope 3 calculations are not covered by existing laws. For example, the Copyright DSM Directive by the European Commission grants copyright protection to information and knowledge representing a creative expression, but not to non-creative factual data (SCDS, 2020). Therefore, firms must decide on a case-by-case basis which data can be shared for what purpose (European Commission, 2018). Furthermore, “[u]nclear messages from govern-

<sup>1</sup> Retrieved November 8, 2022 from <https://digitalstrategy.ec.europa.eu/en/library/data-act-proposal-regulation-harmonized-rules-fair-access-and-use-data>

ments on future regulations or major changes in reporting methodology may discourage early movers from investments in a specific reporting [(and data sharing)] infrastructure and generate a wait-and-see attitude [among firms]” (Kauffmann & Less, 2012, p. 25). In particular, the regulatory body should commit to a clear strategy for future emission reporting requirements (e.g., regarding precision and timing), as without clear legal commitments, firms may be hesitant to report Scope 3 emissions and rather opt to “let sleeping dogs lie,” fearing that reporting might lead to legal implication without urgency (Kauffmann & Less, 2012, p. 29). In sum, these legal and regulatory challenges lead to high uncertainty and costs for supply chain data sharing.

#### **2.4.2 Interoperability**

A study by the European Commission identifies missing interoperability as one of the main obstacles to more data sharing across firms in general (European Commission, 2018). Interoperability can be defined as “a measure of the degree to which diverse systems, organizations, and/or individuals are able to work together to achieve a common goal” (Ide & Pustejovsky, 2010, p. 2). In particular, the lack of interoperability regarding Scope 3 emissions is twofold: First, a lack of harmonized standards to measure GHG emissions, and second, the absence of a common infrastructure to technically exchange emission data across IT systems. Existing standards (e.g., ISO standards, GHG Protocol standards) and sector guidelines (such as the Product Category Rules or Product Environmental Footprint method) leave room for interpretation (ISO, 2019; WBCSD, 2021; European Commission, 2021). Scope 3 emission footprints build up on top of the emission data input of different suppliers but lack an overarching harmonization and integration (Jaeger et al., 2022). With missing data interoperability, i.e., when suppliers within a supply chain use different calculation standards or interpret certain inputs differently, the resulting Scope 3 emission footprint is inconsistent and insufficient. But even if all firms along a supply chain agree to a certain calculation standard, they still face a lack of action interoperability, i.e., most IT systems cannot exchange data with other systems across company boundaries. In a global value chain, including very small producers, a variety of IT systems are used. Current data exchange is realized with high manual efforts and surveys or spreadsheets, leading to high costs (WBCSD, 2021).

### **2.4.3 Data privacy**

Another major obstacle to data sharing is the risk of sharing sensitive data. Product-level emission data can enable reverse engineering to conclude insights into production processes (e.g., information about product composition and supplier networks) and is therefore considered to be competitively relevant (Jaeger et al., 2022). Firms are very hesitant to share such data. For example, a survey among producing firms revealed that 42% fear losing innovative or competitively relevant knowledge if data are shared or locked into one data platform (EFI, 2022). Moreover, Pauer et al. (2018) conducted a survey with executives from large enterprises as well as small and medium-sized firms throughout Germany: 57% of the respondents stated that “anxiety about core data and business secrets being exposed” is a “very big [or] rather a big obstacle” for data and information sharing, 51% expressed that “sharing competitive information is dangerous because it might create advantages for competitors.” Furthermore, 59% of the respondents stated that it is a “very big [or] rather big obstacle” that “we cannot check who is going to read our data if they are uploaded to the platform” (Pauer et al., 2018, p. 25). Those data privacy concerns sum up to a significant obstacle to sharing primary data and thus enabling other firms to calculate their emission footprint.

## **2.5 Overcoming the obstacles to primary data sharing**

There is already a large (and constantly increasing) number of initiatives seeking to address the obstacles reported in the previous section and thus promote the successful sharing of emission data along the supply chain. In this section, we discuss different initiatives and their approaches (focusing on the obstacles as categorized above rather than on individual initiatives).

### **2.5.1 Overcoming legal and regulatory challenges**

Private firms have only limited possibilities to influence legal and regulatory environments. It is the task of policymakers and regulators (e.g., international bodies such as the United Nations and its subsidies, European Commission, or national governments) to remove legal and regulatory barriers to data sharing. First, they could make Scope 3 emission reporting obligatory (regional attempts have been made, e.g., the Corporate Sustainability Reporting Directive of the European Union or the Enhancement and Standardization of

Climate-Related Disclosures of the U.S. Securities and Exchange Commission). Second, policymakers and regulators have the task to ensure “that information provided by [firms] is timely, reliable and relevant” (Kauffmann & Less, 2012, p. 24). The more the precision requirement is emphasized, the more important it becomes to share primary emission data (as it is the only means to provide actual numbers rather than estimates). For example, under the UK Carbon Reduction Commitment Program, the Environment Agency conducts third-party audits of 20% of the participants every year. In Japan, entities that submit false reports or fail to meet the reporting criteria can be fined (Kauffmann & Less, 2012, p. 38). In May 2022, the U.S. Securities and Exchange Commission fined the Bank of New York Mellon USD 1.5 million for misstatements in its sustainability reporting.<sup>2</sup> Third, when governments commit to a future date where Scope 3 emission reporting would be mandatory, they could provide incentives for first movers. Fourth, to meet the global dimension of supply chains, regulatory bodies could create regulations to enable the free movement of data across borders (e.g., regulation on the free flow of non-personal data by the European Commission).

### **2.5.2 Overcoming missing interoperability**

Interoperability in emission data sharing brings a twofold challenge: First, to create a harmonized standard to measure GHG emissions, and second, to create a common infrastructure to technically exchange emission data across IT systems. To address the first challenge of interoperability, standard setters (e.g., Greenhouse Gas Protocol, ISO, European Commission) strive to establish one harmonized standard on how to measure Scope 3 emissions within supply chains in a consistent way. The ISO developed and is still working on various standards to measure emission data (e.g., ISO 14067:2018 defines requirements and guidelines for measuring the carbon footprint of products). Additionally, the European Commission published the Product Environmental Footprint method (PEF), quantifying the environmental impact of products, and the Product Environmental Footprint Category Rules (PEFCR), defining guidance at the level of specific product categories. Private coalitions or public-private partnerships (e.g., World Business Council for Sustainable Development (WBCSD), SME Climate Hub, CEO Alliance, We Mean Business Coalition) can use their market reach to facilitate overarching conversations and to establish a harmonized standard across industries. For example,

---

<sup>2</sup> Retrieved November 7, 2022 from <https://www.sec.gov/news/press-release/2022-86>



the WBCSD has initiated the Partnership for Carbon Transparency (PACT), which brings together leading firms from a range of industries, technology players, industry-focused initiatives, standard-setting organizations, reporting bodies, and regulators seeking to define an overarching standard for Scope 3 emissions. PACT launched a coordinated and open approach to calculate product carbon footprints (PCFs) along the whole value chain. The partnership published a methodology of standards for the calculation of PCFs that leverages and aligns with existing methods and standards, e.g., GHG Protocol, PEF, and PEFCR, and removes the room for interpretation and inconsistency (WRI/WBCSD, 2004, 2011, 2013, 2015; Rekker et al., 2021).

To solve the issue of action interoperability, i.e., the ability of different IT systems to exchange information, PACT, together with the non-profit SINE Foundation, launched a technical infrastructure for data sharing based on open-source technology. This infrastructure defines rules for data sharing, such as data formats and standards describing how to connect the different IT systems used by firms across the value chain (WBCSD, 2021). PACT initiated a collaborative process in which large corporations and software providers developed technical core components of the network together. With the standards defined in the methodology, the infrastructure is open and interoperable, connecting different IT systems. Another example of an open-source infrastructure for sustainability data sharing is OS-Climate, an open-source project backed by the Linux Foundation, building technology and data platforms focusing on data availability, comparability, and reliability. The goal is to provide data about climate change impact and incorporate that data into global financial decision-making and risk management.

As a non-profit organization consisting of multiple firms from the automotive industry, the Catena-X Automotive Network created a uniform standard for data exchange along the entire value chain. It connects all members in the automotive supply chain without lock-in effects, provides a sustainable solution for the digitalization of supply chains, especially for small and medium-sized firms, and supports the cooperation and collaboration of market participants and competitors. Catena-X allows for different kinds of data sharing, including emission data. Finally, the Smart Freight Center was established in 2013 as a non-profit organization and works with the global logistics community to drive transparency and industry action. To harmonize the calculation and reporting of the logistics emission footprints, the Smart Freight Center published the Global Logistics Emissions Council Framework, which

can be implemented by shippers, carriers, and logistics service providers.

To overcome the obstacle of interoperability, the different initiatives need to remain open and collaborative to establish one harmonized standard across countries and industries for consistent data measurement and the exchange of emission data.

### **2.5.3 Overcoming data privacy concerns**

Ensuring data privacy is a necessary condition for successful data sharing between firms. One promising possibility to enable data sharing while protecting the firms' privacy is data trusts, neutral stewards that manage a firm's or person's data on their behalf. A data trust "works within the law to provide ethical, architectural, and governance support for trustworthy data processing" (O'hara, 2019, p. 6). A data trust has two key functions: First, providing data holders the possibility to execute their rights and set limitations on who can do what with their data. Second, providing a technical environment in which data processing and analysis can take place while holding everyone accountable for his or her actions. Organizations such as the SINE Foundation, OS-Climate, or Catena-X were established as non-profit organizations to be able to function as data trusts for the exchange of sustainability data. Besides a data governance, data trusts also need to provide technical tools such as encryption technologies to protect the data.

Academic research on so-called homomorphic encryption technologies has now become possible in practice due to the increase in computing power (Royal Society, 2019; WEF, 2019). Without homomorphic encryption technology, data need to be revealed to be utilized (leading to a decrease in the privacy of the sharing party, i.e., the privacy-personalization paradox (Kaaniche et al., 2020)). Examples of homomorphic encryption are "zero-knowledge proofs" and "secure multi-party computing." With "zero-knowledge proofs" (Goldwasser et al., 1989), a party can prove that it meets certain criteria (e.g., proving that a product carbon footprint is below a certain threshold) without revealing any other information. "Secure multi-party computing" (Yao, 1986) is a protocol that allows computation or analysis of combined data without the different parties revealing their private input (Bogdanov et al., 2012). A large pool of data is always a security risk and will be luring cyber attackers. One solution (used, e.g., by PACT) is to allow the firms to keep their data decentralized and share it peer to peer, which decreases the attractiveness for hackers. Data trusts and homomorphic encryption technologies, in

combination with decentralized data storage, enable firms to overcome the privacy-personalization paradox and thus solve the data privacy concerns as an obstacle for emission data sharing.

Finally, whenever a firm holds confidential data or information, there are issues of cybersecurity involved. Cybersecurity refers to “the set of technologies, processes, and practices designed to protect networks, computers, programs, and data from attack, damage, or unauthorized access, in accordance with the common information security goals: the protection of confidentiality, integrity, and availability of information” (Dunn Cavelty, 2013, p. 105). The global cybersecurity market is constantly increasing. Fortune Business projected it “to grow from USD 155.83 billion in 2022 to USD 376.32 billion by 2029”<sup>3</sup> and private cybersecurity firms will be part of securing the data and information.

#### 2.5.4 Summary

Table. 2.1 presents a short summary of the three main obstacles to primary data sharing and current approaches to overcome them.

**Table 2.1:** Obstacles to data sharing and current approaches

	<b>Legal Clarity and Regulatory Concerns</b>	<b>Missing Data and Action Interoperability</b>	<b>High Risk in Sharing Sensitive Data</b>
<b>Description</b>	Firms lack access to supply chain data due to contractual limitations (e.g., anti-trust); data localization hinders cross-border data flow; high regulatory costs due to uncertainty in emissions reporting.	Lack of harmonized standards for measuring Scope 3 emissions; no common infrastructure for exchanging emission data across IT systems.	Risk of data exposure and reverse engineering (e.g., product composition, supplier networks).
<b>Current Approaches</b>	Transition to mandatory and Scope 3 reporting; focus on reliable and verifiable emission data; commitment to future reporting policy; regulations supporting non-personal data flow.	Cross-industry coalitions, technology players, and standard-setting organizations define unified standards and technical infrastructures.	Neutral data trusts empower data holders and offer secure data processing and analysis; decentralized data storage with homomorphic encryption; cybersecurity firms protect confidential data.

<sup>3</sup> Retrieved March 27, 2022 from <https://www.fortunebusinessinsights.com/industry-reports/cybersecurity-market-101165>

## 2.6 Concluding remarks

This article underlines the benefits of and the obstacles to sharing primary data for the calculation of Scope 3 emissions along the supply chain. We started by defining Scope 3 emissions, their relevance in sustainability reporting, and why their precise calculation (beyond industry averages) requires data sharing among firms. While the ongoing digitization of production is reducing the costs of data measurement, data sharing among firms requires overcoming three main obstacles: legal and regulatory challenges, missing interoperability, and data privacy concerns. The paper reviewed these obstacles and discussed the current progress in overcoming them.

Nowadays, the production of goods is a global and complex process involving several suppliers. The number of sustainability start-ups focusing on the measurement of Scope 3 emissions is increasing every day. Many of them provide estimates based on industry averages. While such data cannot provide precise emission values, they may serve as starting points. However, only primary data sharing provides actual emission numbers and thus has the potential to accelerate the decarbonization of the economy. Primary data sharing may also provide additional benefits to individual firms and enable new business models within digitized supply chains. For example, to account for transportation emissions, a firm needs to share data on the mode of transportation (e.g., road, rail), distance covered, and load specifications (WBCSD, 2021). If these data are shared in real-time, firms could use them to coordinate collective logistic optimization. For example, Procter & Gamble (P&G) and Tupperware shared their logistics data and identified significant potential for load consolidation. They used similar routes, and their truckloads were complementary and thus offered substantial potential gains from coordination: Tupperware's trucks were filled to around 80% of the maximum volume of the vehicle but only 30% of its weight—while P&G's trucks were at 50% of the maximum volume of the vehicle but 95% of its weight. With real-time data sharing, they coordinated a consolidation of their loads, saving 17% of the transportation costs and more than 200 Mt CO<sub>2</sub>.<sup>4</sup> In addition, Pedreira and Melo (2020) suggest a possible reduction of CO<sub>2</sub> emissions of up to 23% if all firms sought similar arrangements as described above. Irannezhad et al. (2018) conducted a study at the Port of Brisbane showing an emission reduction potential of

---

<sup>4</sup> Retrieved March 27, 2022, from [https://www.eknowit.eu/the\\_case\\_studies/case\\_study02.pdf](https://www.eknowit.eu/the_case_studies/case_study02.pdf)

up to 40% if the logistic firms operating at the port would share their data to coordinate their activities (e.g., to reduce the number of empty trucks).

Thus, the potential benefits of data sharing among firms (e.g., logistic consolidation) are even larger than the mere benefits of calculating Scope 3 emissions. In order to utilize the sharing of primary data for the calculation of Scope 3 emissions and beyond, different stakeholders, including regulatory bodies, standard setters, climate initiatives, research institutes, and private firms, need to take a systemic and global perspective to overcome the three major obstacles for data sharing along supply chains.

## Bibliography

- Asswad, J., & Gómez, J. M. (2021). Data ownership: A survey. *Information*, 12(11), 465.
- Barrett, S. (2007). *Why Cooperate? The Incentive to Supply Global Public Goods*. Oxford University Press, USA.
- Barrett, S., & Dannenberg, A. (2012). Climate negotiations under scientific uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, 109(43), 17372–17376.
- Bathiany, S., Dakos, V., Scheffer, M., & Lenton, T. M. (2018). Climate models predict increasing temperature variability in poor countries. *Science Advances*, 4(5).
- Battersby, S. (2017). News feature: Can humankind escape the tragedy of the commons? *Proceedings of the National Academy of Sciences*, 114(1), 7–10.
- Bogdanov, D., Talviste, R., & Willemson, J. (2012). Deploying secure multi-party computation for financial data analysis. In *Financial Cryptography and Data Security: 16th International Conference, FC 2012, Kralendijk, Bonaire, Februray 27-March 2, 2012* (pp. 57–64).
- Bojanova, I. (2014). The digital revolution: What’s on the horizon. *IT Professional*, 16(1), 8–12.
- Boon-Falleur, M., Grandin, A., Baumard, N., & Chevallier, C. (2022). Leveraging social cognition to promote effective climate change mitigation. *Nature Climate Change*, 12, 332–338.
- Brown, T. C., & Kroll, S. (2017). Avoiding an uncertain catastrophe: Climate change mitigation under risk and wealth heterogeneity. *Climatic Change*, 141(2), 155–166.
- Buchholz, W., & Sandler, T. (2021). Global public goods: A survey. *Journal of Economic Literature*, 59(2), 488–545.
- Burton-Chellew, M. N., May, R. M., & West, S. A. (2013). Combined inequality in wealth and risk leads to disaster in the climate change game. *Climatic Change*, 120(4), 815–830.

- Downie, J., & Stubbs, W. (2012). Corporate carbon strategies and greenhouse gas emission assessments: The implications of scope 3 emission factor selection. *Business Strategy and the Environment*, 21(6), 412–422.
- Drexl, J. (2017). Designing competitive markets for industrial data - between proprietisation and access. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 8, 257–292.
- Dunn Cavelty, M. (2013). From cyber-bombs to political fallout: Threat representations with an impact in the cyber-security discourse. *International Studies Review*, 15(1), 105–122.
- EFI. (2022). *Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands*. Retrieved April 24, 2022, from <https://www.e-fi.de/publikationen/gutachten>
- European Commission. (2018). *Study on Emerging Issues of Data Ownership, Interoperability, (Re-)Usability and Access to Data, and Liability. Final Report*. Retrieved August 17, 2023, from <https://op.europa.eu/en/publication-detail/-/publication/74cca30c-4833-11e8-be1d-01aa75ed71a1/language-en>
- European Commission. (2021). *Commission Recommendation of 16.12.2021 on the Use of the Environmental Footprint Methods to Measure and Communicate the Life Cycle Environmental Performance of Products and Organisations*. Retrieved August 17, 2023, from [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=PI\\_COM%3AC%282021%299332](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=PI_COM%3AC%282021%299332)
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.
- FERF. (2014). *Benchmarking: The Accounting and Finance Function*. Morristown N.J.. (Financial Executives Research Foundation)
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397–404.
- Goldwasser, S., Micali, S., & Rackoff, C. (1989). The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1), 186–208.
- Hauser, O. P., Rand, D. G., Peysakhovich, A., & Nowak, M. A. (2014). Cooperating with the future. *Nature*, 511, 220–223.
- Hearnshaw, E. J., & Wilson, M. M. (2013). A complex network approach to supply chain network theory. *International Journal of Operations and Production Management*, 33(4), 442–469.

- Hertwich, E. G., & Wood, R. (2018). The growing importance of scope 3 greenhouse gas emissions from industry. *Environmental Research Letters*, 13(10), 104013.
- Heyward, M. (2007). Equity and international climate change negotiations: A matter of perspective. *Climate Policy*, 7(6), 518–534.
- Huang, Y. A., Weber, C. L., & Matthews, H. S. (2009). Categorization of scope 3 emissions for streamlined enterprise carbon footprinting. *Environmental Science and Technology*, 43(22), 8509–8515.
- Hummel, P., Braun, M., & Dabrock, P. (2021). Own data? Ethical reflections on data ownership. *Philosophy and Technology*, 34(3), 545–572.
- Ide, N., & Pustejovsky, J. (2010). What does interoperability mean, anyway? Toward an operational definition of interoperability for language technology. In *Proceedings of the Second International Conference on Global Interoperability for Language Resources. Hong Kong, China*.
- IPCC. (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (H.-O. Pörtner et al., Eds.). Cambridge, UK and New York, NY, USA: Cambridge University Press.
- Irannezhad, E., Prato, C. G., & Hickman, M. (2018). The effect of cooperation among shipping lines on transport costs and pollutant emissions. *Transportation Research Part D: Transport and Environment*, 65, 312–323.
- ISO. (2019). *Greenhouse Gases – Carbon Footprint of Products – Requirements and Guidelines for Quantification*. International Organization for Standardization (ISO 14067:2018). Retrieved August 17, 2023, from <https://www.iso.org/standard/71206.html>
- Jaeger, F. A., Saling, P., Otte, N., Steidle, R., Bollen, J., Golembiewski, B., ... Wunderlich, J. (2022). Challenges and requirements of exchanging product carbon footprint information in the supply chain. *E3S Web Conf.*, 349, 07005.
- Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, 110(9), 2819–2858.
- Jürkenbeck, K., Spiller, A., & Schulze, M. (2021). Climate change awareness of the young generation and its impact on their diet. *Cleaner and Responsible Consumption*, 3, 100041.
- Kaaniche, N., Laurent, M., & Belguith, S. (2020). Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and



- survey. *Journal of Network and Computer Applications*, 171, 102807.
- Kauffmann, C., & Less, C. T. (2012). *Corporate Greenhouse Gas Emission Reporting - A Stocktaking of Government Schemes, OECD Working Papers on International Investment*. Paris, France: OECD Publishing.
- Keser, C., & Van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, 102(1), 23–39.
- Kitzes, J. (2013). An introduction to environmentally-extended input-output analysis. *Resources*, 2(4), 489–503.
- Korkala, E. A., Hugg, T. T., & Jaakkola, J. J. (2014). Awareness of climate change and the dietary choices of young adults in Finland: A population-based cross-sectional study. *PLoS ONE*, 9(5).
- Kraemer, J., Stüdlein, N., & Zierke, O. (2021). Data as a public good: Experimental insights on the optimal design of B2B data sharing platforms. *SSRN Electronic Journal*.
- Lange, A., Löschel, A., Vogt, C., & Ziegler, A. (2010). On the self-interested use of equity in international climate negotiations. *European Economic Review*, 54(3), 359–375.
- Leistner, M., & Antoine, L. (2022). *IPR and the Use of Open Data and Data Sharing Initiatives by Public and Private Actors*. Study commissioned by the European Parliament’s Policy Department for Citizens’ Rights and Constitutional Affairs at the request of the Committee on Legal Affairs. Retrieved August 17, 2023, from [https://www.europarl.europa.eu/thinktank/en/document/IPOL\\_STU\(2022\)732266](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2022)732266)
- Li, M., Wiedmann, T., & Hadjikakou, M. (2019). Enabling full supply chain corporate responsibility: Scope 3 emissions targets for ambitious climate change Mitigation. *Environmental Science and Technology*, 54(1), 400–411.
- Liu, G., Bangs, C. E., & Müller, D. B. (2013). Stock dynamics and emission pathways of the global aluminium cycle. *Nature Climate Change*, 3(4), 338–342.
- Lohse, J., & Waichman, I. (2020). The effects of contemporaneous peer punishment on cooperation with the future. *Nature Communications*, 11(1815).
- Martens, B., De Streel, A., Graef, I., Tombal, T., & Duch Brown, N. (2020). *Business-to-business data sharing: An economic and legal analysis* (Vol. No. 2020-0). Seville: Joint Research Centre (JRC).

- Matsumura, E. M., Prakash, R., & Vera-Muñoz, S. C. (2014). Firm-value effects of carbon emissions and carbon disclosures. *Accounting Review*, 89(2), 695–724.
- Matthews, H. S., Hendrickson, C. T., & Weber, C. L. (2008). The importance of carbon footprint estimation boundaries. *Environmental Science and Technology*, 42(16), 5839–5842.
- Mendelsohn, R., Dinar, A., & Williams, L. (2006). The distributional impact of climate change on rich and poor countries. *Environment and Development Economics*, 11(2), 159–178.
- Mytton, D. (2020). Hiding greenhouse gas emissions in the cloud. *Nature Climate Change*, 10(8), 701–701.
- Nordhaus, W. (2019). Climate change: The ultimate challenge for economics. *American Economic Review*, 109(6), 1991–2014.
- OECD. (2019). *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*. Paris, France: OECD Publishing.
- O’hara, K. (2019). *Data trusts: Ethics, architecture and governance for trustworthy data stewardship*. University of Southampton.
- Patchell, J. (2018). Can the implications of the GHG Protocol’s scope 3 standard be realized? *Journal of Cleaner Production*, 185, 941–958.
- Pauer, A., Nagel, L., Fedkenhauser, T., Yvonne, F.-S., & Resetko, A. (2018). *Data Exchange as a First Step towards a Data Economy*. Retrieved August 17, 2023, from <https://www.pwc.de/en/digitale-transformation/data-exchange-as-a-first-step-towards-data-economy.pdf>
- Pedreira, H. B., & Melo, T. (2020). *Supply chain coopetition: A simulation model to explore competitive advantages in logistics*. Retrieved March 27, 2022, from <https://ctl.mit.edu/pub/thesis/supply-chain-coopetition-simulation-model-explore-competitive-advantages-logistics>
- Rekker, S. A., Humphrey, J. E., & O’Brien, K. R. (2021). Do sustainability rating schemes capture climate goals? *Business and Society*, 60(1), 125–160.
- Royal Society. (2019). *Protecting Privacy in Practice: The Current Use, Development and Limits of Privacy Enhancing Technologies in Data Analysis*. Retrieved April 24, 2022, from <https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/>
- Rustagi, D., Engel, S., & Kosfeld, M. (2010). Conditional cooperation and costly monitoring explain success in forest commons management. *Sci-*

- ence, 330(6006), 961–965.
- SCDS. (2020). *Analytical Report on EU Law Applicable, to Sharing of Non-Personal Data*. EU Support Centre for Data Sharing.
- Scott, K., Roelich, K., Owen, A., & Barrett, J. (2018). Extending European energy efficiency standards to include material use: An analysis. *Climate Policy*, 18(5), 627–641.
- Tosun, J. (2022). Addressing climate change through climate action. *Climate Action*, 1, 1-4.
- Tyran, J. R., & Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics*, 108(1), 135–156.
- Varian, H. (2018). Artificial intelligence, economics, and industrial organization. In *The Economics of Artificial Intelligence: An Agenda* (pp. 399–419). Chicago, USA and London, UK: University of Chicago Press.
- Waichman, I., Requate, T., Karde, M., & Milinski, M. (2021). Challenging conventional wisdom: experimental evidence on heterogeneity and coordination in avoiding a collective catastrophic event. *Journal of Environmental Economics and Management*, 109, 102502.
- WBCSD. (2021). *Pathfinder Framework: Guidance for the Accounting and Exchange of Product Life Cycle Emissions*. Geneva, Switzerland: World Business Council for Sustainable Development (WBCSD).
- WEF. (2019). *The Next Generation of Data-Sharing in Financial Services: Using Privacy Enhancing Techniques to Unlock New Value*. Retrieved April 24, 2022, from [https://www3.weforum.org/docs/WEF\\_Next\\_Gen\\_Data\\_Sharing\\_Financial\\_Services.pdf](https://www3.weforum.org/docs/WEF_Next_Gen_Data_Sharing_Financial_Services.pdf)
- Wiebe, A. (2017). Protection of industrial data – a new property right for the digital economy? *Journal of Intellectual Property Law and Practice*, 12(1), 62–71.
- Wiedmann, T. (2009). A review of recent multi-region input-output models used for consumption-based emission and resource accounting. *Ecological Economics*, 69(2), 211–222.
- WRI/WBCSD. (2004). *GHG Protocol Corporate Accounting and Reporting Standard, Revised Edition*. Washington DC: World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD).
- WRI/WBCSD. (2011). *Corporate Value Chain (Scope 3) Accounting and Reporting Standard: Supplement to the GHG Protocol Corporate*. Washing-

- ton DC: World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD).
- WRI/WBCSD. (2013). *Technical Guidance for Calculating Scope 3 Emissions: Supplement to Corporate Value Chain (Scope 3) Accounting and Reporting Standard*. Washington DC: World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD).
- WRI/WBCSD. (2015). *GHG Protocol Scope 2 Guidance*. Washington DC: World Resources Institute (WRI) and World Business Council for Sustainable Development (WBCSD).
- Wu, E. (2021). *Sovereignty and Data Localization. Report: Belfer Center for Science and International Affairs, Harvard Kennedy School*. Retrieved November 5, 2022, from <https://www.belfercenter.org/sites/default/files/2021-07/SovereigntyLocalization.pdf>
- Yao, A. C.-C. (1986). How to generate and exchange secrets. *27th Annual Symposium on Foundations of Computer Science*, 162–167.

## **Chapter 3**

# **The Efficacy of (Self-)Verification Instruments in Risky Investments: An Experiment**

### **Bibliographic Information**

This chapter shall be cited as: Stenzel, A., Requate, T., & Waichman, I. (2024). The Efficacy of (Self-) Verification Instruments in Risky Investments: An Experiment.

### **Competing Interest**

The authors declare no competing interests.

### 3.1 Introduction

Truthful information sharing and trust in it are crucial elements for successful economic systems in general and investments in particular (Arrow, 1968; Knack & Keefer, 1997; Zak & Knack, 2001; Guiso et al., 2004). Their absence can lead to adverse selection resulting in suboptimal investments (Akerlof, 1970; Viscusi, 1978; Milgrom, 2008). To overcome this problem, investors and firms have developed trust- and reputation-based mechanisms and instruments to facilitate efficient trading (e.g., Milgrom et al., 1990; Greif, 1993). However, in our digitized and globalized economy, which connects investors and firms from different countries, such mechanisms may not be sufficient. As a result, verification and self-verification (or “quality disclosure,” e.g., Dranove & Jin, 2010; Bernstein et al., 2017; Bapnaa, 2019) institutions and processes (e.g., due diligence, rating systems) have been developed (see Milgrom, 2008). Typically, these processes are very costly, often require a third party, and may not provide accurate or timely information (Hynes, 2017). Recent technological innovations make it possible to share truthful information at a low cost. For example, blockchain-based platforms<sup>1</sup> (when combined with suitable encryption technology) could enable firms to reliably self-reveal financial and non-financial information to investors (e.g., Gephart et al. (2019) propose a blockchain to verify sustainable humane seafood supply chains). The aim of this study is to investigate whether the availability of a low-cost self-verification instrument (emerging with the advances in technology and computational power) improves investment decisions and, thus, the society’s resource allocation compared to a situation that is based on trust alone.

To this end, we use a modified sender-receiver game that mimics key features of the investor-firm relationship. In the “Probabilistic Investment Game,” each investor is provided with an endowment of money that can be invested in a firm. Each firm has either a low (between 10% and 20%) or a high (between 70% and 80%) probability of success, drawn from a discrete uniform distribution (which is common knowledge to all firms and investors).

---

<sup>1</sup> Blockchain can be roughly described as a decentralized ledger in which transactions are simultaneously recorded in a variety of databases (see Tapscott & Tapscott, 2017; Yaga et al., 2019; Ornes, 2019; Rajasekaran et al., 2022). This makes data entries immutable and thus optimal for low-trust environments. Used in combination with encryption technologies, blockchain can save considerable time and cost for certification while providing accurate information in real time (Abbe et al., 2012; Bogdanov et al., 2012; Wang & Kogan, 2018; Gonzalez, 2019; Y. Chen & Bellavitis, 2020; Rajasekaran et al., 2022).

Therefore, the a priori expected probability of success for the investor is 45%. Success means that both the investor and the firm earn a profit equal to the amount invested by the investor (in addition, the investor keeps the initial amount of money), while failure means that the investment is lost. Under these conditions, the investor’s a priori best choice is not to invest. The baseline game has two stages: First, the firm is informed about its probability of success and sends a free message to the investor. The message may reveal the true probability of success or any other value of the firm’s choice (“cheap talk”). Second, the investor receives the message and decides how much of his or her endowment to invest in the firm. Afterward, a lottery based on the firm’s true probability of success determines whether the investment is successful or not. The game is repeated for 15 rounds with fixed roles and random matching (i.e., firms and investors remain in their roles throughout the experiment but do not stay with the same investor and firm).

In addition to the *Baseline* (T1), we conducted three other treatments. In *Costly Verification* (T2), after receiving the free message, investors decide whether to verify the firm’s probability of success at a cost. In *Costly Self-Verification* (T3), firms decide whether to verify their probability of success to investors at a price or instead send a free message that may not reveal the true probability of success. *Self-Verification* (T4) is the same as T3, except that firms can now self-verify their probability of success at a much lower cost. In particular, *Costly Verification* (T2) and *Costly Self-Verification* (T3) mimic the current practice where either investors can undertake costly due diligence (T2), or firms can undertake costly self-verification, e.g., by getting certified through costly intermediaries (T3). *Self-Verification* (T4) mimics the potential use of blockchain technology to reliably self-verify confidential information to investors at a very low cost (Wang & Kogan, 2018).

Our paper is related to several strands of literature. First, it is related to the literature on “cheap talk” communication in sender-receiver games (see e.g., Dickhaut et al., 1995; Crawford, 1998; Forsythe et al., 1999; Cai et al., 2006; Blume et al., 2020). Since the firm decides whether to communicate the true or some other (untrue) probability of success to the investor, our study is inevitably related to the literature on lying and deception in sender-receiver games (see e.g., Gneezy, 2005; Hurkens & Kartik, 2009; Erat & Gneezy, 2012; Gneezy et al., 2013; Vanberg, 2017; Sobel, 2020). Further, it is related to the literature on “quality disclosure” (see e.g., Viscusi, 1978; Dranove & Jin, 2010), on “self-verification,” and on “unraveling” (Dranove & Jin, 2010; Benndorf et

al., 2015; Jin et al., 2021; Burdea et al., 2023). In the next section, we briefly review these strands of literature.

Our main contribution is twofold: First, we contribute conceptually to the literature by modeling a low-cost self-verification instrument and testing it in a “horse race” against a pure trust-based situation as well as against costly verification and self-verification instruments. Second, our baseline game design (the Probabilistic Investment Game) extends the existing literature on sender-receiver games by modeling a realistic and arguably important feature of start-up investments, namely their probabilistic nature, which makes it impossible for the investor to know later whether the firm’s message was true or not, and thus may encourage less truth-telling by the sender than the standard sender-receiver game. Thus, our modified sender-receiver game could be used as a workhorse game in future experiments.

Our main findings are as follows: In all treatments, including *Baseline* (T1), investments in firms with a high probability of success are significantly larger than investments in firms with a low probability of success. The result is consistent with the pattern we predicted for the (self-)verification T2-T4 treatments but contrasts with our prediction for the *Baseline* (T1) treatment. In this regard, we find that firms with a low probability of success often tell the truth, and investors trust the (supposedly “cheap talk”) message and invest accordingly. Compared to the baseline game, (self-)verification instruments increase investors’ efficiency in making payoff-maximizing decisions when investors are matched with firms with a low probability of success. We conclude that (i) a model that accounts for considerable truth-telling by firms and trust by the investors can explain the data much better than assuming no trust in the free message. In particular, our observed behavior is consistent with models of intrinsic costs of lying (e.g., Gneezy et al., 2018). Yet, other considerations such as payoff maximization, inequity aversion, or level- $k$  reasoning do not explain the observed behavior. We further conclude that (ii) relatively expensive (self-)verification is as effective as a cheap self-verification instrument. Our findings have two interpretations: On the one hand, people are more truthful and trustworthy than what is initially hypothesized albeit they cannot be caught in a lie. In fact, the overall efficiency of investment decisions is not lower under trust than under (self-)verification. On the other hand, the availability of a cheap self-verification technology (e.g., blockchain) does not solve the information asymmetry.

The structure of the paper is as follows: The following section reviews the



related literature. Section 3.3 presents the experimental design, the theoretical predictions, and the procedure. In Section 3.4, we report our results. In Section 3.5, we look at alternative models that could explain the observed investment behavior. Finally, Section 3.6 concludes.

## 3.2 Related literature

In the following, we review the related literature, focusing on cheap talk communication and deception in a sender-receiver game, verification (or quality monitoring) by receivers, and self-verification (or quality disclosure) by senders. Since these strands of literature are vast and could be the subject of dedicated review articles, we aim to provide the reader with only a (selective) background relevant to our own experiment.

### 3.2.1 Cheap talk communication

Our *Baseline* treatment is related to the extensive literature on “cheap talk” communication in sender-receiver games. The seminal theoretical work of Crawford and Sobel (1982) lays the foundation for an extensive experimental literature on sender-receiver games with partial conflicts of interest (for a survey of the theoretical literature, see Sobel (2013)). A key related theoretical finding is that when sender and receiver preferences are misaligned, and the receiver has a best-response strategy according to the prior distribution of types, cheap talk games have a “babbling” equilibrium, where the sender’s message is uninformative and ignored by the receiver (Sobel, 2013). Cheap talk communication experiments have been reviewed by Crawford (1998) and Blume et al. (2020). In the following, we briefly review the studies most relevant to our experiment. Dickhaut et al. (1995) conducted an experimental investigation of Crawford and Sobel’s (1982) model, varying the degree to which the interests of senders and receivers are aligned. This experiment consisted of 12 rounds with random matching. They show that when preferences are closely aligned, senders disclose their type more truthfully. Moreover, although the information content decreases as interests diverge, the message still contains more information than theoretically predicted. Cai et al. (2006) further investigate Crawford and Sobel’s (1982) model. In an experiment with 20-31 rounds and random matching, they observe that participants communicate more than the most informative equilibrium predicts. Forsythe et al. (1999) conducted an experiment with 22 rounds and random matching with three treatments, with

and without communication (as well as an “anti-fraud” treatment, see below). They find that sellers overcommunicate when they can send cheap talk messages and that the gullibility of buyers leads to a higher transaction rate than predicted in equilibrium.<sup>2</sup>

### 3.2.2 Lying and deception in sender-receiver games

Sobel (2020) provides a definition and theoretical framework for lying and deception. In short, a lie is a message that the sender believes to be false. A deception is a message that is intended to cause the receiver to hold false beliefs. We have adopted this definition in the review below, although the studies below may not have originally been written in this terminology.

In this context, lying or cheating games describe non-strategic situations in which participants report an outcome to an experimenter (which only they observe but not the experimenter). The report, in turn, determines the participant’s payoff. “Cheating games” allow one to study the pure motives for lying without considering strategic effects. In such a setting, Fischbacher and Föllmi-Heusi (2013) report that about 20% of participants act in line with the assumption of payoff maximization, 39% remain truth-telling, and an additional 20% neither tell the truth nor maximize their payoff from lying. Gneezy et al. (2013) show that the individual tendency to lie increases with the gains from lying. They also find that when gains are high, the frequency of lying increases with experience. However, they also observe a considerable aversion to lying that cannot be explained by strategic considerations. Kajackaite and Gneezy (2017) show that when participants are concerned about being exposed as a liar, it influences their decision to lie. The study finds that when this concern is present, increased payout incentives do not increase lying behavior. However, when this concern is absent, people are more likely to lie when offered greater incentives.

In the context of a “cheating game,” Gneezy et al. (2018) developed a theoretical model of the intrinsic costs of lying. It includes three types of costs: (i) related to the monetary gains from lying, (ii) related to the distance between the outcome and the reported message, and (iii) related to the probability that a statement will be perceived as a lie. They find that all of these types of costs matter and that the highest proportion of lies report the maximum possible

---

<sup>2</sup> Each participant alternates between being a buyer and a seller. An interesting result by Forsythe et al. (1999) is that dishonesty and gullibility are positively correlated across participants (i.e., those sellers who lie more are also more inclined to believe sellers when acting as buyers).

outcome. However, on average, 25% of participants refrain from choosing the maximum lie (referred to as a “partial lie”). Also, more participants partially lie when the experimenter cannot observe the outcomes than when the experimenter can verify the observed outcome, and partial lying increases as the prior probability of the highest outcome decreases. Similarly, Mazar et al. (2008) show that when people had the ability to lie, they lied, but the magnitude of dishonesty was relatively small (relative to the possible maximum amount). Participants’ behavior was insensitive to the expected external costs and benefits associated with the dishonest acts but was sensitive to contextual manipulations related to self-concept.

Abeler et al. (2019) conducted a study containing both a meta-analysis of cheating game experiments and a novel experiment. Their meta-study consists of 429 “cheating game” treatments. They find that people refrain from lying maximally, giving up, on average, about three-quarters of the potential gains from lying. When the distribution of the states of the world is uniform, senders report all possible states with positive probability. However, states that lead to higher payoffs are reported more often than states that lead to lower payoffs. Their original experiment was designed to test for plausible theoretical motives for truth-telling. They find that the main motivation for being honest is a preference for both “being honest” and “being seen as honest.”

We now turn to games with deception as defined by Sobel (2020). Gneezy (2005) conducted an experiment in which the sender has private information about the payoffs associated with the receiver’s decisions. The sender sends a message to the receiver that can be true or false, and the receiver decides whether to follow the message, which determines the payoffs for both players. Gneezy (2005) finds that participants consider their own payoff gain and the receiver’s loss when deciding whether or not to lie. In a treatment where both the sender’s gain from lying and the receiver’s loss were small, only 36% of senders lied. When the gain to the sender was small, but the loss to the receiver was large, only 17% lied. In a treatment where the gain to the sender was large and equal to the loss to the receiver, 52% of the senders lied. Hurkens and Kartik (2009) show that Gneezy’s (2005) results may be driven by the participants’ social preferences rather than their truth-telling preferences. Sutter (2009) distinguishes between “benevolent truth-tellers” and “sophisticated truth-tellers” (the latter tell the truth but expect others not to follow their message, which is considered as deception according to Sobel (2020)). He observed more “sophisticated truth-tellers” in teams than

among individuals and that sender teams are less truthful than individuals.<sup>3</sup> Erat and Gneezy (2012) find that a significant fraction of senders do not lie even when lying results in a Pareto improvement for the sender and receiver (white lies), demonstrating a pure aversion to lying independent of any social preferences for outcomes. Vanberg (2017) further tests whether participants actually have strong preferences for telling the truth, but he finds little evidence for unconditional truth-telling.

### 3.2.3 Verification by receivers and self-verification by senders

The remaining treatments in our study (T2-T4) feature “quality disclosure” (Viscusi, 1978; Dranove & Jin, 2010) either through monitoring by recipients (verification treatment T2) or by the senders (self-verification treatments T3 and T4). For self-verification by senders, Dranove and Jin (2010) survey the theoretical and empirical literature on quality disclosure and certification. A related finding is that firms are more likely to disclose their quality when disclosure costs are lower, product quality is higher, and the expected benefits from disclosure are higher. In this context, Forsythe et al. (1999) also conducted an “anti-fraud” treatment in which senders (sellers) send a message that must contain the true value (but can also contain additional values). In the anti-fraud treatment, when the receivers (buyers) receive multiple values (which must include the true value), the authors find that participants infer the lowest value from the message. Consequently, the anti-fraud treatment increases receiver efficiency relative to the cheap talk treatment. Jin et al. (2021) conducted an experiment (45 rounds with random matching) in which senders could either reveal a truthful signal or not. Their design consists of two main factors: Participants either received feedback after each round or no feedback at all, and participants either remained in their roles as sender and receiver throughout the experiment or the roles were randomly assigned after each round. Jin et al. (2021) find that receivers are not sufficiently skeptical when senders do not reveal the signal. However, this changes with repetition and feedback. Over time, receivers become more skeptical when senders do not disclose. Without feedback, the rate of convergence to the theoretical predictions is much slower. Benndorf et al. (2015) conducted a 15-round, fixed-matching experiment in which participants (“workers”) could decide whether

<sup>3</sup> In this regard, Cohen et al. (2009, p. 1321) find that “[teams] are more strategic than individuals in that they will adopt the course of action that best serves their economic interest.”

to truthfully disclose their productivity, which was determined at the beginning of each experiment (and be paid according to their productivity), or not to disclose (and be paid according to the average productivity of workers who did not disclose).<sup>4</sup> Disclosure is costly. The authors find that disclosure rates are lower than predicted in equilibrium. In terms of equilibrium choices, workers who should disclose their productivity in equilibrium deviate from the equilibrium more often than those workers who should conceal in equilibrium. Over time, the rate of disclosure moderately converges towards equilibrium.<sup>5</sup>

A specialized strand within the verification and self-verification literature examines how sustainability aspects are communicated in product decisions and how buyers perceive and respond to them. For example, Etilé and Teyssier (2016) analyze the effectiveness of third-party certification versus brand-based signaling in a market for corporate social responsibility (CSR) attributes. Their experimental results show that only third-party certification leads to a separating equilibrium, increasing CSR investments. They observed buyers' gullibility paying more without guaranteeing higher CSR standards. Therefore, unverifiable claims used as part of a brand-building strategy led to higher sellers' profits as compared to the benchmark situation where labeling is not possible. Cason and Gangadharan (2002) investigate environmental labeling in markets with incomplete consumer information. Their laboratory experiment finds that while reputations and unverified "cheap talk" signals sometimes increase the provision of environmentally friendly goods, the only reliable mechanism to ensure higher product quality is third-party certification, which effectively mitigates information asymmetry.

Finally, Burdea et al. (2023) provide an experimental investigation (30 rounds with random matching) of the models of Glazer and Rubinstein (2004, 2006). Their study combines cheap talk, partial verification by receivers, and self-verification by senders. To our knowledge, Burdea et al. (2023) is the first experimental study (and the only study prior to ours) to combine receiver verification and sender self-verification. However, their experimental design is fundamentally different from ours. In their study, quality is deterministic and consists of two dimensions (implemented by two cards with numbers). Verification (or self-verification) is mandatory, costless, and only partial (i.e., only

<sup>4</sup> Each group consists of 6 members. The 15 periods are divided into 3 subperiods of 5 rounds with different productivity distributions.

<sup>5</sup> Hagenbach and Perez-Richet (2018) experimentally investigated a disclosure game in which the sender can only send true information (but not necessarily all information). However, unlike in our study, the sender's preferences are non-monotonic, and thus, their results are not comparable to our current study.

one of the cards can be verified). Burdea et al. (2023) observe that participants inflate their quality (i.e., lie about the true state) in both treatments, verification (by the receiver) and self-verification (by the sender). Only three participants (out of 192) never lied about their quality. Although receivers are more skeptical in the sender self-verification treatments, receivers' payoffs do not differ significantly between the verification and self-verification treatments.

There is also a growing interest in Bayesian persuasion games (Kamenica & Gentzkow, 2011). As in sender-receiver games, the sender wants to persuade the receiver to take a certain action (as opposed to the receiver's a priori best action), while the receiver wants to align its action with the (unknown) true state of the world. The difference is that the sender commits to a signaling strategy before learning the true state of the world. The receiver then observes the realization of the signal and takes action. In a recent experimental study, Fréchette et al. (2022) investigate a model of Bayesian persuasion that includes additional elements of partial commitment and verification (25 rounds of random matching). For any given level of commitment, the authors distinguish between verifiable and unverifiable information. When messages are unverifiable, senders are free to misreport their private information. When messages are verifiable, information cannot be misreported, but it can be hidden. A finding relevant to our study is that in low-commitment treatments, the information content of the message increases when the message is verifiable.

### 3.3 Design, predictions, and experimental procedure

First, we present the baseline game and the treatments. We then outline the theoretical predictions and describe the experimental procedure.

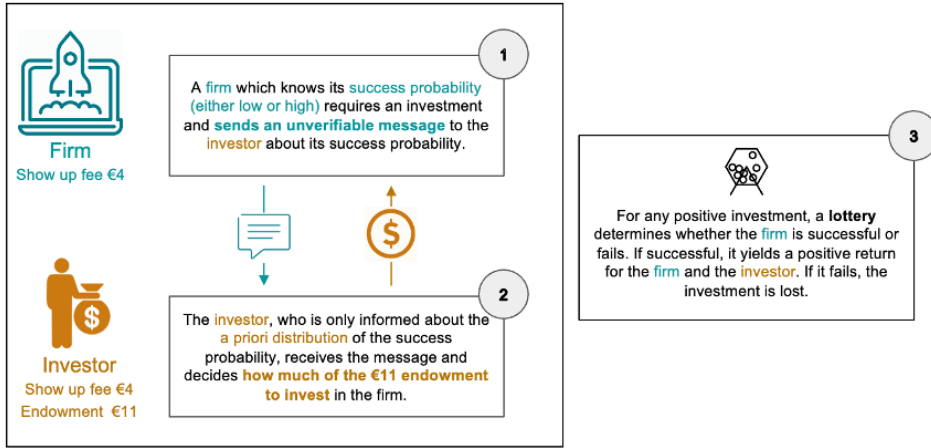
#### 3.3.1 Experimental design and workhorse game

The Probabilistic Investment Game is shown in Figure 3.1.<sup>6</sup> It is an extension of sender-receiver games by adding risk to the receiver's decision. The game

---

<sup>6</sup> A previous version of the Probabilistic Investment Game was designed by Stenzel et al. (2022) in an unpublished manuscript. This game is somewhat similar to the famous "trust game" (or "investment game") of Berg et al. (1995) and Kreps (1990). In both games, the first mover takes a risk by sending money to a second mover. In Berg et al. (1995), the investment always succeeds in tripling the amount invested, but the investor's return depends on the amount the second mover returns to her. In contrast, in our game, the investment may succeed or fail depending on the probability of success of the firm, but the investor's return on her investment is already determined by a perfect contract. Also related are the probabilistic versions of the trust game (see e.g., Vranceanu et al., 2012; Güth et al., 2014).

consists of two players: an investor (referred to with female pronouns) and a firm (referred to with neutral pronouns). The investor starts with a monetary endowment  $w \in \mathbb{N}$  and can invest an amount  $i \in [0, w]$  in the firm. For our experiment, we set  $w = \text{€}11$ . The firm owns no capital but contributes a potential innovation that can yield a positive return for the investor and the firm. The probability that the innovation will be successful is  $\theta \in \Theta = \Theta_l \cup \Theta_h$ , where  $\Theta_l = \{10\%, 11\%, \dots, 20\%\}$  and  $\Theta_h = \{70\%, 71\%, \dots, 80\%\}$ . We assume that all values of  $\theta$  within  $\Theta_l$  and  $\Theta_h$  are equally likely, i.e.,  $\theta$  is drawn from a uniform distribution over  $\Theta$ . Note that  $\theta$  can only take the discrete values explicitly listed in  $\Theta_l$  and  $\Theta_h$ , such as 10%, 11%, 12%, etc., and not intermediate values like 11.5%. That is, the investor's a priori expected probability of success is 45%. The interpretation of our selection of possible probabilities of success is that we consider two types of firms: firms with a low probability of success and firms with a high probability of success. The given range of success probabilities increases the space of possible messages for the firms. To avoid any experimenter demand effect, we aimed to prevent participants from feeling less justified in lying by having only one high and one low value.



**Figure 3.1:** The Probabilistic Investment Game (repeated design)

Each round in the baseline game then proceeds as follows: In the first stage, the firm is informed about its probability of success  $\theta$  and sends a free (“cheap talk”) message to the investor. The message can contain the true probability of success or any other value chosen by the firm (within  $\Theta$ ). In the second stage, after receiving the message, the investor chooses how much of the endowment to invest in the firm  $i \in [0, w]$ . Then, a lottery with two outcomes determines whether the innovation succeeds (with probability  $\theta$ ) or

fails (with probability  $1 - \theta$ ). If it succeeds, both the investor and the firm receive a positive return. In particular, if the investment is successful (with a probability of  $\theta$ ), the investor will receive a payoff of  $w + i$ , while the firm will receive a payoff of  $i$ . If the investment fails (with a probability of  $1 - \theta$ ), the investor will receive a payoff of  $w - i$ , while the firm will receive no additional payoff. Therefore, for a given probability of success  $\theta$ , the investor’s expected payoff is  $2\theta i + w - i$ , and the firm’s expected payoff is  $\theta i$ . The game’s payout structure mimics the characteristics of investing in start-ups, where companies make a case for high returns and investors bear the financial risk.

After the lottery determines the success of the innovation, a new round begins with a new probability of success for the firm. While the firms and investors maintain their roles throughout the experiment, the firms are randomly matched with different investors in each round. In addition to the payoffs from the experiment, each participant (investors and firms) receives a show-up fee of €4. The treatment structure is shown in Table 3.1. In addition to the *Baseline* (T1) treatment described above, we ran three other treatments in which the firm’s message can be verified or self-verified. The *Costly Verification* (T2) treatment is analogous to *Baseline* (T1), but now, after receiving the free message (and before making an investment decision), the investor can verify the firm’s true probability of success at a cost of  $c$  (in our experiment,  $c$  equals €1). The investor then receives the message: “It is certified by the experimental team that the success probability of the project is  $\theta$ .”

**Table 3.1:** Treatments

Name	Verification? (Cost)	Paid by	Sessions	Particip.
T1: Baseline	No	–	6	124
T2: Costly Verification	Yes, high (1 €)	Investor	5	98
T3: Costly Self-Verification	Yes, high (1 €)	Firm	5	98
T4: Self-Verification	Yes, low (0.1 €)	Firm	6	118
			22	438

Next, the *Costly Self-Verification* (T3) and the *Self-Verification* (T4) treatments model the case where a firm can reliably reveal its probability of success to the investor. In this case, the investor receives the message: “It is certified by the experimental team that the success probability of the project is  $\theta$ .” In *Costly Self-Verification* (T3), the firm decides whether to verify its true probability of success  $\theta$  to the investor at a cost  $c$  equal to €1. This treatment mimics firms’ options for costly self-verification through a third party (e.g.,



costly labeling). Finally, *Self-Verification* (T4) is the same as *Costly Self-Verification* (T3) with the only difference that now the cost of self-verification is low, i.e.,  $\tilde{c} < c$ . In our experiment,  $\tilde{c}$  is equal to €0.1. Note that the cost of acquiring information for the investor in *Costly Verification* (T2) and the firm's cost in the self-verification treatments (T3 and T4) are deducted from the show-up fee.<sup>7</sup> In the self-verification treatments (T3 and T4), if the firm decides not to send the certified message (with its true probability of success), it must instead send a free ("cheap talk") message (as in treatments T1-T2). Thus, the investor receives either the free message (which may or may not be true) or the true probability of success  $\theta$ , and then chooses the amount to invest  $i \in [0, w]$ . This is another design novelty. In previous self-verification studies, self-verification could either be used or not. But if it is not used, a free (non-binding) message cannot be sent instead.

As a final note, our study was not originally designed to investigate the ethical aspects of truth-telling or the reasons for it, but free communication between senders and receivers is inevitably related to the issue of lying and deception by senders.<sup>8</sup> In our treatments without (self-)verification opportunity (or when firms do not use it), firms have to send a free message containing the true or any other value of the probability of success. We also believe that an untrue message sent by a firm may be viewed by the participants as a legitimate strategy (and even expected by investors) in the sense that "someone who does not 'bluff' in poker is a bad player" (Erat & Gneezy, 2012, p. 730).

### 3.3.2 Predictions

In this section, we apply a solution concept for Bayesian games with communication by Myerson (1991) to our game. In the following, we present our predictions and give intuition around them. More details and formal proofs can be found in the Appendix 3.A.

<sup>7</sup> Because the firm does not receive an initial endowment, in *Costly Self-Verification* (T3) and *Self-Verification* (T4), the cost had to be deducted from the show-up fee. For comparability, in *Costly Verification* (T2), the verification costs are also deducted from the investor's show-up fee.

<sup>8</sup> We wrote the instructions in such a way as to avoid emphasizing any norm with respect to the messages (neither a norm of truth-telling nor a norm of lying). We wrote that "Participant B may reveal the true probability of success or any other value of his/her choice."

### Baseline (T1)

First, let us consider the Probabilistic Investment Game without communication. That is, the firm does not send a message, and the investor has only her a priori knowledge about the firm's probability of success. In a Bayesian equilibrium, the investor responds optimally, taking into account the implications in light of her probabilistic beliefs and maximizing her expected utility over her possible strategy choices. If both the investor and the firm are risk neutral, their expected utilities are equal to their expected payoffs.<sup>9</sup> Given the distribution of probabilities of success and the payoff structure, not investing leads to a higher expected payoff for the investor than any positive amount of investment (as  $2\bar{\theta}i + w - i = -0.1i + w < w$  for all  $0 < i \leq w$ ). Therefore, not investing is the best choice for a risk-neutral investor in the Probabilistic Investment Game without communication.

Now, let us consider the messaging stage of the Probabilistic Investment Game. Can there be a messaging strategy for the firm leading to an equilibrium with a positive amount of investment? According to the revelation principle for general Bayesian games, the answer is “no.” In such an equilibrium, the investor would invest a positive amount after receiving a specific free message. Therefore, the equilibrium strategy for any firm (including firms with a low probability of success) must be to send that particular message in order to optimize its payoff. As a consequence, the investor wants to deviate, ignore the message, and not invest, contradicting the equilibrium assumption. Therefore, any equilibrium candidate with a positive investment amount can be ruled out, which makes the firm's messaging strategy irrelevant.

### Costly Verification (T2)

In *Costly Verification* (T2), the investor must choose: She can either verify the firm's probability of success at a cost or not, which leaves both players in a game equal to the baseline treatment. The investor purchases the information only if the expected payoff from it is higher than not investing at all (the equilibrium in *Baseline*). We set the parameters so that purchasing the information leads to a higher payoff for the investor. Therefore, the equilibrium for *Costly Verification* is as follows: The investor purchases the verification; once the investor learns the firm's probability of success, she will invest according to her risk preference if the firm's probability of success is high and invest

---

<sup>9</sup> We discuss the case of risk aversion in Appendix 3.B.1.

nothing if the firm’s probability of success is low.

### **Costly Self-Verification (T3) and Self-Verification (T4)**

In *Costly Self-Verification* (T3), the firm must choose: It can either self-verify the probability of success at a cost or not, the latter leaving both players in a game equal to the baseline treatment. A firm sends a self-verified message if its expected payoff is greater than 0 (its equilibrium payoff in *Baseline*). For a firm with a low probability of success, the investor is expected not to invest (regardless of her risk preferences), and therefore, the equilibrium strategy for the firm is to save the cost (i.e., not to send the self-verified message) and send a cheap talk message instead. This is a separating (fully informative) equilibrium in T3 and T4: Only firms with a high probability of success self-verify, and the investor invests accordingly. Firms with a low probability of success do not self-verify, and the investor does not invest.

### **3.3.3 Experimental procedure**

The experiment was conducted between January and February 2022.<sup>10</sup> Participants were recruited from different faculties of a large German university using hroot (Bock et al., 2014). It was conducted online using oTree version 5.6 (D. Chen et al., 2016). A total of 438 participants (58% women) participated in four treatments (124, 98, 98, and 118 participants in the T1-T4 treatments, respectively). We conducted 5-6 sessions per treatment, usually attended by 18-22 participants (see Table 3.1).<sup>11</sup> Each session consisted of three stages: First, participants received instructions (including quiz questions) upon logging into the experiment. The instructions were identical for both investors and firms, as the roles had not yet been determined at this point. Second, participants were randomly assigned to investors and firms (these roles were fixed throughout the experiment), and the 15-round experiment started. Third, before each round, investors and firms were randomly matched, and a new probability of success was randomly drawn for each firm. Finally, after fin-

---

<sup>10</sup> One pilot session of *Self-Verification* (T4) was conducted on November 4, 2021. We conducted the pilot with only 10 periods, and after successfully completing the pilot in a reasonable amount of time, we decided to increase the length of the experiment to 15 periods. Therefore, we exclude the pilot from the analyzed dataset.

<sup>11</sup> We aimed for 24 participants per session. Due to no-shows and since not wanting to cancel an already booked session at the last minute, we usually had a lower number of participants (typically 18, 20, or 22). Out of a total of 22 sessions, we had one session with 16 participants, 7 sessions with 18 participants, 8 sessions with 20 participants, 4 sessions with 22 participants, and 2 sessions with 24 participants.

ishing the last round of the experiment, the participants completed a short post-experimental questionnaire that collected demographic data as well as information about their beliefs, risk attitudes, and perceived difficulty of the experiment. Participants reported an average difficulty level of 1.83 (“Easy”), with a standard deviation of 0.77 (ranked on a scale from 1 = Very Easy to 5 = Very Difficult). The earnings were transferred to the participants’ bank accounts within two business days.

### 3.4 Results

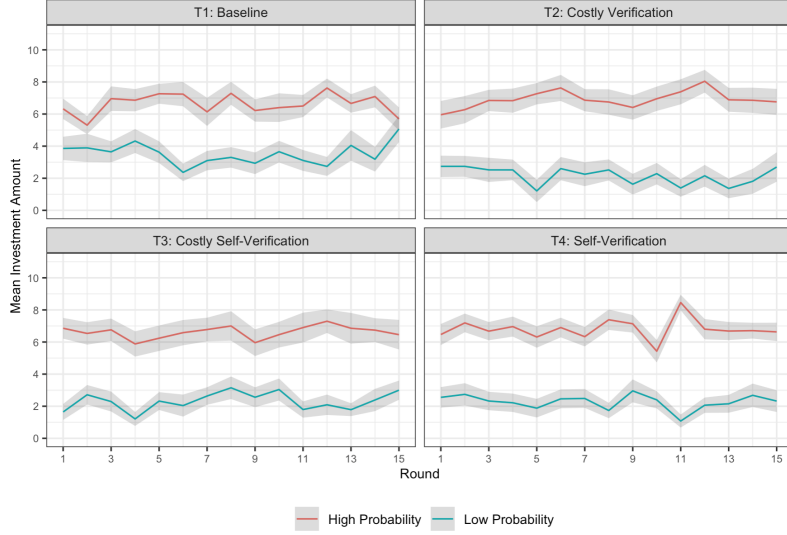
Our analysis first examines the *effectiveness* of the (self-)verification instruments (i.e., the dependence of the investment on the firm’s unknown probability of success), but without considering whether verification or self-verification was used. We proceed by investigating whether firms report their probability of success truthfully to the investor. We then examine the investor’s trust in the free (supposedly “cheap talk”) message. Next, we investigate verification and self-verification behavior in the T2-T4 treatments, respectively. Finally, we measure the Investor Efficiency in making payoff-maximizing decisions with and without (self-)verification instruments.

Our formal statistical analysis presented in this section is based on random-effects panel data regressions. In addition, as a robustness check and to account for the random matching between investors and firms within a session, we also estimated linear regressions with two-way clustered standard errors (i) across *investors* and (ii) across *session*  $\times$  *period* (Cameron et al., 2011). The results of these latter regressions are presented in Appendix 3.E and are similar to the panel data regressions in the body of the paper.

#### 3.4.1 Investment per firm’s probability of success

As a measure of the effectiveness of the verification and self-verification instruments, we consider the impact of a firm’s probability of success on the amount of investment (still without considering the specific communication and (self-)verification possibilities). Figure 3.2 shows that in all treatments (including *Baseline*), investments in firms with a high probability of success (hSP, between 70% and 80%) are significantly higher than in firms with a low probability of success (lSP, between 10% and 20%). These results are consistent with the predictions for the (self-)verification treatments T2-T4. However, the finding of higher investment amounts in firms with a high probability of

success in *Baseline* (which relies only on truthfulness and trust) contradicts our prediction.



Note: The shaded areas indicate one standard error above and below the average investment amount.

**Figure 3.2:** Average investments in firms with a high and low probability of success

Although we observe a similar general pattern across treatments, we want to measure quantitatively the strength of the relationship between a firm's probability of success and the amount invested. The effectiveness of the treatment from the investor's perspective increases if investments in firms with a high probability of success are higher than investments in firms with a low probability of success. Here, we expect a significant *higher* coefficient in the (self-)verification treatments T2-T4 than in the *Baseline*. Table 3.2 shows the corresponding random-effects panel data regression models. Model 1 shows the treatment effects without controlling for low and high probabilities of success. Models 2 and 3 (model 3 also includes the investor's risk propensity as measured in the post-experimental questionnaire) show that being matched with an hSP firm significantly increases investment in all treatments (Wald tests for the coefficient of hPS yield  $p = 0.000$  for each of the treatments).<sup>12</sup> In particular, in *Baseline*, hSP increases investment by €3.21. However, this effect is much larger in *Costly Verification* (T2), *Costly Self-Verification* (T3), and *Self-Verification* (T4) than in *Baseline* (T1) (investments are on average €1.33, €1.06, and €1.41 higher for firms with a high probability of success

<sup>12</sup> We test the null hypothesis that the sum of the coefficients  $\beta_{[\text{hSP}]}$  and  $\beta_{[\text{hSP} \times \text{respective treatment}]}$  equals zero.

in T2, T3, and T4 than in *Baseline*, with  $p = 0.017, p = 0.026, p = 0.001$ , respectively, Model 3). We also find that the less risk-averse an investor is, the more she will invest in the risky project (t-test,  $p = 0.000$ ). Moreover, investments do not change over time in any of the treatments. The increase in investment in firms with a high probability of success compared to firms with a low probability of success is, on average, larger in the (self-)verification treatments than in *Baseline*. In terms of investment levels, there are no overall differences between firms with a high probability of success in *Baseline* and in the (self-)verification treatments (Wald tests across treatments yield p-values between 0.1784 and 0.973, Model 3).

**Table 3.2:** Investment per probability of success (random-effects estimations)

	Model 1	Model 2	Model 3
Treatment T2	-1.01 (1.14)	-1.42 (1.06)	-1.22 (1.12)
Treatment T3	-1.48 (1.14)	-2.00* (1.09)	-2.16* (1.11)
Treatment T4	-0.84 (1.11)	-1.29 (1.01)	-1.33 (0.86)
high probability of success (hSP)	-	3.21*** (0.33)	3.21*** (0.33)
hSP $\times$ Treatment T2	-	1.33** (0.56)	1.33** (0.56)
hSP $\times$ Treatment T3	-	1.07** (0.48)	1.06** (0.48)
hSP $\times$ Treatment T4	-	1.41*** (0.43)	1.41*** (0.43)
Period	0.02 (0.03)	0.01 (0.03)	0.01 (0.03)
Treatment T2 $\times$ Period	-0.00 (0.05)	-0.02 (0.04)	-0.02 (0.04)
Treatment T3 $\times$ Period	0.02 (0.04)	0.02 (0.04)	0.02 (0.04)
Treatment T4 $\times$ Period	-0.04 (0.04)	-0.02 (0.04)	-0.02 (0.04)
Risk test	-	-	0.34*** (0.05)
Constant	5.54*** (0.86)	3.93*** (0.83)	1.74** (0.83)
Session FE	yes	yes	yes
Observations	3,285	3,285	3,285
R-squared	0.014	0.276	0.324

Note: Random-effects panel data estimations. Dependent variable: Investment (a value from €0 to €11). “high probability of success (hSP)” indicates success probability values of 70%-80% (otherwise values of 10%-20%). “Risk test” is taken from the post-experimental questionnaire: It is the amount invested (between €0 and €10) in a risky lottery, where a lower amount indicates a higher degree of risk aversion. The standard errors are clustered across investors. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

**Result 1.** *In all treatments, investment is substantially higher for firms with a high probability of success than for firms with a low probability of success. Moreover, the level of investment in firms with a high probability of success does not differ between the Baseline and (self-verification) treatments.*

### 3.4.2 Truth-telling and strategic behavior

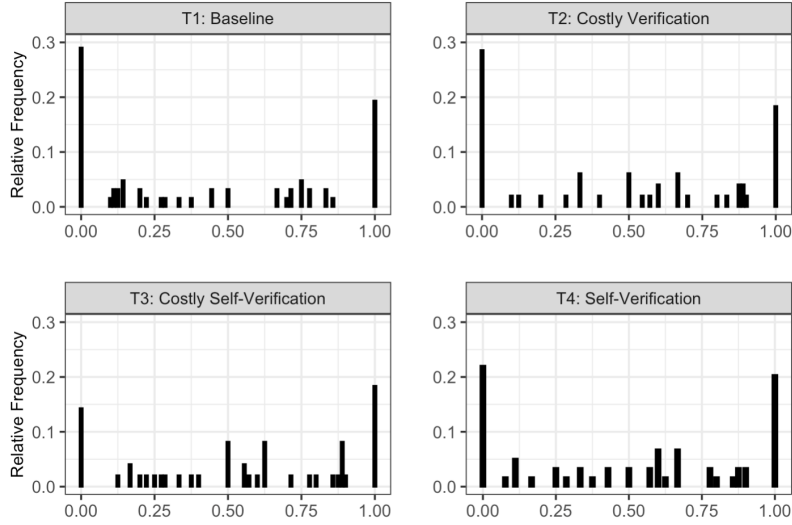
Given the possibility of sending a free message to the investor, firms might try to persuade the investor to invest. In the following, we test whether firms reveal their true state or behave strategically and send a high message to increase their chance of receiving the investment, regardless of their actual probability of success. We define “deception” as a binary variable that takes the value 1 when a firm with a low (high) probability of success sends a message with a high (low) probability of success and zero otherwise. This also implies that a “small” lie by a firm with a low or high probability of success (e.g., a firm at 71% sending a message indicating a 78% probability of success) is not considered deceptive.<sup>13</sup>

Figure 3.3 shows that a substantial fraction of low-probability firms tell the truth; in *Baseline*, only 19% of firms always sent a high message (“deception”) after receiving a low probability of success. In contrast, 28% of firms are always truthful (i.e., they sent a low message every time they had a low probability of success). The remaining firms are sometimes truthful. Interestingly, the availability of a verification instrument in T2 does not significantly reduce the firms’ strategic behavior ( $p = 0.594$ , robust rank order test). Furthermore, we find no differences when comparing the deception distributions across the four treatments ( $p = 0.434$ , Kruskal–Wallis test). Similar to *Baseline*, there is a considerable share of firms that always tell the truth (never send a high message after receiving a low probability of success).<sup>14</sup>

**Result 2.** *We observe a considerable amount of truth-telling by firms. In fact, in Baseline (T1) without verification possibility, 28% of the low probability of success firms never sent a deceptive message. In contrast, 19% always sent a high message. These numbers are similar to those in the Costly Verification (T2) treatment, where investors can verify the firms’ true probabilities of*

<sup>13</sup> This definition is consistent with Sobel (2020) in the sense that a small lie (within the range of high or low probabilities of success) does not change the optimal decision of a gullible investor from not investing to investing.

<sup>14</sup> In T3 and T4, firms have the option of either sending a free message or self-verifying at a cost. Therefore, the firms’ messaging strategies in the self-verification treatments (T3 and T4) cannot be fully compared to those in T1 and T2.



Note: The x-axis indicates the consistency of deception versus truth-telling. A value of 0 indicates a firm that never deceived after receiving a low probability of success, while a value of 1 indicates a firm that always deceived (i.e., sent a message with a high probability of success even though its actual probability of success was low). Self-verification is considered to be truth-telling in T3-T4.

**Figure 3.3:** Deception by firms (with a low probability of success)

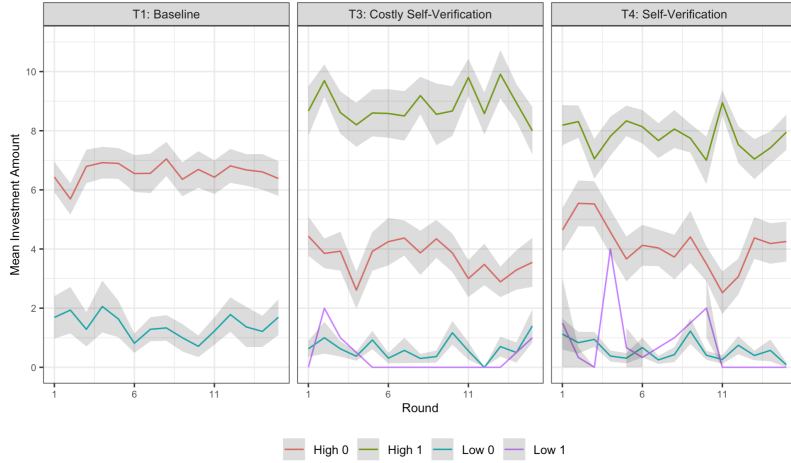
*success.*

### 3.4.3 Investment per free message

Result 1 indicates a separating investment pattern in *Baseline* (T1). One reason may be that investors trust the high free messages much more than if they were just “cheap talk.” In fact, Result 2 shows that firms are telling the truth to a significant extent. Now, we want to know how much investors trust a free message and their resulting investment behavior.

Figure 3.4 shows the investment per information received in *Baseline* (T1) and in the self-verification (T3-T4) treatments. This figure shows that in *Baseline* (T1), investments seem to be largely influenced by the corresponding free message. Interestingly, in the self-verification treatments T3-T4 (where firms could self-verify instead of sending a free message), the free message also has a considerable effect on investment, but smaller than in *Baseline* (T1). That is, investors become more skeptical when they receive a free message indicating a high probability of success instead of a self-verified one. However, they are not fully skeptical, which is particularly surprising in T4, where the cost of self-verification is very low. Finally, Figure 3.4 shows that a free message stating a low probability of success is treated in the same way as self-verification of a low probability of success. Investors assume that no firm would send a low





Note: “High 0” indicates unverified free messages claiming a high probability of success, while “High 1” indicates self-verified high probabilities of success. “Low 0” indicates unverified free messages claiming a low probability of success, and “Low 1” denotes a self-verified low probability of success. We omitted Costly Verification (T2) from the figure because it could contain up to eight different lines not fully comparable with the Baseline (T1) and the self-verification treatments T3 and T4 (since there are now 2x2x2 parameters: the free message value, the true probability of success, and whether it was verified or not) – see Figure 3.D.1 in the appendix.

**Figure 3.4:** Average investment per free and self-verified message

message if it had a high probability of success. Thus, there is no “truth-telling premium” for low probability of success firms that self-verify their type.

The random-effects panel data regression in Table 3.3 shows that investment is significantly affected by a high free message (hfM) in all treatments (Wald tests yield  $p = 0.0000$  for all treatments).<sup>15</sup> In *Baseline*, a high free message increases investment on average by €5.30. However, the effect is much smaller in size than in the self-verification treatments T3 and T4, where it increases investment by only €3.89 and €3.76, respectively. There is no difference in the effects between the latter two treatments (Wald test yields  $p = 0.2603$ ). Comparing investments after receiving a high free message, we find that investments are higher in *Baseline* than in T3 ( $p = 0.0112$ ) and T4 ( $p = 0.0263$ ), but no significant difference is found between T3 and T4 ( $p = 0.3515$ ).

**Result 3.** *Investment amounts are affected by free messages indicating either high or low probabilities of success. However, in the self-verification treatments T3 and T4, the amount of investment after receiving a high free message is*

<sup>15</sup> This regression model includes only the Baseline T1, and the self-verification treatments (T3-T4). The reason is that in T2, the investor could verify the message, and thus, it is not directly comparable to the free message in the other treatments (where the investor can believe the message or not).

**Table 3.3:** Investment per free message

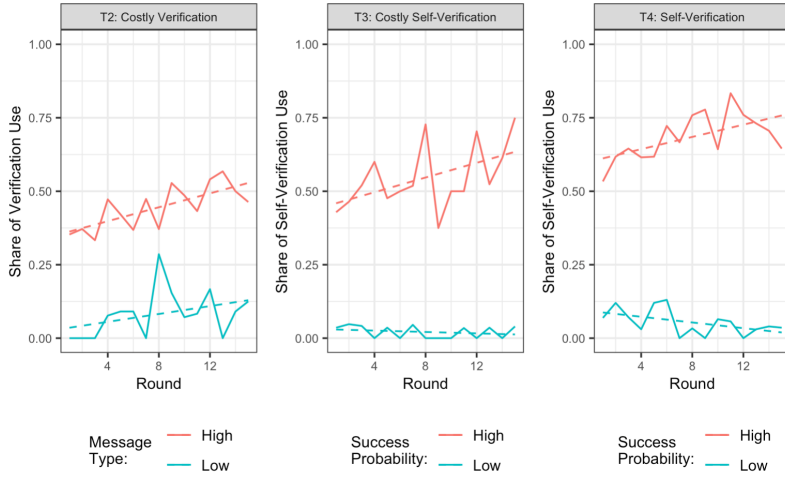
	Model (1)
Treatment T3	-1.81 (1.17)
Treatment T4	-0.64 (0.92)
High free Message (hfM)	5.30*** (0.42)
hfM $\times$ Treatment T3	-1.41** (0.60)
hfM $\times$ Treatment T4	-1.54*** (0.58)
Period	0.00 (0.03)
Treatment T3 $\times$ Period	-0.05 (0.03)
Treatment T4 $\times$ Period	-0.04 (0.04)
Risk test	0.27*** (0.06)
Constant	0.16 (0.87)
Session FE	Yes
Observations	2,012
R-squared (overall)	0.4211

Note: Random-effects panel data estimation. The regression includes all observations in the Baseline (T1) but only free messages in treatments T3 and T4. Dependent variable: Investment (a value from €0 to €11). “High free message (hfM)” indicates a free message (i.e., a message that is not self-verified) with a probability of success value of 70%-80% (otherwise values of 10%-20%). “Risk test” is taken from the post-experimental questionnaire: It is the amount invested (between €0 and €10) in a risky lottery, where a lower amount indicates a higher degree of risk aversion. Standard errors are clustered across investors. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

*not as high as in Baseline (T1).*

#### 3.4.4 Verification use

The left panel of Figure 3.5 shows the use of verification in *Costly Verification* (T2) by message type, while the two right panels illustrate the use of self-verification according to the firm’s actual probability of success in T3 and T4. In T2, the figure shows that low messages are rarely verified, while high messages are frequently verified. Verifications of high messages seem to increase over time. Regarding T3 and T4, the figure shows that firms with a low probability of success (ISP) tend not to self-verify. However, hSP firms often self-verify, and this increases over time. Thus, it appears that in T3 and T4, firms learn over time that it is in their best interest to self-verify when their probability of success is high.



Note: Left panel: the red (blue) line shows the share of verified (by investors) messages indicating a high (low) probability of success. Right panels: The red (blue) line shows the share of self-verified (by firms) messages indicating a high (low) probability of success. The dashed lines show the average time trend.

**Figure 3.5:** (Self-)verification of messages

Table 3.4 shows linear probability model regressions of the panel data for the self-verification treatments (T3-T4). Model 1 shows the treatment effects without controlling for the probability of success. Model 2 shows that the probability of using self-verification is considerably higher for hSP firms than for lSP firms. In particular, Model 2 shows that the probability of self-verification is 55% and 63% higher for hSP firms than for lSP firms in *Costly Self-Verification* (T3) and *Self-Verification* (T4), respectively (Wald test yields  $p = 0.000$  for T3 and T4).<sup>16</sup> This is to be expected since, in theory, only hSP firms should self-verify their type. Moreover, we find that in both T3 and T4, self-verification increases over time for hSP firms (Wald test for hSP in T3 and T4 yields  $p = 0.0004$  and  $p = 0.0001$ ) but not for lSP firms in T3 (Wald test for lSP firms in T3 yields  $p = 0.3488$ ). For T4 we find marginal significance (Wald test for lSP firms T4 yields  $p = 0.0594$ ).<sup>17</sup> We also estimated linear probability models for the *Costly Verification* (T2) treatment, regressing the use of verification on the receipt of a message with a high probability of success (hSP-message) and the time period.<sup>18</sup> The regressions are shown

<sup>16</sup> We test the null hypothesis that the sum of the coefficients  $\beta_{[\text{hSP}]}$  and  $\beta_{[\text{hSP} \times \text{respective treatment}]}$  equals zero.

<sup>17</sup> For hSP firms, we test the null hypothesis that the sum of the coefficients  $\beta_{[\text{hSP} \times \text{period}]}$  and  $\beta_{[\text{hSP} \times \text{period} \times \text{respective treatment}]}$  equals zero. For lSP firms, we test the null hypothesis that the sum of the coefficients  $\beta_{[\text{period}]}$  and  $\beta_{[\text{period} \times \text{respective treatment}]}$  equals zero.

<sup>18</sup> We estimated separate models for T2 and T3-T4 because in T2, verification is done by investors, whereas in T3 and T4, it is done by firms.

**Table 3.4:** Firms' self-verification use

	Model (1)	Model (2)	Model (3)
Treatment T4	0.09 (0.09)	0.08 (0.11)	0.10 (0.11)
high probability of success (hSP)	-	0.55*** (0.06)	0.43*** (0.07)
hSP $\times$ Treatment T4	-	0.08 (0.08)	0.06 (0.10)
Period	0.01** (0.00)	0.01** (0.00)	-0.00 (0.00)
Period $\times$ hSP	-	-	0.02*** (0.00)
Period $\times$ Treatment T4	-0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)
hSP $\times$ Treatment T4 $\times$ Period	-	-	0.00 (0.01)
Constant	0.29*** (0.06)	0.02 (0.07)	0.08 (0.07)
Session FE	Yes	Yes	Yes
Observations	1,620	1,620	1,620
R-squared (overall)	0.043	0.434	0.438

Note: Random-effects panel data regressions including only the self-verification treatments T3 and T4. Dependent variable: verification use (1 = verification use, 0 = otherwise). “High probability of success (hSP)” indicates success probability values of 70%-80% (otherwise values of 10%-20%). Standard errors are clustered across firms. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

in Appendix 3.D.3, and their results can be summarized as follows: In T2, a high message significantly increases the probability of verification (t-test,  $p = 0.000$ ). Consistent with the result in the self-verification treatments T3 and T4, we also find that the use of verification in T2 increases over time only with a high message (Wald test yields  $p = 0.0092$ ) but not with a low message ( $p = 0.876$ ).

**Result 4.** *In the self-verification treatments (T3-T4), firms with a high probability of success self-verify considerably more than their low probability counterparts. Moreover, self-verification by firms with a high probability of success increases over time, while the self-verification of firms with a low probability of success does not change over time. In the Costly Verification (T2) treatment, verification is higher after receiving a high free message (this increases over time). However, the use of the (self-)verification instruments is considerably lower than theoretically predicted.*

### 3.4.5 Investor Efficiency

Finally, we are interested in how the (self-)verification instruments affect the efficiency of investors (making payoff-maximizing decisions). For this purpose, we define the *Investor Efficiency* as

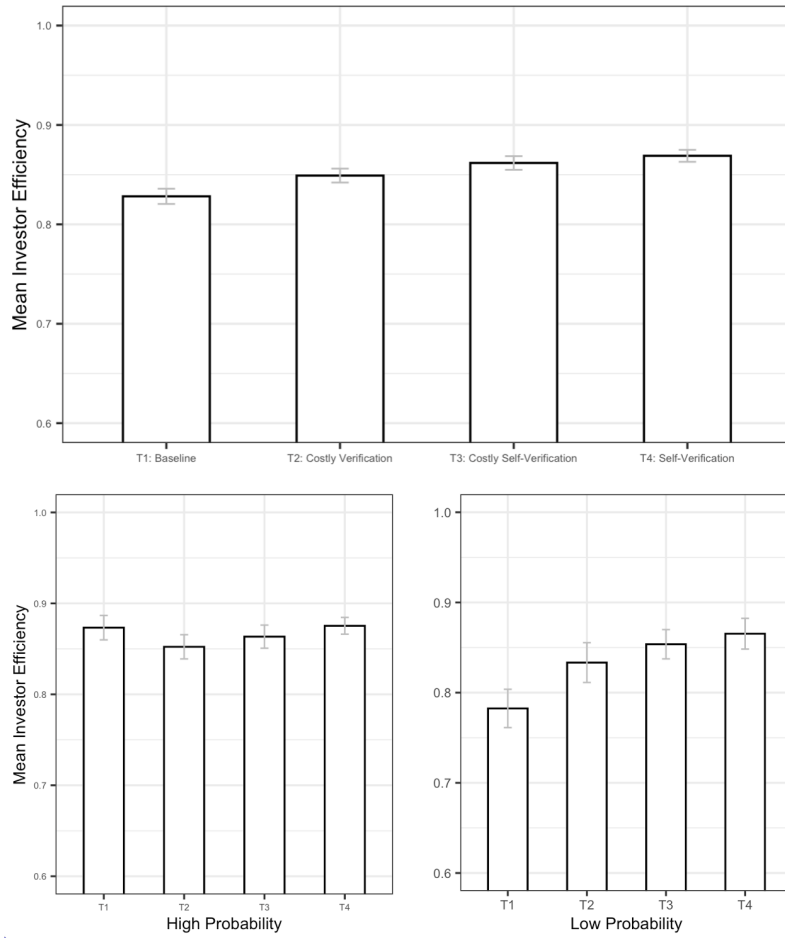
$$\frac{\text{Observed expected investor payoff}}{\text{Maximum expected investor payoff (under perfect information)}}$$

where *observed expected investor payoff* is the expected payoff given the observed investment and the probability of success (but before the realization of success or failure). *Maximum expected investor payoff* is equal to  $2\theta w$  if the investor is matched with an hSP firm and equal to  $w$  if the investor is matched with a lSP firm. The top panel of Figure 3.6 shows the average Investor Efficiency in periods 3 – 13. The average Investor Efficiency shows that investors earn 83%, 85%, 86%, and 87% of the maximum payoffs in the respective T1-T4 treatments.

Table 3.5 shows random-effects panel data regressions. Consistent with the bottom panel of Figure 3.6, Models 2 and 3 in Table 3.5 show that Investor Efficiency is higher when matched with hSP firms than with lSP firms only in *Baseline* (t-test,  $p = 0.004$  in both Models 2 and 3), but not in treatments T2-T4 (e.g., in Model 3 the Wald test yields  $p = 0.5688$ ,  $p = 0.5360$ , and  $p = 0.4378$  in T2, T3, and T4).<sup>19</sup> Investor Efficiency of hSP is also not different when pairwise compared between Costly Verification, Costly Self-Verification, and Self-Verification. This suggests that under the verification treatment (T2) and under the self-verification treatments (T3-T4), investors make equally efficient investments regardless of whether the firm is “good” or “bad.” In contrast, in the *Baseline* (T1), the efficiency of investors’ decisions depends on their “luck” in being matched with a “good” firm. Moreover, Investor Efficiency does not change over time in any treatment. Finally, for lSP firms, the verification and self-verification treatments (T2-T3) yield significantly higher Investor Efficiency than in *Baseline* (T1), but not in T4 ( $p = 0.030$ ,  $p = 0.028$ , and  $p = 0.188$  comparing T1 with T2, T3, and T4, respectively, in Model 3).

Therefore, when investors can make informed decisions (in T2-T4), the Investor Efficiency does not depend on the randomly determined probability

<sup>19</sup> We test the null hypothesis that the sum of the coefficients  $\beta_{[\text{hSP}]}$  and  $\beta_{[\text{hSP} \times \text{respective treatment}]}$  equals zero.



The top panel shows the average Investor Efficiency across all probabilities of success. The bottom panel shows the average Investor Efficiency per high (left-hand panel) and low (right-hand panel) probability of success. The error bars indicate 95% confidence intervals.

**Figure 3.6:** Average Investor Efficiency

of success (high or low). Finally, in *Costly Verification* (T2), we expected investors to verify each message, which reduces Investor Efficiency. Interestingly, we find no differences in Investor Efficiency in treatment T2 when compared to the self-verification treatments T3-T4.

**Result 5.** *Baseline (T1) is the only treatment where Investor Efficiency depends on the firms' probability of success. In the (self-)verification treatments (T2-T4), investors have similar efficiency for low and high probability firms.*

**Table 3.5:** Investor Efficiency

	Model (1)	Model (2)	Model (3)
Treatment T2	0.04 (0.03)	0.07** (0.03)	0.07** (0.03)
Treatment T3	0.04 (0.03)	0.07** (0.03)	0.07** (0.03)
Treatment T4	0.01 (0.03)	0.05 (0.03)	0.05 (0.03)
high probability of success (hSP)	-	0.09*** (0.03)	0.09*** (0.03)
hSP $\times$ Treatment T2	-	-0.07 (0.04)	-0.07 (0.04)
hSP $\times$ Treatment T3	-	-0.07** (0.04)	-0.07** (0.04)
hSP $\times$ Treatment T4	-	-0.07* (0.04)	-0.07* (0.04)
Period	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Treatment T2 $\times$ Period	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Treatment T3 $\times$ Period	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Treatment T4 $\times$ Period	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Risk test	-	-	0.00 (0.00)
Constant	0.84*** (0.02)	0.79*** (0.02)	0.78*** (0.03)
Session FE	Yes	Yes	Yes
Observations	3,285	3,285	3,285
R-squared (overall)	0.022	0.041	0.042

Note: Random-effects panel data regressions. Dependent variable: Investor Efficiency. Standard errors are clustered across investors. “Risk test” is taken from the post-experimental questionnaire: It is the amount invested (between €0 and €10) in a risky lottery, where a lower amount indicates a higher degree of risk aversion. Standard errors are clustered across investors. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

### 3.5 Revisiting our predictions: Alternative explanations for the observed investment behavior

Our results suggest that free communication is more informative than theoretically predicted and that a substantial fraction of investors trust the (unverified) free messages leading to higher average investments for firms with a high probability of success than for their low probability counterparts in all treatments (see Figure 3.2 in Section 3.4.1). However, our prediction for *Baseline* was a babbling equilibrium with no investment. We consider three alternative models that could potentially explain the observed behavior. The

first is the model of inequity aversion (Fehr & Schmidt, 1999), which suggests that investors might invest to reduce the difference between the investors' and firms' expected payoffs. The second is the model of level- $k$  reasoning (e.g., Nagel, 1995), which gives rise to the conjecture that the players start with naive beliefs about their counterparts and adjust their beliefs accordingly by  $k$  iterations. In Appendices 3.C.1 and 3.C.2, we show that these models do not give a different prediction than our standard model.

The third is the model of intrinsic cost of lying (Gneezy et al., 2018). If we adopt our model by assuming that if a firm has an intrinsic cost of lying  $c_l$ , it would only lie if the expected payoff, based on its belief in how its message changes the investor's belief in the probability of success, is greater than  $c_l$ . We define  $i^*(m)$  as the amount of investment expected by the firm after sending message  $m$ . That is, a firm with probability of success  $\theta$  sends a deceptive message  $m$  if

$$i^*(m)\theta \geq c_l.$$

Given our parameters, each investment amount  $i$  is less or equal to €11.00 and  $\theta \leq 20\%$  for all  $\theta \in \Theta_l$ . Therefore, for all firms with a low probability of success, we have

$$i^*(m)\theta \leq €11.00 \times 20\% = €2.20.$$

That is, if  $c_l > €2.20$ , a firm with a low probability of success does not lie (even if it assumes that the message affects the investor's belief), and the investor can trust the message in equilibrium.

We have not specifically tested for the intrinsic cost of lying. However, in Section 3.4.2, we find that firms often tell the truth, a result consistent with a model assuming the intrinsic cost of lying. In particular, in T1 and T2, 58.52% and 50.28% of the free messages sent by firms with low probability of success are truthful. In T3 and T4, 41.78% and 48.85% of the LSP firms either send a free message with low probability or self-verify. High-probability firms almost never report a low probability of success.<sup>20</sup> If the investor believes that a certain share of messages indicating a high probability of success are true, such messages may influence the investor's belief about the expected probability of success in such a way that the investor invests a positive investment amount. In Appendix 3.C.3, we show with a numerical example that if the investor believes that at least 29% of firms are truth-telling (which is considerably

---

<sup>20</sup> The latter behavior occurred in 0.7%, 0.0%, 0.8%, and 0.2% of the cases in T1, T2, T3, and T4).



lower than the actually observed truthfulness), a message indicating a high probability of success leads to a positive investment amount and would create a separating investment pattern between firms with low and high probability of success.

### 3.6 Conclusions and discussion

In this paper, we introduced the Probabilistic Investment Game to study the effect of costly verification, costly self-verification, and cheap self-verification instruments on investment in risky firms compared to a situation based on trust alone (our baseline game). Our main finding is that investments in firms with a high probability of success are substantially higher than in firms with a low probability of success in all treatments, even in the *Baseline* treatment (where the firm’s probability of success remains unknown to the investor). Such a segregated investment pattern is consistent with models that assume an intrinsic cost of lying. This finding that free messages contain more information than theoretically predicted (“cheap talk”) and that investors respond to this information with positive investments extends previous studies of “cheap talk” communication (e.g., Dickhaut et al., 1995; Cai et al., 2006; Jin et al., 2021; Fréchette et al., 2022) to a probabilistic game situation. This result is far from trivial, as Kajackaite and Gneezy (2017) show that people lie more when they do not fear being exposed as liars. The probabilistic nature of our experiment, compared to a deterministic setting, may encourage deceptive behavior because investors cannot infer the probability of success from the outcome of the lottery (failure can be attributed to bad luck). In spite of this, we find high levels of truthfulness and trustworthiness.

In addition, we expected in the self-verification treatments (T3 and T4) that any free message (i.e., a message that is not self-verified) would be treated as a revelation of a low probability of success. Consequently, this should have led to a fully informative equilibrium. In fact, we find that firms with a low probability of success almost never self-verify, and firms with a high probability of success self-verify less than expected. These results resemble the findings of Benndorf et al. (2015), although they use a very different design and context.<sup>21</sup> But unlike Benndorf et al. (2015), our findings cannot be explained

<sup>21</sup> In their labor market experiment, workers can reveal their productivity but then impose a negative externality on others and also put pressure on the other workers to reveal. They found that participants predicted to disclose in equilibrium were significantly more likely to violate the prediction than participants predicted to conceal.

by level- $k$  reasoning. In particular, we observe that investors, on average, invest a positive amount even when firms do not self-verify and instead send a free message claiming a high probability of success in the self-verification treatments (T3 and T4). This indicates that (some) investors must believe that a firm with a high probability of success will not necessarily self-verify. Even in Treatment T4, where self-verification is very cheap, investors seem to hold such beliefs. This slows down the unraveling effect needed to overcome the overall information asymmetry.

Two features of our design may have affected the substantial truth-telling and trust that we already observed in *Baseline*: First, in our online experiment, the experimenter could observe the firms' probability of success and their free messages. Thus, participants who are sensitive to their image in the eyes of the experimenter may avoid lying. Second, under the game's parameters, deception causes considerable harm to the investor but provides little benefit to the firm (see Gneezy et al., 2018). For example, if a firm has a low probability of success of 15% and sends a free message claiming a high probability of success of 75%, investing her full endowment will reduce the investor's expected payoff from €11.00 to €1.65, whereas the firm's expected payoff will only increase from €0.00 to €1.65. Future research should test the self-verification instruments in situations where the gain from lying is considerably larger.

In summary, our findings have both optimistic and pessimistic interpretations: On the positive side, people are more truth-telling and more trusting than initially assumed. In fact, the overall efficiency of investment decisions based on truth-telling and trust alone is not lower than when (self-)verification is used. On the negative side, the availability of a cheap self-verification technology does not eliminate the information asymmetry. This undermines a potentially important benefit of blockchain technology, or the like (e.g., Gephart et al., 2019; Rajasekaran et al., 2022). Thus, merely creating a technology is not enough to solve the problem; its adoption must become widespread to realize its potential.

## Bibliography

- Abbe, E., Khandani, A., & Lo, A. (2012). Privacy-preserving methods for sharing financial risk exposures. *American Economic Review*, 102(3), 65–70.
- Abeler, J., Nosenzo, D., & Roymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4), 1115–1153.
- Akerlof, G. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488–500.
- Arrow, K. (1968). The economics of moral hazard: Further comment. *The American Economic Review*, 58(3), 537–539.
- Bapnaa, S. (2019). Complementarity of signals in early-stage equity investment decisions: Evidence from a randomized field experiment. *Management Science*, 65(2), 933–952.
- Benndorf, V., Kübler, D., & Normann, H. T. (2015). Privacy concerns, voluntary disclosure of information, and unraveling: An experiment. *European Economic Review*, 75, 43–59.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1), 122–142.
- Bernstein, S., Korteweg, A., & Laws, K. (2017). Attracting early-stage investors: Evidence from a randomized field experiment. *Journal of Finance*, 72(2), 509–538.
- Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 72(2), 321–338.
- Blume, A., Lai, E., & Lim, W. (2020). Strategic information transmission: A survey of experiments and theoretical foundations. In *Handbook of Experimental Game Theory* (pp. 311–347). Cheltenham, UK: Edward Elgar Publishing Ltd.
- Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71, 117–120.
- Bogdanov, D., Talviste, R., & Willemson, J. (2012). Deploying secure multi-

- party computation for financial data analysis. In *International Conference on Financial Cryptography and Data Security* (pp. 57–64).
- Burdea, V., Montero, M., & Sefton, M. (2023). Communication with partially verifiable information: An experiment. *Games and Economic Behavior*, *142*, 113–149.
- Cai, H., Tao, J., & Wang, Y. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior*, *56*(1), 7–36.
- Cameron, C., Gelbach, J., & Miller, D. (2011). Robust inference with multiway clustering. *Journal of Business and Economic Statistics*, *29*(2).
- Cason, T. N., & Gangadharan, L. (2002). Environmental labeling and incomplete consumer information in laboratory markets. *Journal of Environmental Economics and Management*, *43*(1), 113–134.
- Chen, D., Schonger, M., & Wickens, C. (2016). oTree - An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.
- Chen, Y., & Bellavitis, C. (2020). Blockchain disruption and decentralized finance: The rise of decentralized business models. *Journal of Business Venturing Insights*, *13*, 88–97.
- Cohen, T., Gunia, B., Kim-Jun, S. Y., & Murnighan, K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology*, *45*(6), 1321–1324.
- Correia, S. (2019). *Reghdfe: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects*. Boston, USA: Boston College Department of Economics.
- Crawford, V. (1998). A survey of experiments on communication via cheap talk. *Journal of Economic Theory*, *78*(2), 286–298.
- Crawford, V., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*(6), 1431–1451.
- Dickhaut, J., McCabe, K., & Mukherji, A. (1995). An experimental study of strategic information transmission. *Economic Theory*, *6*(3), 389–403.
- Dranove, D., & Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, *48*(4), 935–963.
- Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723–733.
- Etilé, F., & Teyssier, S. (2016). Signaling corporate social responsibility:

- Third-party certification versus brands. *The Scandinavian Journal of Economics*, 118(3), 397–432.
- Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114(3), 817–868.
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525–547.
- Forsythe, R., Lundholm, R., & Rietz, T. (1999). Cheap talk, fraud, and adverse selection in financial markets: Some experimental evidence. *The Review of Financial Studies*, 12(3), 481–518.
- Fréchette, G., Lizzeri, A., & Perego, J. (2022). Rules and commitment in communication: An experimental analysis. *Econometrica*, 90(5), 2283–2318.
- Gephart, J., Froehlich, H., & Branch, T. (2019). Opinion: To create sustainable seafood industries, the United States needs a better accounting of imports and exports. *Proceedings of the National Academy of Sciences*, 116(19), 9142–9146.
- Ghidoni, R., Cleave, B., & Suetens, S. (2019). Perfect and imperfect strangers in social dilemmas. *European Economic Review*, 116, 148–159.
- Glazer, J., & Rubinstein, A. (2004, nov). On optimal rules of persuasion. *Econometrica*, 72(6), 1715–1736.
- Glazer, J., & Rubinstein, A. (2006). A study in the pragmatics of persuasion: a game theoretical approach. *Theoretical Economics*, 1, 395–410.
- Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1), 384–394.
- Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, 108(2), 419–453.
- Gneezy, U., Leonard, K., & List, J. (2009). Gender Differences in competition: evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5), 1637–1664.
- Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, 112(2), 631–645.
- Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization*, 93, 293–300.
- Gonzalez, L. (2019). Blockchain, herding and trust in peer-to-peer lending. *Managerial Finance*, 46(6), 815–831.

- Greif, A. (1993). Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. *The American Economic Review*, 83(3), 525–548.
- Guiso, L., Sapienza, P., & Zingales, L. (2004, June). The role of social capital in financial development. *American Economic Review*, 94(3), 526–556.
- Güth, W., Muegera, H., Musau, A., & Ploner, M. (2014). Deterministic versus probabilistic consequences of trust and trustworthiness: An experimental investigation. *Journal of Economic Psychology*, 42, 28–40.
- Hagenbach, J., & Perez-Richet, E. (2018). Communication with evidence in the lab. *Games and Economic Behavior*, 112, 139–165.
- Holt, C., & Laury, S. (2002). Risk aversion and incentive effects. *The American Economic Review*, 92(5), 1644–1655.
- Hurkens, S., & Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. *Experimental Economics*, 12(2), 180–192.
- Hynes, R. (2017). Maximum possible accuracy in credit reports. *Law and Contemporary Problems*, 80(3), 87–115.
- Jin, G. Z., Luca, M., & Martin, D. (2021). Is no news (perceived as) bad news? An experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, 13(2), 141–173.
- Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, 102, 433–444.
- Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6), 2590–2615.
- Knack, S., & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics*, 112(4), 1251–1288.
- Kreps, D. (1990). Corporate culture and economic theory. *Perspectives on Positive Political Economy*, 90(8), 90–143.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644.
- Milgrom, P. (2008). What the seller won't tell you: Persuasion and disclosure in markets. *Journal of Economic Perspectives*, 22(2), 115–131.
- Milgrom, P., North, D., & Weingast, B. (1990). The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics & Politics*, 2(1), 1–23.
- Myerson, R. B. (1991). *Game theory: Analysis of conflict*. Cambridge, USA:

Harvard University Press.

- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, 85(5), 1313–1326.
- Ornes, S. (2019). Blockchain offers applications well beyond Bitcoin but faces its own limitations. *Proceedings of the National Academy of Sciences of the United States of America*, 116(42), 20800–20803.
- Rajasekaran, A. S., Azees, M., & Al-Turjman, F. (2022). A comprehensive survey on blockchain technology. *Sustainable Energy Technologies and Assessments*, 52, 102039.
- Sobel, J. (2013). Giving and receiving advice. *Advances in Economics and Econometrics*, 1, 305–341.
- Sobel, J. (2020). Lying and deception in games. *Journal of Political Economy*, 128(3), 907–947.
- Stenzel, A., Requate, T., & Waichman, I. (2022). *The probabilistic investment game*. (Unpublished manuscript)
- Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal*, 119(534), 47–60.
- Tapscott, A., & Tapscott, D. (2017). How Blockchain is changing finance. *Harvard Business Review*, 1(9), 2–5.
- Vanberg, C. (2017). Who never tells a lie? *Experimental Economics*, 20(2), 448–459.
- Viscusi, K. (1978). A note on "lemons" markets with quality certification. *The Bell Journal of Economics*, 9(1), 277–279.
- Vranceanu, R., Sutan, A., & Dubart, D. (2012). Trust and financial trades: Lessons from an investment game where reciprocators can hide behind probabilities. *Journal of Socio-Economics*, 41(1), 72–78.
- Wang, Y., & Kogan, A. (2018). Designing confidentiality-preserving Blockchain-based transaction processing systems. *International Journal of Accounting Information Systems*, 30, 1–18.
- Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2019). Blockchain technology overview. *arXiv preprint arXiv:1906.11078*.
- Zak, P., & Knack, S. (2001). Trust and growth. *The Economic Journal*, 111(470), 295–321.

## Appendix

### 3.A Theoretical predictions

In the following, we present our predictions and formal proofs for the different treatments. Given the finite number of rounds (15), players can use backward induction to solve the game. Reputation building is unlikely to occur in random matching, and cooperative coordination can be ruled out (Ghidoni et al., 2019). Therefore, our predictions are the same as those for a one-shot game. We apply a solution concept for a Bayesian game with communication from Myerson (1991). In general, it is a difficult problem to derive all equilibria of a Bayesian game with communication. It simplifies the analysis to start by analyzing the Bayesian game without communication (i.e., the firm does not send a free message) and then add communication in a second step. Using the revelation principle for Bayesian games, we then derive the equilibria of the Probabilistic Investment Game.

#### A Bayesian game without communication

In general, a Bayesian game (without communication) consists of a set of players and, for each player, a set of possible actions, a set of possible types, a probability function representing what a player believes about the other types given its own type, and a utility function. In the Probabilistic Investment Game, we have two players – an investor and a firm. Only the firm can be of different types (the success probabilities), and only the investor has different actions to choose from (the different investment amounts). Let  $C = \{0, 1, \dots, w\}$  be the set of possible actions for the investor, and  $\Theta = \Theta_l \cup \Theta_h$  be the set of possible types the firm can adopt. For each  $\theta \in \Theta$ , let  $p(\theta) = \frac{1}{22}$  be the investor's probability function describing her belief about the firm's type (probability of success). If investors and firms are risk neutral (we consider the case of risk aversion in section 3.B.1), then given a probability of success  $\theta$ , their utility equals their expected payoffs. We denote the investor's and the



firm's utility functions as

$$u_F(i, \theta) = \theta i, \quad (3.1)$$

$$u_I(i, \theta) = 2\theta i + w - i. \quad (3.2)$$

Note that, given the parameters explained above, we have

$$\sum_{\theta \in \Theta} p(\theta) u_I(i, \theta) = \frac{1}{22} \sum_{\theta \in \Theta} (2\theta i + w - i) = -0.1i + w$$

which we will use repeatedly in the following proofs. This structure defines a Bayesian game without communication according to Myerson (1991). A *randomized-strategy* (for the investor)  $\sigma$  in the set  $\Delta(C)$  (the set of probability distributions over  $C$ ) satisfies

$$\begin{aligned} \sigma(i) &\geq 0, \forall i \in C, \\ \sum_{i \in C} \sigma(i) &= 1 \end{aligned}$$

where  $\sigma(i)$  is the probability that the investor will invest an amount  $i$ . Since there is nothing for the firm to choose, a Bayesian equilibrium of the Probabilistic Investment Game without communication consists only of a randomized strategy  $\sigma$  that maximizes the investor's expected payoff, i.e.,

$$\sigma(\cdot) \in \operatorname{argmax}_{\tau \in \Delta(C)} \sum_{\theta \in \Theta} p(\theta) \sum_{i \in C} \tau(i) u_I(i, \theta). \quad (3.3)$$

**Theorem 1.** *The unique Bayesian equilibrium of the Probabilistic Investment Game without communication is  $\sigma^*(0) = 1$ ; i.e., the investor does not invest.*

*Proof.* For any (mixed) strategy  $\tau \in \Delta(C)$  suggesting a positive investment amount, there exists at least one  $i \in C$  with  $i > 0$  and  $\tau(i) > 0$ . For all such

strategies, we have

$$\begin{aligned}
\sum_{\theta \in \Theta} p(\theta) \sum_{i \in C} \tau(i) u_I(i, \theta) &= \sum_{i \in C} \tau(i) \sum_{\theta \in \Theta} p(\theta) u_I(i, \theta) \\
&= \sum_{i \in C} \tau(i) \sum_{\theta \in \Theta} p(\theta) (2\theta i - i + w) \\
&= \sum_{i \in C} \tau(i) (-0.1i + w) \\
&< w \\
&= \sum_{\theta \in \Theta} p(\theta) u_I(0, \theta) \\
&= \sum_{\theta \in \Theta} p(\theta) \sum_{i \in C} \sigma^*(i) u_I(i, \theta).
\end{aligned}$$

Therefore, the only strategy optimising (3.3) is  $\sigma^*(0) = 1$ . I.e., the investor does not invest.  $\square$

## Baseline

We will now use the revelation principle for Bayesian games to derive the equilibria for the Probabilistic Investment Game (with communication). Consider a mediated communication system of the following form: First, the firm confidentially reports its probability of success to a mediator (the investor has only one type and therefore nothing to report). After receiving the report, the mediator confidentially recommends an investment amount to the investor (the firm has only one action and therefore receives no recommendation). The agent's recommendation depends on the firm's report either deterministically or randomly. Let  $\mu(i \mid m)$  denote the conditional probability that the agent would recommend the investment amount  $i$  to the investor after the firm reports the probability of success  $m \in \Theta$ . Obviously,  $\mu(i \mid m)$  must satisfy the probability constraints  $\sum_{c \in C} \mu(c \mid m) = 1$  and  $\mu(c \mid m) \geq 0$  for all  $m \in \Theta$ . In general, any such function  $\mu : T \rightarrow \Delta(C)$  is called a *mediation plan* or *mechanism* for a Bayesian game with communication.

If the firm honestly reports the probability of success to the mediator, and the investor follows the mediator's recommendation, then the expected utility for the firm with the probability of success  $\theta$  and the investor from the mechanism  $\mu$  would be

$$U_F(\mu \mid \theta) = \sum_{i \in C} \mu(i \mid \theta) u_F(i, \theta), \quad (3.4)$$

$$U_I(\mu) = \sum_{\theta \in \Theta} \sum_{i \in C} p(\theta) \mu(i \mid \theta) u_I(i, \theta). \quad (3.5)$$

However, we must recognize that firms may lie about their type, or investors may disobey the mediator's recommendation. Thus, a mechanism  $\mu$  induces a communication game in which firms must choose their report, and investors must choose an action in  $C$  as a function of the mediator's recommendation. Formally, for the induced communication game, we denote the action space of the firm and the investor as

$$B = \{(m, \delta) \mid m \in \Theta, \delta : C \rightarrow C\}.$$

A strategy  $(m, \delta)$  in  $B$  represents a combined plan for the firm to report the probability of success  $m$  to the mediator and for the investor to choose an amount to invest in  $C$  as a function of the mediator's recommendation according to  $\delta$ . I.e., the investor would invest  $\delta(i)$  if the mediator recommended  $i$ . The action cannot depend on the firm's reported probability of success because the firm communicates separately and confidentially with the mediator.

Now, suppose that the firm reports its probability of success  $\theta$  truthfully to the mediator. Then the investor's expected utility payoff of the mechanism  $\mu$  would be

$$U_I^*(\mu, \delta) = \sum_{\theta \in \Theta} \sum_{i \in C} p(\theta) \mu(i \mid \theta) u_I(\delta(i), \theta).$$

Assuming that the investor follows the mediator's investment recommendation, the firm's expected utility payoff would be

$$U_F^*(\mu, m \mid \theta) = \sum_{i \in C} \mu(i \mid m) u_F(i, \theta).$$

We say that a mechanism  $\mu$  is *incentive compatible* if and only if it is a Bayesian equilibrium for the firms to report their types truthfully and for the investors to follow the mediator's recommendation when the mediator uses the mechanism  $\mu$ . Thus  $\mu$  is incentive compatible if and only if it satisfies the following general incentive constraints:

$$U_I(\mu) \geq U_I^*(\mu, \delta), \forall \delta : C \rightarrow C, \quad (3.6)$$

$$U_F(\mu \mid \theta) \geq U_F^*(\mu, m \mid \theta), \forall \theta \in \Theta, \forall m \in \Theta. \quad (3.7)$$

If the mediator is using an incentive-compatible mediation plan and the firm communicates independently and confidentially with the mediator, then the firm cannot gain by lying, and the investor cannot gain by disregarding the investment recommendation.

By the revelation principle for general Bayesian games (see e.g., Myerson, 1991), for any Bayesian equilibrium of a communication game, we can construct an equivalent (i.e., each type of player gets the same expected utility) incentive-compatible mechanism in the following way: The mediator asks the firm about its probability of success. Then, the mediator computes the probability of success that the firm would report in equilibrium. Then, the mediator computes the investment amount that the investor would choose as a function of that report and sends it as a recommendation to the investor. We now test whether different equilibrium candidates for the Probabilistic Investment Game are incentive-compatible.

**Lemma 1.** *The mechanism  $\mu^*(0 \mid \theta) = 1$  for any  $\theta \in \Theta$  is incentive compatible.*

*Proof.* First, we show that the investor's incentive constraint (3.6) is satisfied. We have  $\mu^*(0 \mid \theta) = 1$  and  $\mu^*(i \mid \theta) = 0$  for all  $i \in C$  with  $i > 0$  and therefore

$$\begin{aligned} U_I^*(\mu^*, \delta) &= \sum_{\theta \in \Theta} \sum_{i \in C} p(\theta) \mu^*(i \mid \theta) u_I(\delta(i), \theta) \\ &= \sum_{\theta \in \Theta} p(\theta) u_I(\delta(0), \theta) \\ &= \sum_{\theta \in \Theta} p(\theta) (2\theta\delta(0) + w - \delta(0)) \\ &= -0.1\delta(0) + w. \end{aligned}$$

For all  $\delta : C \rightarrow C$ , we have  $\delta(0) \geq 0$  and therefore

$$U_I^*(\mu^*, \delta) = -0.1\delta(0) + w \leq w = U_I(\mu^*).$$

This means that the investor cannot gain by disobeying the mediator's recommendation, and the incentive constraint (3.6) is satisfied.

Second, we show that the firm's incentive constraint (3.7) is satisfied. If the investor follows the mediator's recommendation and does not invest, the firm's payoff is 0 for all probabilities of success. That is,

$$U_F(\mu^* \mid \theta) = U_F^*(\mu^*, m \mid \theta) = 0$$

for all  $\theta, m \in \Theta$  and (3.7) is satisfied.  $\square$

Therefore, not investing is also an equilibrium of our game with communication. Can any communication by the firm lead to a positive investment equilibrium? According to the revelation principle, the answer is "no."

**Lemma 2.** *Every mechanism recommending a positive investment amount to the investor is not incentive-compatible.*

*Proof.* We will show that any mechanism recommending a positive investment amount that satisfies the firm's incentive constraint (3.7) does not satisfy the investor's incentive constraint (3.6). To show this, we assume that there is an incentive-compatible mechanism  $\mu$  recommending a positive investment amount and satisfying the firm's incentive constraint (3.7). If the investor follows the mediator's recommendation  $\mu(i \mid \cdot)$ , the firm cannot gain by lying to the mediator. Therefore, the mediator's recommendation must be independent of the message. I.e., there exists an  $i^* > 0$  with

$$\sum_{i \in C} \mu(i \mid \theta) i \theta = i^* \theta \quad (3.8)$$

for any  $\theta \in \Theta$ . Otherwise, some firms with a specific probability of success could gain by lying to the mediator.

We now show that if (3.8) is satisfied, the mechanism  $\mu$  cannot satisfy the investor's incentive constraint (3.6). For the investor, we have

$$U_I(\mu) = \sum_{\theta \in \Theta} \sum_{i \in C} p(\theta) \mu(i \mid \theta) u_I(i, \theta) \quad (3.9)$$

$$= \sum_{\theta \in \Theta} p(\theta) \sum_{i \in C} \mu(i \mid \theta) (2\theta i - i + w) \quad (3.10)$$

$$= \sum_{\theta \in \Theta} p(\theta) \sum_{i \in C} (2\theta \mu(i \mid \theta) i - \mu(i \mid \theta) i + \mu(i \mid \theta) w) \quad (3.11)$$

$$= \sum_{\theta \in \Theta} p(\theta) (2\theta i^* - i^* + w) \quad (3.12)$$

$$= -0.1i^* + w < w = U_I^*(\mu, \delta) \quad (3.13)$$

with  $\delta(i) = 0$  for all  $i \in C$ . This contradicts the incentive constraint for the investor and proves the Lemma.  $\square$

The last lemma means that there cannot be an equilibrium for the Probabilistic Investment Game with a positive amount of investment. Therefore, the equilibrium in the Probabilistic Investment Game does not change compared to the game without communication.

**Theorem 2.** *There is a unique equilibrium for the Probabilistic Investment Game characterized by the incentive-compatible mechanism  $\mu^*(0 | \theta) = 1$  for any  $\theta \in \Theta$ .*

### Costly Verification (T2)

First, we consider *Costly Verification* (T2) without communication. The investor's action space is now extended by the decision to purchase information about the firm's probability of success before making an investment decision. Assuming risk neutrality, given a probability of success  $\theta$ , the investor's utility function depends on her decision whether to purchase the information at a fixed cost  $c$  ( $v = 1$ ) or not ( $v = 0$ ) and the investment amount  $i$  as follows

$$u_I(v, i, \theta) = 2\theta i + w - i - vc. \quad (3.14)$$

**Lemma 3.** *The equilibrium for Costly Verification (T2) without communication is as follows: The investor buys the verification and invests according to her risk preference if the firm's probability of success is high and does not invest if it is low.*

*Proof.* If the investor does not buy the verification, the game is equivalent to the *Baseline* treatment, and the investor does not invest. To prove the equilibrium, we need to show that the expected payoff of the investor who buys the verification is higher than that of the investor who does not invest.

If the verification is bought, the investor is informed about the firm's probability of success  $\theta$ . We define the investor's best action  $i^*(\theta) \in C$  depending on the received information about the firm's probability of success  $\theta$ . The investor buys the information if

$$\sum_{\theta \in \Theta} p(\theta) u_I(1, i^*(\theta), \theta) \geq w. \quad (3.15)$$

For a risk-neutral investor, the strategy that maximizes her payoff is to invest the full endowment  $i^*(\theta) = w$  if  $\theta \in \Theta_h$  and to invest nothing otherwise, i.e.,  $i^*(\theta) = 0$  if  $\theta \in \Theta_l$ . We chose the parameter  $c = \text{€}1$  so that equation (3.15) holds. Consequently, buying the verification optimizes the equilibrium condition (3.3).  $\square$

Can there be an equilibrium with higher payoffs for the firm and the investor if we add the free message stage? The only way to increase the investor's payoff would be to save on the cost of verification (i.e., to not buy the verification). However, with the revelation principle, there is no equilibrium with a positive investment amount without verification. Hence, the equilibrium theorem follows.

**Theorem 3.** *The unique equilibrium of Costly Verification (T2) is as follows: The investor buys the verification and invests according to her risk preference when the firm's probability of success is high and does not invest when it is low.*

### Costly Self-Verification (T3) and Self-Verification (T4)

Finally, we turn to the treatments where the firm chooses to send either a self-verified message (with its true probability of success) or a free message to the investor. The firm's action space is now the self-verification decision, and its utility function (assuming risk neutrality) depends on the probability of success  $\theta$ , the decision whether to send the certified message at a fixed cost  $c$  ( $v = 1$ ) or not ( $v = 0$ ), and the amount invested  $i$ . The utility function is then given by:

$$u_F(i, \theta, v) = \theta i - vc. \quad (3.16)$$

**Theorem 4.** *The unique equilibrium for the self-verification treatments T3 and T4 is as follows: When the probability of success is high, the firm self-verifies and does not otherwise. If the investor receives a self-verified message, she invests according to her risk preference and does not invest otherwise.*

*Proof.* If the firm does not self-verify its probability of success, the investor does not invest. For a firm with a low probability of success ( $\theta \in \Theta_l$ ), the investor is expected not to invest (regardless of her risk preferences), and therefore, the equilibrium strategy for the firm is to save the cost of verification

(i.e., not to send the certified message) and to send a cheap talk message instead. For firms with a high probability of success  $\theta_h$ , an investor's best action is to invest  $i^*(\theta_h)$  according to her risk preferences given by equation (3.19). We have  $\theta i^*(\theta) \geq c$ , and therefore, for a firm with a high probability of success, it is payoff-optimizing to self-verify. Note that even for highly risk-averse investors,  $\theta i^*(\theta) \geq c$  holds for  $\theta \in \Theta_h$ .

As in the previous treatments, it is easy to show that any mechanism in which the mediator recommends a positive investment amount without self-verification is not incentive-compatible. *Self-Verification* (T4) is analogous to *Costly Self-Verification* (T3) but with a lower cost  $0 < \tilde{c} < c$  for sending the certified message containing the true probability of success. However, the theoretical predictions remain the same.  $\square$

### 3.B Risk aversion

In the following, we first extend the utility functions of the theory section 3.A to show that the predictions for risk-neutral and risk-averse investors are largely the same. Second, we show estimates of risk aversion according to our post-experimental questionnaire.

#### 3.B.1 Reflecting risk aversion

To reflect investors' risk aversion, we adapt the utility function used by Holt and Laury (2002), i.e.,  $u(x) = x^{1-r}/(1-r)$  where  $r$  is the constant relative degree of risk aversion. For our Probabilistic Investment Game, the investor's expected utility, its first-order condition, and the resulting optimal investment amount given a probability of success  $\theta$  are obtained from Equations 3.17 - 3.19:

$$u_I(i, \theta, r) = \theta \frac{(w+i)^{1-r}}{1-r} + (1-\theta) \frac{(w-i)^{1-r}}{1-r}, \quad (3.17)$$

$$\frac{\partial u_I(i, \theta, r)}{\partial i} = \theta (w+i)^{-r} - (1-\theta) (w-i)^{-r} = 0, \quad (3.18)$$

$$i = w \frac{(1-\theta)^{\frac{-1}{r}} - \theta^{\frac{-1}{r}}}{(1-\theta)^{\frac{-1}{r}} + \theta^{\frac{-1}{r}}}. \quad (3.19)$$

According to Equation 3.19, a risk-averse investor will invest a positive amount when the firm's probability of success is high ( $\theta \in \Theta_h$ ), and nothing (as  $i \geq 0$ )



when it is low ( $\theta \in \Theta_l$ ). In our post-experiment survey, we elicited investors' risk aversion and found a slight risk aversion of  $r = 0.26$  on average (see Appendix 3.B.2). This means that, for example, for  $\theta = 75\%$ , such a slightly risk-averse investor ( $r = 0.26$ ) invests €10.68 (compared to a risk-neutral investor who invests €11.00).

Given a very similar investment behavior, the predictions for risk-averse investors remain the same as for risk-neutral investors. In particular, equation (3.15) also holds for the average risk aversion of  $r = 0.26$  that we observed in our post-experimental questionnaire.<sup>22</sup> Thus, in *Costly Verification* (T2), a risk-averse investor will also verify the firm's probability of success and invest according to equation (3.19) when the probability of success is high.

### 3.B.2 Estimating the risk aversion of our participants

In our post-experimental questionnaire, we asked all participants the following question (following Gneezy & Potters, 1997; Gneezy et al., 2009):

*“Imagine that you have to decide how much of €10 to invest in a risky project. There is an equal chance that the investment will fail (50%) or succeed (50%). If the investment fails, you lose the amount you invested. If the investment succeeds, you receive 2.5 times the amount invested. How much do you invest?”*

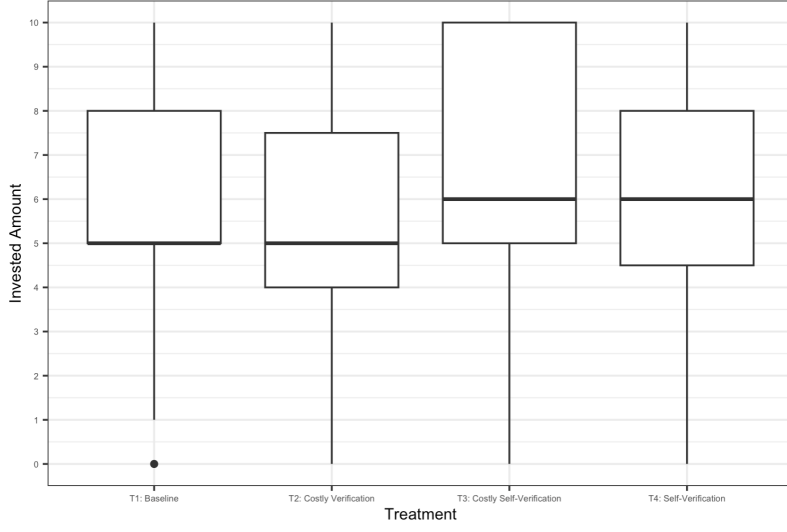
The average (median) investment across all treatments is €6.01 (€5.00). Figure 3.B.1 shows box plots of the investment decisions in the different treatments.<sup>23</sup>

We use the utility function with constant relative risk aversion by Holt and Laury (2002):  $u(x) = x^{1-r}/(1-r)$ , where  $r$  indicates the (constant) level of risk aversion. However, we adopt them to our risk testing task. In particular, we define investors with  $0 < r < 0.15$  as risk loving to risk neutral, with  $0.15 \leq r < 0.41$  as slightly risk averse, and with  $0.41 \leq r < 1$  as risk averse.<sup>24</sup> The expected utility function for the lottery in our survey and its first-order condition (depending on the amount invested  $i$ ) for the investor are:

<sup>22</sup> Equation (3.15) even holds for a very risk-averse investor with  $r \rightarrow 1$ .

<sup>23</sup> Regressing with clustered robust two-way standard errors (i) across *investors* and (ii) across *session*, we do not find differences across treatments.

<sup>24</sup> Holt and Laury (2002) used a different risk test. They classify participants with  $r \leq -0.15$  as risk loving,  $-0.15 \leq r < 0.15$  as risk neutral,  $0.15 \leq r < 0.41$  as slightly risk averse,  $0.41 \leq r < 0.68$  as risk averse, and  $0.68 \leq r < 1$  as very risk averse. Our risk test cannot account for risk-loving investors. Thus, our highest category combines risk-loving and risk-neutral individuals.



**Figure 3.B.1:** Investor's self-indicated risk aversion (box plot)

$$U(i) = \frac{1}{2} \frac{(10-i)^{1-r}}{1-r} + \frac{1}{2} \frac{(10+1.5i)^{1-r}}{1-r},$$

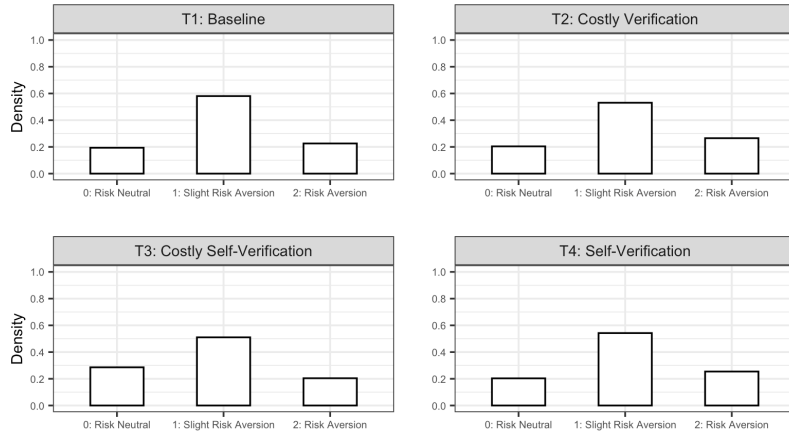
$$\frac{\partial U(i)}{\partial i} = -\frac{1}{2} (10-i)^{-r} + \frac{3}{4} (10+1.5i)^{-r} = 0,$$

with  $0 < r < 1$  as the risk aversion parameter. Solving the first-order condition for the risk aversion parameter as a function of the amount invested, we obtain

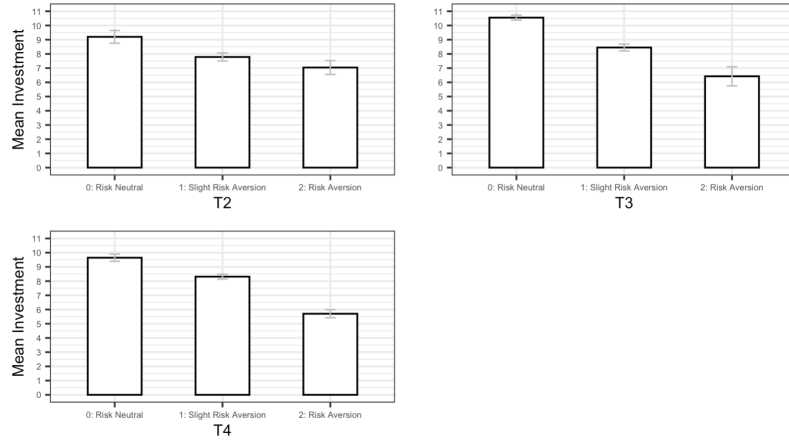
$$r = \frac{\ln\left(\frac{3}{2}\right)}{\ln\left(\frac{10+1.5i}{10-i}\right)}.$$

Plugging in the average (median) investment gives  $r = 0.26$  ( $r = 0.32$ ), indicating a slight risk aversion. Figure 3.B.2 shows the distribution of risk preferences, with a similar pattern across treatments. The risk test in the post-experimental questionnaire is not incentivized and could therefore suffer from hypothetical bias.

As a robustness check, we use the investment decisions for firms with a verified high probability of success and see if they are consistent with the classification shown in Figure 3.B.2. Figure 3.B.3 shows the average investment in (self-)verified firms with a high probability of success for the different types of risk aversion shown in Figure 3.B.2. It can be seen that the higher the risk aversion of the investor, the less she invests. This supports our classification of risk preferences from the post-experimental questionnaire.



**Figure 3.B.2:** Distribution of risk aversion across treatments



**Figure 3.B.3:** Average investments after (self-)verification of firms with a high probability of success

### 3.C Alternative models

In section 3.5, we examined whether three alternative models could explain the observed behavior better than our original model. In the following, we present in more detail the predictions based on inequity aversion, level- $k$  reasoning, and the intrinsic cost of lying.

#### 3.C.1 Inequity aversion

Fehr and Schmidt (1999) use the following utility function to model inequity aversion:

$$u_i(\mathbf{x}) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max[x_j - x_i, 0] - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max[x_i - x_j, 0],$$

where  $\mathbf{x} = (x_1, \dots, x_n)$  denotes the vector of earnings of the  $n$  group members. In addition,  $\alpha_i$  is the “disadvantageous inequality” parameter (in the case where a group member  $j$  earns more than the reference individual  $i$ ), while  $\beta_i$  is the “advantageous inequality” parameter (in the case where a group member  $j$  earns less than the reference individual). Applying this utility function to investors in our Probabilistic Investment Game, we obtain the following expected utility:<sup>25</sup>

$$\begin{aligned} E[u^I(i, l); \theta] &= \theta(w + i) + (1 - \theta)(w - i) - \beta_i \theta(w + i - i) - \beta_i(1 - \theta)(w - i) \\ &= i(2\theta - 1 + \beta_i(1 - \theta)) + w(1 - \beta_i). \end{aligned}$$

Under our parameters, for any high probability of success ( $70\% \leq \theta \leq 80\%$ ), the investor has a dominant strategy to invest the whole endowment. For any low probability of success ( $10\% \leq \theta \leq 20\%$ ), the investor’s dominant strategy depends on the advantageous inequality parameter  $\beta_i$ . If  $\beta_i \geq (1-2\theta)/(1-\theta)$ , the investor invests the whole endowment and does not invest otherwise. However, this condition is likely to be met by only a small fraction of the investors and does not characterize the average investor. In this respect, the supplementary material in Blanco et al. (2011) shows that only  $10/61 \approx 16\%$  of individuals have a  $\beta_i > 0.75$  (the condition for investing when  $\theta = 20\%$ ) and only  $2/61 \approx 3\%$  have a  $\beta_i > 0.89$  (the condition for investing when  $\theta = 10\%$ ). Thus, by and large, the inequity-version model gives the same predictions as under payoff maximization, as shown in Section 3.3.2. Thus, the inequity aversion model cannot explain the observed behavior any better than the payoff maximization model.

### 3.C.2 Level- $k$ reasoning

Similar to Benndorf et al. (2015), we apply level- $k$  reasoning to the Probabilistic Investment Game. We assume that level-0 firms send a randomized message independently of their actual probability of success, and level-0 investors in-

<sup>25</sup> Note that our game involves only two players (an investor and a firm), and the firm cannot earn more than the investor (so the  $\alpha$  parameter is omitted).

vest a random amount independently of the message. We then calculate the best replies for  $k \geq 1$  assuming that a level- $k$  player believes that the other player is a level- $k'$  player with  $k' = k - 1$ .

A level-1 investor ignores the message (given its randomness) and sticks to her a priori best choice (i.e., does not invest). The strategy of a level-1 firm is somewhat irrelevant, as the level-0 investor randomizes over her possible investment amounts. A level-2 investor still ignores the message as a level-1 firm's strategy does not necessarily include any truth-telling and still does not invest. For a level-2 firm, its strategy still remains irrelevant (as a level-1 investor does not invest). For  $k > 2$ , the best response of level- $k$  firms and investors remains the same as for  $k = 2$ .

A different but plausible assumption for level-0 firms is that they do not fully randomize. For example, firms with a high probability of success never send a message with a low probability of success. However, since the investor does not know the function the firm uses to randomize its messages, a level-1 investor will still ignore the message. Applying level- $k$  reasoning to the other treatments leads to the same predictions as under "payoff-maximization" described in Section 3.3.2.

### 3.C.3 Intrinsic costs of lying and belief updating

Here, we use a numerical example to show that even if the investor believes that only a small fraction of firms ( $2/7$ ) have sufficiently high intrinsic costs of lying – such that these firms always reveal the true probability of success – a message indicating a high probability of success can lead to a positive investment amount. We assume that the investor updates her belief according to Bayes rule

$$P(\theta|m) = \frac{P(m|\theta)P(\theta)}{P(m)}$$

where  $P(\theta|m)$  expresses the investor's belief that the actual probability of success is  $\theta$  after receiving the message  $m$ ,  $P(m|\theta)$  expresses the investor's belief in the probability of receiving a message  $m$  if the actual probability of success is  $\theta$ , where  $P(\theta) = 1/22$  is the belief that the probability of success is  $\theta$ , and  $P(m)$  is the investor's belief in the overall probability of receiving the message  $m$ .

To simplify the notation and reduce the number of necessary assumptions, we assume  $\Theta_l = \{15\%\}$  and  $\Theta_h = \{75\%\}$  with  $P(\theta) = 50\%$  for both prob-

abilities of success. Note that we assume only two values of the probability of success instead of a range, while keeping the probability of getting a low or high probability of success and the average of them exactly as in the experiment. We assume that an investor believes that  $2/7$  of all firms have an intrinsic cost of lying greater than €2.20 (i.e., they always report their probability of success truthfully) and that the remaining share has no intrinsic cost of lying. In the latter case, firms with  $\theta = 15\%$  send a deceptive message  $m = 75\%$ ; i.e.,  $P(m = 75\%|\theta = 15\%) = 5/7$ . The investor assumes that a firm with a high probability of success would never send a low message, i.e.,  $P(m = 75\%|\theta = 75\%) = 1$ . Consequently, the investor's belief in the overall probability of receiving a message  $m = 75\%$  is

$$P(m = 75\%) = 50\% \times 1 + 50\% \times \frac{5}{7} = 86\%.$$

After receiving a free message  $m = 75\%$ , the investor updates her beliefs in the following way

$$\begin{aligned} P(\theta = 15\%|m = 75\%) &= \frac{5/7 \times 50\%}{86\%} = 42\% \\ P(\theta = 75\%|m = 75\%) &= \frac{1 \times 50\%}{86\%} = 58\% \end{aligned}$$

affecting the investor's belief about the expected probability of success in such a way that

$$\begin{aligned} \mu(\theta|m = 75\%) &= P(\theta = 75\%|m = 75\%) 75\% + P(\theta = 15\%|m = 75\%) 15\% \\ &= 50\% \end{aligned}$$

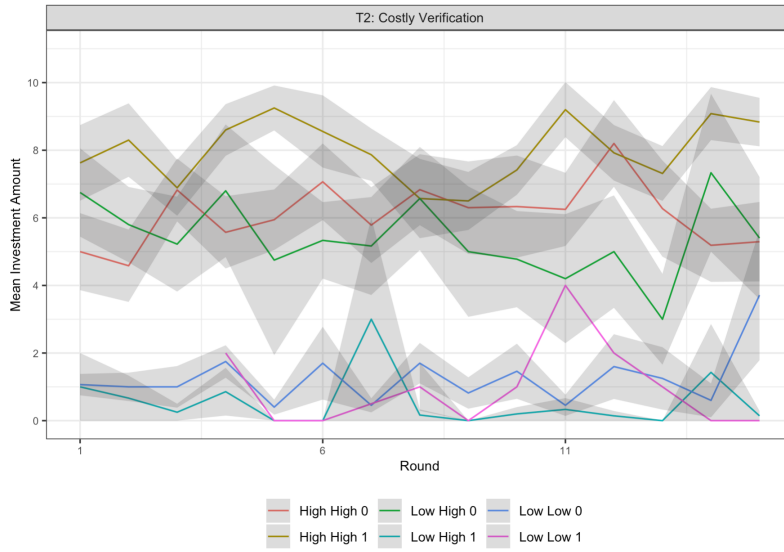
which is the cut-off probability of success for investing a positive amount for risk-neutral and risk-averse investors. This means that if the investor believes that at least  $2/7$  of all firms have an intrinsic cost of lying greater than €2.20, then a message indicating a high probability of success will lead to a positive investment amount – even in *Baseline*.

### 3.D Further analysis

In this section, we show additional figures and tables that support the analysis in the main text. First, we show the average investment amounts for verified and unverified messages in *Costly Verification* (T2), which we omitted from

Figure 3.4 (because the treatment is not fully comparable to the *Baseline* (T1) and the self-verification treatments T3 and T4). Second, we extend the result of Section 3.4.2 by looking at the evolution of truth-telling and strategic behavior over the 15 periods. Third, we complement the analysis of Section 3.4.4 in also estimating linear probability models for the *Costly Verification* (T2) treatment.

### 3.D.1 Investment per information available in Costly Verification (T2)

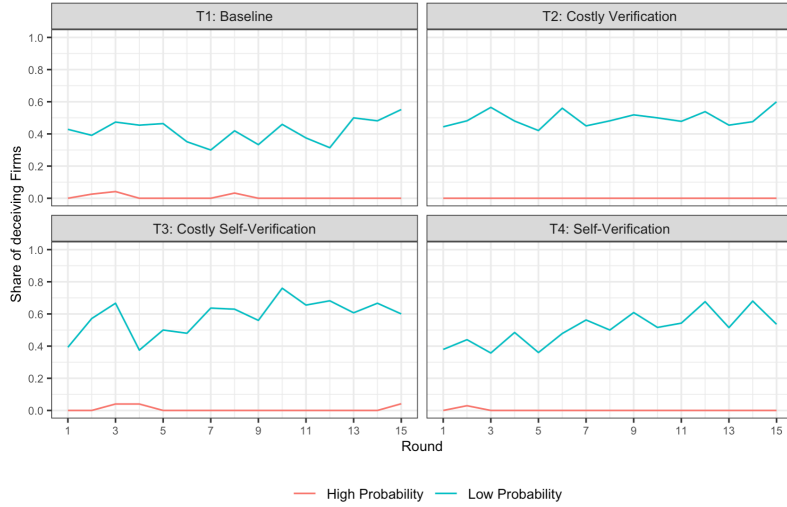


**Figure 3.D.1:** Investment per free and verified messages in Costly Verification (T2)

Figure 3.D.1 shows the average investment amounts for verified and unverified messages in T2. “High High 0” denotes a firm with a high probability of success sending a high message that is not verified. “High High 1” denotes a firm with a high probability of success sending a high message that is verified. “Low High 0” denotes a firm with a low probability of success sending a high message that is not verified. “Low High 1” denotes a firm with a low probability of success sending a high message that is verified. “Low Low 0” indicates a firm with a low probability of success sending a low message that is not verified. “Low Low 1” indicates a firm with a low probability of success sending a low message that is verified.

### 3.D.2 Deceptive behavior by firms

Figure 3.D.2 shows the percentage of deception over time, where deception is defined as firms with a low (high) probability of success sending a high (low) message. This figure shows hSP firms almost never deceive, but lSP firms deceive to a considerable extent. As we have seen in Figure 3.3, even in *Baseline*, where the true probability of success is never known to the investor, the proportion of deception is far below the maximum.



Note: Deception is a binary variable receiving a value of 1 when a low (high) probability of success firm sends a message indicating that it has a high (low) probability of success, and otherwise 0. Moreover, self-verification is considered as truth-telling.

**Figure 3.D.2:** Share of deception over time

Table 3.D.1 shows regressions from random-effects panel data linear probability models. Model 1 shows the treatment effects without controlling for the low and high probability of success. Model 2 shows that in all treatments, the deception of firms with a high probability of success (hSP) is significantly lower than by firms with a low probability of success (lSP) (at  $p = 0.000$ ). In particular, hSP firms deceive 41% less than their lSP counterparts in the baseline treatment. However, in *Costly Verification* (T2), *Costly Self-Verification* (T3), and *Self-Verification* (T4), hSP firms deceive 49%, 54%, and 51% less than their low probability of success (lSP) counterparts. However, we observe no differences in deception when comparing the (self-)verification treatments T2-T4 pairwise.



**Table 3.D.1:** Deception per firm and probability of success

	Model (1)	Model (2)
Treatment T2	0.03 (0.10)	0.05 (0.11)
Treatment T3	0.06 (0.09)	0.13 (0.10)
Treatment T4	0.00 (0.10)	0.02 (0.12)
high probability of success (hSP)	-	-0.41*** (0.05)
hSP $\times$ Treatment T2	-	-0.08 (0.08)
hSP $\times$ Treatment T3	-	-0.14** (0.07)
hSP $\times$ Treatment T4	-	-0.10 (0.07)
Period	-0.00 (0.00)	0.00 (0.00)
Treatment T2 $\times$ Period	-0.00 (0.00)	0.00 (0.00)
Treatment T3 $\times$ Period	0.01 (0.00)	0.01 (0.00)
Treatment T4 $\times$ Period	0.01** (0.00)	0.01** (0.00)
Constant	0.19*** (0.05)	0.40*** (0.07)
Session FE	Yes	Yes
Observations	3,285	3,285
R-squared (overall)	0.029	0.353

Note: Random-effects panel data linear probability model estimations. “high probability of success (hSP)” indicates success probability values of 70% – 80% (otherwise values of 10% – 20%). The standard errors are two-way clustered across firms. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

### 3.D.3 Verification in Costly Verification (T2)

Table 3.D.2 shows a panel-data random-effects linear probability regression. Model 2 shows that the verification probability increases significantly when the investor receives a free (“cheap talk”) message indicating a high probability of success (t-test,  $p = 0.000$ ). In addition, the verification probability does not change over time when the investor receives a message with a low probability of success (t-test,  $p = 0.876$ ), but it increases for firms with high probability of success (Wald test yields  $p = 0.0092$ ).

## 3.E Robustness

This section presents alternative regressions to the panel data regressions presented in the main body of the paper, in order to account for random matching

**Table 3.D.2:** Verification use in Costly Verification (T2)

	Model (1)	Model (2)
Period	0.01** (0.00)	0.00 (0.01)
high probability of success (hSP-) message	-	0.31*** (0.07)
hSP-message $\times$ Period	-	0.01 (0.01)
Constant	0.18** (0.09)	-0.07 (0.09)
Session FE	Yes	Yes
Observations	735	735
Overall R-squared	0.035	0.145

Note: Random-effects panel data linear probability model estimations. Dependent variable: verification use in the Costly Verification (T2) treatments. “high probability of success (hSP-) message” indicates a message indicating a high probability of success value of 70%-80% (otherwise values of 10%-20%). The standard errors are clustered across investors. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

between investors and firms within a session. In particular, it presents linear regressions with a two-way clustered standard error (i) over *investors* and (ii) over *session  $\times$  period* (Cameron et al., 2011).<sup>26</sup>

### Robustness to Table 3.2

Although we observe a similar pattern across treatments (that investment is higher for high probability firms than for low probability firms), we want to measure quantitatively how strong the relationship is between a firm’s probability of success and investment. The more investment depends on hSP (i.e., investment amounts are higher for firms with a high probability of success than for firms with a low probability of success), the more effective the treatment is from the investor’s perspective. Here we expect stronger relationships in the (self-)verification treatments T2-T4 (where the theoretical prediction is to invest only in firms with a high probability of success) than in the *Baseline* (with a predicted babbling equilibrium and hence no investment). Table 3.E.1 shows regression models with two-way standard errors clustered over (i) *investors* and (ii) over *session  $\times$  period*. Model 1 shows treatment effects without controlling for low and high probability of success. Models 2 and 3 (Model 3 also includes the investor’s risk propensity as measured in the post-experimental questionnaire) show that being matched with a hSP firm significantly increases investment in all treatments (a joint significance test  $F(1, 218)$  yields  $p = 0.000$

<sup>26</sup> For this analysis, we used Stata 17 with the module `reghdfe` (Correia, 2019).

**Table 3.E.1:** Investment per probability of success

	Model (1)	Model (2)	Model (3)
Treatment T2	-1.01 (1.09)	-1.56 (1.01)	-1.33 (1.07)
Treatment T3	-1.48 (1.09)	-2.06** (1.01)	-2.19** (1.05)
Treatment T4	-0.84 (1.07)	-1.30 (0.95)	-1.36* (0.80)
high probability of success (hSP)	-	3.14*** (0.34)	3.15*** (0.33)
hSP $\times$ Treatment T2	-	1.63*** (0.56)	1.58*** (0.56)
hSP $\times$ Treatment T3	-	1.20** (0.48)	1.12** (0.47)
hSP $\times$ Treatment T4	-	1.44*** (0.44)	1.47*** (0.42)
Period	0.02 (0.02)	0.01 (0.02)	0.01 (0.02)
Treatment T2 $\times$ Period	-0.00 (0.04)	-0.02 (0.04)	-0.02 (0.04)
Treatment T3 $\times$ Period	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)
Treatment T4 $\times$ Period	-0.04 (0.04)	-0.02 (0.03)	-0.02 (0.03)
Risk test	-	-	0.34*** (0.05)
Constant	5.54*** (0.83)	3.96*** (0.77)	1.78** (0.78)
Session FE	Yes	Yes	Yes
Observations	3,285	3,285	3285
R-squared	0.01	0.28	0.32

Note: Linear model estimations. Dependent variable: Investment (a value from €0 to €11). “high probability of success (hSP)” indicates a success probability of 70%-80% (otherwise values of 10%-20%). “Risk test” is taken from the post-experimental questionnaire: It is the amount invested (between €0 and €10) in a risky lottery, where a lower amount indicates a higher degree of risk aversion. The Standard errors are two-way clustered (across (i) *investors* and (ii) *session  $\times$  period*). Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

for each of the treatments).<sup>27</sup> In particular, in *Baseline*, hSP increases investment by €3.15. However, this effect is much larger in *Costly Verification* (T2), *Costly Self-Verification* (T3), and *Self-Verification* (T4) than in *Baseline* (T1) (investments are on average €1.58, €1.12, and €1.47 higher in T2, T3, and T4 than in *Baseline* at  $p = 0.005$ ,  $p = 0.018$ , and  $p = 0.001$  respectively). There are no differences in the effect of high probability between the verification/self-verification treatments (in Model 3, Wald test yields  $p = 0.4128$ ,  $p = 0.8259$ , and  $p = 0.4218$  in pairwise comparison of T2 and T3, T2 and T4, and T3

<sup>27</sup> We test the null hypothesis that the sum of the coefficients  $\beta_{[\text{hSP}]}$  and  $\beta_{[\text{hSP} \times \text{respective treatment}]}$  equals zero.

and T4, respectively). As expected, we also find that the less risk-averse an investor is, the more she will invest in the risky project (t-test,  $p = 0.000$ ). Moreover, investments do not change over time in any of the treatments. The variable “Period” is neither significant in *Baseline* (t-test,  $p = 0.537$ ), nor in any other treatment (Wald test yields  $p = 0.8346$ ,  $p = 0.2390$ , and  $p = 0.7087$  for T2, T3, and T4).

### Robustness to Table 3.3

**Table 3.E.2:** Investment per free communication treatment

	Model (1)
Treatment T3	−1.66 (1.20)
Treatment T4	−0.43 (0.86)
high free message (hfM)	5.27*** (0.43)
hfM × Treatment T3	−1.89*** (0.61)
hfM × Treatment T4	−1.57*** (0.56)
Period	0.00 (0.02)
Treatment T3 × Period	−0.05 (0.03)
Treatment T4 × Period	−0.07** (0.03)
Risk test	0.29*** (0.07)
Constant	0.06 (0.85)
Session FE	Yes
Observations	2,012
R-squared	0.42

Note: Linear model estimation. The regression includes all observations in the Baseline (T1) but only free messages in T3 and T4. Dependent variable: Investment (a value from €0 to €11). “high free message (hfM)” indicates a free message (i.e., a message that is not self-verified) indicating success probability values of 70%-80% (otherwise values of 10%-20%). “Risk test” is taken from the post-experimental questionnaire: It is the amount invested (between €0 to €11) in a risky lottery, where a lower amount indicates a higher degree of risk aversion. The standard errors are two-way clustered across (i) investors and (ii) session × period. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

Table 3.E.2 shows a regression with a two-way standard errors clustered over (i) *investors* and (ii) *session × period*.<sup>28</sup> It shows that investment is

<sup>28</sup> This regression model only includes the Baseline T1, and self-verification (T3-T4) treatments. In T2, the investor was able to verify the message. Thus, it is not directly

significantly affected by a high free message (hfM) in all treatments (Wald test yields  $p < 0.0001$  for all treatments). In *Baseline*, a high free message increases investment on average by €5.27. However, the effect is considerably smaller in the self-verification T3 and T4 treatments, where it increases investment only by €3.38 and €3.70 (no difference in the effects between these two treatments (Wald test yields  $p = 0.2564$ ). This means that compared to T3, the lower self-verification cost in T4 does not seem to reduce the investors' level of skepticism when they receive a free high message (instead of a self-verified one).

### Robustness to Table 3.4

**Table 3.E.3:** Firms' self-verification use

	Model (1)	Model (2)	Model (3)
Treatment T4	0.15 (0.10)	0.04 (0.09)	0.06 (0.09)
high probability of success (hSP)	-	0.54*** (0.06)	0.44*** (0.08)
hSP $\times$ Treatment T4	-	0.10 (0.09)	0.08 (0.11)
Period	0.01* (0.00)	0.01** (0.00)	-0.00 (0.00)
Period $\times$ hSP	-	-	0.01** (0.01)
Period $\times$ Treatment T4	-0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)
hSP $\times$ Treatment T4 $\times$ Period	-	-	0.00 (0.01)
Constant	0.29*** (0.07)	0.03 (0.07)	0.08 (0.07)
Session FE	Yes	Yes	Yes
Observations	1,620	1,620	1,620
R-squared	0.04	0.43	0.44

Note: Linear probability model estimations only include the self-verification (T3 and T4) treatments. Dependent variable: verification use (1 = verification use, 0 = otherwise). "high probability of success (hSP)" indicates success probability values of 70%-80% (otherwise values of 10%-20%). The standard errors are two-way clustered across (i) firms and (ii) session  $\times$  period. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

Table 3.E.3 shows the results of linear probability model regressions for the self-verification treatments (T3-T4). As self-verification is carried out by firms, we use two-way clustered standard errors over (i) *firms* and (ii) *session*  $\times$  *period*. Model 1 shows the treatment effects without controlling for the low and high probability of success. The results from Model 2 show that

---

comparable to the free message in the other treatments (where the investor can believe it or not).

the probability for self-verification use is considerably higher for hSP firms than for lSP firms. In particular, Model 2 indicates that the probability for self-verification is 54% and 64% higher for hSP than for lSP firms in *Costly Self-Verification* (T3) and (low cost) *Self-Verification* (T4), respectively (Wald tests yield  $p = 0.000$  for T3 and T4).<sup>29</sup> This is to be expected since, in theory, only hSP firms should self-verify their type. Moreover, we find that in both T3 and T4 self-verification increases over time for hSP firms but not for lSP firms (Wald test for hSP in T3 and T4 yields  $p = 0.0316$  and  $p = 0.0435$ , for lSP firms  $p = 0.8787$  and  $p = 0.2147$ ).<sup>30</sup>

### Robustness to Table 3.5

Table 3.E.4 shows regression models with a two-way standard errors clustered over (i) *investors* and (ii) *session × period*. Model 1 shows that Investor Efficiency is only marginally higher in *Costly Verification* (T2) and *Costly Self-Verification* (T3) than in *Baseline* (T1), but not in the *Self-Verification* (T4).<sup>31</sup> In the pairwise comparisons between the (self-)verification treatments, we find no differences in Investor Efficiency.<sup>32</sup> Consistent with the bottom panel of Figure 3.6, Models 2 and 3 in Table 3.E.4 show that Investor Efficiency is higher when matched with hSP firms than with lSP firms only in *Baseline* (t-test,  $p = 0.003$  in both Models 2 and 3), but not in the other treatments (Wald tests (Model 2) yield  $p = 0.602$ ,  $p = 0.584$ , and  $p = 0.459$  in T2, T3, and T4, respectively).<sup>33</sup> In Model 3 the p-values are  $p = 0.606$ ,  $p = 0.594$ , and  $p = 0.456$  in T2, T3, and T4, respectively. Investor Efficiency of hSP is also not different when pairwise compared between T2, T3, and T4 (Wald test yields p-values between 0.928 and 0.990 in Model 2 and between 0.918 and 0.997 in Model 3). Also, Investor Efficiency does not change over time in any treatment. Finally, for lSP firms, the self-verification treatments (T3-T4) yield significantly higher Investor Efficiency than in *Baseline* ( $p = 0.05$ ,  $p = 0.06$  in

<sup>29</sup> We test the null hypothesis that the sum of the coefficients  $\beta_{[\text{hSP}]}$  and  $\beta_{[\text{hSP} \times \text{respective treatment}]}$  equals zero.

<sup>30</sup> For the hSP firms, we test the null hypothesis that the sum of the coefficients  $\beta_{[\text{hSP} \times \text{period}]}$  and  $\beta_{[\text{hSP} \times \text{period} \times \text{respective treatment}]}$  equals zero. For the lSP firms, we test the null hypothesis that the sum of the coefficients  $\beta_{[\text{period}]}$  and  $\beta_{[\text{period} \times \text{respective treatment}]}$  equals zero.

<sup>31</sup> The t-tests comparing Baseline (T1) with treatments T2, T3, and T4 yield the following respective p-values:  $p = 0.07$ ,  $p = 0.100$ , and  $p = 0.736$ .

<sup>32</sup> Wald test yields  $p = 0.991$ ,  $p = 0.245$ , and  $p = 0.279$  in pairwise comparisons of T2 and T3, T2 and T4, and T3 and T4, respectively.

<sup>33</sup> We test the null hypothesis that the sum of the coefficients  $\beta_{[\text{hSP}]}$  and  $\beta_{[\text{hSP} \times \text{respective treatment}]}$  equals zero.

**Table 3.E.4: Investor Efficiency**

	Model (1)	Model (2)	Model (3)
Treatment T2	0.04* (0.02)	0.07** (0.03)	0.08** (0.03)
Treatment T3	0.04* (0.02)	0.08** (0.03)	0.07** (0.03)
Treatment T4	0.01 (0.02)	0.05 (0.03)	0.05 (0.03)
high probability of success (hSP)	-	0.09*** (0.03)	0.09*** (0.03)
hSP $\times$ Treatment T2	-	-0.07* (0.04)	-0.07* (0.04)
hSP $\times$ Treatment T3	-	-0.08** (0.04)	-0.08** (0.04)
hSP $\times$ Treatment T4	-	-0.07** (0.04)	-0.07** (0.04)
Period	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Treatment T2 x Period	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Treatment T3 x Period	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Treatment T4 x Period	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Risk test	-	-	0.00 (0.00)
Constant	0.84*** (0.01)	0.79*** (0.02)	0.78*** (0.03)
Session FE	Yes	yes	yes
Observations	3,285	3,285	3,285
R-squared	0.02	0.04	0.04

Note: Linear model estimations. Dependent variable: Investor Efficiency. The standard errors are two-way clustered across (i) investors and (ii) session  $\times$  period. “Risk test” is taken from the post-experimental questionnaire: It is the amount invested (between €0 and €10) in a risky lottery, where a lower amount indicates a higher degree of risk aversion. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

T3 and T4), but not in the *Costly Verification* (T2) ( $p = 0.103$  in T2).

### Robustness to Table 3.D.1

Table 3.E.5 shows linear probability model regressions (with two-way clustered standard errors across (i) *firms* and (ii) *session  $\times$  period*). Model 1 shows the treatment effects without controlling for the low and high probability of success. Model 2 shows that in all treatments, deception by firms with high probability of success (hSP) is significantly lower than that of firms with low probability of success (lSP). However, they are significantly lower in *Baseline* than in the other treatments (at  $p < 0.0001$  in each comparison). In particular, hSP firms deceive 41% less than their lSP counterparts in the baseline

**Table 3.E.5:** Deception per firm and probability of success

	Model (1)	Model (2)
Treatment T2	0.03 (0.09)	0.08 (0.07)
Treatment T3	0.06 (0.08)	0.12 (0.07)
Treatment T4	0.00 (0.09)	0.04 (0.07)
high probability of success (hSP)	-	-0.41*** (0.05)
hSP $\times$ Treatment T2	-	-0.09 (0.07)
hSP $\times$ Treatment T3	-	-0.17** (0.07)
hSP $\times$ Treatment T4	-	-0.10 (0.07)
Period	-0.00 (0.00)	0.00 (0.00)
Treatment T2 $\times$ Period	-0.00 (0.00)	0.00 (0.00)
Treatment T3 $\times$ Period	0.01 (0.00)	0.01 (0.00)
Treatment T4 $\times$ Period	0.01** (0.00)	0.01* (0.00)
Constant	0.19*** (0.04)	0.41*** (0.05)
Session FE	Yes	Yes
Observations	3,285	3,285
R-squared	0.03	0.34

Note: Linear model estimations. Dependent variable: deception (1 = deception, 0 = otherwise). “high probability of success (hSP)” indicates success probability values of 70%-80% (otherwise values of 10%-20%). The standard errors are two-way clustered (across (i) firms and (ii) session  $\times$  period). Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

treatment. However, in *Costly Verification* (T2), *Costly Self-Verification* (T3), and *Self-Verification* (T4), hSP firms deceive 50%, 58%, and 51% less than their lsp counterparts.<sup>34</sup> Since some firms never deceive, this result formally establishes that, in principle, only firms with a low probability of success (sometimes) deceive.



**Table 3.E.6:** Verification use in Costly Verification (T2)

	Model (1)	Model (2)
Period		0.34*** (0.07)
high probability of success (hSP-) message	0.01*** (0.00)	0.01 (0.01)
hSP-message $\times$ Period		0.00 (0.01)
Constant	0.18** (0.08)	-0.10 (0.09)
Session FE	Yes	yes
Observations	735	735
R-squared	0.03	0.15

Note: Linear probability model estimations. Dependent variable: verification use in the Costly Verification (T2) treatments. “high probability of success (hSP-) message” indicates a message with a value of 70%-80% (otherwise values of 10%-20%). The standard errors are two-way clustered across (i) investors and (ii) session  $\times$  period. Finally, \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%, respectively.

### Robustness to Table 3.D.2

Table 3.E.6 shows a panel-data random-effects linear probability regression. Model 2 shows that the verification probability increases significantly when the investor receives a free (“cheap talk”) message indicating a high probability of success (t-test,  $p = 0.000$ ). In addition, the verification probability does not change over time when the investor receives a message with a low probability of success (t-test,  $p = 0.575$ ), but it increases for firms with a high probability of success (Wald test yields  $p = 0.0086$ ).

## 3.F Translated instructions (computer screens)

This section includes the translated instructions for the *Baseline*, *Costly Verification* (T2), and *Costly Self-Verification* (T3). The instructions of *Self-Verification* (T4) are identical to T3, but with a lower price for the self-verification. Each subsection represents a single screen presented to the participants.

<sup>34</sup> A statistically significant reduction in deception for hSP relative to the Baseline (T1) is only observed in Costly Self-Verification (T3) ( $p = 0.020$ ). We observe no differences in deception when pairwise comparing between the (self-)verification treatments (F(1, 218), for hSP firms p-values are between 0.3006 and 0.8987, for lSP firms p-values are between 0.3115 and 0.6168).

### **3.F.1 Baseline**

#### **Welcome**

Welcome and thank you for your participation. You have been asked to participate in an online economic experiment. Please activate your webcam and mute your microphone. From now on, you are not allowed to communicate with any of the other participants. If you have any question, please write us privately via the chat function in Zoom. The experiment is conducted under strict compliance to the rules of conduct for economic experiments of the University of Kiel. That means amongst other things that all information is truthfully provided to you. You can end the experiment at any time and without giving a reason. Your data will be treated with strict confidentiality and processed and published anonymously by the researchers involved. The instructions are identical for all participants in the sessions.

Each of you earns 4€ as a show-up fee for participating in this experiment. During the experiment, you can earn additional money depending on your decision and the decision of another participant you are matched with. Your payout from the experiment will be transferred to your bank account within 2-3 days after the experiment. We ask you about your account information at the end of the experiment.

The experiment is expected to last about 45 minutes. It consists of 15 rounds and ends with a short questionnaire. At the beginning of the experiment, each of you is assigned a role, either Participant A or Participant B. These roles remain fixed throughout the 15 rounds (people who are assigned as Participants A or B will stay in their roles during the whole experiment). Before each round, there will be a random matching of participants from different roles (one Participant A is matched with one Participant B). Hence, in each round you are to assume that you play with a different person (it is very unlikely that you are matched with the same person more than twice). The identity of you or any of the other participants is never revealed, neither during nor after the experiment.

#### **The Structure of one Round**

In each round, Participant A is endowed with 11€ and can invest any amount between 0€ and 11€ in a project owned by Participant B. The project can

succeed or fail. It succeeds with a given probability ( $p$ ) and fails with a probability ( $1 - p$ ). For example, let us assume that the success probability is 80%. Ergo, the probability that the project fails is 20%. This is comparable with an urn containing 100 balls: 80 white balls and 20 black balls. There is one random draw from the urn. If the ball that comes up is white, the project succeeds and yields a positive return for both, Participant A and Participant B. If the ball is black, the project fails, and the investment made by Participant A is lost.

If Participant A invests nothing, Participant A earns the endowment of 11€ and Participant B earns 0€. If Participant A decides to invest any amount larger than 0€ in the project and if the project succeeds, Participant A retains the amount invested and additionally earns the same amount on top of it. In this case, Participant B also earns the amount invested. If the project fails, the amount invested by Participant A is lost. See the following examples:

1. If Participant A invests the full endowment of 11€ and the project succeeds, then Participant A retains the investment of 11€ and earns additional 11€, that means that Participant A earns 22€ in total, whereas Participant B earns 11€. However, if the project fails, Participant A earns 0€ and Participant B earns 0€. That means that Participant A lost the 11€ investment.
2. If Participant A invests 10€ (out of the 11€ endowment) and the project succeeds, then Participant A retains the investment of 10€ and earns an additional 10€. Please recall that Participant A has still 1€ that (s)he did not invest in the project. That means that Participant A earns 21€ in total, whereas Participant B earns 10€. However, if the project fails, Participant A earns 1€ (that was not invested), and Participant B earns 0€. That means that Participant A lost the 10€ investment.

### Verification Questions

If Participant A invests 9€ out of the 11€ endowment, and the project succeeds:

1. What is the additional payoff for Participant A: [9€]
2. What is the total payoff (ignoring show-up fee) for Participant A: [20€]
3. What is the total payoff (ignoring show-up fee) for Participant B: [9€]

## Success Probability

It is common knowledge to Participant A and B that the success probability ( $p$ ) is randomly selected from two intervals  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$  with an equal chance for each probability in those intervals. This means that the average probability of success is 45%. Neither the participants nor the experimental team can influence the selection of the success probability. It is randomly determined by the computer. Please note that the success probability is drawn randomly and independently for each of the ten rounds; i.e., it is very unlikely that the success probability is the same for two rounds.

For example, if the given success probability is 80%. This is comparable with an urn containing 100 balls: 80 white balls and 20 black balls. There is one random draw from the urn. If the ball that comes up is white, the project succeeds and yields a positive return for both, Participant A and Participant B. If the ball is black, the project fails, and the investment made by Participant A is lost.

## The Sequence of Decisions in Every Round

In every round, each Participant (A and B) makes one decision. Participant B learns about the true probability of success of the project, while Participant A does not. Please note that the success probability is randomly and independently determined every round. Participant A never learns the actual success probability (neither during nor after the experiment). The first decision is made by Participant B who sends a free message to Participant A.

In the free message, Participant B can reveal the true probability of success or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$ .* Participant B can choose any number in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ . Participant A never learns the true probability of success (neither during nor after the experiment). Next, Participant A receives the message from Participant B and makes a decision how much to invest in the project (between 0€ and 11€).

## Your Payout

As a reminder: In each round and independently of the decisions and outcomes in previous rounds, Participant A is endowed with 11€. For each round, the payoffs of Participant A and Participant B are determined based on the amount invested by Participant A and the outcome of the project. If Participant A invests nothing, Participant A receives the endowment of 11€ and Participant B receives 0€ for this round. If Participant A decides to invest any amount larger than 0€ in the project and if the project succeeds, Participant A retains the amount invested and additionally receives the same amount on top of it. In this case, Participant B also receives the amount invested. If the project fails, the amount invested by Participant A is lost. For your payout in the experiment, only one random round will be determined. The computer randomly chooses one of the 15 rounds (i.e., each round has a 6.7% chance to be selected for your payout). Neither the participants nor the experimental team can influence the selection of the round. Thus, your payout from the experiment includes the show-up fee of 4€ plus the payoff from the one round that is randomly selected by the computer.

## Instructions Summary

1. The roles of Participant A and Participant B are randomly determined by the computer.
2. The first round starts: Participants assigned as Participant A are endowed with 11€. The success probability ( $p$ ) of the project is randomly determined by the computer and is uniformly distributed in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ . The average success probability is 45%.
3. Participant B is informed about the true probability of success of the project, while Participant A is not.
4. At this point, Participant B sends a free message to Participant A. In the free message, Participant B can reveal the true value or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$ .* Participant B can choose any number in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ .
5. Participant A receives the free message and decides how much to invest in the project (between 0€ and 11€).

6. The payoff for this round is determined according to the decisions made by the participants and the success probability of the project that was determined in Step 2. The success of the project is determined by a random lottery made by the computer: With a probability ( $p$ ), the project succeeds. With a probability ( $1 - p$ ), the project fails. The payoffs for this round are determined accordingly.
7. The participants are informed about the result of the lottery and the payoff for the round.
8. Now the first round is over, and the next round starts again at step 2.
9. We repeat steps 2-8 for 15 rounds. After 15 rounds, the decisions are over, and you are asked to fill a short questionnaire.
10. Finally, the computer randomly determines which of the 15 rounds will be selected for payout. Your payout from the experiment is the payoff of this randomly selected round plus the show-up fee. The experiment ends.

### **Role Determination**

This is Round  $[r]$  out of 15 rounds. You are randomly determined the role of Player  $[A/B]$ .

### **Free Message**

You are Participant B. This is Round  $[r]$  out of 15 rounds. Your success probability in this round is  $[\Theta]$ .

At this stage, you send a free message to Participant A. The free message can contain the true value or any other value of your choice that represents the probability of success. You can choose any number in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ . You are not obliged to send the true probability. Please note that Participant A never learns the true probability of success. Participant A might use your free message to decide whether and how much to invest in your project. Message to be sent: *Participant B claims that the success probability of the project is  $m\%$ .*

### **Message Delivered**

You are Participant  $[A/B]$ . This is Round  $[r]$  out of 15 rounds. Participant B sent Participant A the free message: *Participant B claims that the success*

*probability of the project is  $[m]\%$ .* This means, if the message is true, then the project is comparable to an urn containing 100 balls:  $[m]$  white balls and  $(100-[m])$  black balls. There is one random draw from the urn. If the ball that comes up is white, the project succeeds and yields a positive return for both, Participant A and Participant B. If the ball is black, the project fails, and the investment made by Participant A is lost.

### **Investment Decision**

You are Participant A. This is Round  $[r]$  out of 15 rounds. From Participant B, you received a free message: *Participant B claims that the success probability of the project in this round is  $[m]\%$ .* Please consider that the free message might be true or not. Make a decision how much you want to invest (between 0€ and 11€).

Reminder: Your payoff depends on your investment amount and whether the project succeeds or fails. The success of the project is determined by a random lottery made by the computer: With the success probability ( $p$ ), the project succeeds. With a probability  $(1 - p)$ , the project fails. If the project succeeds: You receive 11€ plus the amount invested and Player B receives the amount invested. If the project fails: You lose the amount invested and Player B receives nothing.

### **Payoffs for this Round**

You are Participant  $[A/B]$ . This is Round  $[r]$  out of 15 rounds. Given the actual success probability, the computer randomly determined whether the firm's project was successful or not: The project  $[succeeded / failed]$ . Participant A invested  $[i]$ € in the project. Therefore, you receive  $[e_r]$ € from the project for this round.

Recall that at the end of the experiment, only one randomly determined round will be selected for your payout. If this round will be determined for your final payout, your final payout would be the payoff from the project of this round  $[e_r]$ € plus the show-up fee of 4€; i.e.  $[e_r+4]$ € would be your final payout.

We will now start the next round  $[r+1]$  out of 15 rounds. We will now re-match Participants A and B to determine with whom you will be matched in the next round. Please note that you will always remain in the role of Participant A /

B until the last round.

### 3.F.2 Costly Verification (T2)

In this section, we show all instructions (computer screens) that are different from *Baseline*.

#### The Sequence of Decisions in Every Round

In every round, Participant A makes two decisions and Participant B makes one decision. Participant B learns about the true probability of success of the project, while Participant A does not. Please note that success probability is randomly and independently determined every round. The first decision is made by Participant B who sends a free message to Participant A. In the free message, Participant B can reveal the true probability of success or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$* . Participant B can choose any number in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ .

Next, Participant A receives the free message from Participant B and makes two decisions: First decision: whether or not to purchase a certified message from the experimental team containing the true probability of success. The certified message costs Participant A 1€ from the show-up fee. Please note that Participant B never learns whether Participant A purchased the certified message or not. Second decision (independently on purchasing a certified information or not): how much to invest in the project (between 0€ and 11€).

#### Instructions Summary

1. The roles of Participant A and Participant B are randomly determined by the computer.
2. The first round starts: Participants assigned as Participant A are endowed with 11€. The success probability ( $p$ ) of the project is randomly determined by the computer and is uniformly distributed in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ . The average success probability is 45%.
3. Participant B is informed about the true probability of success of the project, while Participant A is not.



4. At this point, Participant B sends a free message to Participant A. In the free message, Participant B can reveal the true value or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$* . Participant B can choose any number in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$
5. Participant A receives the free message and makes two decisions: First, whether or not to purchase a certified message from the experimental team containing the true probability of success. The certified message costs Participant A 1€ from the show-up fee. Please note that Participant B never learns whether Participant A purchased the certified message or not. Second, how much to invest in the project (between 0€ and 11€).
6. The payoff for this round is determined according to the decisions made by the participants and the success probability of the project that was determined in Step 2. The success of the project is determined by a random lottery made by the computer: With a probability ( $p$ ), the project succeeds. With a probability ( $1 - p$ ), the project fails. The payoffs for this round are determined accordingly.
7. The participants are informed about the result of the lottery and the payoff for the round.
8. Now the first round is over, and the next round starts again at step 2.
9. We repeat steps 2-8 for 15 rounds. After 15 rounds, the decisions are over, and you are asked to fill a short questionnaire.
10. Finally, the computer randomly determines which of the 15 rounds will be selected for payout. Your payout from the experiment is the payoff of this randomly selected round plus the show-up fee. The experiment ends.

### Free Message

You are Participant B. This is Round  $[r]$  out of 15 rounds. Your success probability in this round is  $[\Theta]$ . At this stage, you send a free message to Participant A. The free message can contain the true value or any other value of your choice that represents the probability of success. You can choose any number between in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ . You are not obliged to send the true probability. Please note that Participant A could purchase a certified message from the experimental team for 1€ that contains

the true probability of success of the project. Participant A might use your free message to decide whether and how much to invest in your project.

### **Certified Message**

You are Participant A. This is Round  $[r]$  out of 15 rounds. At this stage, you need to decide whether or not to purchase a certified message from the experimental team about the success probability of the project. The certified message costs you 1€ from the show-up fee and contains the true probability of success.

### **Investment Decision**

**If Verified:** You are Participant A. This is Round  $[r]$  out of 15 rounds. You purchased a certified message from the experimental team: *It is certified by the experimental team that the success probability of the project in this round is  $[\Theta]$ .* Please consider that this is the true probability of success. Additionally, you received a free message from Participant B: *Participant B claims that the success probability of the project is  $[m]\%$ .* Please consider that the free message might be true or not. Make a decision how much you want to invest (between 0€ and 11€).

Reminder: Your payoff depends on your investment amount and whether the project succeeds or fails. The success of the project is determined by a random lottery made by the computer: With a probability of  $[\Theta]$ , the project succeeds: You receive 11€ plus the amount invested and Participant B receives the amount invested. With a probability of  $1 - [\Theta]$ , the project fails: You lose the amount invested and Participant B receives nothing.

### **3.F.3 Costly Self-Verification (T3)**

In this section, we show all instructions (computer screens) different from *Baseline*.

### **The Sequence of Decisions in Every Round**

In every round, each Participant (A and B) makes one decision. Participant B learns about the true probability of success of the project, while Participant A does not. Please note that success probability is randomly and independently determined every round. The first decision is made by Participant B: whether

to send either a free message or a certified message to Participant A. In the free message, Participant B can reveal the true probability of success or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $[m]\%$* . Participant B can choose any number in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ . The certified message costs Participant B 1€ from the show-up fee and contains the true probability of success. The message is the following: *It is certified by the experimental team that the success probability of the project is  $[\Theta]$* . Next, Participant A is informed whether a free message or a certified message was sent. Consequently, after receiving the message from Participant B, Participant A makes a decision how much to invest in the project (between 0€ and 11€).

### Instructions Summary

1. The roles of Participant A and Participant B are randomly determined by the computer.
2. The first round starts: Participants assigned as Participant A are endowed with 11€. The success probability ( $p$ ) of the project is randomly determined by the computer and is uniformly distributed in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ . The average probability of success is 45%.
3. Participant B is informed about the true probability of success of the project, while Participant A is not.
4. At this point, Participant B decides to send either a free message or a certified message to Participant A.
5. In the free message, Participant B can reveal the true value or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$* . Participant B can choose any number in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ .
6. The certified message costs Participant B 1€ from the show-up fee and contains the true probability of success. The message is the following: *It is certified by the experimental team that the success probability of the project is  $[\Theta]$* .
7. Participant A learns whether Participant B sent a free or a certified message and decides how much to invest in the project (between 0€ and 11€).

8. The payoff for this round is determined according to the decisions made by the participants and the success probability of the project that was determined in Step 2. The success of the project is determined by a random lottery made by the computer: With a probability ( $p$ ), the project succeeds. With a probability ( $1 - p$ ), the project fails. The payoffs for this round are determined accordingly.
9. The participants are informed about the result of the lottery and the payoff for the round.
10. Now the first round is over, and the next round starts again at step 2.
11. We repeat steps 2-10 for 15 rounds. After 15 rounds, the decisions are over, and you are asked to fill a short questionnaire.
12. Finally, the computer randomly determines which of the 15 rounds will be selected for payout. Your payout from the experiment is the payoff of this randomly selected round plus the show-up fee. The experiment ends.

### Message Decision

You are Participant B. This is Round  $[r]$  out of 15 rounds. The true probability of success in this round is  $[\Theta]$ . At this stage, you can decide between either sending a free message to Participant A or buying a certified message containing the true probability of success. The free message can contain the true value or any other value of your choice that represents the probability of success. You can choose any number in  $[10\%, 11\%, \dots, 20\%]$  and  $[70\%, 71\%, \dots, 80\%]$ . You are not obliged to send the true probability. Please note that Participant A never learns the true probability of success. The certified message provides Participant A with the true probability of success. The certified message costs 1€ which is deducted from your show-up fee. Note that Participant is informed on the type of message (whether it is free message or certified message) that you sent. Do you want to purchase the certified message for 1€?

### Certified Message

**If self-verified:** You are Participant  $[A/B]$ . This is Round  $[r]$  out of 15 rounds. It is certified by the experimental team that the success probability of the project is  $[\Theta]$ . Participant B had to pay 1€ to send the certified message. Please consider that this is the true probability of success.

## Investment Decision

**If self-verified:** You are Participant A. This is Round  $[r]$  out of 15 rounds. From Participant B, you received a certified message: *It is certified by the experimental team that the success probability of the project is in this round  $[\Theta]$ .* Participant B had to pay 1€ to send you the certified message. Please consider that this is the true probability of success. Make a decision how much you want to invest (between 0€ and 11€).

Reminder: Your payoff depends on your investment amount and whether the project succeeds or fails. The success of the project is determined by a random lottery made by the computer. With a probability of  $[\Theta]$ , the project succeeds: You receive 11€ plus the amount invested and Participant B receives the amount invested. With a probability of  $1 - [\Theta]$ , the project fails: You lose the amount invested and Participant B receives nothing.

## **Chapter 4**

# **Verification and Blockchain-Based Self-Verification Instruments for Risky Investment Decisions**

### **Bibliographic Information**

This chapter shall be cited as: Stenzel, A. (2022). Verification and Blockchain-Based Self-Verification Instruments for Risky Investment Decisions: An Experiment.

### **Acknowledgements**

I would like to thank Israel Waichman and Till Requate for their valuable comments and support in the development of this work.

### **Competing Interest**

The author declares no competing interest.

## 4.1 Introduction

The sharing of financial information is a crucial element in facilitating business and trade, but faithful and complete disclosure is not always in the best interest of a given stakeholder. For example, in order to enhance their probability of securing funding, firms may elect to withhold pertinent information from prospective investors. This may result in adverse selection, which, in turn, may lead to suboptimal investments (Akerlof, 1970; Viscusi, 1978; Milgrom, 2008). In order to circumvent such unfavorable outcomes, a variety of formal and informal verification institutions and processes (e.g., due diligence, rating systems) have been established (for further insight, see Milgrom, 2008).<sup>1</sup> Typically, these come at substantial costs and may not reveal the most up-to-date or precise information (Hynes, 2017).<sup>2</sup> A recent technological innovation, Blockchain combined with suitable encryption technology, promises to solve the asymmetric information problem at considerably lower costs by allowing firms to reliably self-reveal information to investors.<sup>3</sup> The present experimental study compares the efficacy of a novel self-verification instrument with that of traditional verification institutions in the context of risky investments characterized by information asymmetry.

Blockchain can be defined as a decentralized accounting ledger where transactions are recorded simultaneously to a multitude of databases (Tapscott & Tapscott, 2017; Yaga et al., 2019; Ornes, 2019). This renders data entries immutable, thereby rendering it optimal for low-trust environments. However, it may also prevent individuals and organizations from sharing data, as they fear exposing sensitive information. Thus, to use Blockchain for sharing sensitive data, it needs to be accompanied by suitable encryption technology. One such promising invention is “zero-knowledge proofs” (Goldwasser et al., 1989): A party can prove that it meets certain criteria (e.g., funding or sustainability requirements) without revealing any other information. Another invention is “secure multi-party computing” (Yao, 1986), a protocol that allows computation or analysis of combined data without the different parties revealing their

---

<sup>1</sup> For “quality disclosure,” see, e.g., Dranove and Jin (2010) and in the special context of early-stage startups see Bernstein et al. (2017) and Bapnaa (2019).

<sup>2</sup> Furthermore, third-party certifiers may have a conflict of interest, which can result in scandals such as the Enron case (Dranove & Jin, 2010) or the recent cases involving Wirecard and Volkswagen in Germany.

<sup>3</sup> Blockchain-based applications could be used for various verification purposes, e.g., to verify sustainable and humane seafood supply chains (Gephart et al., 2019).

private input (Bogdanov et al., 2012).<sup>4</sup> In this regard, Abbe et al. (2012) provide an approach combining Blockchain and secure multi-party computing to enable financial audits without an intermediary. Moreover, Wang and Kogan (2018) combine Blockchain and zero-knowledge proofs to build a transaction processing system for real-time accounting, continuous monitoring, and fraud prevention. In combination, Blockchain with encryption can save considerable time and cost of due diligence while providing accurate and real-time information (see discussions by Gonzalez (2019) and Y. Chen and Bellavitis (2020)).

To study the possible effects of (self-)verification instruments on risky investments, we use a stylized two-player game mimicking essential features of the investor-firm relationship: the Probabilistic Investment Game (Stenzel et al., 2022). The baseline treatment is depicted in Figure 4.1: An investor starts with an €8 endowment that can be invested in a firm. Each firm has a success probability drawn from a discrete uniform distribution  $\theta \sim U(20\%, 70\%)$  (i.e., the expected value is 45%).<sup>5</sup> Success means that both, the investor and the firm, earn a profit equal to the investor’s investment (in addition, the investor keeps the initial amount of money), whereas failure means that the invested amount is lost. The baseline treatment includes two stages: First, the firm is informed about its success probability and sends an unverifiable message to the investor. The message may contain the true probability of success or any other value of the firm’s choice. Second, the investor who is only informed about the a priori distribution of the success probability decides how much of an initial endowment to invest in the firm. After that, a lottery based on the true probability of success will decide on the success or failure of the project.

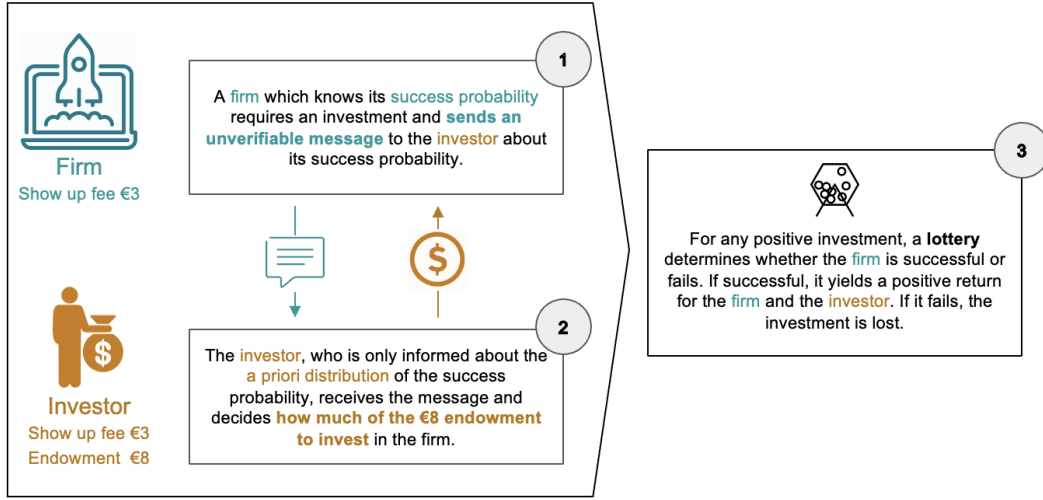
This study adapts the Probabilistic Investment Game (Stenzel et al., 2022) and includes five treatments. We first model the current standard practice of investors using third parties to obtain information as part of a costly due diligence process. In the treatments *Costly Noisy Verification* (T2) and *Costly Verification* (T3), the investor can purchase either noisy (in T2)<sup>6</sup> or accurate (in T3) information about the firm’s true probability of success after receiv-

<sup>4</sup> The recent development in computational power allows the implementation of the theoretical ideas by Goldwasser et al. (1989) and Yao (1986). For an overview of the latest innovations of encryption technologies and their applications in the finance industry, see Blake et al. (2019).

<sup>5</sup> We assume that all values of  $\theta$  within  $[20\%, 21\%, \dots, 70\%]$  are equally likely. Note that  $\theta$  can only take the discrete values explicitly listed above, such as 20%, 21%, 22%, etc., and not intermediate values like 21.5%.

<sup>6</sup> In T2, the information is represented as inaccurate or outdated, as is often the case in practice (Hynes, 2017).





**Figure 4.1:** The Probabilistic Investment Game (one-shot design)

ing the firm's message and before deciding how much to invest in the firm. *Costly Self-Verification* (T4) represents the current practice whereby firms may undertake costly self-verification, for example, by obtaining certification through costly intermediaries. Then, in *Self-Verification* (T5), we model using Blockchain or other self-verification technology to ensure the reliable dissemination of confidential information to investors. The distinction of T4 and T5 lies in the differing costs associated with self-verification. In the case of *Costly Self-Verification* (T4), the act of self-verification is accompanied by a significant financial burden. Conversely, in *Self-Verification* (T5), the cost of self-verification is low, as is the promise of the emerging technologies (Wang & Kogan, 2018).

The paper proceeds as follows: In the next section, we give an overview of the relevant literature. Section 4.3 presents the design, our predictions, and the experimental procedure. We then present our results in Section 4.4. Section 4.5 is a summary and conclusion.

## 4.2 Related literature

The literature referenced in this chapter has significant overlap with that presented in the previous chapter. To avoid repetition, a brief summary is provided here. A comprehensive literature review can be found in Section 3.2. Our *Baseline* (T1) is related to the extensive body of experimental literature on cheap talk communication in a sender-receiver signaling game (Berg et al., 1995; Crawford, 1998; Forsythe et al., 1999; Cai et al., 2006; Abeler et al.,

2019; Blume et al., 2020). Treatments T2-T3 can be considered cheap talk with monitoring (Abeler et al., 2019; Burdea et al., 2023). Treatments T4-T5 are related to “persuasion games” where senders can disclose information (Forsythe et al., 1999; Milgrom, 2008; Benndorf et al., 2015; Jin et al., 2021; Fréchette et al., 2022; Burdea et al., 2023). The workhorse game extends the existing literature by modeling the probabilistic nature of investments (the firm’s true type is a probability figure instead of a deterministic figure, e.g., “good” and “bad”). This probabilistic nature makes it impossible for the investor to later know whether the firm’s message was true or not, and thus may encourage less truth-telling from the sender than in a standard sender-receiver game.

### 4.3 Design, predictions, and experimental procedure

This section introduces the Probabilistic Investment Game, before outlining the theoretical predictions and experimental procedure.

#### 4.3.1 Experimental design

The one-shot variation of the Probabilistic Investment Game is shown in Figure 4.1. It is an extension of sender-receiver games by adding risk to the receiver’s decision. The game consists of two players: an investor (referred to with female pronouns) and a firm (referred to with neutral pronouns). The investor receives an investment endowment of  $w \in \mathbb{N}$ , and can invest an amount  $i \in [0, w]$  into the firm. In our experiment, we set  $w = \text{€}8$ . The firm does not own any capital but contributes a potential innovation that could yield a positive return to the investor and the firm. The probability for the innovation to be successful is  $\theta \in \Theta$ , where  $\Theta = [a\%, b\%]$  with  $a, b \in \mathbb{N}$  and  $0 \leq a < b \leq 100$ . In our experiment, we set  $a = 20$  and  $b = 70$ . The success probability  $\theta \sim \{a\%, b\%\}$  is randomly drawn following a discrete uniform distribution, which is common knowledge.

The baseline game then proceeds as follows: In the first stage, the firm is informed about its success probability  $\theta$ , and sends an unverifiable message  $m$  to the investor. The message can contain the true probability of success or any other value of the firm’s choice (between  $a\%$  and  $b\%$ ). In the second stage, after receiving  $m$ , the investor chooses how much of the endowment to invest in the firm  $i \in [0, w]$ . Afterward, a random lottery determines whether the innovation succeeds (with probability  $\theta$ ) or fails (with probability  $1 - \theta$ ). If

successful, the investor will receive a payoff of  $w + i$ , while the firm will receive a payoff of  $i$ . Otherwise, investor's payoff is  $w - i$  and the firm's payoff is 0. Therefore, for a given probability of success  $\theta$ , the investor's expected payoff is  $2\theta i + w - i$ , and the firm's expected payoff is  $\theta i$ . Note that, in addition to these payoffs from the experiment, a show-up fee of €3 is paid to each participant (investors and firms).

**Table 4.1:** Treatments

Name	Verification?	Verification cost	Paid by	Pairs (participants)
T1: Baseline	No	–	–	54 (108)
T2: Costly Noisy Verification	Noisy	high, $c$ (1 €)	Investor	51 (102)
T3: Costly Verification	Accurate	high, $c$ (1 €)	Investor	51 (102)
T4: Costly Self-Verification	Accurate	high, $c$ (1 €)	Firm	51 (102)
T5: Self-Verification	Accurate	low, $\tilde{c}$ (0.1 €)	Firm	53 (106)
				260 (520)

Besides the *Baseline* (T1) treatment described above, we conducted four additional treatments where the firm's message can be partially or fully verified or self-verified. The treatments are outlined in Table 4.1. The *Costly Noisy Verification* (T2) and the *Costly Verification* (T3) treatments are analogous to T1, but now after the message  $m$  is received (and before making an investment decision), the investor can purchase information about the firm's true probability of success at a cost  $c$ . In the *Costly Noisy Verification* (T2) treatment, the information is noisy. The purchased information contains an interval of 20 percentage points that includes the true probability of success, but it does not need to be in the middle of the interval. By contrast, in the *Costly Verification* (T3) treatment, the exact true probability of success is revealed.

In *Costly Self-Verification* (T4) and the *Self-Verification* (T5) treatments, the investor cannot purchase additional information (as in T2 and T3). However, the firm can reliably send the investor (through a certified message) its true probability of success  $\theta$  at a cost  $c$  in T4 and  $\tilde{c} < c$  in T5. If the firm decides not to send its true probability of success in the certified message, it sends a cheap talk message  $m$  instead (as in treatments T1-T3). Thus, the investor receives either  $m$  (which may or may not correspond to the true value) or the true probability of success  $\theta$ , and chooses the investment amount  $i \in [0, w]$ . Note that the investor's cost of purchasing information in T2-T3 and the firm's cost of providing reliable information in T4-T5 is deducted from

the show up-fee.<sup>7</sup> The payoff functions for T2-T5 are amended accordingly.

### 4.3.2 Predictions

Our theoretical predictions are as follows: In the *Baseline* (T1), given that the a priori expectation of the true probability of success is 45%, the investor does not invest. Following the revelation principle for general Bayesian games, there cannot be any messaging strategy for the firm leading to an equilibrium with a positive amount of investment. In *Costly Noisy Verification* (T2) and *Costly Verification* (T3), the verification cost is set to outweigh the expected benefit. Hence, in equilibrium, the investor does not purchase information and does not invest. In *Costly Self-Verification* (T4) and *Self-Verification* (T5), we expect a separating investment pattern. If a firm does not send a self-verified message with its true probability of success, the investor concludes that the success probability is low and refrains from investing. Consequently, a firm with a sufficiently high probability of success is expected to self-verify even when it is costly to do so (Grossman & Hart, 1980). In this case, investors invest the whole endowment. The proofs are analogous to the proofs represented in Section 3.3.2.

Please note two distinctions between the one-shot and the repeated design that have implications for the predictions. Firstly, in the one-shot scenario, the costs associated with verification in T2 and T3 are set to exceed the potential benefits, resulting in the investor opting not to verify. Secondly, in the repeated Probabilistic Investment Game, a noisy verification treatment was not conducted. The arguments presented in the proof for noisy verification remain consistent with those employed for the purpose of accurate verification.

### 4.3.3 Experimental procedure

The experiment was conducted between January and February 2021. We conducted 5 sessions for T5 and 6 sessions for the other treatments. Each session was usually attended by 16-20 participants.<sup>8</sup> The experiment was conducted using oTree version 3.2.5 (D. Chen et al., 2016) with 520 participants (58%

---

<sup>7</sup> As the firms do not receive initial endowments, in *Costly Self-Verification* (T4) and *Self-Verification* (T5) the costs are deducted from the show-up fee. To keep these former treatments comparable with *Costly Noisy Verification* (T2) and *Costly Verification* (T3), we also deduct the verification cost from the investor's show-up fee.

<sup>8</sup> Out of a total of 29 sessions, we had one session with 12 and one session with 14 participants, 7 sessions with 16 participants, 9 sessions with 18 participants, 7 sessions with 20 participants, 4 sessions with 22 participants.

women) in 5 treatments (54, 51, 51, 51, and 53 investor-firm pairs in the respective treatments).<sup>9</sup> Participants were recruited using hroot (Bock et al., 2014) from different faculties of a large German University. Participants indicated, on average, a difficulty level of 2.02 (“Easy”), with a standard deviation of 0.82 (ranked on a scale from 1 = Very Easy to 5 = Very Difficult). The average earning was €8.06 and transferred to the participants’ bank accounts within 2 days.

## 4.4 Results

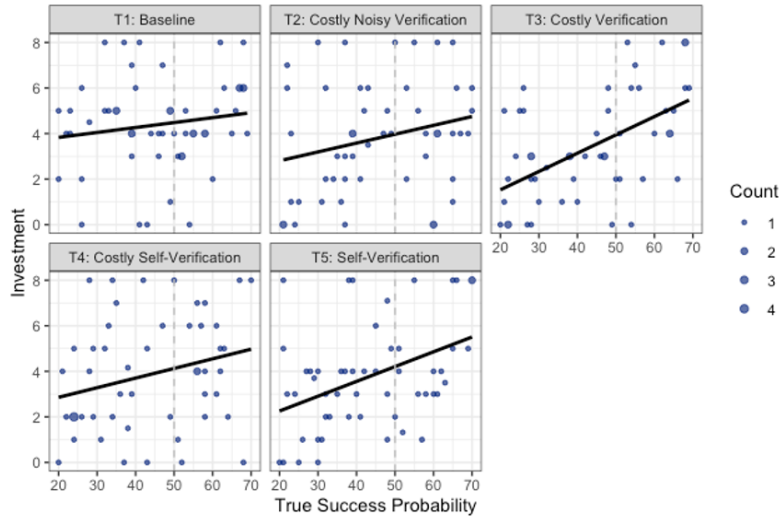
### 4.4.1 Effectiveness in conditioning investment on true probability of success

Figure 4.2 shows scatter plots of investments per true probability of success, indicating that the theoretical predictions do not provide a good description of the observed investment patterns in any treatment. Recall that no investment is predicted in T1-T3, and in T4-T5 investments are only predicted for self-verified probabilities of at least 50%. However, between 87% and 93% of investors invested a positive amount in the treatments. One measure of investment effectiveness is whether investments increase with the true probabilities of success (which are a priori unknown to the investors). Using a linear regression ( $n = 260$ ; R-squared = 0.14, see Table 4.2), only in *Costly Verification* (T3) and *Self-Verification* (T5) we find that the true probability of success positively and significantly affects the amount invested (Wald tests, see Table 4.A.2, yield  $p = 0.0000$  and  $p = 0.0026$ , respectively). The true probability of success does not significantly affect investment in the *Baseline* (T1) (Wald test yields  $p = 0.2043$ ), and only marginally in *Costly Noisy Verification* (T2) and *Costly Self-Verification* (T4) (Wald tests,  $p = 0.0954$  and  $p = 0.0542$ , respectively).<sup>10</sup>

Table 4.2 indicates that *Costly Verification* (T3) and *Self-Verification* (T5) are effective in conditioning investments on their respective true probability of success. However, what leads to this effectiveness? Is it caused by the actual use of (self-)verification? Or rather by the option value of verification (i.e.,

<sup>9</sup> Four pairs were excluded from the data analysis because of participants registering for the experiment with a different email and participating more than once. Because of the online payment procedure, we are sure to detect all such cases.

<sup>10</sup> Since a positive relationship between investment and true probability of success is more important for higher levels of the latter, Table 4.A.1 shows also a model regressing “investments” on “squared success probability.” All Wald tests are reported in Table 4.A.2.



Note: “Count” indicates cases of multiple identical observations in a given treatment. A regression line with a flat slope indicates no significant relationship between firms’ true probability of success and investment amounts, whereas a line with a steep slope implies a positive relationship between firms’ success probabilities and investment amounts.

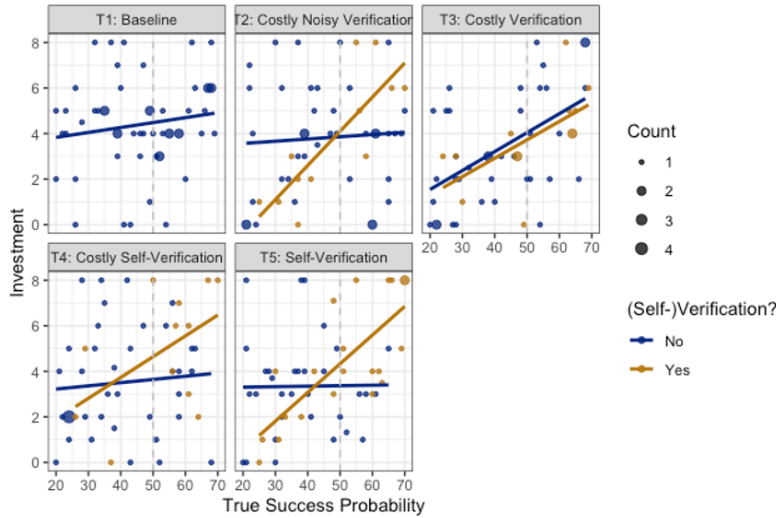
**Figure 4.2:** Scatter plot and linear prediction: Investments per true probability of success

**Table 4.2:** Investment per true probability of success

	(Model 1)
Success Prob.	0.02 (0.02)
Treatment T2	−1.37 (1.44)
Treatment T3	−3.48*** (1.20)
Treatment T4	−1.39 (1.30)
Treatment T5	−2.44* (1.32)
Treatment T2 × Success Prob.	0.02 (0.03)
Treatment T3 × Success Prob.	0.06** (0.02)
Treatment T4 × Success Prob.	0.02 (0.03)
Treatment T5 × Success Prob.	0.04 (0.03)
Constant	3.40*** (0.87)
Observations	260
R-squared	0.14

Note: Robust standard errors in parentheses and \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%.

the fact that the investor could verify the firm's true probability of success)? Figure 4.3 shows linear relationships between investments and the true probability of success with and without actual use of (self-)verification. This figure indicates that only in T3 effectiveness is attained without actual use of verification. In the other treatments, effectiveness is only achieved when verification or self-verification are actually used and, thus, when investors' decisions are fully informed.<sup>11</sup> This is supported by a linear regression model ( $n = 206$ ; R-squared = 0.17, see Table 4.3): If (self-)verification is not used, only in *Costly Verification* (T3), the investment is significantly affected by the true probability of success (Wald test,  $p = 0.0001$ ; in Treatments T2, T4, and T5: Wald test yields  $p = 0.1319$ ,  $p = 0.3175$ , and  $p = 0.0731$ , respectively). However, when (self-)verification is used, then the true probability of success significantly affects the amount invested in *Costly Verification* (T3), *Costly Self-Verification* (T4), and *Self-Verification* (T5) (Wald test,  $p = 0.0000$ ,  $p = 0.0222$ , and  $p = 0.0026$ , respectively). The true probability of success is only marginally significant in *Costly Noisy Verification* (T2) (Wald test,  $p = 0.0717$ ).<sup>12</sup>



Note: “Count” indicates cases of multiple identical observations in a given treatment. “(Self-)Verification” indicates actual instrument use (“Yes,” indicated by orange dots and lines) or not (“No,” indicated by blue dots and regression lines).

**Figure 4.3:** Investment per true probability of success with and without using (self-)verification

<sup>11</sup> The only exception is Treatment T2, where after the noisy verification, investors are better (but not fully) informed about the true probability of success.

<sup>12</sup> Since a positive relationship between investment and true probability of success is more important for higher levels of the latter, Table 4.A.3 shows also a model regressing “investments” on “squared success probability.” All Wald tests are reported in Tables 4.A.4 and 4.A.5.

**Table 4.3:** Investment per true probability of success and instrument use

	(Model 1)
Success Prob.	0.04 (0.02)
Verified $\times$ Success Prob.	0.05* (0.02)
Treatment T3	-2.21 (1.46)
Treatment T4	0.44 (1.53)
Treatment T5	-0.60 (1.57)
Treatment T3 $\times$ Success Prob.	0.05 (0.03)
Treatment T4 $\times$ Success Prob.	-0.01 (0.03)
Treatment T5 $\times$ Success Prob.	0.01 (0.03)
Verified $\times$ Success Prob $\times$ Treatment T3	0.03 (0.03)
Verified $\times$ Success Prob $\times$ Treatment T4	0.00 (0.03)
Verified $\times$ Success Prob $\times$ Treatment T5	0.02 (0.03)
Constant	2.05* (1.17)
Observations	206
R-squared	0.17

Note: Robust standard errors in parentheses and \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%. Regression only includes treatment with (self-)verification T2-T5.

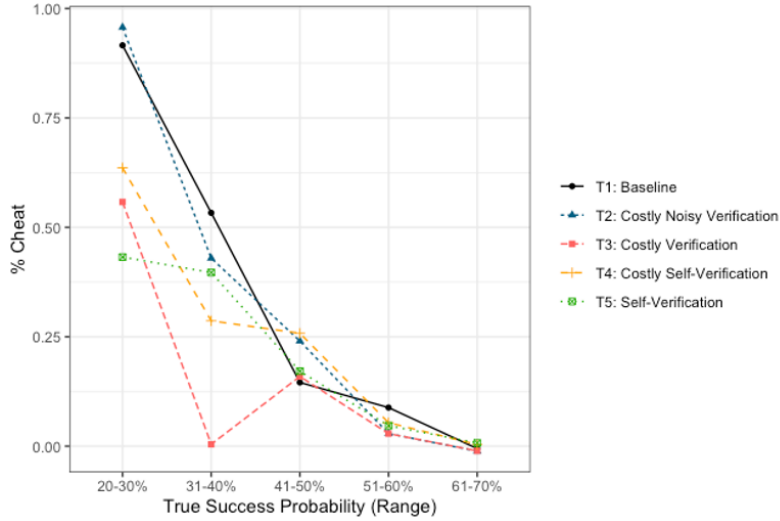
We conclude that *Costly Verification* (T3) attains effectiveness because of the option value of investors to verify the success probability. By contrast, in *Self-Verification* (T5), effectiveness is only achieved when the firms send a self-verified message with their true probability of success. Notably, 40% of the firms use self-verification in T5, but only 24% and 25% of the investors verify in T2 and T3, respectively, and 24% of the firms self-verify in T4. Thus, the effectiveness of *Self-Verification* (T5) is due to the frequent use of self-verification when it is sufficiently cheap, leading to more informed and thus effective investors' decisions.

#### 4.4.2 High effectiveness in Costly Verification even without instrument use

Now, we investigate how *Costly Verification* (T3) attains high effectiveness even without investors actually using verification. Figure 4.4 shows the aver-



age “percentage cheating” per true probability of success range. “Percentage cheating” measures the deviation of the cheap talk message from the true probability of success.<sup>13</sup> This figure indicates that (i) firms do not communicate the highest possible probability of success to the investors and (ii) levels of percentage cheating are lowest in *Costly Verification* (T3). A regression analysis supports the aforementioned results (see Table 4.4).<sup>14</sup>



Note: “% cheat” is defined by the difference between the message received and the true probability of success divided by the true probability of success. In T4 and T5, we consider a self-verified message as a message containing the true probability of success.

**Figure 4.4:** Average percentage cheating per true probability of success range

Lower cheating of firms in T3 would lead to effectiveness in conditioning investment on the true probability of success only if investors believe the cheap talk message. After making their investment decision (and in case of no verification), we asked investors to indicate their guesses for the firm’s true probability of success. Then, we regress the investor guesses about the true probability of success on the cheap talk messages received from the firms. This regression ( $n = 198$ ; R-squared = 0.43, see Table 4.5) shows that the cheap talk communication positively affects the investor’s guess about the true probability of success (Wald tests yield  $p \leq 0.0001$  in all treatments).<sup>15</sup> Thus, the effectiveness of *Costly Verification* (T3) in conditioning investment on the true probability of success can be explained by less cheating merely because of the

<sup>13</sup> Note that self-verified messages in T4 and T5 are considered as truth-telling.

<sup>14</sup> Table 4.4 presents a log-linear estimation. Given the exponential decay of percentage cheating, this estimation provides the best fit of the data (see Section 4.A.3 in the Appendix).

<sup>15</sup> All Wald tests are reported in Table 4.A.11 and 4.A.12.

**Table 4.4:** “Log(percentage cheating)” per true probability of success

	(Model 1) Log(% cheat)
Success Prob.	−0.09*** (0.01)
Treatment T2	−0.47 (0.43)
Treatment T3	−0.93** (0.45)
Treatment T4	−1.26* (0.73)
Treatment T5	−0.74 (0.49)
Treatment T2 × Success Prob.	0.01 (0.01)
Treatment T3 × Success Prob.	0.01 (0.01)
Treatment T4 × Success Prob.	0.03* (0.02)
Treatment T5 × Success Prob.	0.02 (0.01)
Constant	2.52*** (0.18)
Observations	133
R-squared	0.67

Note: Robust standard errors in parentheses and \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%. Dependent variable: log(percentage cheating).

option of verification (albeit the actual verification cost is high!) and the investors’ trust in the message.

#### 4.4.3 Measuring Investor and Societal Efficiency

Other evaluation criteria (besides effectiveness which is the effect of true probability of success on the investment decision) are (i) Investor Efficiency and (ii) Societal (joint) Efficiency. We define these efficiencies as follows:

$$\text{Investor Efficiency} = \frac{\text{Observed expected investor payoff}}{\text{Max. expected investor payoff (under perfect information)}},$$

$$\text{Societal Efficiency} = \frac{\text{Observed expected joint payoff}}{\text{Max. expected joint payoff (under perfect information)}}.$$

Because of the random nature of the lottery, we look at the expected payoffs before the lottery has been conducted (taking into account the investment

**Table 4.5:** Investor's guess and the cheap talk message received

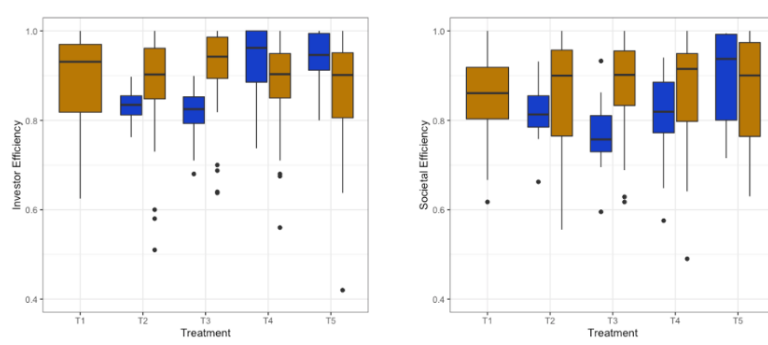
	(Model 1)
Message	0.62*** (0.09)
Treatment T2	-0.57 (6.75)
Treatment T3	5.94 (7.42)
Treatment T4	6.67 (7.80)
Treatment T5	0.53 (8.60)
Treatment T2 $\times$ Message	0.08 (0.14)
Treatment T3 $\times$ Message	-0.09 (0.15)
Treatment T4 $\times$ Message	-0.14 (0.15)
Treatment T5 $\times$ Message	-0.04 (0.16)
Constant	12.32** (5.05)
Observations	198
R-squared	0.43

Note: Robust standard errors in parentheses and \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%; excluding observations with (self-)verification use.

decision and the true probability of success). The maximal expected investor payoff is defined as investing the whole endowment at a true probability of success above 50% and investing nothing otherwise (and without paying for verification). The maximal expected joint payoff is defined as investing the whole endowment at a true probability of success above 33% and investing nothing otherwise. Neither the investor nor the firm pays for (self-)verification.

The theoretical predictions (under risk neutrality) are as follows: Investor Efficiency is maximized in T4 and T5 ( $= 1$ ) and is larger than in T1-T3. Societal Efficiency in T4 and T5 is not maximized as we hypothesize no investments for true probability of success between 33% and 50% but is nevertheless larger than in T1-T3. A drawback of these measures is that they do not produce considerably different values between a priori best and worst investment decisions when the success probability is close to 50%. For instance, under a success probability of 45%, the investor efficiency is 0.9 even when investing the whole endowment. Investment rates (i.e., the share of investors investing a positive amount) in the different treatments are between 86% and 93%, leading to low variability in the efficiency measures between treatments. See Section 4.A.6

in the Appendix for more examples.



Note: The utilisation of verification (in T2 and T3) and self-verification (in T4 and T5) is indicated by the use of blue shading. Conversely, instances where verification or self-verification was not employed are denoted by brown shading.

**Figure 4.5:** Investor and Societal Efficiency per (self-)verification instrument use in the different treatments (box plot)

We pairwise compare sequential treatments (T1 and T2, T2 and T3, etc.) with respect to Investor Efficiency and Societal Efficiency. We find no differences when pairwise comparing treatments with respect to either Investor Efficiency (two-sided tests, p-values between  $p = 0.1349$  and  $p = 0.8015$ ) or Societal Efficiency (p-values between  $p = 0.2861$  and  $p = 0.8644$ ). Next, we test for differences in efficiencies within treatments. In the verification treatments, Investor Efficiency is significantly higher for the investor when not paying for verification (two-sided MWU-test, in T2:  $n = 39$  vs.  $n = 12$ ,  $p = 0.0100$ ; in T3:  $n = 38$  vs.  $n = 13$ ,  $p = 0.0001$ ). In the self-verification treatments, Investor Efficiency is significantly higher when firms use self-verification (MWU-test, in T4:  $n = 39$  vs.  $n = 12$ ,  $p = 0.0454$ ; in T5:  $n = 32$  vs.  $n = 21$ ,  $p = 0.0101$ ). Turning to Societal Efficiency, in *Costly Verification* (T3) and *Costly Self-Verification* (T4), efficiency is reduced with actual verification or self-verification (MWU-test, in T3:  $n = 38$  vs.  $n = 13$ ,  $p = 0.0004$ , in T4:  $n = 39$  vs.  $n = 12$ ,  $p = 0.0480$ ), but not in *Self-Verification* (T5) ( $n = 32$  vs.  $n = 21$ ,  $p = 0.5852$ ). As full unraveling does not occur, it seems that only when a (self-)verification instrument is sufficiently cheap it has the potential to be efficient from a social viewpoint.

## 4.5 Conclusions

Information asymmetry between investors and firms is a prominent example of a market failure leading to inefficient resource allocation. We experimentally

investigate whether a recently available self-verification technology could improve investment decisions and, thus, resource allocation. Our stylized game models a situation between an informed firm and an uninformed investor. In the current study, we employed a one-shot design. In one-shot interactions, there are no future consequences or externalities to a firm’s cheating behavior. Therefore, verification and self-verification instruments are more important than in repeated interactions.

Our findings are as follows: Without verification or self-verification possibility, investments are uninformed and thus ineffective. We then model the current practice where investors attain relevant information about firms using third parties. We find that such a verification instrument is effective in conditioning investment decisions on the firms’ true (but unknown) probability of success when the verification information is accurate but not when the verification information is noisy, albeit the latter is arguably a better representation of the current practice (Hynes, 2017). Moreover, we find that firms do not fully lie in their cheap talk messages. This finding is consistent with those of previous studies on cheap talk and extends this result to games with a probabilistic nature (e.g. Dickhaut et al., 1995; Forsythe et al., 1999; Cai et al., 2006; Abeler et al., 2019). Yet, firms lie less when investors have the option to accurately verify their true probability of success (although verification is highly costly).

Furthermore, in contrast to the theoretical predictions, we find in all our treatments that cheap talk communication does affect the investors’ beliefs on the true probability of success. Such investors’ trust in cheap talk messages is in line with Forsythe et al. (1999) and Benndorf et al. (2015). Thus, less cheating of firms, together with investors’ trust into the cheap talk messages, leads to more effective investment decisions in T3 (even without the actual use of verification).

Next, Blockchain with suitable encryption technology provides new possibilities to reliably communicate up-to-date confidential information to investors at low costs (Wang & Kogan, 2018; Faccia & Mosteanu, 2019). We observe, however, that these potentials are not fully realized in the self-verification treatments, as unraveling does not completely occur (in line with Dranove and Jin (2010) and Jin et al. (2021)). Yet, when a self-verification instrument is sufficiently cheap, it is effective in conditioning investment decisions on the firms’ true probability of success. In this case, firms use self-verification more often than in the other treatments, allowing more investors to be informed

and, therefore, to make more effective decisions. Finally, our data indicate that when (self-)verification instruments are costly, they may not be efficient from a social viewpoint. Hence, recent technological advances allowing the use of Blockchain may have the potential for long-term efficiency increase only when they become available at sufficiently low costs.

## Bibliography

- Abbe, E., Khandani, A., & Lo, A. (2012). Privacy-preserving methods for sharing financial risk exposures. *American Economic Review*, 102(3), 65–70.
- Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4), 1115–1153.
- Akerlof, G. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488–500.
- Bapnaa, S. (2019). Complementarity of signals in early-stage equity investment decisions: Evidence from a randomized field experiment. *Management Science*, 65(2), 933–952.
- Benndorf, V., Kübler, D., & Normann, H. T. (2015). Privacy concerns, voluntary disclosure of information, and unraveling: An experiment. *European Economic Review*, 75, 43–59.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1), 122–142.
- Bernstein, S., Korteweg, A., & Laws, K. (2017). Attracting early-stage investors: Evidence from a randomized field experiment. *Journal of Finance*, 72(2), 509–538.
- Blake, M., McWaters, J., & Galaski, R. (2019). The Next Generation of Data-Sharing in Financial Services: Using Privacy Enhancing Techniques to Unlock New Value. *World Economic Forum*.
- Blume, A., Lai, E., & Lim, W. (2020). Strategic information transmission: A survey of experiments and theoretical foundations. In *Handbook of Experimental Game Theory* (pp. 311–347). Edward Elgar Publishing Ltd.
- Bock, O., Baetge, I., & Nicklisch, A. (2014). Hroot: Hamburg Registration and Organization Online Tool. *European Economic Review*, 71, 117–120.
- Bogdanov, D., Talviste, R., & Willemson, J. (2012). Deploying secure multi-party computation for financial data analysis. In *International Conference on Financial Cryptography and Data Security* (pp. 57–64).

- Burdea, V., Montero, M., & Sefton, M. (2023). Communication with partially verifiable information: An experiment. *Games and Economic Behavior*, 142, 113–149.
- Cai, H., Tao, J., & Wang, Y. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior*, 56(1), 7–36.
- Chen, D., Schonger, M., & Wickens, C. (2016). oTree - An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- Chen, Y., & Bellavitis, C. (2020). Blockchain disruption and decentralized finance: The rise of decentralized business models. *Journal of Business Venturing Insights*, 13, 88–97.
- Crawford, V. (1998). A survey of experiments on communication via cheap talk. *Journal of Economic Theory*, 78(2), 286–298.
- Dickhaut, J., McCabe, K., & Mukherji, A. (1995). An experimental study of strategic information transmission. *Economic Theory*, 6(3), 389–403.
- Dranove, D., & Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4), 935–963.
- Faccia, A., & Mosteanu, N. (2019). Accounting and blockchain technology: From double-entry to triple-entry. *The Business and Management Review*, 10(2), 108–116.
- Forsythe, R., Lundholm, R., & Rietz, T. (1999). Cheap talk, fraud, and adverse selection in financial markets: Some experimental evidence. *The Review of Financial Studies*, 12(3), 481–518.
- Fréchette, G., Lizzeri, A., & Perego, J. (2022). Rules and commitment in communication: An experimental analysis. *Econometrica*, 90(5), 2283–2318.
- Gephart, J., Froehlich, H., & Branch, T. (2019). Opinion: To create sustainable seafood industries, the United States needs a better accounting of imports and exports. *Proceedings of the National Academy of Sciences*, 116(19), 9142–9146.
- Goldwasser, S., Micali, S., & Rackoff, C. (1989). The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1), 186–208.
- Gonzalez, L. (2019). Blockchain, herding and trust in peer-to-peer lending. *Managerial Finance*, 46(6), 815–831.
- Grossman, S., & Hart, O. (1980). Disclosure laws and takeover bids. *The*



- Journal of Finance*, 35(2), 323–334.
- Hynes, R. (2017). Maximum possible accuracy in credit reports. *Law and Contemporary Problems*, 80(3), 87–115.
- Jin, G. Z., Luca, M., & Martin, D. (2021). Is no news (perceived as) bad news? An experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, 13(2), 141–173.
- Milgrom, P. (2008). What the seller won’t tell you: Persuasion and disclosure in markets. *Journal of Economic Perspectives*, 22(2), 115–131.
- Ornes, S. (2019). Blockchain offers applications well beyond Bitcoin but faces its own limitations. *Proceedings of the National Academy of Sciences of the United States of America*, 116(42), 20800–20803.
- Stenzel, A., Requate, T., & Waichman, I. (2022). *The probabilistic investment game*. (Unpublished manuscript)
- Tapscott, A., & Tapscott, D. (2017). How Blockchain is changing finance. *Harvard Business Review*, 1(9), 2–5.
- Viscusi, K. (1978). A note on “lemons” markets with quality certification. *The Bell Journal of Economics*, 9(1), 277–279.
- Wang, Y., & Kogan, A. (2018). Designing confidentiality-preserving Blockchain-based transaction processing systems. *International Journal of Accounting Information Systems*, 30, 1–18.
- Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2019). Blockchain technology overview. *arXiv preprint arXiv:1906.11078*.
- Yao, A. (1986). How to generate and exchange secrets. *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*, 162–167.

## Appendix

### 4.A Further analysis

#### 4.A.1 Investment per true probability of success

A major analysis criterion in our study is whether the investment is positively and significantly affected by the true probability of success (which, without the use of (self-)verification, is unknown to the investor). The stronger the dependency of investment on the true probability of success is, the more effective the treatment condition (note that in the extreme case, uninformed investors make investment decisions that are completely unaffected by the true probability of success). Figure 4.2 shows that investment appears to be most strongly affected by true probability of success (i.e., highest slopes) in *Costly Verification* (T3) and *Self-Verification* (T5). By contrast, in *Baseline* (T1), there seems to be little dependency between the true probability of success and investment (almost flat slope). To formally establish this result, we use linear regression models to regress investment on true probability of success.

In the regressions shown in Table 4.A.1, the effect of the success probability on investment is captured by the joint effects of “Success Prob.” plus “Success Prob. in the respective treatment.” Since a positive relationship is more important for higher levels of true probability of success, Model 2 regresses “investments” on “squared success probability.” The results of the joint tests in Model 2 are similar to Model 1. In both models, only in *Costly Verification* (T3) and *Self-Verification* (T5), true probability of success affects investment at a  $p \leq 0.05$  significance level, whereas in the *Baseline* (T1), such significant effect is not observed. All tests are two-sided. Wald tests whether true probability of success affects investment in each of the treatments are shown in Table 4.A.2.

**Table 4.A.1:** Investment per true success probability, Model 1 and 2

	(Model 1)	(Model 2)
Success Prob.	0.02 (0.02)	-
Treatment T2	-1.37 (1.44)	-0.87 (0.89)
Treatment T3	-3.48*** (1.20)	-2.34*** (0.74)
Treatment T4	-1.39 (1.30)	-0.88 (0.83)
Treatment T5	-2.44* (1.32)	-1.56* (0.80)
Treatment T2 $\times$ Success Prob.	0.02 (0.03)	-
Treatment T3 $\times$ Success Prob.	0.06** (0.02)	-
Treatment T4 $\times$ Success Prob.	0.02 (0.03)	-
Treatment T5 $\times$ Success Prob.	0.04 (0.03)	-
(Success Prob.) <sup>2</sup>	-	0.00 (0.00)
Treatment T2 $\times$ (Success Prob.) <sup>2</sup>	-	0.00 (0.00)
Treatment T3 $\times$ (Success Prob.) <sup>2</sup>	-	0.00** (0.00)
Treatment T4 $\times$ (Success Prob.) <sup>2</sup>	-	0.00 (0.00)
Treatment T5 $\times$ (Success Prob.) <sup>2</sup>	-	0.00 (0.00)
Constant	3.40*** (0.87)	3.77*** (0.56)
Observations	260	260
R-squared	0.14	0.14

Note: Robust standard errors in parentheses and \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%. Model 2 regresses “investments” on “squared success probability.”

**Table 4.A.2:** Wald tests whether true probability of success affects investment in each of the treatments

<b>Model 1:</b>
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T1}]} = 0$ $F(1, 250) = 1.62, p = 0.2043$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T2}]} = 0$ $F(1, 250) = 2.80, p = 0.0954$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T3}]} = 0$ $F(1, 250) = 19.42, p = 0.0000$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T4}]} = 0$ $F(1, 250) = 3.74, p = 0.0542$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T5}]} = 0$ $F(1, 250) = 9.23, p = 0.0026$
<b>Model 2:</b>
$\text{Ho: } \beta_{[\text{Theta}^2]} + \beta_{[\text{Theta}^2 \times \text{Treatment T1}]} = 0$ $F(1, 250) = 2.13, p = 0.1457$
$\text{Ho: } \beta_{[\text{Theta}^2]} + \beta_{[\text{Theta}^2 \times \text{Treatment T2}]} = 0$ $F(1, 250) = 2.61, p = 0.1072$
$\text{Ho: } \beta_{[\text{Theta}^2]} + \beta_{[\text{Theta}^2 \times \text{Treatment T3}]} = 0$ $F(1, 250) = 22.40, p = 0.0000$
$\text{Ho: } \beta_{[\text{Theta}^2]} + \beta_{[\text{Theta}^2 \times \text{Treatment T4}]} = 0$ $F(1, 250) = 3.01, p = 0.0838$
$\text{Ho: } \beta_{[\text{Theta}^2]} + \beta_{[\text{Theta}^2 \times \text{Treatment T5}]} = 0$ $F(1, 250) = 11.49, p = 0.0008$

#### 4.A.2 Investment per true probability of success with and without (self-)verification

Next, we would like to learn how instrument use (i.e., verification or self-verification) affects the effectiveness of investments (in causing investment to be positively and significantly affected by the true probability of success). A scatter plot showing the investment per true probability of success with and without (self-)verification use is shown in Figure 4.3. Table 4.A.3 displays linear regression models that allow testing whether the investment is effective depending on the instrument use. The results of Model 1 are reported in the main paper. As in Table 4.A.1 above, Model 2 regresses “investments” on the “squared success probability.” The results of the Wald tests are reported in Tables 4.A.4 and 4.A.5.

In Model 1, without the use of the (self-)verification instrument, only *Costly Verification* (T3) is effective in that true probability of success significantly affects investment (at  $p \leq 0.05$ ) but not any of the other treatments. When using (self-)verification instruments, *Costly Verification* (T3), *Costly Self-Verification* (T4), and *Self-Verification* (T5) are effective in that the true probability of success significantly and positively affects investment (at  $p \leq 0.05$ ). Model 2, which regresses investment on “squared success probability,” yields a similar result to Model 1 except that the investment in *Costly Noisy Verification* (T2) turns significant (compared to marginally significant in Model 1).

**Table 4.A.3:** Investment per true probability of success with and without (self-)verification, Model 1 and 2

	(Model 1)	(Model 2)
Success Prob.	0.04 (0.02)	-
Verified $\times$ Success Prob.	0.05* (0.02)	-
Treatment T3	-2.21 (1.46)	-1.54* (0.87)
Treatment T4	0.44 (1.53)	0.23 (0.94)
Treatment T5	-0.60 (1.57)	-0.38 (0.93)
Treatment T3 $\times$ Success Prob.	0.05 (0.03)	-
Treatment T4 $\times$ Success Prob.	-0.01 (0.03)	-
Treatment T5 $\times$ Success Prob.	0.01 (0.03)	-
Verified $\times$ Success Prob $\times$ Treatment T3	0.03 (0.03)	-
Verified $\times$ Success Prob $\times$ Treatment T4	0.00 (0.03)	-
Verified $\times$ Success Prob $\times$ Treatment T5	0.02 (0.03)	-
(Success Prob.) <sup>2</sup>	-	0.00 (0.00)
Verified $\times$ (Success Prob.) <sup>2</sup>	-	0.00** (0.00)
Treatment T3 $\times$ (Success Prob.) <sup>2</sup>	-	0.00* (0.00)
Treatment T4 $\times$ (Success Prob.) <sup>2</sup>	-	-0.00 (0.00)
Treatment T5 $\times$ (Success Prob.) <sup>2</sup>	-	0.00 (0.00)
Verified $\times$ (Success Prob.) <sup>2</sup> $\times$ Treatment T3	-	0.00 (0.00)
Verified $\times$ (Success Prob.) <sup>2</sup> $\times$ Treatment T4	-	0.00 (0.00)
Verified $\times$ (Success Prob.) <sup>2</sup> $\times$ Treatment T5	-	0.00 (0.00)
Constant	2.05* (1.17)	2.92*** (0.70)
Observations	206	206
R-squared	0,17	0,18

Note: Robust standard errors in parentheses. Regression only includes treatment with (self-)verification T2-T5 and \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%. Model 2 regresses “investments” on “squared success probability.”

**Table 4.A.4:** Wald tests whether true probability of success affects investment in each of the treatments with and without (self-)verification, Model 1

Model 1:
<p><b>Without the use of verification:</b></p> <p>Ho: <math>\beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T2}]} = 0</math>  <math>F(1, 194) = 2.29, p = 0.1319</math></p> <p>Ho: <math>\beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T3}]} = 0</math>  <math>F(1, 194) = 15.31, p = 0.0001</math></p> <p>Ho: <math>\beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T4}]} = 0</math>  <math>F(1, 194) = 1.00, p = 0.3175</math></p> <p>Ho: <math>\beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T5}]} = 0</math>  <math>F(1, 194) = 3.25, p = 0.0731</math></p> <p><b>With the use of (self-)verification:</b></p> <p>Ho: <math>\beta_{[\text{Verified}]} + \beta_{[\text{Theta}]} + \beta_{[\text{Verified} \times \text{Theta} \times \text{Treatment T2}]} = 0</math>  <math>F(1, 194) = 3.28, p = 0.0717</math></p> <p>Ho: <math>\beta_{[\text{Verified}]} + \beta_{[\text{Theta}]} + \beta_{[\text{Verified} \times \text{Theta} \times \text{Treatment T3}]} = 0</math>  <math>F(1, 194) = 18.99, p = 0.0000</math></p> <p>Ho: <math>\beta_{[\text{Verified}]} + \beta_{[\text{Theta}]} + \beta_{[\text{Verified} \times \text{Theta} \times \text{Treatment T4}]} = 0</math>  <math>F(1, 194) = 5.31, p = 0.0222</math></p> <p>Ho: <math>\beta_{[\text{Verified}]} + \beta_{[\text{Theta}]} + \beta_{[\text{Verified} \times \text{Theta} \times \text{Treatment T5}]} = 0</math>  <math>F(1, 194) = 9.30, p = 0.0026</math></p>

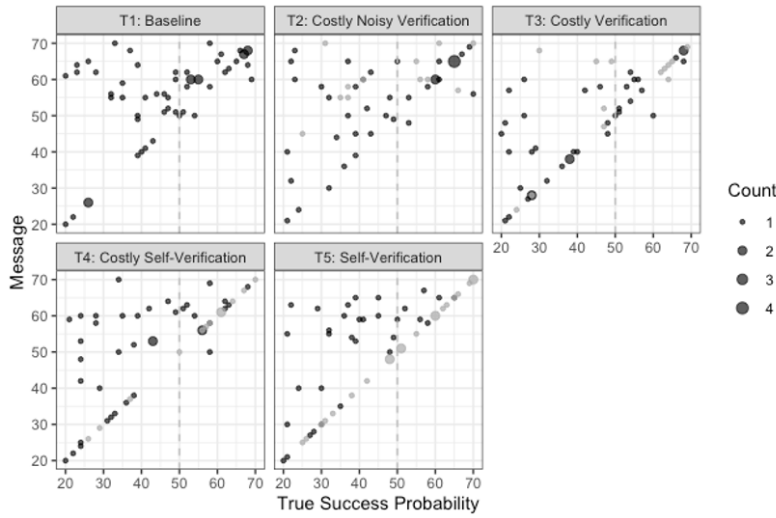
**Table 4.A.5:** Wald tests whether true probability of success affects investment in each of the treatments with and without (self-)verification, Model 2

<b>Model 2:</b>
<b>Without use of verification:</b>
Ho: $\beta_{[\text{Theta}^2]} + \beta_{[\text{Theta}^2 \times \text{Treatment T2}]} = 0$ F(1, 194) = 1.39, p = 0.2397
Ho: $\beta_{[\text{Theta}^2]} + \beta_{[\text{Theta}^2 \times \text{Treatment T3}]} = 0$ F(1, 194) = 15.37, p = 0.0001
Ho: $\beta_{[\text{Theta}^2]} + \beta_{[\text{Theta}^2 \times \text{Treatment T4}]} = 0$ F(1, 194) = 0.31, p = 0.5767
Ho: $\beta_{[\text{Theta}^2]} + \beta_{[\text{Theta}^2 \times \text{Treatment T5}]} = 0$ F(1, 194) = 1.58, p = 0.2106
<b>With use of (self-)verification:</b>
Ho: $\beta_{[\text{Verified}]} + \beta_{[\text{Theta}^2]} + \beta_{[\text{Verified} \times \text{Theta}^2 \times \text{Treatment T2}]} = 0$ F(1, 194) = 5.43, p = 0.0208
Ho: $\beta_{[\text{Verified}]} + \beta_{[\text{Theta}^2]} + \beta_{[\text{Verified} \times \text{Theta}^2 \times \text{Treatment T3}]} = 0$ F(1, 194) = 20.29, p = 0.0000
Ho: $\beta_{[\text{Verified}]} + \beta_{[\text{Theta}^2]} + \beta_{[\text{Verified} \times \text{Theta}^2 \times \text{Treatment T4}]} = 0$ F(1, 194) = 6.24, p = 0.0133
Ho: $\beta_{[\text{Verified}]} + \beta_{[\text{Theta}^2]} + \beta_{[\text{Verified} \times \text{Theta}^2 \times \text{Treatment T5}]} = 0$ F(1, 194) = 14.07, p = 0.0002



#### 4.A.3 Percentage cheating per true probability of success

Figure 4.A.1 shows scatter plots of the true probability of success (x-axis) and the communicated success probability by the firm (y-axis). Observations on the 45°-line are cases of either truth-telling (in all treatments) or self-verification (in T4 and T5). This figure indicates that there is a considerable amount of truth-telling (especially in T3, but even in T1 and T2) and self-verification in T4 and T5. Moreover, there is very little “reverse cheating” (i.e., a firm rarely reports a lower probability of success than its true value).



Note: Observations with (without) (self-)verification are indicated by light (dark) grey shading.

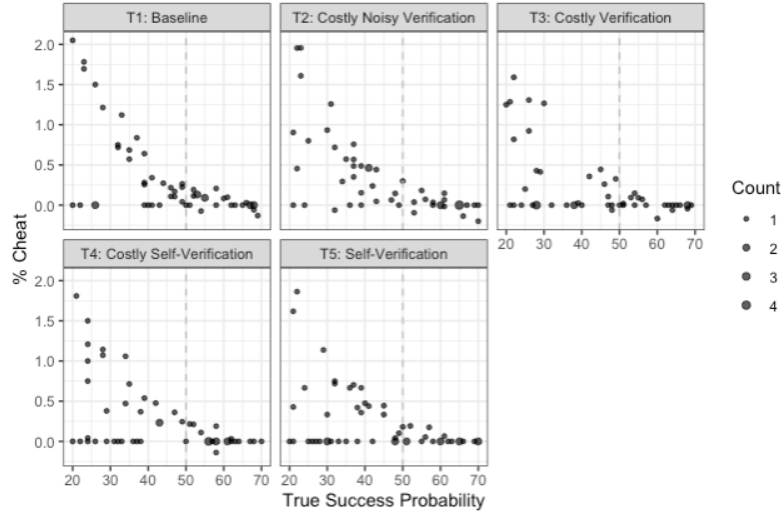
**Figure 4.A.1:** Messages per true probability of success (scatter plot)

Now, we would like to learn why *Costly Verification* (T3) is effective in conditioning investment on true probability of success even when verification is not used, while in all the other treatments, effectiveness is only achieved under actual verification or self-verification. To this end, we construct a “percentage cheating” variable, which is defined as follows:

$$\text{Percentage cheating} = \frac{\text{Message by the firm} - \text{true probability of success}}{\text{true probability of success}}.$$

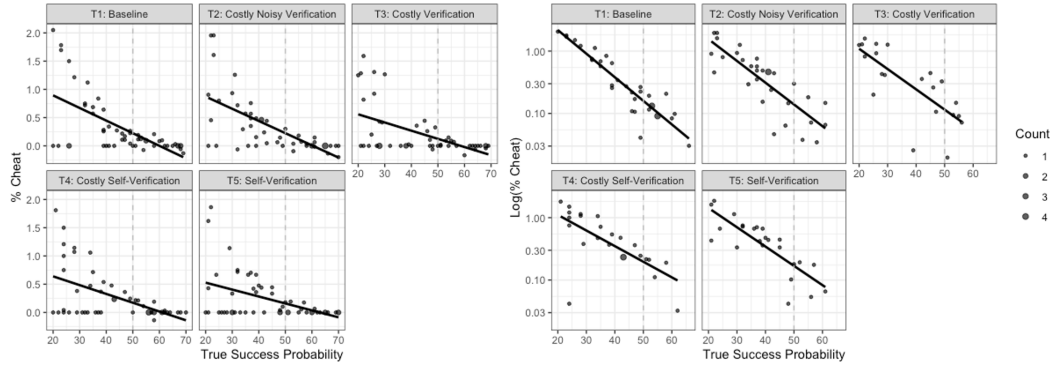
Importantly, we consider the self-verification in treatments T4 and T5 as truth-telling (i.e., “percentage cheating = 0”). Figure 4.A.2 shows a scatter plot of “percentage cheating” and true probability of success. This figure indicates that (i) “percentage cheating” is subject to an exponential decay (by design) with respect to the true probability of success in all treatments, and (ii) there

is less cheating in *Costly Verification* (T3) than in the other treatments. This latter feature is more easily observed in Figure 4.4 showing the mean percentage cheating per true probability of success range.



**Figure 4.A.2:** “Percentage cheating” per true probability of success (scatter plot)

Table 4.A.6 shows an estimation of two regression models. Model 2 is a standard estimation regressing “percentage cheating” on the true probability of success. However, because of the exponential decay of investment per true probability of success, it appears that Model 1, where the dependent variable is “log (percentage cheating),” fits the data considerably better than Model 2 (the R-squared in Model 1 is 0.67 compared to 0.34 in Model 2). This is also visualized in Figure 4.A.3. Taking Model 2 as the correct estimation model, we establish the result of lower cheating in *Costly Verification* (T3) than in the baseline game. Joint significance tests are shown in Table 4.A.7.



Note: “Percentage cheat” (Model 2) is the left figure and “Log(percentage cheat)” (Model 1) is the right figure.

**Figure 4.A.3:** “Percentage cheat” and “Log(percentage cheat)” as dependent variables

**Table 4.A.6:** “Percentage cheating” per true probability of success

	(Model 1) Log(% cheat)	(Model 2) % cheat
Success Prob.	−0.09*** (0.01)	−0.02*** (0.01)
Treatment T2	−0.47 (0.43)	−0.04 (0.40)
Treatment T3	−0.93** (0.45)	−0.49 (0.38)
Treatment T4	−1.26* (0.73)	−0.39 (0.39)
Treatment T5	−0.74 (0.49)	−0.57 (0.38)
Treatment T2 × Success Prob.	0.01 (0.01)	0.00 (0.01)
Treatment T3 × Success Prob.	0.01 (0.01)	0.01 (0.01)
Treatment T4 × Success Prob.	0.03* (0.02)	0.01 (0.01)
Treatment T5 × Success Prob.	0.02 (0.01)	0.01 (0.01)
Constant	2.52*** (0.18)	1.34*** (0.31)
Observations	133	260
R-squared	0.67	0.34

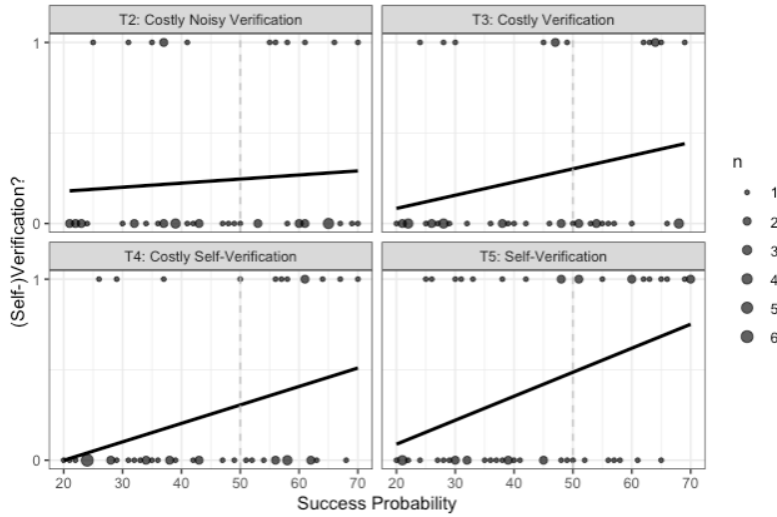
Note: Robust standard errors in parentheses and \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%. Model 2 uses “Percentage cheat” and Model 1 “Log(percentage cheat)”.

**Table 4.A.7:** Wald tests whether true probability of success affects percent-age cheating in each of the treatments

<b>Model 1 Log(% cheat):</b>
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T1}]} = 0$ $F(1, 123) = 297.70, p = 0.0000$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T2}]} = 0$ $F(1, 123) = 62.28, p = 0.0000$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T3}]} = 0$ $F(1, 123) = 45.51, p = 0.0000$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T4}]} = 0$ $F(1, 123) = 12.76, p = 0.0005$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T5}]} = 0$ $F(1, 123) = 40.75, p = 0.0000$
<b>Model 2 % cheat:</b>
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T1}]} = 0$ $F(1, 250) = 15.66, p = 0.0001$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T2}]} = 0$ $F(1, 250) = 22.23, p = 0.0000$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T3}]} = 0$ $F(1, 250) = 14.17, p = 0.0002$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T4}]} = 0$ $F(1, 250) = 13.52, p = 0.0003$
$\text{Ho: } \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T5}]} = 0$ $F(1, 250) = 8.86, p = 0.0032$

#### 4.A.4 Use of (self-)verification per true probability of success

Next, we analyze the use of verification (in T2 and T3) and self-verification (in T4 and T5) per true probability of success. Recall that we originally predicted that in *Costly Noisy Verification* (T2) and *Costly Verification* (T3), there will be no use of the instrument (i.e., investors would not verify), irrespective of the probability. Moreover, in *Costly Self-Verification* (T4) and *Self-Verification* (T5), we predicted a separating pattern; below a true probability of success of 50%, there will be no use of self-verification, and somewhere above it (the exact threshold depends on the firm's believe on the investor's risk aversion)<sup>16</sup>, firms will self-verify their true probability of success. Figure 4.A.4 shows a scatter plot of use of (self-)verification instrument (a binary variable: takes 1 for instrument use and 0 otherwise) per true probability of success. This figure indicates that our theoretical predictions are not supported by the data: Some firms with high probability of success did not send self-verified messages even in the *Self-Verification* (T5) treatment where it is very cheap, whereas some firms with low probability of success sent self-verified messages even in the *Costly Self-Verification* (T4) treatment where it is very costly.



Note: Verification and self-verification instrument use receives the value of 1, while no use of the instruments receives the value of 0.

**Figure 4.A.4:** (Self-)verification use per true probability of success (scatter plot and linear prediction)

<sup>16</sup> If a firm believes that the investor is slightly risk-averse ( $r = 0.26$ ), it should self-verify with a probability of success that is at least equal to or greater than 51% (54%) in T5 (T4), see Equation 3.19.

**Table 4.A.8:** (Self-)verification use depending on true probability of success (linear probability model)

	(Model 1)
Success Prob.	0.00 (0.00)
Treatment T3	-0.20 (0.24)
Treatment T4	-0.34 (0.23)
Treatment T5	-0.31 (0.25)
Treatment T3 $\times$ Success Prob.	0.01 (0.01)
Treatment T4 $\times$ Success Prob.	0.01 (0.01)
Treatment T5 $\times$ Success Prob.	0.01** (0.01)
Constant	0.13 (0.18)
Observations	206
R-squared	0.12

Note: Linear probability model estimating the likelihood of (self-)verification use as a function of a firm's true probability of success. The dependent variable is a binary indicator of verification choice (1 = verification used, 0 = verification not used). The coefficients represent marginal effects, showing how an increase in true probability of success affects verification likelihood. Regression only includes treatment with (self-)verification T2-T5. Robust standard errors in parentheses and \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%.

**Table 4.A.9:** Wald tests whether true probability of success affects (self-)verification

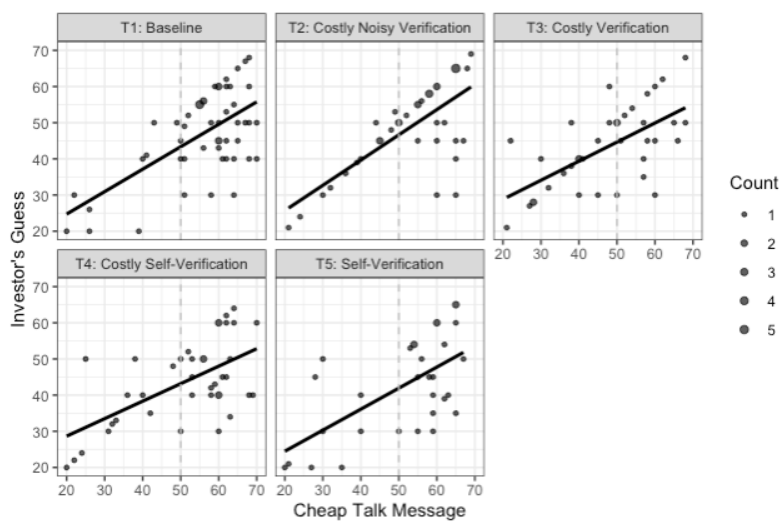
Model 1:
$H_0: \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T2}]} = 0$ $F(1, 250) = 0.35, p = 0.5544$
$H_0: \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T3}]} = 0$ $F(1, 250) = 3.61, p = 0.0588$
$H_0: \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T4}]} = 0$ $F(1, 250) = 6.99, p = 0.0089$
$H_0: \beta_{[\text{Theta}]} + \beta_{[\text{Theta} \times \text{Treatment T5}]} = 0$ $F(1, 250) = 12.33, p = 0.0006$

Note: Wald tests assessing whether true probability of success significantly influences the decision to (self-)verify. The test checks the null hypothesis that the probability of (self-)verification is independent of the firm's actual probability of success. Rejection of the null suggests that firms with higher true success probabilities are more (or less) likely to opt for verification.

The linear probability model shown in Table 4.A.8 indicates that true probability of success does indeed significantly affect purchasing the self-verified message in the T4 and T5 treatments (at  $p \leq 0.05$  significance level). In T3, we find a marginally significant effect of the true probability of success on verification. One plausible explanation is that investors only verify cheap talk messages claiming a high probabilities of success. Since we learned that there is less cheating in T3, firms stating high probability of success are more likely to actually have a high probability of success and are therefore verified by investors more often. Wald tests whether true probability of success affects (self-)verification are shown in Table 4.A.9.

#### 4.A.5 Investors' beliefs per cheap talk message

Our analysis so far indicates that firms cheat less in *Costly Verification* (T3). But how does it affect investors in T3 to make more effective decisions that depend on the true probability of success without actually using the verification instrument? The answer is illustrated in Figure 4.A.5 and in the corresponding regression table (Table 4.A.10). In the absence of instrument use, cheap talk messages significantly affect the investors' guesses about the true probability of success (at  $p \leq 0.01$  for all treatments, in all model estimations). Thus, in *Costly Verification* (T3), firms cheat less, leading to investors' guesses that are closer to the true probability of success than in the other treatments, and consequently to the high effectiveness in this treatment even without actual verification by investors. Joint significance Wald tests whether the cheap talk message affects the investor's guess of true probability of success (excluding (self-)verification instances) are shown in Table 4.A.11 and 4.A.12.



Note: Excluding observations with (self-)verification

**Figure 4.A.5:** Investor's guess and the cheap talk message received (scatter plot and linear prediction)



**Table 4.A.10:** Investor's guess and the cheap talk message received

	(Model 1)	(Model 2)	(Model 3) Log(Guess)
Message	0.62*** (0.09)	-	0.02*** (0.00)
Treatment T2	-0.57 (6.75)	1.26 (4.98)	0.09 (0.21)
Treatment T3	5.94 (7.42)	2.71 (4.91)	0.24 (0.22)
Treatment T4	6.67 (7.80)	2.61 (5.41)	0.20 (0.23)
Treatment T5	0.53 (8.60)	-2.49 (6.09)	-0.02 (0.26)
Treatment T2 $\times$ Message	0.08 (0.14)	-	-0.00 (0.00)
Treatment T3 $\times$ Message	-0.09 (0.15)	-	-0.00 (0.00)
Treatment T4 $\times$ Message	-0.14 (0.15)	-	-0.00 (0.00)
Treatment T5 $\times$ Message	-0.04 (0.16)	-	-0.00 (0.00)
(Message) <sup>2</sup>	-	0.01*** (0.00)	-
Treatment T2 $\times$ (Message) <sup>2</sup>	-	0.00 (0.00)	-
Treatment T3 $\times$ (Message) <sup>2</sup>	-	-0.00 (0.00)	-
Treatment T4 $\times$ (Message) <sup>2</sup>	-	-0.00 (0.00)	-
Treatment T5 $\times$ (Message) <sup>2</sup>	-	0.00 (0.00)	-
Constant	12.32** (5.05)	26.80*** (3.82)	2.88*** (0.16)
Observations	198	198	198
R-squared	0.43	0.41	0.47

Note: Robust standard errors in parentheses and \*, \*\*, and \*\*\* indicate significance levels of at least 10%, 5%, and 1%; excluding observations with (self-)verification use. Model 2 regresses the investor's guess on the squared success probability. Model 3 uses Log(Guess).

**Table 4.A.11:** Wald tests whether the cheap talk message affects the investor's guess of true probability of success, Model 1 and 2

<b>Model 1:</b>
$H_0: \beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T1}]} = 0$ $F(1, 188) = 43.43, p = 0.0000$
$H_0: \beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T2}]} = 0$ $F(1, 188) = 47.03, p = 0.0000$
$H_0: \beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T3}]} = 0$ $F(1, 188) = 20.55, p = 0.0000$
$H_0: \beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T4}]} = 0$ $F(1, 188) = 17.47, p = 0.0000$
$H_0: \beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T5}]} = 0$ $F(1, 188) = 19.16, p = 0.0000$
<b>Model 2:</b>
$H_0: \beta_{[\text{Message}^2]} + \beta_{[\text{Message}^2 \times \text{Treatment T1}]} = 0$ $F(1, 188) = 27.71, p = 0.0000$
$H_0: \beta_{[\text{Message}^2]} + \beta_{[\text{Message}^2 \times \text{Treatment T2}]} = 0$ $F(1, 188) = 31.13, p = 0.0000$
$H_0: \beta_{[\text{Message}^2]} + \beta_{[\text{Message}^2 \times \text{Treatment T3}]} = 0$ $F(1, 188) = 20.77, p = 0.0000$
$H_0: \beta_{[\text{Message}^2]} + \beta_{[\text{Message}^2 \times \text{Treatment T4}]} = 0$ $F(1, 188) = 15.03, p = 0.0001$
$H_0: \beta_{[\text{Message}^2]} + \beta_{[\text{Message}^2 \times \text{Treatment T5}]} = 0$ $F(1, 188) = 17.42, p = 0.0000$

Note: Excluding observations with (self-)verification

**Table 4.A.12:** Wald tests (whether the cheap talk message affects the investor's guess of true probability of success, Model 3)

Model 3:
Ho: $\beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T1}]} = 0$ F(1, 188) = 36.71, p = 0.0000
Ho: $\beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T2}]} = 0$ F(1, 188) = 37.33, p = 0.0000
Ho: $\beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T3}]} = 0$ F(1, 188) = 18.82, p = 0.0000
Ho: $\beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T4}]} = 0$ F(1, 188) = 16.63, p = 0.0001
Ho: $\beta_{[\text{Message}]} + \beta_{[\text{Message} \times \text{Treatment T5}]} = 0$ F(1, 188) = 19.02, p = 0.0000

Note: Excluding observations with (self-)verification

#### 4.A.6 Examples for Investor Efficiency and Societal Efficiency

*Example 1: If  $\theta = 52\%$ , and the investor purchases the certified information for €1 and invests €6.*

- Actual expected investor payoff =  $0.52 \times 12 + 2 - 1 = 7.24$  and maximal expected investor payoff =  $0.52 \times 16 = 8.32$
- Investor Efficiency =  $\frac{7.24}{8.32} = 0.87$
- Actual expected joint payoffs =  $7.24 + 3.12 = 10.36$  and maximal expected joint-payoffs =  $8.32 + 4.16 = 12.48$
- Societal Efficiency =  $\frac{10.36}{12.48} = 0.83$

*Example 2: If  $\theta = 45\%$ , and the investor purchases the certified information for €1 and does not invest.*

- Actual expected investor payoff =  $8 - 1 = 7$  and maximal expected investor payoff = 8
- Investor Efficiency =  $\frac{7}{8} = 0.875$
- Actual expected joint payoffs = 7 and maximal expected joint-payoffs =  $7.2 + 3.6 = 10.8$

- Societal Efficiency =  $\frac{7}{10.8} = 0.64$

*Example 3: If  $\theta = 45\%$ , and the investor invests the full endowment of €8.*

- Actual expected investor payoff = 7.2 and maximal expected investor payoff = 8
- Investor Efficiency =  $\frac{7.2}{8} = 0.9$
- Actual expected joint payoffs =  $7.2 + 3.6 = 10.8$  and maximal expected joint-payoffs =  $7.2 + 3.6 = 10.8$
- Societal Efficiency = 1

#### 4.A.7 Mann-Whitney U-tests

We pairwise compare across treatments with respect to Investor Efficiency and Societal Efficiency. However, the pairwise tests reported below indicate no significant difference in efficiencies between the treatments.

**Table 4.A.13:** Mann-Whitney U-test for stepwise difference in Investor Efficiency

Investor Efficiency
T1 ( $n = 54$ ) vs. T2 ( $n = 51$ ): $z = 1.495, p = 0.1349$
T2 ( $n = 51$ ) vs. T3 ( $n = 51$ ): $z = -1.404, p = 0.1603$
T3 ( $n = 51$ ) vs. T4 ( $n = 51$ ): $z = 0.251, p = 0.8015$
T4 ( $n = 51$ ) vs. T5 ( $n = 53$ ): $z = -1.013, p = 0.3110$

**Table 4.A.14:** Mann-Whitney U-test for stepwise difference in Societal Efficiency

Societal Efficiency
T1 ( $n = 54$ ) vs. T2 ( $n = 51$ ): $z = 0.520, p = 0.6033$
T2 ( $n = 51$ ) vs. T3 ( $n = 51$ ): $z = -0.261, p = 0.7938$
T3 ( $n = 51$ ) vs. T4 ( $n = 51$ ): $z = -0.171, p = 0.8644$
T4 ( $n = 51$ ) vs. T5 ( $n = 53$ ): $z = -1.067, p = 0.2861$

## 4.B Translated instructions (computer screens)

This section includes the translated instructions for *Baseline*, *Costly Noisy Verification* (T2), and *Self-Verification* (T5). The instructions of *Costly Verification* (T3) are identical to T2, but with an accurate verification. The instructions of *Costly Self-Verification* (T4) are identical to T5, but with a higher price for the self-verification. Each subsection represents a single screen presented to the participants.

### 4.B.1 Baseline

#### Introduction

Welcome and thank you for your participation. You have been asked to participate in an online economic experiment. Please activate your webcam and mute your microphone. From now on, you are not allowed to communicate with any of the other participants. If you have any question, please write us privately via the chat function in Zoom. The experiment is conducted under strict compliance to the rules of conduct for economic experiments of the University of Kiel. That means amongst other things that all information is truthfully provided to you. You can end the experiment at any time and without giving a reason. Your data will be treated with strict confidentiality and processed and published anonymously by the researchers involved. Each of you earns 3€ as a show-up fee for participating in this experiment. During the course of the experiment, you can earn additional money depending on your decision and the decision of another participant you are matched with. Your total earning from the experiment is transferred to your bank account in 2-3 days after the experiment. We ask you about your account information at the end of the experiment. The experiment lasts about 15 minutes and ends with a short questionnaire. At the beginning of the experiment, each of you is assigned a role – either Participant A or Participant B. Each of you is matched with one other Participant of the opposite role (i.e., Participant A is matched with Participant B). The identity of you or any of the other participants is never revealed, neither during nor after the experiment. The instructions are identical for Participant A and Participant B.

## The structure of the experiment

Participant A is endowed with 8€ and can invest any amount between 0€ and 8€ in a project owned by Participant B. The project can succeed or fail. It succeeds with a given probability ( $p$ ) and fails with a probability  $(1 - p)$ . For example, let us assume that the success probability is 70%. Ergo, the probability that the project fails is 30%. This is comparable with an urn containing 100 balls: 70 white balls and 30 black balls. There is one random draw from the urn. If the ball that comes up is white, the project succeeds and yields a positive return for both, Participant A and Participant B. If the ball is black, the project fails, and the investment made by Participant A is lost. If Participant A invests nothing, Participant A earns the endowment of 8€ and Participant B earns 0€. If Participant A decides to invest any amount larger than 0€ in the project and if the project succeeds, Participant A retains the amount invested and additionally earns the same amount on top of it. In this case, Participant B also earns the amount invested. If the project fails, the amount invested by Participant A is lost. See the following examples:

- If Participant A invests the full endowment of 8€ and the project succeeds, then Participant A retains the investment of 8€ and earns additional 8€, that means that Participant A earns 16€ in total, whereas Participant B earns 8€. However, if the project fails, Participant A earns 0€ and Participant B earns 0€. That means that Participant A lost the 8€ investment.
- If Participant A invests 7€ (out of the 8€ endowment) and the project succeeds, then Participant A retains the investment of 7€ and earns an additional 7€. Please recall that Participant A has still 1€ that (s)he did not invest in the project. That means that Participant A earns 15€ in total, whereas Participant B earns 7€. However, if the project fails, Participant A earns 1€ (that was not invested), and Participant B earns 0€. That means that Participant A lost the 7€ investment.

## Verification Questions

If Participant A invests 5€ out of the 8€ endowment, and the project succeeds:

1. What is the additional earning for Participant A: [5€]
2. What is the total earning (ignoring show-up fee) for Participant A: [13€]
3. What is the total earning (ignoring show-up fee) for Participant B: [5€]

## Success Probability

It is common knowledge to Participant A and B that the success probability ( $p$ ) is randomly selected from an interval between 20% and 70% with an equal chance for each probability in this interval. This means that the average success probability is 45%. Neither the participants nor the experimental team can influence the selection of the success probability. It is selected randomly by the computer.

For example, if the given success probability is 70%. This is comparable with an urn containing 100 balls: 70 white balls and 30 black balls. There is one random draw from the urn. If the ball that comes up is white, the project succeeds and yields a positive return for both, Participant A and Participant B. If the ball is black, the project fails, and the investment made by Participant A is lost.

## The Sequence of Decisions in the Experiment

In this experiment each Participant (A and B) makes one decision. Participant B learns about the true success probability of the project, while Participant A does not. The first decision is made by Participant B who sends a free message to Participant A. In the free message, Participant B can reveal the true success probability or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$ .* Participant B can choose any number between 20% and 70%. Participant A never learns the true success probability (neither during nor after the experiment). Next, Participant A receives the message from Participant B and makes a decision how much to invest in the project (between 0€ and 8€).

## Instructions Summary

1. The roles of Participant A and Participant B are randomly determined by the computer. Those participants assigned as Participant A are endowed with 8€. The success probability ( $p$ ) of the project is randomly determined by the computer and is uniformly distributed in an interval between 20% and 70%. The average success probability is 45%.
2. Participant B is informed about the true success probability of the project, while Participant A is not.

3. At this point, Participant B sends a free message to Participant A. In the free message, Participant B can reveal the true value or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$* . Participant B can choose any number between 20% and 70%.
4. Participant A receives the free message and decides how much to invest in the project (between 0€ and 8€).
5. At this stage, the decisions are over, and you are asked to fill a short questionnaire.
6. The payoff is determined according to the decisions made by the participants and the success probability of the project that was determined in Step 1. The success of the project is determined by a random lottery made by the computer: With a probability ( $p$ ), the project succeeds. With a probability ( $1 - p$ ), the project fails. The payoffs from the project are determined accordingly.
7. The participants are informed about the result of the lottery, the payoff from the project, and their total earning from the experiments. The experiment ends.

### **Role Determination**

You are randomly determined the role of Player [A/B].

### **Free Message**

You are Participant [B]. Your success probability is [Theta]. At this stage, you send a free message to Participant A. The free message can contain the true value or any other value of your choice that represents the probability of success. You can choose any number between 20% and 70%. You are not obliged to send the true probability. Please note that Participant A never learns the true success probability. Participant A might use your free message to decide whether and how much to invest in your project. Message to be sent: *Participant B claims that the success probability of the project is  $[m]\%$* .

### **Message Delivered**

Participant B sent Participant A the free message: *Participant B claims that the success probability of the project is  $[m]\%$* . This means, if the message is



true, then the project is comparable to an urn containing 100 balls:  $[m]$  white balls and  $(100-[m])$  black balls. There is one random draw from the urn. If the ball that comes up is white, the project succeeds and yields a positive return for both, Participant A and Participant B. If the ball is black, the project fails, and the investment made by Participant A is lost.

### Investment Decision

You are Participant A. From Participant B, you received a free message: *Participant B claims that the success probability of the project is  $[m]\%$* . Please consider that the free message might be true or not. Make a decision how much you want to invest (between 0€ and 8€). Reminder: Your payoff depends on your investment amount and whether the project succeeds or fails. The success of the project is determined by a random lottery made by the computer: With the success probability ( $p$ ), the project succeeds. With a probability  $(1 - p)$ , the project fails. If the project succeeds: You receive 8€ plus the amount invested and Player B receives the amount invested. If the project fails: You lose the amount invested and Player B receives nothing.

#### 4.B.2 Costly Noisy Verification (T2)

In this section, we show all instructions (computer screens) that are different from the baseline treatment.

#### The sequence of decisions in the experiment

In this experiment, Participant A makes two decisions and Participant B makes one decision. Participant B learns about the true probability of success of the project, while Participant A does not. Participant A never learns the actual probability of success (neither during nor after the experiment). The first decision is made by Participant B who sends a free message to Participant A. In the free message, Participant B can reveal the true probability of success or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$* . Participant B can choose any number between 20% and 70%. Next, Participant A receives the free message from Participant B and makes two decisions:

1. First decision: whether or not to purchase a certified message from the experimental team with some information on the success probability (see more details below). The certified message costs Participant A 1€ from

the show-up fee. Please note that Participant B never learns whether Participant A purchased the certified message or not.

2. Second decision (independently of purchasing a certified information or not): how much to invest in the project (between 0€ and 8€).

The certified message costs Participant A 1€ from the show-up fee and contains an interval of 20 percentage points that includes the true probability of success. Please note that the true probability of success does not need to be in the middle of the interval. For example, if the true probability of success is 60%, one possible certified message could be: *It is certified by the experimental team that the true probability of success lies between 45% and 65%*. Another possible certified message could be: *It is certified by the experimental team that the true probability of success lies between 50% and 70%*. And there are numerous other possible certified messages containing the true probability of success. Please note that Participant B never learns whether Participant A purchased the certified information or not.

### Instructions Summary

1. The roles of Participant A and Participant B are randomly determined by the computer. Those participants assigned as Participant A are endowed with 8€. The success probability ( $p$ ) of the project is randomly determined by the computer and is uniformly distributed in an interval between 20% and 70%. The average success probability is 45%.
2. Participant B is informed about the true probability of success of the project, while Participant A is not.
3. At this point, Participant B sends a free message to Participant A. In the free message, Participant B can reveal the true value or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$* . Participant B can choose any number between 20% and 70%.
4. Participant A receives the free message and makes two decisions: First, whether or not to purchase a certified message from the experimental team about the success probability of the project. The certified message costs Participant A 1€ from the show-up fee and contains an interval of 20 percentage points that includes the true probability of success. Please

note that the true probability of success does not need to be in the middle of the interval. Participant B never learns whether Participant A purchased the certified message or not. Second, how much to invest in the project (between 0€ and 8€).

5. At this stage, the decisions are over, and you are asked to fill a short questionnaire.
6. The payoff is determined according to the decisions made by the participants and the success probability of the project that was determined in Step 1. The success of the project is determined by a random lottery made by the computer: With a probability ( $p$ ), the project succeeds. With a probability ( $1 - p$ ), the project fails. The payoffs from the project are determined accordingly.
7. The participants are informed about the result of the random lottery made by the computer, the payoff from the project, and their total earning from the experiments. The experiment ends.

### Free Message

You are Participant B, and your success probability is  $[\Theta]$ . At this stage, you send a free message to Participant A. The free message can contain the true value or any other value of your choice that represents the success probability. You can choose any number between 20% and 70%. You are not obliged to send the true probability of success. Please note that Participant A could purchase a certified message from the experimental team for 1€ that contains an interval of 20 percentage points that includes the true probability of success. Even after receiving the message, Participant A does not learn the true probability of success for certain. Participant A might use your free message to decide whether and how much to invest in your project. Message to be sent: *Participant B claims that the success probability of the project is  $[m]\%$ .*

### Certified Message

You are Participant A. At this stage, you need to decide whether or not to purchase a certified message from the experimental team about the success probability of the project. The certified message costs you 1€ from the show-up fee and contains an interval of 20 percentage points that includes the true

probability of success. For example, if the true probability of success is 60%, one possible certified message could be: *It is certified by the experimental team that the true probability of success lies between 45% and 65%*. Another possible certified message could be: *It is certified by the experimental team that the true probability of success lies between 50% and 70%*. And there are numerous other possible certified messages containing the true probability of success.

It is common knowledge to both Participant A and B that the success probability ( $p$ ) is randomly selected from an interval between 20% and 70% with an equal chance for each probability in this interval.

### Investment Decision

**If verified:** You are Participant A, and you purchased a certified message from the experimental team: *It is certified by the experimental team that the true probability of success lies between  $[l]\%$  and  $[u]\%$* . Additionally, you received a free message from Participant B: *Participant B claims that the success probability of the project is  $[m]\%$* . Please consider that the free message might be true or not. Make a decision how much you want to invest (between 0€ and 8€). Reminder: Your payoff depends on your investment amount and whether the project succeeds or fails. The success of the project is determined by a random lottery made by the computer: With the success probability ( $p$ ), the project succeeds. With a probability  $(1 - p)$ , the project fails. If the project succeeds: You receive 8€ plus the amount invested and Participant B receives the amount invested. If the project fails: You lose the amount invested and Participant B receives nothing.

#### 4.B.3 Self-Verification (T5)

In this section, we show all instructions (computer screens) that are different from the baseline treatment.

#### The sequence of decisions in the experiment

In this experiment each Participant (A and B) makes one decision. Participant B learns about the true probability of success of the project, while Participant A does not. The first decision is made by Participant B: whether to send either a free message or a certified message to Participant A. In the free message, Participant B can reveal the true probability of success or any other value

of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$ .* Participant B can choose any number between 20% and 70%. The certified message costs Participant B 10 Cents from the show-up fee and contains the true probability of success. The message is the following: *It is certified by the experimental team that the success probability of the project is  $[\Theta]$ .* Next, Participant A is informed whether a free message or a certified message was sent. After receiving the message from Participant B, Participant A makes a decision how much to invest in the project (between 0€ and 8€).

### Instructions Summary

1. The roles of Participant A and Participant B are randomly determined by the computer. Those participants assigned as Participant A are endowed with 8€. The success probability ( $p$ ) of the project is randomly determined by the computer and is uniformly distributed in an interval between 20% and 70%. The average success probability is 45%.
2. Participant B is informed about the true probability of success of the project, while Participant A is not.
3. At this point Participant B decides to send either a free message or a certified message to Participant A. In the free message, Participant B can reveal the true value or any other value of his/her choice. The message is the following: *Participant B claims that the success probability of the project is  $m\%$ .* Participant B can choose any number between 20% and 70%. The certified message costs Participant B 0,10€ from the show-up fee and contains the true probability of success. The message is the following: *It is certified by the experimental team that the success probability of the project is  $\Theta$ .*
4. Participant A learns whether Participant B sent a free or a certified message and decides how much to invest in the project (between 0€ and 8€).
5. At this stage, the decisions are over, and you are asked to fill a short questionnaire.
6. The payoff is determined according to the decisions made by the participants and the success probability of the project that was determined in

Step 1. The success of the project is determined by a random lottery made by the computer: With a probability ( $p$ ), the project succeeds. With a probability ( $1 - p$ ), the project fails. The payoffs from the project are determined accordingly.

7. The participants are informed about the result of the random lottery made by the computer, the payoff from the project, and their total earning from the experiments. The experiment ends.

### Message Decision

You are Participant B and the true probability of success is  $[\Theta]$ . At this stage you can decide between either sending a free message to Participant A or buying a certified message containing the true probability of success. The free message can contain the true value or any other value of your choice that represents the probability of success. You can choose any number between 20% and 70%. You are not obliged to send the true probability. Please note that Participant A never learns the true probability of success. The certified message provides Participant A with the true probability of success. The certified message costs 10 Cents which is deducted from your show-up fee. Note that Participant A is informed on the type of message (free message or certified message) you sent. Do you want to purchase the certified message for 0,10€?

### Investment Decision

**If self-verified:** You are Participant A. From Participant B, you received a certified message: *It is certified by the experimental team that the success probability of the project is  $[\Theta]$ .* Participant B had to pay 0,10€ to send you the certified message. Please consider that this is the true probability of success. Make a decision how much you want to invest (between 0€ and 8€). Reminder: Your payoff depends on your investment amount and whether the project succeeds or fails. The success of the project is determined by a random lottery made by the computer: With the success probability ( $p$ ), the project succeeds. With a probability ( $1 - p$ ), the project fails. If the project succeeds: You receive 8€ plus the amount invested and Participant B receives the amount invested. If the project fails: You lose the amount invested and Participant B receives nothing.

## **Chapter 5**

# **Punishment with Communication Sustains Altruistic Cooperation Across Generations**

### **Bibliographic Information**

This chapter shall be cited as: Stenzel, A., Lohse, J., Requate, T., & Waichman, I. (2024). Punishment with Communication Sustains Altruistic Cooperation Across Generations.

### **Competing Interest**

The authors declare no competing interest.

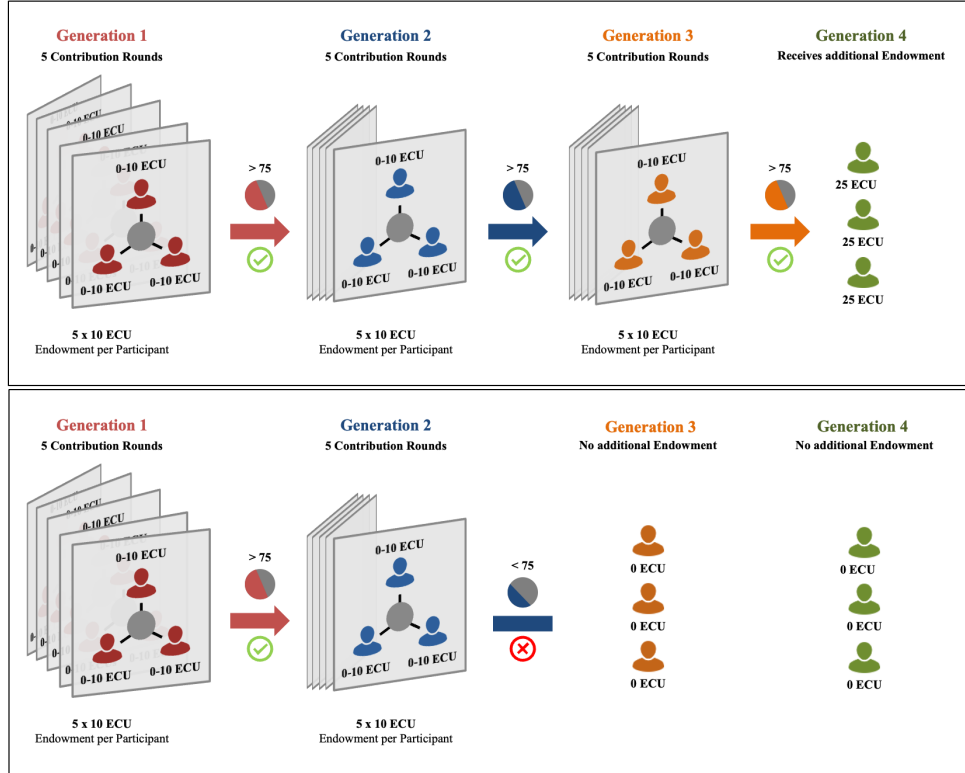
## 5.1 Introduction

Tackling global challenges such as climate change or biodiversity loss requires current generations to give up wealth for the benefit of future generations (Weiss, 1990). While humans are unique in their abilities to cooperate with unrelated strangers, these abilities have evolved early in human history to support mutually beneficial cooperation within small groups through reciprocity and reputation (Axelrod, 1981; Fehr & Gächter, 2000, 2002; Milinski et al., 2002; Fehr & Fischbacher, 2003; Nowak, 2006; Van Lange et al., 2013; Rand & Nowak, 2013). Addressing the challenges above instead requires a different form of cooperation where current generations are willing to bear the full cost of cooperation while future strangers reap all the benefits from it. Whether such altruistic cooperation is common and which institutions can support it remains an open question. Here, we investigate whether decentralized institutions based on costly punishment with and without communication can support altruistic cooperation in an intergenerational cooperation problem.

Human cooperation has been mostly studied in “classical social dilemmas,” where cooperation is mutually beneficial. That is, if all decision makers of a given generation cooperate, their payoffs are higher than if all decision makers defect. However, each decision maker achieves the highest payoff when they unilaterally defect, and all other decision makers cooperate (Dawes, 1980). Despite these individual incentives to free ride, cooperation in classical social dilemmas can arise because of altruistic motives but also out of strategic payoff-maximizing motives, especially when individuals interact repeatedly such that reciprocity and reputation matter (Kreps et al., 1982; Fehr et al., 2002). There are studies where cooperation benefits both current and future generations (Chermak & Krause, 2002; Fischer et al., 2004; Jacquet et al., 2013; Duffy & Laffky, 2016; Sherstyuk et al., 2016). Pure intergenerational cooperation, on the other hand, poses the sternest test of human cooperation because here, cooperation only benefits later generations, and therefore, there is a more limited set of motives to cooperate. To succeed, it requires altruistic cooperation, i.e., cooperating solely for the benefit of others. Unlike individual altruism, altruistic cooperation is special as its success also hinges on collective action.

The Intergenerational Goods Game (IGG; see Figure 5.1) has been designed to investigate this edge case empirically (Hauser et al., 2014; Lohse & Waichman, 2020). Established theories and corresponding empirical evi-





Note: Participants are divided into cohorts of four generations with three members each. In Gen 1, each member is endowed with 10 Experimental Currency Units (ECU) in each of the 5 periods (i.e., a total of 50 ECU per member). In each period, each of the three members simultaneously makes a decision to contribute to a group account. If, at the end of the 5 periods, the members have contributed at least 75 ECU (i.e., half of the total available endowment to the three members), the experiment continues with the next generation, otherwise, the experiment ends. If the contributions of Gen 3 reach the threshold, each member of Gen 4 receives an endowment of 25 ECU, and the game ends without any further decision being made. Upper panel: a case with full intergenerational cooperation (i.e., Gen 4 is reached). Lower panel: a case where contributions of Gen 2 have failed to reach the threshold, and the experiment ends. Members of the subsequent Gens 3 and 4 are endowed with 0 ECU.

**Figure 5.1:** Repeated Intergenerational Goods Game (IGG)

dence from classical social dilemmas can only provide a limited understanding of the institutions that effectively support altruistic cooperation in the IGG. Analyzing the IGG through the lens of game theory shows that successful intergenerational cooperation requires a number of stringent conditions to be met. In particular, it depends on the presence of cooperative types within a generation: Altruists (those who contribute unconditionally) and strong reciprocators (those who contribute conditionally depending on the behavior of other group members) are inclined to cooperate, even though their cooperation only benefits members of the next generation (Gintis, 2000; Fehr & Gächter,

2002). However, altruists and strong reciprocators would still only contribute if they expected sufficient cooperation from their peers to reach the threshold; otherwise, their sacrifice would be in vain, not benefiting later generations. In contrast, payoff-maximizing participants would not sacrifice their resources for strangers in any case.

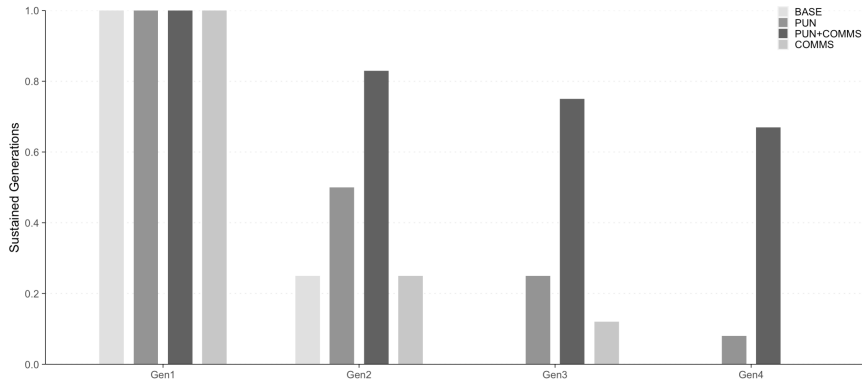
Previous experimental evidence on the IGG suggests that while a majority of individuals are willing to cooperate, their contributions are insufficient to sustain intergenerational cooperation (Hauser et al., 2014; Camerer et al., 2018; Lohse & Waichman, 2020). In this respect, (Hauser et al., 2014; Balmford et al., 2023) instills some optimism by showing that a centralized institution, median voting, achieves virtually full intergenerational cooperation. Yet, in a less centralized setting, where voting is only partially binding, it no longer supports intergenerational cooperation (Hauser et al., 2014). Notably, implementing the (full) voting outcome requires centralized enforcement, often lacking in international settings where votes are non-binding. The insufficiency of voting without enforcement is particularly evident in the context of climate negotiations such as the Paris Agreement, where many attribute the failure to meet the agreed-upon 2°C target to the lack of enforcement (Barrett, 2008, 2016; Roelfsema et al., 2020). Addressing this shortcoming, Lohse and Waichman (2020) show that a decentralized enforcement institution, costly peer punishment (which has been shown to increase cooperation in classical social dilemmas (Fehr & Gächter, 2000, 2002)), does increase intergenerational cooperation in the IGG, albeit still at most at a moderate level.

One design feature that may have reduced the scope for intergenerational cooperation by Lohse and Waichman (2020) is that in each generation, members make only one contribution decision. Allowing for repeated contribution decisions reduces the financial risk of signalling cooperativeness and thus may increase (intergenerational) cooperation (Schelling, 1960). Moreover, repetition can potentially increase the effectiveness of punishment by allowing for strategic motives for punishment. The use of repetition as a deliberate design element has been employed to model elements of climate change negotiations in intra-generational settings (Milinski et al., 2008; Tavoni et al., 2011).

We study the effectiveness of three different decentralized institutions based on costly punishment with and without communication using a repeated IGG game. There are four treatments: BASE, PUN, PUN&COMMS, and COMMS. Figure 5.1 describes the baseline (BASE). PUN is identical to BASE, except that after each of the five contribution decisions, members can reduce their

peers' payoffs at a cost (with a 3:1 fine-to-fee ratio). PUN&COMMS is identical to PUN, except that before each of the five contribution decisions, members of the current generation can freely communicate with each other. We first ran BASE, PUN, and PUN&COMMS. Given our results, we wanted to disentangle whether the effect of PUN&COMMS was due to communication alone or whether it was a joint effect of punishment and communication. Therefore, we added the COMMS treatment, which included only the communication and contribution stages. This treatment is hence akin to international diplomacy where future covenants, as outlined in Ostrom et al. (1992), must be enforced without swords.

In the following, we present our predictions based on a simple theoretical framework introduced in Section 5.B. In BASE, payoff-maximizing members never contribute. In contrast, strong reciprocators or altruists would cooperate if they believed they were in a generation with the potential to reach the threshold. However, given this uncertainty, all generations (even a generation with three altruists) may fail if they do not have sufficient confidence in the cooperativeness of their peers. In PUN, only strong reciprocators are willing to punish. Consequently, payoff-maximizing members will cooperate if they believe that the punishment will reduce their payoff more than their actual contribution. Therefore, the belief of altruists and strong reciprocators that the threshold will be reached increases, and so does their willingness to cooperate. Thus, we expect higher cooperation in PUN than in BASE. In COMMS, participants can reveal their types, and in generations containing only cooperative participants, communication could facilitate cooperation by helping to coordinate on reaching the threshold (Tavoni et al., 2011). Therefore, we expect higher intergenerational cooperation in COMMS than in BASE. In PUN&COMMS, strong reciprocators can reveal their type and induce cooperation even by payoff maximizing members of their generation. Consequently, we expect higher intergenerational cooperation in PUN&COMMS than in COMMS. Whether intergenerational cooperation is higher in PUN&COMMS than in PUN depends on the prevalence (in PUN&COMMS) or beliefs of prevalence (in PUN) of strong reciprocators in the decision making generation.



In BASE, 12 cohorts started in Gen 1, and 3, 0, and 0 cohorts continued to Gen 2, Gen 3, and Gen 4. In PUN, 12 cohorts started in Gen 1, and 6, 3, and 1 cohorts continued to Gen 2, Gen 3, and Gen 4. In PUN&COMMS, 12 cohorts started in Gen 1, and 10, 9, and 8 continued to Gen 2, Gen 3, and Gen 4. In COMMS, 8 cohorts started in Gen 1, and 2 and 1 cohorts reached Gen 2 and 3. Then, one participant in Gen 3 dropped out of the experiment, and we had to stop the experiment for this cohort, although the cohort could have still reached Gen 4. Section 5.F shows that the inclusion/exclusion of this generation has no significant effects on the results.

**Figure 5.2:** Sustained generations across treatments

## 5.2 Results

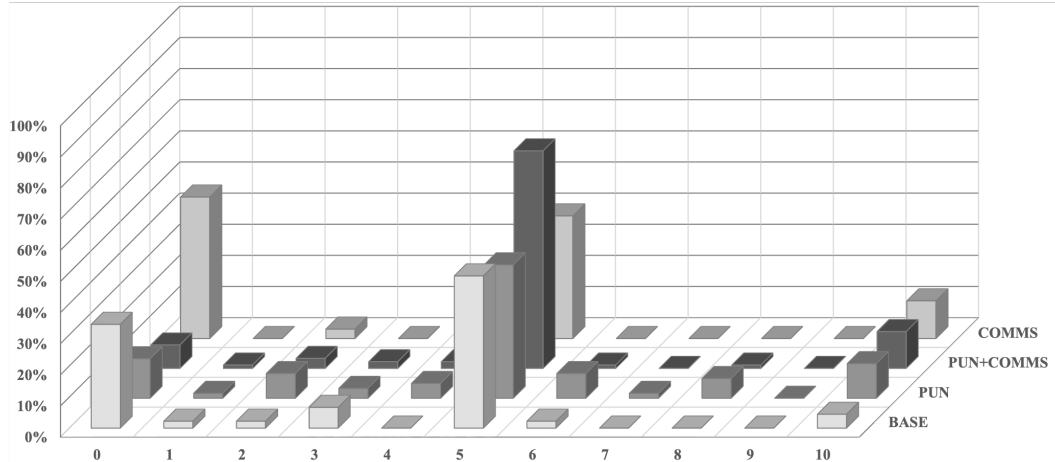
### 5.2.1 Intergenerational cooperation

Figure 5.2 shows that more generations are sustained in PUN&COMMS than in any of the other treatments (at  $p < .001$  or lower in each pairwise comparison using a robust rank order test). In BASE, only 25% of the cohorts reached the second generation, and no cohort reached the third (and fourth) generation. In PUN, 50% reached the second generation, 25% the third, and 8.3% the fourth generation. These figures closely resemble the results of Lohse and Waichman (2020).<sup>1</sup> Punishment increases the number of sustained generations but at most to a moderate level. Thus, even when given a better chance of success through repetition within a generation, punishment alone remains insufficient to maintain cooperation for the sake of future generations. This changes dramatically when punishment is combined with communication: in PUN&COMMS, 83%, 75%, and 66% of the cohorts reached generations 2, 3, and 4, respectively. This is not due to the effect of communication alone, since in COMMS, only 25% and 12.5% reached generations 2 and 3, respectively. Overall, only 8.3%, 12.5%, and 27.7% of possible future generations are reached in BASE, COMMS, and PUN, whereas 75% of future generations

<sup>1</sup> BASE: 29%, 12%, and 0%; PUN: 56%, 28%, and 12% reached generations 2, 3, and 4, respectively.

are reached in PUN&COMMS (Chi-squared test;  $p < 0.001$  for each pairwise comparison between PUN&COMMS and the other treatments, see Table 5.C.2).

As is common in the literature (Hauser et al., 2014; Lohse & Waichman, 2020), we complement our results with a bootstrapping simulation with 40,000 contribution decisions drawn from the observed contribution decisions in the first period (see Section 5.G). This simulation provides further evidence that our results are not driven by random group assignment within a sample of a limited size.



Note: Distribution of individual contributions (ranging from 0 to 10 ECU) in the first period of all generations across four experimental conditions: BASE, PUN, PUN+COMMS, and COMMS. The x-axis represents the contribution levels, while the y-axis shows the percentage of participants choosing each level. The bars indicate the frequency of contributions in each treatment. A total of 45, 63, 93, and 33 observations in BASE, PUN, PUN&COMMS, and COMMS.

**Figure 5.3:** Frequency of individual contributions in the first periods of all generations

### 5.2.2 Individual contributions

To examine how individual contribution behavior explains the treatment differences in sustained generations, we compare contributions between treatments either in the first periods only or in all periods of all generations. The reason for looking at contributions in the first and all periods separately is that in later periods even those who were initially inclined to cooperate would not contribute if they observed that the threshold could not be reached. The first periods' contribution is based on participants' expectations without yet observing the actual contributions. Figure 5.3 shows that in all treatments except for COMMS, the most frequent individual contribution in the first periods

is 5 Experimental Currency Units (ECU, i.e., the focal contribution for symmetric cooperation). In COMMS, the modal contribution is zero (although a considerable share of individuals also contributes 5 ECU). This underlines that communication without punishment may undermine cooperation; it accentuates non-cooperative intentions and deters potential cooperators from spending on a lost cause. These observations are further supported by linear regressions (see Table 5.D.1). Contributions in the first periods do neither differ between BASE and COMMS, nor between PUN and PUN&COMMS (average individual contribution in the first periods in all generations are 3.29 and 3.24 ECU in BASE and COMMS compared to 4.81 and 5.05 ECU in PUN and PUN&COMMS). However, contributions in PUN and PUN&COMMS are higher than in BASE and COMMS. These results also hold for individual contributions across all periods and generations, where average individual contributions are 2.65 and 2.38 ECU in BASE and COMMS compared to 4.21 and 4.63 ECU in PUN and PUN&COMMS. Thus, punishment increases individual contributions relative to the baseline, while communication does not. Without enforcement through punishment, cooperation fails because payoff-maximizing members cannot be persuaded to contribute. Put differently, even with the opportunity to communicate, strong reciprocators cannot sway their group members with words alone; those words need to be sharpened by the presence of swords.

If contributions in PUN&COMMS are not significantly higher than in PUN, how can the former sustain considerably more intergenerational cooperation than the latter? In PUN&COMMS, participants use the chat to coordinate on the focal point contribution of 5 ECU (where equal contributions equalize earnings within a generation) considerably more often than in PUN. The percentage of focal point contribution overall in all periods in BASE, COMM, and PUN are 32%, 35%, and 37% compared to 73% in PUN&COMMS. Using linear probability model regressions (see Table 5.D.2), we find that there are significantly more contributions of exactly 5 ECU in PUN&COMMS than in PUN (in the first periods of all generations,  $F(1, 230)=11.55$ ,  $p=0.0008$ ; in all periods of all generations,  $F(1, 230)=39.25$ ,  $p < 0.0001$ ). A comparison of generations narrowly failing to reach the threshold between PUN and PUN&COMMS demonstrates how communication is effectively used by members to coordinate on reaching the threshold: In PUN, 4 out of 14 generations (29%) contributed at least 90% (68 ECU) of the threshold but failed to reach the threshold. In PUN&COMMS, only one out of 28

generations (4%) did so.

### 5.2.3 Determinants of received punishment

PUN&COMMS achieves high levels of intergenerational cooperation without executing much punishment or expressing explicit threats thereof in the chat. 17% of decisions were punished in PUN, while only 5% were punished in PUN&COMMS (see Table 5.D.4). This suggests that peer punishment facilitates cooperation mainly through deterrence (Fudenberg & Pathak, 2010). Also, punished members do not increase their contribution in the following period. The median period-to-period change in an individual's contribution after receiving punishment is 0 ECU.

### 5.2.4 Relative benefits of different institutions

Finally, we examine how different institutions benefit a generational cohort as a whole. Comparing average payoffs within a cohort across treatments, we find that payoffs are notably higher in PUN&COMMS than in any of the other treatments (using a robust rank order test,  $p < 0.01$  for each pairwise comparison, see Table 5.C.3). This outcome aligns with our prior observations: (i) a significantly higher proportion of generations is sustained in PUN&COMMS, and (ii) fewer resources are expended on punishment due to (iii) improved coordination. Coordination allows groups to reach the threshold without overusing resources or to decide early on to avoid wasting resources by not attempting to meet the threshold when there is insufficient willingness to contribute.

## 5.3 Discussion

At first glance, our central result, that a combination of communication and punishment leads to increased intergenerational cooperation while either institution alone does not, may appear unsurprising to those familiar with studies of communication and punishment in classical social dilemmas. But as we explain below, it is far from obvious why findings from the classical social dilemma literature should extend to altruistic cooperation studied here. In linear public goods games, both punishment (Fehr & Gächter, 2000, 2002; Nikiforakis & Normann, 2008) and communication (Isaac & Walker, 1988) alone have been each shown to deter free riding. It is hence noteworthy that neither of these institutions is sufficient to support cooperation in the IGG. To

better understand this result, it is important to recall the fundamental difference between the types of cooperation required in the IGG and classical social dilemmas and their evolutionary antecedents. In classical social dilemmas, the underlying principle is that *mutual cooperation* leads to the maximization of the group payoff. This core principle of mutually beneficial cooperation also underlies non-linear or otherwise more complex social dilemmas where only punishment combined with communication has been shown to lead to high levels of cooperation (Ostrom et al., 1992; Janssen et al., 2010; Cason & Gangadharan, 2016). However, the IGG models a fundamentally different situation requiring altruistic cooperation rather than mutually beneficial cooperation: The current generation has the most to gain from non-cooperation because the losses from non-cooperation are borne entirely by the next generation, leaving each member of the current generation unaffected by the immediate consequences of defection. Thus, for any given generation, *mutual defection* leads to maximization of the generational payoff. In line with this theoretical consideration, we find that communication alone even reduces altruistic cooperation by enhancing the incidence of zero contributions (pairwise comparison between COMMS and either PUN or PUN&COMMS yields  $p \leq 0.001$ , see regression in Table 5.D.3). By contrast, communication when backed by possible punishment helps coordinate an outcome desired by the group members. The observation that altruistic cooperation can be sustained through a combination of punishment and communication hence not only raises novel questions for evolutionary models of human altruism but also has wider consequences for the situations modeled by the IGG.

Although stylized in nature, the IGG provides some insights into current debates on tackling climate change or biodiversity loss. Centralized coordination and enforcement are politically difficult to establish internationally. We provide the first evidence that a decentralized institution, which is arguably more feasible at the international level, can largely overcome the IGG dilemma. However, this requires that words are “backed up” by swords. Communication alone, a common institution from small groups to high-level climate diplomacy at COP meetings, is not sufficient to overcome this dilemma. Our results provide grounds for cautious optimism about solving intergenerational goods dilemmas through appropriate institutional design.



## 5.4 Methods

### 5.4.1 Design

Figure 5.1 illustrates the structure and procedures of the repeated Intergenerational Goods Game (IGG). A session consists of 1 or 2 independent cohorts of 12 participants. All participants in a session belonged to the same treatment and were given identical instructions. A cohort consists of 12 participants divided into 4 generations of 3 members each. Participants are informed about the order of their generation and that the assignment of participants to the generations is random. Decisions in each generation are made in 5 periods. In the baseline (BASE) condition, each member of Gen 1 is endowed with 10 ECU (Experimental Currency Units) in each period and must decide how many ECU to allocate to the group account (while the rest of the endowment automatically remains in their private account). If the sum of all individual contributions to the group account at the end of the 5 periods is less than 75 ECU, the experiment ends immediately, and the payoff for all participants of later generations is zero. If the sum of all individual contributions to the group account is 75 ECU or more (i.e., at least 50% of the total generation endowment), the game continues and moves to Gen 2. Regardless of the sum of total contributions to the group account, members of Gen 1 earn the corresponding amounts in their private accounts (i.e., the total endowment of 50 ECU minus the sum of their individual contributions to the group account over the 5 periods). The same procedure applies to members of Gens 2 and 3. The experiment ends with the decisions of the members of Gen 3. If they contribute 75 ECU or more, the members of Gen 4 earn 25 ECU each. Otherwise, the members of Gen 4 earn nothing. For Gens 1 and 2, reaching the threshold costs the current generation exactly 75 ECU but provides the subsequent generation(s) with an endowment of 150 ECU. Therefore, there is a gain in total earnings (within a cohort) from reaching the threshold. We choose the parameters such that there is a symmetric focal point for cooperation (each member contributes a total of 25 ECU over the five periods). COMMS is identical to BASE, except that prior to the contribution stage there is a communication stage, where members of the current generation can chat freely with each other for 3 minutes in periods 1 and 2 and 1.5 minutes in periods 3, 4, and 5 (in an anonymous chat window; i.e., identification is not allowed).

PUN is identical to BASE except that after the contribution stage, there is a punishment stage, where members see the individual contributions and

can assign 0, 1, or 2 negative points to their peers. Punishment costs 1 ECU and reduces the punished participant’s payoff by 3 ECU per negative point. PUN&COMMS includes all three stages, communication (as in COMMS), contribution, and punishment (as in PUN).

#### 5.4.2 Procedure

The experiment took place between December 2022 and May 2023. Participants were recruited from the experimental laboratories of two major German universities using hroot (Bock et al., 2014). The study was conducted online using oTree version 5.6 (Chen et al., 2016). A total of 528 participants (55% identify as female) participated in four treatments (144, 144, 144, and 96 participants in the BASE, PUN, PUN&COMMS, and COMMS treatments, respectively). We initially randomized between BASE, PUN, and PUN&COMMS. But given our results, we could not disentangle whether the effect of PUN&COMMS was due to communication alone or whether it was a joint effect of punishment and communication. We therefore added COMMS, which includes only the communication and contribution stages (without punishment). In total, we have 12 independent cohorts in each of the first 3 treatments and 8 cohorts of COMMS. All participants received 5€ as a show-up fee regardless of their payoff in the game. We ran 7 sessions per treatment with one or two cohorts (12 or 24 participants). Each session consisted of three parts: First, after logging in to the experiment, participants received instructions, including quiz questions. The instructions were the same for all participants. Second, participants were randomly assigned to a generation, and the experiment began. Finally, the participants completed a short post-experiment questionnaire that elicited demographic data and information about their motivation and perceived difficulty with the experiment. Participants reported a mean difficulty level of 2.2 (“Easy”) with a standard deviation of 0.92 (ranked on a scale from 1=Very Easy to 5=Very Difficult). Earnings were converted to euro (at an ECU/euro exchange rate of 2.5:1) and transferred to the participants’ bank accounts within two working days. We provide a detailed design and procedure protocol in Section 5.A.

## Bibliography

- Axelrod, R. (1981). The evolution of cooperation. *Science*, 212(4489), 1390–1396.
- Balmford, B., Marino, M., & Hauser, O. (2023). *Voting in Intergenerational Goods Averts ‘Tipping Point’ Ambiguity Crises* (Tech. Rep.). Exeter: Available at SSRN: <https://ssrn.com/abstract=4410361>.
- Barrett, S. (2008). Climate treaties and the imperative of enforcement. *Oxford Review of Economic Policy*, 24(2), 239–258.
- Barrett, S. (2016). Coordination vs. voluntarism and enforcement in sustaining international environmental cooperation. *Proceedings of the National Academy of Sciences*, 113(51), 14515–14522.
- Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71, 117–120.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., ... others (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644.
- Cason, T. N., & Gangadharan, L. (2016). Swords without covenants do not lead to self-governance. *Journal of Theoretical Politics*, 28(1), 44–73.
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- Chermak, J. M., & Krause, K. (2002). Individual response, information, and intergenerational common pool problems. *Journal of Environmental Economics and Management*, 43(1), 47–70.
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, 31(1), 169–193.
- de Matos Fernandes, C. A., Bakker, D. M., & Dijkstra, J. (2022). Assessing the test-retest reliability of the social value orientation slider measure. *Judgment and Decision Making*, 17(1), 31–49.

- Duffy, J., & Laffky, J. (2016). Birth, death and public good provision. *Experimental Economics*, 19, 317–341.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785–791.
- Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13, 1–25.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980–994.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137–140.
- Feltoich, N. (2005). Critical values for the robust rank-order test. *Communications in Statistics - Simulation and Computation*, 34(3), 525–547.
- Fischer, M.-E., Irlenbusch, B., & Sadrieh, A. (2004). An intergenerational common pool resource experiment. *Journal of Environmental Economics and Management*, 48(2), 811–836.
- Fudenberg, D., & Pathak, P. A. (2010). Unobserved punishment supports cooperation. *Journal of Public Economics*, 94(1-2), 78–86.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206(2), 169–179.
- Hauser, O. P., Rand, D. G., Peysakhovich, A., & Nowak, M. A. (2014). Cooperating with the future. *Nature*, 511(7508), 220–223.
- Isaac, R. M., & Walker, J. M. (1988). Communication and free-riding behavior: The voluntary contribution mechanism. *Economic Inquiry*, 26(4), 585–608.
- Jacquet, J., Hagel, K., Hauert, C., Marotzke, J., Röhl, T., & Milinski, M. (2013). Intra-and intergenerational discounting in the climate game. *Nature Climate Change*, 3(12), 1025–1028.
- Janssen, M. A., Holahan, R., Lee, A., & Ostrom, E. (2010). Lab experiments for the study of social-ecological systems. *Science*, 328(5978), 613–617.
- Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners’ dilemma. *Journal of Economic Theory*, 27(2), 245–252.
- Lohse, J., & Waichman, I. (2020). The effects of contemporaneous peer punishment on cooperation with the future. *Nature Communications*, 11(1), 1815.
- Milinski, M., Semmann, D., & Krambeck, H.-J. (2002). Reputation helps

- solve the ‘tragedy of the commons’. *Nature*, 415(6870), 424–426.
- Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A., & Marotzke, J. (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences*, 105(7), 2291–2294.
- Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771–781.
- Nikiforakis, N., & Normann, H.-T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11, 358–369.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805), 1560–1563.
- Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86(2), 404–417.
- Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, 17(8), 413–425.
- Roelfsema, M., van Soest, H. L., Harmsen, M., van Vuuren, D. P., Bertram, C., den Elzen, M., ... others (2020). Taking stock of national climate policies to evaluate implementation of the Paris agreement. *Nature Communications*, 11(1), 2096.
- Schelling, T. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Sherstyuk, K., Tarui, N., Ravago, M.-L. V., & Saijo, T. (2016). Intergenerational games with dynamic externalities and climate change experiments. *Journal of the Association of Environmental and Resource Economists*, 3(2), 247–281.
- Tavoni, A., Dannenberg, A., Kallis, G., & Löschel, A. (2011). Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences*, 108(29), 11825–11829.
- Van Lange, P. A., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, 120(2), 125–141.
- Weiss, E. B. (1990). Our rights and obligations to future generations for the environment. *American Journal of International Law*, 84(1), 198–207.

## Appendix

### 5.A Experimental procedure and treatments

#### 5.A.1 Terminology and experimental design

First, we describe our terminology. Treatment is the experimental condition (BASE, PUN, PUN&COMMS, or COMMS), and Session refers to a specific date and time. Each session consists of either one or two cohorts. One cohort consists of 12 participants divided into 4 generations of 3 members. In each generation, members make decisions in 5 consecutive periods.

Each participant receives a €5.00 show-up fee and can earn additional money depending on whether his/her generation is reached and his/her decisions (and in PUN and PUN&COMMS, also on the punishment decisions of his/her group members). In BASE, in each period, each member of Gen 1 is endowed with 10 ECU (Experimental Currency Units; the exchange rate in the experiment is 2.5:1, i.e., 2.5 ECU are exchanged for €1.00) per period. I.e., each member receives 50 ECU in total (10 ECU in each period for five periods). In every period, each member decides how many ECU to allocate to a group account (while the rest of the endowment remains (automatically) in their private account). If the sum of all individual contributions to the group account at the end of the five periods is less than 75 ECU, the experiment ends instantly, and the payoff for all participants of later generations is zero. If the sum of all individual contributions to the group account amounts to 75 ECU or more (i.e., at least 50% of the total generation endowment), the game continues to Gen 2. Independent from the decisions of later generations, members of Gen 1 earn the respective amounts in their private accounts (i.e., the total endowment of 50 ECU minus the sum of their individual contributions to the group account in the 5 periods). The same procedure applies to members of Gens 2 and 3. The experiment ends with the decisions of Gen 3. If they contribute 75 ECU or more, then each member of Gen 4 earns 25 ECU. Otherwise, members of Gen 4 earn nothing. For Gens 1 and 2, reaching the

threshold costs 75 ECU to the current generation but provides an endowment of 150 ECU to the subsequent generation(s). Hence, there is a gain in total earnings (across multiple generations) from reaching the threshold. We choose the parameters such that there is a symmetric focal point for cooperation (each member contributes a total of 25 ECU in the five periods). Note that in this game, any form of payoff-maximizing behavior results in contributing nothing. In any case, each participant earned the €5.00 show-up fee.

### 5.A.2 Procedure and treatments

We conducted 28 sessions between December 7, 2022, and May 8, 2023. Of these, 12 sessions consisted of one, and 16 sessions consisted of two cohorts. We treat each cohort as an independent observation. After running 22 sessions with randomized order of the three treatments BASE, PUN, and PUN&COMMS, we found that intergenerational cooperation (sustained generations) is considerably higher in PUN&COMMS than in BASE and PUN. At this point, we could not disentangle whether the high occurrence of intergenerational cooperation is because of the combined effects of punishment and communication or whether it is due to communication alone. Therefore, we added COMMS as a separate treatment, and Sessions #20, 23-28 included only the COMMS treatment. We conducted the experiment online using oTree version 5.6 (Chen et al., 2016). The 528 participants were recruited using hroot (Bock et al., 2014) from two large German universities' experimental economics participant pools. Participants belonged to various faculties (thereof 116 economic students). 340 participants were from one university and 188 from the other university. In total, 17 and 7 sessions included students from either the first or the second university, and 4 sessions included students from both universities.

Treatment	Sessions	Cohorts	Participants	Avg. Payout (€)	Avg. Duration (Mins)
BASE	7	12	144	9.59	27.86
PUN	7	12	144	9.44	34.86
PUN&COMMS	7	12	144	13.25	59.14
COMMS	7	8	96	10.24	35.00
Total	28	44	528	10.67	39.21

Note: "Sessions" = experimental date and time (includes 1 or 2 cohorts), "Cohort" = independent sequence of 12 participants divided into four generations. Average Payout includes a show-up fee of €5.00.

**Table 5.A.1:** Treatments

Each session consisted of three stages: First, participants received instruc-

tions (including quiz questions) upon logging in to the experiment. The instructions were identical for all participants. Second, participants were randomly assigned to a specific generation, and the experiment started.<sup>2</sup> After making their contribution decision in each period, each member saw a screen with his/her contributions per period, the contributions per period of each of the two other members in their generation, the sum of the periods' contributions, and the total contribution up to this point.

In COMMS and PUN&COMMS, members of the current generation could communicate via a chat window before making contribution decisions in each period. They had 3 minutes for communication in the initial two periods and 1.5 minutes in the final three. Participants were not allowed to identify themselves or other group members in the chat. In PUN and PUN&COMMS, after making the contribution decisions and learning about the individual contribution decisions, members of the current generation could assign 0, 1, or at most 2 negative points to each of the two other members in their generation after learning about the individual contribution decisions. Each negative point costs 1 ECU to the punisher and reduces the punishee's payoff by 3 ECU. After the punishment decision, members of this current generation received feedback on the total of punishment points received.

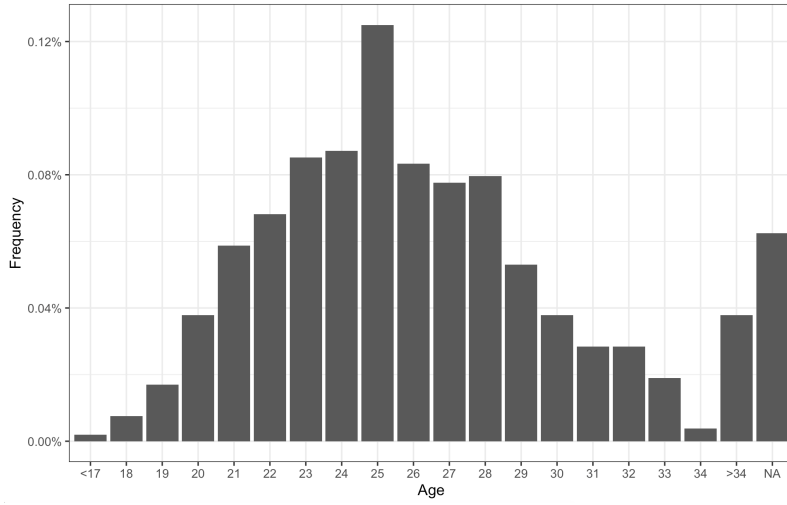
We had 28 dropouts (i.e., participants not finishing the experiment). In compliance with ethics procedures, participants were informed that they could leave the experiment at any time without giving any reason. As we conducted the experiment online, some participants might have experienced technical difficulties and dropped out. Others may have been frustrated waiting and left. In all but one case, the drop-outs were from generations that were not reached. Thus, these dropouts do not affect the performance in the experiment. However, in one cohort of COMMS in Gen 3, one of the active participants dropped out during the experiment. This particular cohort reached Gen 3, but because of the dropout, it could not continue (we know that the cohort reached Gen 3 and could have reached Gen 4). As shown in Section 5.F, this does not affect the statistical analysis (as all other 7 cohorts in COMMS failed in Gen 1).

After all participants in the session had made their decisions, they were asked to complete a short post-experimental questionnaire that provided demographic data and information about their motivation and perceived difficulty of the experiment. Out of 528 participants in the experiment, 33 did not

---

<sup>2</sup> "Per period" means in the current period as well as previous periods. Each member of an active generation participates in five periods.





**Figure 5.A.1:** Age distribution of participants

fill in the questionnaire (28 drop-outs and an additional 5 who had technical problems with the questionnaire and contacted the experimenter). Regarding gender, of the 495 participants who filled in the questionnaire, 55.35%, 43.23%, and 1.41% self-identified their gender as female, male, and diverse, respectively. In particular, of the 495 participants, one indicated an age of “less than 17,” and 20 indicated an age of “larger than 34.” The remaining 474 participants self-reported an average age of 25.44 years (std. dev = 3.44). For more details, see the distribution in Figure 5.A.1. As to the subject of study, out of the 495 observations, 23.43% reported studying economics/business. Finally, participants indicated, on average, a difficulty level of 2.2 (“Easy”), with a standard deviation of 0.92 (ranked on a scale from 1=Very Easy to 5=Very Difficult). The average duration of one session was 39 minutes. The average payout was €10.67 per participant (minimum was €5.00 and maximum €25.00). We transferred the earnings to the participants’ bank accounts within 2 days.

## 5.B Theoretical framework

### 5.B.1 Introduction

This section introduces a theoretical framework for cooperation in the repeated Intergenerational Goods Game (IGG). This framework provides reasoning on how much cooperation to expect in the baseline treatment and how the different decentralized institutions we test in our experiment (punishment, com-

munication, and both institutions combined) could facilitate intergenerational cooperation in the IGG. Given the experimental nature of our paper, we only present a stylized illustration of our framework and describe possible outcomes without a formal mathematical model. In the repeated IGG, participants are randomly divided into cohorts of four generations with three members each. Members of the first generation are endowed with 10 Experimental Currency Units (ECU) in each of the 5 periods (i.e., a total of 50 ECU per member). Each member makes simultaneous contribution decisions to a group account for each period in the BASE treatment. After each period of contribution decisions, members see the individual contributions of all group (generation) members and the group account balance. If, at the end of the 5 periods, members contribute at least 75 ECU (i.e., half of the total available endowment to the three members), the experiment continues with the next generation; otherwise, the experiment ends. If contributions of Gen 3 reach the threshold, each member of Gen 4 receives an endowment of 25 ECU, and the game ends without any further decision being made. See Figure 5.1 in the main part of the paper for a detailed description.

We conducted three additional treatments: PUN is identical to BASE, except that members can reduce their peers' payoffs after each of the five contribution decisions at a cost (with a 3:1 fine-to-fee ratio). PUN&COMMS is identical to PUN, except that members of the current generation can freely communicate with each other before each of the five contribution decisions (i.e., every period includes one communication stage before and one punishment stage after the contribution stage). COMMS includes only the communication and contribution decisions.

To predict outcomes in the repeated IGG, we make assumptions about the presence of certain cooperative and non-cooperative types in a representative generation. First, from the extensive literature on human cooperation, we assume three archetypical players: *Payoff Maximizers* (abbreviated with P), *Strong Reciprocators* (R), and *Altruists* (A).<sup>3</sup> Second, we make assumptions about the different types. *Payoff Maximizers* optimize their own payoff independent of the consequence for other participants. *Strong Reciprocators* are willing to cooperate conditionally and to punish unfair behavior depending on

<sup>3</sup> We assume only three types, but the classification could be extended. The type of *Payoff Maximizers* includes myopic payoff maximizers as well as "reciprocal altruists" who are willing to pay a cost for cooperation for a future gain (Gintis, 2000) and "competitive" types (Murphy et al., 2011) who aim at maximizing the difference between their own and other members' payoffs (in the IGG, to maximize the difference, one needs to maximize his/her own payoff).

the other group members' behavior. The game parameters of our experiment were chosen to facilitate one focal cooperation norm: the contribution of 5 ECU per period (i.e., it requires participants to contribute half of their endowment (thus, equal contributions), leading to equal earnings in their group).<sup>4</sup> In this regard, we define cooperation as contributing at least 5 ECU. We assume that *Strong Reciprocators* are willing to contribute 5 ECU if they expect others to do so (and to costly punish others who contribute less than the focal cooperation norm). *Altruists'* willingness to cooperate is unconditional to the other participants' behavior. While we assume that *Strong Reciprocators* are only willing to contribute to the focal contribution of 5 ECU, *Altruists* are even willing to spend more than the 5 ECU to compensate for a non-cooperative participant in the group. Note that despite their inclination to cooperate, neither a *Strong Reciprocator* nor an *Altruist* is willing to contribute to a lost cause (if they expect the threshold not to be reached). Thus, their willingness to cooperate depends on their belief in the type-composition of their generation. The distinction between these two "cooperative" types becomes obvious under punishment.

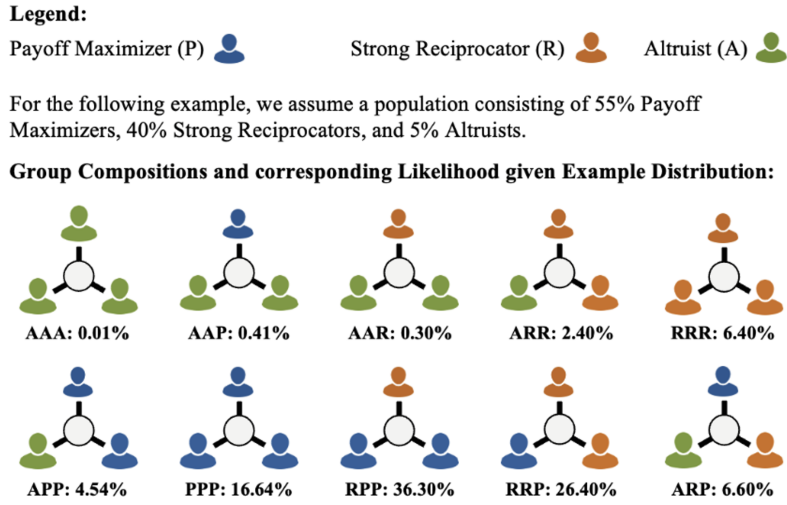
While *Strong Reciprocators* are willing to punish what they perceive as unfair behavior, *Altruists* never punish (see below). Given the three types, ten different group compositions are possible. Each composition has a different likelihood of being randomly drawn, which depends on a third assumption: the distribution of types in the population (yet, this is based on past literature). Let us illustrate the substantial differences in the likelihood of specific group compositions to materialize using an example: Assume a population with 55% *Payoff Maximizers*, 40% *Strong Reciprocators*, and 5% *Altruists*. Under this assumption, the group consisting only of *Altruists* AAA has the lowest (0.01%) and the group RPP the highest (36.30%) likelihood to materialize, see Figure 5.B.1. Each group composition has a different chance of reaching the threshold. In the following, we explain our theoretical reasoning and show the effects of punishment and communication.

## 5.B.2 Theoretical framework

In BASE, participants cannot communicate and thus cannot identify the types in their group/generation.<sup>5</sup> Their contributions depend on their own type and

<sup>4</sup> Moreover, if this focal contribution norm is followed across all generations, the earnings of all 12 cohort members are equal.

<sup>5</sup> After the first period, participants can see the contribution decisions of their group members. However, concluding a participant's type from his/her contribution is diffi-



**Figure 5.B.1:** Possible group compositions and example likelihoods

beliefs about the other types in their group. In COMMS and PUN&COMMS, we assume that the group composition becomes common knowledge among all participants in the communication stage (i.e., no participant miscommunicates their own type in the chat). As this simplifies the analysis, we start with treatments with communication (i.e., COMMS and then PUN&COMMS). We proceed by illustrating the considerable impact of punishment using a numerical example before turning to BASE and PUN, where the type composition of the group is unknown.

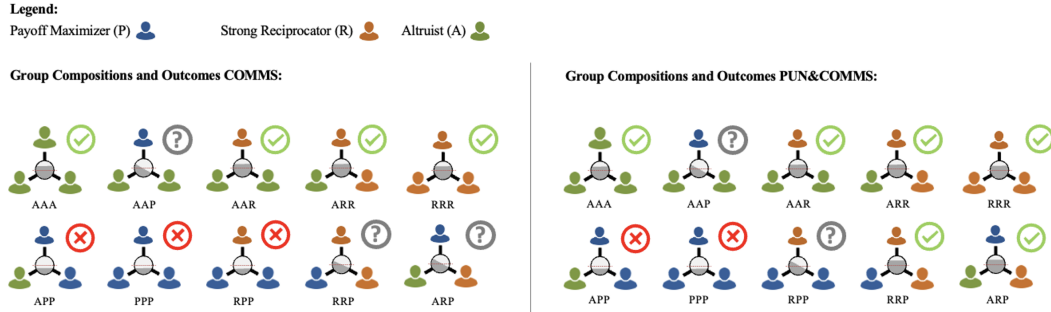
### Communication (COMMS)

Given our definition of *Altruists* and *Strong Reciprocators*, groups of only those types reach the threshold (see Figure 5.B.2). *Payoff Maximizers* never contribute as any contribution reduces their payoff. Despite their willingness to cooperate, neither a *Strong Reciprocator* nor an *Altruist* is willing to contribute to a lost cause. As no one can reach the threshold unilaterally, all groups with at least two *Payoff Maximizers* cannot sustain the generation. Most interesting are groups containing one *Payoff Maximiser*. Whether the generation can be sustained depends on the *Strong Reciprocator's* or *Altruists's* willingness to compensate for the missing contributions. The group AAP can only be sustained if the *Altruists* are willing to spend at least 75% of their endowment. In ARP, if *Strong Reciprocators* are only willing to contribute 50% of their endowment, an *Altruist* must spend 100% to sustain the group. Therefore,

---

cult. For example, a contribution of 0 ECU might indicate a *Payoff Maximiser* or an *Altruist* who doubts that the group can reach the threshold.

while the outcome of some groups may appear obvious, the outcome of others depends on the strength of the expression of individual preferences of the cooperative types.



Note: A green check sign indicates that we expect the group to be sustained, and a failure sign (red X) indicates that we expect the group to be not sustained. Groups with an unclear outcome are marked with a gray question mark.

**Figure 5.B.2:** Group composition and possible outcomes

## Punishment and communication (PUN&COMMS)

The advantage of punishment is that it allows *Strong Reciprocators* to induce *Payoff Maximizers* to cooperate and thereby increases intergenerational cooperation. Importantly, neither *Payoff Maximizers* (as it reduces their payoff) nor *Altruists* (as punishment harms others) ever punish. *Strong Reciprocators* punish unfair behavior, even when this behavior cannot be justified in terms of present or future self-interest (Gintis, 2000; Fehr & Gächter, 2002). Therefore, in the groups RRP and ARP, *Payoff Maximizers* fear punishment and cooperate to maximize their payoff (see Figure 5.B.2). The success of the group RPP depends on the *Strong Reciprocator's* willingness to punish unfair behavior. Theoretically, the *Strong Reciprocator* has enough endowment to punish two *Payoff Maximizers*. If the participant can make a credible threat to the *Payoff Maximizers*, even the group RPP can be sustained in PUN&COMMS. Note that in COMMS (without punishment), this group composition is set up for failure without a chance to reach the threshold. This means punishment allows the groups RRP and ARP and gives RPP a chance to sustain the generation – all groups with a relatively high likelihood to be drawn if the type composition within a population contains a sufficient number of *Strong Reciprocators* as in our example (see Figure 5.B.1).

## A numerical example for COMMS and PUN&COMMS

To illustrate the considerable impact of punishment on the likelihood of a generation to be sustained, we want to extend the example given in Figure 5.B.2 by two more assumptions: First, we assume that *Altruists* are willing to contribute up to 75% of their endowment to sustain a generation (i.e., the group AAP is sustained in COMMS and PUN&COMMS) whereas *Strong Reciprocators* will not contribute more than 50% (i.e., the group RRP and ARP is not sustained in COMMS). Second, a *Strong Reciprocator* is willing to punish up to two *Payoff Maximizers* (i.e., the group RPP can be sustained in PUN&COMMS). Given those assumptions and the distribution example in Figure 5.B.1, Figure 5.B.3 shows that allowing for punishment increases the likelihood of sustaining a generation by 728%.<sup>6</sup>

	Type	Distribution			
A	Altruists	5%			
R	Strong Reciprocator	40%			
P	Payoff Maximizers	55%			

	Sustained Generation ( 1 = yes, 0 = no )		Group Composition		Likelihood for Sustained Generation	
	COMMS	PUN&COMMS		Likelihood	COMMS	PUN&COMMS
AAA	1	1		0,01%	0,0%	0,0%
AAR	1	1		0,30%	0,3%	0,3%
AAP	1	1		0,41%	0,4%	0,4%
ARR	1	1		2,40%	2,4%	2,4%
ARP	0	1		6,60%	0,0%	6,6%
APP	0	0		4,54%	0,0%	0,0%
RRR	1	1		6,40%	6,4%	6,4%
RRP	0	1		26,40%	0,0%	26,4%
RPP	0	1		36,30%	0,0%	36,3%
PPP	0	0		16,64%	0,0%	0,0%
Weighted Likelihood to sustain a Generation:					9,5%	78,8%

Note: The left column shows the possible group compositions. In the columns COMMS and PUN&COMMS, 1 denotes groups sustaining the generation, and 0 represents failed groups. The column likelihood shows the likelihood that this specific group is drawn (same as in Figure 5.B.1). Finally, we calculate the total weighted likelihood to sustain a generation in COMMS and PUN&COMMS.

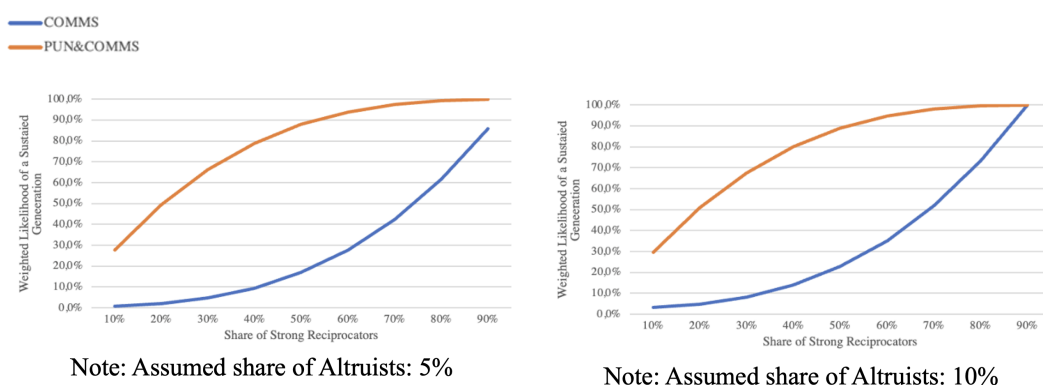
**Figure 5.B.3:** Numerical example for the likelihood of a sustained generation in COMMS and PUN&COMMS

Figure 5.B.4 provides an intuition for how the likelihood of a sustained generation depends on the distribution of types in COMMS and PUN&COMMS.<sup>7</sup>

<sup>6</sup> Even if we assume that *Altruists* are only willing to contribute 50% of their endowment and that a *Strong Reciprocator* is willing to punish a maximum of one *Payoff Maximizer*, punishment in PUN&COMMS would still increase intergenerational cooperation by 362% compared with COMMS.

<sup>7</sup> While we do not know the types of our subjects, Murphy et al. (2011) present a slider measure of social value orientation. In their test (Table 1, p.775), they observe that 61% of the classified participants are prosocial (in our framework such classification is

In line with the literature (e.g., Murphy et al., 2011; de Matos Fernandes et al., 2022), *Altruists* are assumed to make out only a minor share of the population. Therefore, we consider two scenarios: In the left figure, we assume a population share of 5% *Altruists*, and in the right figure, we assume a population share of 10% *Altruists*. The x-axis represents the variable share of *Strong Reciprocators*. For the three types, the remaining share of *Payoff Maximizers* is 100% minus the other two types (*Altruists* and *Strong Reciprocators*). Both subfigures, with different shares of *Altruists*, show very similar patterns. This illustrates that the presence of *Altruists* has only a marginal impact on the success of a generation which reflects this type's unwillingness to use the punishment option. In COMMS, intergenerational cooperation can only be sustained if the population has a very high share of reciprocal types. In contrast, in PUN&COMMS, a moderate population share of 20% *Strong Reciprocators* (i.e., 75% *Payoff Maximisers*) is already sufficient to increase the weighted likelihood of sustaining a generation to 50%.



**Figure 5.B.4:** Impact of change in the distribution of types on the likelihood of a sustained generation in COMMS and PUN&COMMS

### Baseline (BASE)

In BASE, participants cannot reveal their types, i.e., BASE equals COMMS without perfect knowledge about the participants' types. A *Payoff Maximizer* never contributes. Contrary, a *Strong Reciprocator* and an *Altruist* contribute if they believe they are in a group with a chance to reach the threshold. I.e., their strategy is based on their beliefs about the group composition. If *Altruists*

---

in line with both *Altruist* and *Strong Reciprocators* – however, if *Altruists* are required to contribute more than 5 ECU than it is only consistent with *Strong Reciprocators*) and that 39% are classified as individualistic (36%) and competitive (3%), both these definitions map into our *Payoff Maximizers* types.

and *Strong Reciprocators* doubt that their group can reach the threshold, they refrain from contributing, and all groups (even AAA) have the potential to fail. Thus, we expect lower chances of reaching the threshold in BASE compared to COMMS.

### **Punishment (PUN)**

In PUN, group composition is unknown to the participants. Payoff-maximizing participants condition their contribution decision on their belief about the presence of *Strong Reciprocators* in their group and their willingness to punish for not contributing. If they believe there is a sufficiently high likelihood that they receive punishment for deviating from the focal contribution of 5 ECU and this punishment is more costly than contributing, cooperation becomes a payoff-maximizing strategy. *Altruists* and *Strong Reciprocators* form a belief about the presence of *Payoff Maximizers* and whether they would cooperate under the threat of punishment (i.e., this is a second- or higher-degree belief, a belief about the belief of the payoff-maximizing participants that a *Strong Reciprocator* will punish them). Thus, depending on the individual beliefs of the participants, even the group PPP can be sustained. Contrary, if they do not believe that the effect of punishment is sufficient, all groups may fail (as in BASE).

### **5.B.3 Predictions**

Based on these theoretical considerations, we expect BASE to be the lower bound of intergenerational cooperation because *Payoff Maximizers* will never cooperate, and it may be difficult for cooperative types to reveal their type. In PUN, even *Payoff Maximizers* would cooperate (if they expect punishment to reduce their payoff more than the focal norm of cooperation (5 ECU)). Thus, we expect higher cooperation in PUN than in BASE. In COMMS, the different types reveal themselves, leading to intergenerational cooperation in groups consisting of cooperative types (such as AAA, RRR, AAR, and RRA). In PUN&COMMS, *Strong Reciprocators* can reveal their type, and cooperation can become a payoff-maximizing strategy for *Payoff Maximizers*, leading to a considerably higher number of group compositions reaching the threshold



compared to COMMS. To sum it up, our predictions are:

$$BASE \leq COMMS \leq PUN\&COMMS \text{ and} \\ BASE \leq PUN.$$

Whether intergenerational cooperation is higher in PUN&COMMS than in PUN depends on the participants' beliefs about the group composition and their risk aversion. A risk-loving *Payoff Maximiser* might "test" the group by not contributing in the first period to see whether a *Strong Reciprocator* is in the group (punishing the uncooperative behavior). After receiving the punishment, it is still uncertain whether the group can compensate for the shortage of the first period without the possibility of communication. Assuming such behavior, intergenerational cooperation in PUN would be lower than in PUN&COMMS. On the other hand, if risk-averse *Payoff Maximisers* contribute in any case as they fear punishment, even a generation consisting of only *Payoff Maximisers* would reach the threshold in PUN. In this case, intergenerational cooperation in PUN would be higher than in PUN&COMMS.

## 5.C Pairwise tests

In this section, we provide details on the pairwise tests reported in the article.

**Table 5.C.1:** Occurrence of sustained generations in the different treatments

	Observations	U-statistic
BASE vs. COMMS:	12 vs. 8	-0.106
BASE vs. PUN:	12 vs. 12	-1.299
BASE vs. PUN&COMMS:	12 vs. 12	-3.774***
COMMS vs. PUN:	8 vs. 12	-0.974
COMMS vs. PUN&COMMS:	8 vs. 12	-3.321**
PUN vs. PUN&COMMS:	12 vs. 12	-2.743**

Note: Two-sample Fligner-Policello robust rank order tests. A variable between 1-4 indicating the number of sustained generations. \*, \*\*, and \*\*\* denote 10%, 5%, and 1% significance levels. Two-sided small-sample using Feltovich (2005).

**Table 5.C.2:** Success rate in sustaining generations across all possible generations

	Observations	$\chi^2(1)$ -statistic	P-value (two-sided)
BASE vs. COMMS:	36 vs. 24	0.2778	0.598
BASE vs. PUN:	36 vs. 36	4.5997**	0.032
BASE vs. PUN&COMMS:	36 vs. 36	32.9143***	0.000
COMMS vs. PUN:	24 vs. 36	1.9804	0.159
COMMS vs. PUN&COMMS:	24 vs. 36	22.5000***	0.000
PUN vs. PUN&COMMS:	36 vs. 36	16.06800***	0.000

Note: Pairwise Chi-squared tests. We assume that any cohort has 3 generations. For example, if a cohort reached Gen 4, the success is (1,1,1), whereas if a cohort failed in Gen 1, the data are recorded as (0,0,0). Finally, \*, \*\*, and \*\*\* denote 10%, 5%, and 1% significance levels.

**Table 5.C.3:** Total payoffs per cohort

	Observations	U-statistic
BASE vs. COMMS:	12 vs. 8	-0.909
BASE vs. PUN:	12 vs. 12	0.534
BASE vs. PUN&COMMS:	12 vs. 12	-3.714***
COMMS vs. PUN:	8 vs. 12	0.991
COMMS vs. PUN&COMMS:	8 vs. 12	-2.796**
PUN vs. PUN&COMMS:	12 vs. 12	-4.466***

Note: Two-sample Fligner-Policello robust rank order tests. \*, \*\*, and \*\*\* denote 10%, 5%, and 1% significance levels. Two-sided small-sample using Feltovich (2005).

## 5.D Regressions and their tests

In this section, we provide details on the regressions and coefficient tests reported in the article.

**Table 5.D.1:** Linear regressions: Individual contributions

	Model 1 Period 1: Gen 1	Model 2 Period 1: All gen	Model 3 All Periods: Gen 1	Model 4 All Periods: All gen
COMMS	−0.92 (0.89)	−0.05 (0.74)	−0.57 (0.69)	−0.28 (0.56)
PUN	1.06 (0.65)	1.52*** (0.55)	1.44*** (0.51)	1.55*** (0.42)
PUN&COMMS	1.72*** (0.58)	1.76*** (0.48)	1.97*** (0.43)	1.97*** (0.37)
Constant	3.42*** (0.48)	3.29*** (0.41)	2.73*** (0.38)	2.65*** (0.33)
Observations	132	234	132	234
R-squared	0.11	0.08	0.19	0.18
Pairwise comparisons of coefficients tests				
BASE vs. COMMS	t-statistics = −1.03, p = 0.307	t-statistics = −0.06, p = 0.950	t-statistics = −0.82, p = 0.415	t-statistics = −0.50, p = 0.619
BASE vs. PUN	t-statistics = 1.62, p = 0.108	t-statistics = 2.78, p = 0.006	t-statistics = 2.83, p = 0.005	t-statistics = 3.69, p = 0.000
BASE vs. PUN&COMMS	t-statistics = 2.95, p = 0.004	t-statistics = 3.67, p = 0.000	t-statistics = 4.56, p = 0.000	t-statistics = 5.38, p = 0.000
COMMS vs. PUN	F(1, 128) = 5.13, p = 0.0253	F(1, 230) = 4.91, p = 0.0277	F(1, 128) = 8.98, p = 0.0033	F(1, 230) = 12.53, p = 0.0005
COMMS vs. PUN&COMMS	F(1, 128) = 10.31, p = 0.0017	F(1, 230) = 7.57, p = 0.0064	F(1, 128) = 17.08, p = 0.0001	F(1, 230) = 22.59, p = 0.0000
PUN vs. PUN&COMMS	F(1, 128) = 1.48, p = 0.2256	F(1, 230) = 0.32, p = 0.5738	F(1, 128) = 1.82, p = 0.1799	F(1, 230) = 1.92, p = 0.1674

Note: The dependent variable is the contribution in the initial period in the first generation (first column), contribution in the initial period in all generations (second column), averaged over the five periods in the first generation (third column), and averaged over the five periods in all generations (fourth column). Robust stand errors are given in parentheses. BASE is the default treatment. \*, \*\*, and \*\*\* denote 10%, 5%, and 1% significance levels (p-values are two-sided).

**Table 5.D.2:** Linear probability model regressions: Individual contributions of 5 ECU (focal point cooperation)

	Model 1 Period 1: Gen 1	Model 2 Period 1: All gen	Model 3 All Periods: Gen 1	Model 4 All Periods: All gen
COMM	-0.250** (0.126)	-0.095 (0.116)	0.194* (0.100)	0.176** (0.086)
PUN	-0.056 (0.121)	-0.060 (0.099)	-0.000 (0.056)	0.029 (0.053)
PUN&COMMS	0.222* (0.115)	0.210** (0.090)	0.500*** (0.094)	0.535*** (0.064)
Constant	0.500*** (0.086)	0.489*** (0.076)	0.056 (0.039)	0.067* (0.038)
Observations	132	234	132	234
R-squared	0.103	0.068	0.259	0.272
Pairwise comparisons of coefficients tests				
BASE vs. COMMS	t-statistics = -1.99, p = 0.049	t-statistics = -0.82, p = 0.415	t-statistics = 1.94, p = 0.055	t-statistics = 2.05, p = 0.042
BASE vs. PUN	t-statistics = -0.46, p = 0.646	t-statistics = -0.61, p = 0.543	t-statistics = -0.00, p = 1.000	t-statistics = 0.53, p = 0.594
BASE vs. PUN&COMMS	t-statistics = 1.93, p = 0.056	t-statistics = 2.33, p = 0.021	t-statistics = 5.33, p = 0.000	t-statistics = 8.38, p = 0.000
COMMS vs. PUN	F(1, 128) = 2.40, p = 0.1239	F(1, 230) = 0.10, p = 0.7492	F(1, 128) = 3.76, p = 0.0546	F(1, 230) = 2.96, p = 0.0869
COMMS vs. PUN&COMMS	F(1, 128) = 15.48, p = 0.0001	F(1, 230) = 9.30, p = 0.0026	F(1, 128) = 5.92, p = 0.0163	F(1, 230) = 15.13, p = 0.0001
PUN vs. PUN&COMMS	F(1, 128) = 5.87, p = 0.0168	F(1, 230) = 11.55, p = 0.0008	F(1, 128) = 28.41, p = 0.0000	F(1, 230) = 63.53, p = 0.0000

Note: The dependent variable is a binary indicator that takes the value of 1 if the participant contributes 5 ECU in the initial period (first and second columns) or if they contribute 5 ECU in each of the five periods (third and fourth columns); otherwise, it is set to 0. Robust standard errors are given in parentheses. BASE is the default treatment. \*, \*\*, and \*\*\* denote 10%, 5%, and 1% significance levels (p-values are two-sided).

**Table 5.D.3:** Linear probability model regressions: Individual contributions of 0 ECU

	Model 1 Period 1: Gen 1	Model 2 Period 1: All gen	Model 3 All Periods: Gen 1	Model 4 All Periods: All gen
COMM	0.292** (0.131)	0.121 (0.115)	0.361*** (0.127)	0.202* (0.109)
PUN	-0.222** (0.097)	-0.206** (0.084)	-0.194** (0.077)	-0.159** (0.071)
PUN&COMM	-0.306*** (0.086)	-0.258*** (0.077)	-0.222*** (0.071)	-0.158** (0.068)
Constant	0.333*** (0.081)	0.333*** (0.072)	0.222*** (0.071)	0.222*** (0.063)
Observations	132	234	132	234
R-squared	0.25	0.13	0.29	0.23
Pairwise comparisons of coefficients tests				
BASE vs. COMMS	t-statistics = 2.23, p = 0.028	t-statistics = 1.06, p = 0.292	t-statistics = 2.85, p = 0.005	t-statistics = 1.85, p = 0.065
BASE vs. PUN	t-statistics = -2.29, p = 0.024	t-statistics = -2.47, p = 0.014	t-statistics = -2.54, p = 0.012	t-statistics = -2.25, p = 0.026
BASE vs. PUN&COMMS	t-statistics = -3.57, p = 0.001	t-statistics = -3.35, p = 0.001	t-statistics = -3.12, p = 0.002	t-statistics = -2.31, p = 0.022
COMMS vs. PUN	F(1, 128) = 19.51, p = 0.0000	F(1, 230) = 10.94, p = 0.0011	F(1, 128) = 26.11, p = 0.0000	F(1, 230) = 14.71, p = 0.0002
COMMS vs. PUN&COMMS	F(1, 128) = 31.21, p = 0.0000	F(1, 230) = 16.43, p = 0.0001	F(1, 128) = 30.86, p = 0.0000	F(1, 230) = 15.16, p = 0.0001
PUN vs. PUN&COMMS	F(1, 128) = 1.88, p = 0.1729	F(1, 230) = 1.04, p = 0.3098	F(1, 128) = 0.97, p = 0.3260	F(1, 230) = 0.00, p = 0.9798

Note: The dependent variable is a binary indicator that takes the value of 1 if the participant contributes 0 ECU in the initial period (first and second columns) or if they contribute 0 ECU in each of the five periods (third and fourth columns); otherwise, it is set to 0. Robust standard errors are given in parentheses. COMMS is the default treatment. \*, \*\*, and \*\*\* denote 10%, 5%, and 1% significance levels (p-values are two-sided).

**Table 5.D.4:** Linear probability model regressions: Punishment received

	Model 1	Model 2	Model 3
PUN&COMMS	-0.12*** (0.02)	-0.27*** (0.06)	-0.22*** (0.06)
Deviation from the others' average contribution	0.03*** (0.01)	0.05*** (0.01)	0.04*** (0.01)
PUN&COMMS x Deviation from the others' average contribution	-	-0.06*** (0.01)	-0.06*** (0.01)
Average others' contribution	-0.01* (0.01)	-0.03*** (0.01)	0.01 (0.02)
PUN&COMMS x Average others' contribution	-	0.04*** (0.01)	0.04* (0.02)
Contribute at least 15 in the period	-	-	-0.21*** (0.06)
PUN&COMMS x Contribute at least 15 in the period	-	-	-0.03 (0.09)
Constant	0.23*** (0.05)	0.31*** (0.06)	0.23*** (0.06)
Period FE	Yes	Yes	Yes
Generation FE	Yes	Yes	Yes
Observations	780	780	780
R-squared	0.12	0.16	0.20

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10%, respectively.

### Interpretation of regression models in Table 5.D.4: Why are participants punished?

Using a linear probability model in Table 5.D.4, we find that participants were punished significantly less in PUN&COMMS than in PUN (17% of decisions were punished in PUN, whereas only 5% in PUN&COMMS). Our results remain similar using a random effect panel model instead (shown in Table 5.D.6), although coefficient sizes are slightly smaller and turn insignificant in Model 3. Contributing less than the others' average increases punishment only in PUN but not in PUN&COMMS. Yet, the higher the average other members' contribution, the higher the punishment in PUN&COMMS but not in PUN. Finally, there is significantly less punishment in both treatments in periods with generations contributing 15 ECU or more (i.e., the focal contribution for cooperation). These latter three results are robust with respect to the estimation method (attained in both Tables 5.D.4 and 5.D.6). In PUN, only 24% of all members contribute exactly as many ECU as the average of the other two members compared to 78% in PUN&COMMS. Thus, it appears that PUN&COMMS attains high intergenerational cooperation without executing much punishment.

**Table 5.D.5:** Pairwise tests of Table 5.D.4

<b>Model 2:</b>	
Deviation from others' average contribution + PUN&COMMS x Deviation from others' = 0	
F(1, 768) = 0.08, p=0.7710	
Average others' contribution + PUN&COMMS x Average others' contribution	
F(1, 768) = 1.67, p=0.1972	
<b>Model 3:</b>	
Deviation from others' average contribution + PUN&COMMS x Deviation from others' = 0	
F(1, 766) = 2.53, p=0.1123	
Average others' contribution + PUN&COMMS x Average others' contribution	
F(1, 766) = 13.66, P= 0.0002	
Contribute at least 15 in the period + PUN&COMMS x Contribute at least 15 in the period	
F(1, 766) = 13.35, p=0.0003	

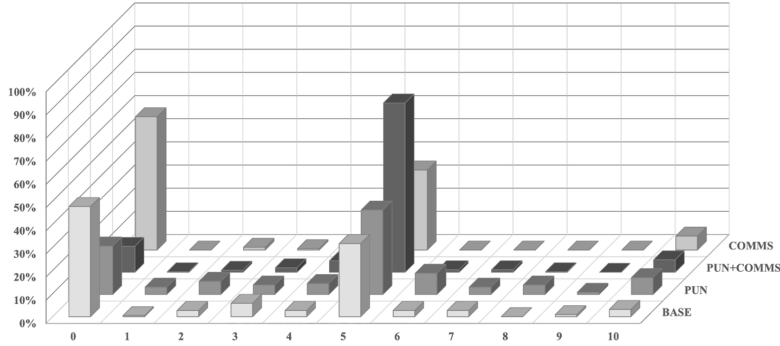
**Table 5.D.6:** Random-effects panel data linear probability model regressions:  
Punishment received

	(1) Model 1	(2) Model 2	(3) Model 3
PUN & COMS	-0.12** (0.06)	-0.21* (0.12)	-0.18 (0.12)
Deviation from the others' average contribution	0.02*** (0.00)	0.04*** (0.00)	0.03*** (0.00)
PUN&COMMS x Deviation from the others' average contribution	-	-0.04***	-0.05***
Average others' contribution	-0.01*** (0.00)	(0.01) -0.02* (0.01)	(0.01) 0.01 (0.02)
PUN&COMMS x Average other's contribution		0.02 (0.02)	0.03** (0.01)
Contribute at least 15 in the period			-0.15*** (0.05)
PUN&COMMS x Contribute at least 15 in the period			-0.03 (0.02)
Constant	0.23*** (0.06)	0.28*** (0.08)	0.22** (0.09)
Period FE	YES	YES	YES
Generation FE	YES	YES	YES
Observations (5 Rounds)	780	780	780
Participants	156	156	156
R-squared	0.12	0.15	0.21

**Table 5.D.7:** Pairwise tests of Table 5.D.6

<b>Model 2:</b>
Deviation from others' average contribution + PUN&COMMS x Deviation from others' = 0 Chi-squared(1) = 0.05 , p=0.8266
Average others' contribution + PUN&COMMS x Average others' contribution Chi-squared(1) = 0.01 , p=0.9116
<b>Model 3:</b>
Deviation from others' average contribution + PUN&COMMS x Deviation from others' = 0 Chi-squared(1) = 1.15 , p=0.2833
Average others' contribution + PUN&COMMS x Average others' contribution Chi-squared(1) = 7.45 , p=0.0064
Contribute at least 15 in the period + PUN&COMMS x Contribute at least 15 in the period Chi-squared(1) = 6.59, p=0.0102

## 5.E Individual contributions in all periods



Note: Distribution of individual contributions (ranging from 0 to 10 ECU) in all periods of all generations across four experimental conditions: BASE, PUN, PUN+COMMS, and COMMS. The x-axis represents the contribution levels, while the y-axis shows the percentage of participants choosing each level. The bars indicate the frequency of contributions in each treatment. A total of 225, 315, 465, and 165 observations in BASE, PUN, PUN&COMMS, and COMMS.

**Figure 5.E.1:** Frequency of individual contributions in all periods of all generations

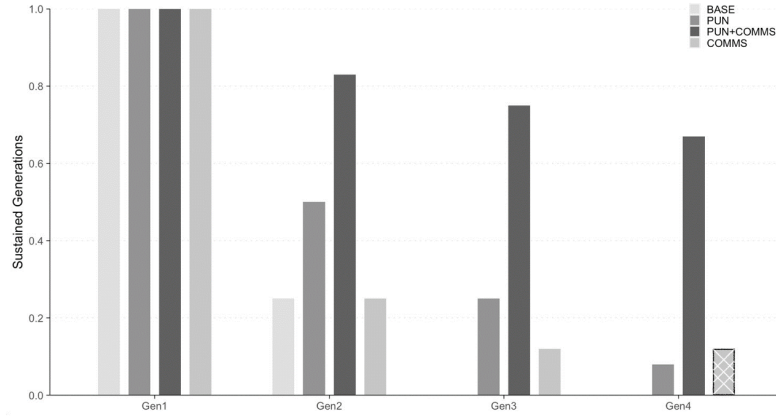
## 5.F Robustness test regarding the lost observation

In COMMS, there was one cohort that reached Gen 3, but at this point, a member of this generation left the experiment. Thus, we had to stop the experiment at Gen 3 without knowing what would have happened if the participant had stayed. In a sense, leaving the experiment may be interpreted as a form of free riding, so it may be reasonable to assume that the cohort reached Gen 3. However, we need to study whether this assumption affects the results.

When inspecting the sequence of maximal generations reached, Table 5.F.1 shows that when considering the lost observation in the COMMS treatment, the average maximal generation slightly increases from 1.375 to 1.5.

However, the addition of one successful generation in one COMMS cohort does not change the results of the robust rank order tests or the Chi-squared tests. Still, there are significant differences at less than 1% level between PUN&COMMS and any of the other treatments (see Tables 5.F.1 and 5.F.2).





Note: In BASE, 12 cohorts started in Gen 1, and 3, 0, and 0 cohorts continued to Gen 2, Gen 3, and Gen 4. In PUN, 12 cohorts started in Gen 1, and 6, 3, and 1 cohorts continued to Gen 2, Gen 3, and Gen 4. In PUN&COMMS, 12 cohorts started in Gen 1, and 10, 9, and 8 continued to Gen 2, Gen 3, and Gen 4. In COMMS, 8 cohorts started in Gen 1, and 2 and 1 cohorts reached Gen 2 and 3. Then, one participant in Gen 3 dropped off the experiment, and we had to end the experiment for this cohort while the cohort could have reached Gen 4. This is marked in the figure.

**Figure 5.F.1:** Sustained generations across treatments considering the lost observation

**Table 5.F.1:** Descriptive statistics for sustained generations

	Average	Median	SD	Min	Max	Observations
BASE	1.25	1	0.45	1	2	12
PUN	1.833	1.5	1.03	1	4	12
PUN&COMMS	3.25	4	1.22	1	4	12
COMMS	1.375	1	0.74	1	3	8
COMMS*	1.5	1	1.07	1	4	8

Note: A variable between 1-4 indicating the number of sustained generations. COMMS does not consider that the particular cohort reached Gen 4, while COMMS\* considers it.

## 5.G Bootstrap simulations

Because our experiment requires an independent cohort of 12 participants, we have a limited number of cohorts (12 in BASE, PUN, and PUN&COMMS and 8 in COMMS). The number of realizations (affected by the random assignment of participants into generations) is small compared to all possible combinations. Assigning participants to generations and placing them at a specific position of the generational sequence can further impact the number of sustained generations. Imagine that we have a cohort of 12 individuals in the BASE treatment, where 6 of them are *Payoff Maximizers*, and 6 are *Strong Reciprocators* who would like to cooperate. If the *Strong Reciprocators*

**Table 5.F.2:** Occurrence of sustained generations in the different treatments, based on Table 5.C.1

	Observations	U - statistic
BASE vs. COMMS:	12 vs. 8	-0.106
BASE vs. COMMS*:	12 vs. 8	-0.106
BASE vs. PUN:	12 vs. 12	-1.299*
BASE vs. PUN&COMMS:	12 vs. 12	-3.774***
COMMS vs. PUN:	8 vs. 12	-0.974
COMMS vs. PUN&COMMS:	8 vs. 12	-3.321***
COMMS* vs. PUN:	8 vs. 12	-0.828
COMMS* vs. PUN&COMMS:	8 vs. 12	-2.585**
PUN vs. PUN&COMMS:	12 vs. 12	-2.743**

Note: Two-sample Fligner-Policello robust rank order test. A variable between 1-4 indicating the number of sustained generations. \*, \*\*, and \*\*\* denote 10%, 5%, and 1% significance levels. Two-sided small-sample using Feltovich (2005). COMMS does not consider that the particular cohort reached Gen 4, while COMMS\* considers it.

**Table 5.F.3:** Success rate in sustaining generations across all possible generations, based on Table 5.C.2

	Observations	$\chi^2(1)$ -statistic	P-value (two-sided)
BASE vs. COMMS:	36 vs. 24	0.2778	0.598
BASE vs. COMMS*:	36 vs. 24	0.9704	0.325
BASE vs. PUN:	36 vs. 36	4.5997**	0.032
BASE vs. PUN&COMMS:	36 vs. 36	32.9143***	0.000
COMMS vs. PUN:	24 vs. 36	1.9804	0.159
COMMS vs. PUN&COMMS:	24 vs. 36	22.5000***	0.000
COMMS* vs. PUN:	24 vs. 36	0.9938	0.319
COMMS* vs. PUN&COMMS:	24 vs. 36	19.6218***	0.000
PUN vs. PUN&COMMS:	36 vs. 36	16.0680***	0.000

Note: Pairwise Chi-squared tests. We assume that any cohort has 3 generations. For example, if a cohort reached Gen 4, the success is (1,1,1), whereas if a cohort failed in Gen 1, the data are recorded as (0,0,0). Finally, \*, \*\*, and \*\*\* denote 10%, 5%, and 1% significance levels(two sided). COMMS does not consider that the particular cohort reached Gen 4, while COMMS\* considers it.

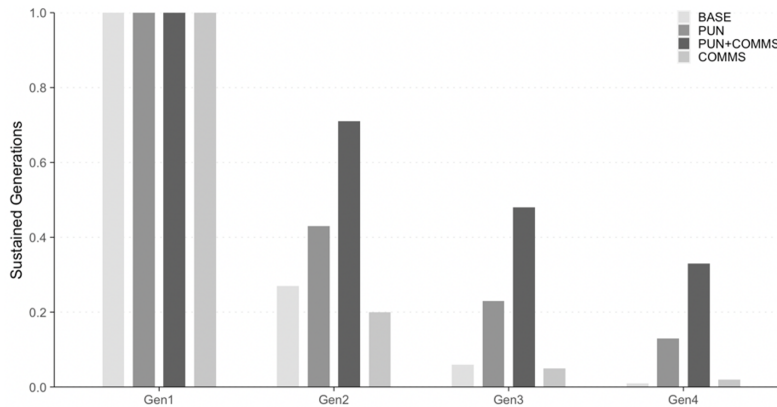
are (randomly) assigned to the first two generations, then the cohort reaches the third generation. However, if at least one *Payoff Maximizer* is assigned to the first generation, the group fails to reach the second generation.

Therefore, to test the robustness of our main result, we conducted a bootstrap simulation similar to Hauser et al. (2014) and Lohse and Waichman (2020). We simulated 10,000 randomly generated generational sequences for each treatment (in total, 40,000). In each cohort, we randomly sample (with replacement) a series of participants' contribution decisions for each generation, calculate the total amount contributed to the group account, and observe whether the generation was sustained. In particular, our procedure works as

follows (for each treatment):

1. For the first generation, we consider only participants of Gen 1 in the respective treatment. We randomly sample (with replacement) three of their contribution decisions in the first period. Based on the contribution decisions, we calculate the total amount contributed to the group account by multiplying the sum of the three contribution decisions by 5 (due to the 5 periods) and observe whether the sum is above the threshold.
2. For the second and third generations, we consider all generations in the respective treatment. We randomly sample (with replacement) three of their contribution decisions in the first period. Based on the contribution decisions, we calculate the total amount contributed to the group account by multiplying the sum of the three contribution decisions by 5 (due to the 5 periods) and learn whether the sum is above the threshold.
3. Finally, we determine the point at which the generational sequence would break down (if any) and start again at step 1.

Using this procedure, we simulated 10,000 groups of 4 generations. For BASE, COMMS, and PUN, the simulated mean value of sustained generations is very close to the results of our experiment (BASE: 1.4 in the simulation vs. 1.3 in the experiment, COMMS: 1.3 vs. 1.4, and PUN: 1.8 vs. 1.8). Even though, PUN&COMMS still reaches considerably higher sustained generations than the other treatments, the mean of the simulation is considerably lower than the mean in our experiment (2.6 vs. 3.3).



**Figure 5.G.1:** Simulation result for sustained generations across treatments

However, while the simulation approach may a priori be particularly useful for BASE and PUN, it is not the case under communication. In COMMS and

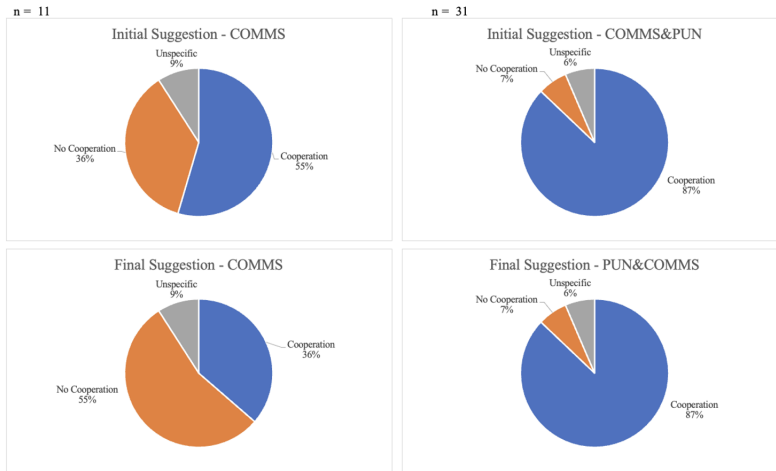
PUN&COMMS, there is a communication stage before the first period, and thus, randomly sampling 3 participants from different groups will likely not reflect the effect of communication, which allows all three group members to draw inferences on group composition (whether to cooperate or to defect fully). Indeed, in COMMS and PUN&COMMS, first-period members coordinated well to contribute 25 each (when they wanted to reach the threshold) or 0 (when they did not want to reach the threshold). Note that another weakness of the simulation for our game is that we can only use the first-period (i.e., independent) contribution decisions for our simulation. If most cooperation or coordination happens at later periods, it would not be reflected in the simulation. Yet, the simulation results in Figure 5.G.1 demonstrate that first-period contributions provide a rather good prediction of success.

## 5.H Analysis of communication

We analyzed the chat messages to learn how communication was used with and without punishment. Four annotators received annotation guidelines (see Section 5.H.2) to evaluate the content of the chat messages. The annotators received one log file for each generation with all chat messages written in all five periods (see an example below). For each logfile, they answered two questions about whether there was a suggestion “to cooperate,” “not to cooperate,” or “no suggestion,” and three questions regarding punishment (only for PUN&COMMS). In 71% of all cases, all four annotators came to the same conclusion. In 26% of the cases, three annotators came to the same conclusion. In this case, we applied a majority rule (i.e., we selected the interpretation of the three annotators). Finally, in 3% of the cases, only two annotators came to the same conclusion. We marked those chats as “unspecified.”

### 5.H.1 Results of the communication analysis

The annotators’ conclusions about the content of the chat messages are as follows: In PUN&COMMS, the first suggestion in the chat was more often to cooperate than in COMMS. This could indicate that in the presence of punishment, *Payoff Maximizers* are more hesitant to reveal their type early in the chat. In COMMS, the annotators recognized that some payoff-maximizing participants used the chat to reveal their strategy. Without punishment, *Strong Reciprocators* and *Altruists* do not have any leverage against them (and in fact should not cooperate if they believe that the threshold is not



Note: The top panel answers question 1: What was the initial suggestion in the chat? (“Initial Suggestion”). The bottom panel answers question 2: What was the final suggestion at the end of the chat? (“Final Suggestion”).

**Figure 5.H.1:** Communication analysis

met), leading to an increase in uncooperative messages throughout the chat (see the final suggestions in the bottom panel of Figure 5.H.1). Contrary, the annotators did not find a change in the generation’s intent to cooperate in PUN&COMMS. This might be because payoff-maximizing participants were more hesitant (as they might have feared punishment by *Strong Reciprocators*) to influence the discussions.

Next, in PUN&COMMS, the annotators found only very few occasions (6%) in which participants explicitly threatened other participants with punishment. Instead, on many more occasions (19%), participants agreed not to punish each other if they stuck to what the group agreed on. In PUN&COMMS, the participants rather used the chat to agree on the focal point of cooperation (5 ECU per participant per period; also see Figure 5.3 in the main article). It seems that the presence of punishment leads to more cooperative behavior as communication makes the presence of *Strong Reciprocators* sufficiently prominent to force *Payoff Maximizers* into cooperation.

## 5.H.2 Annotation guidelines

The following guidelines were given to the annotators:

### Guidelines

Your task is to read through the chat messages and answer a few questions for every group. The answer to each question might not always be clear, and even

with the same annotation guideline, annotators can come to different answers. To counterbalance these inconsistencies, the chat messages are annotated by 4 annotators. After the annotation process, in case of disagreement, we use the majority opinion to determine the outcome.

## **The experiment**

A cohort of 12 participants is divided into four generations (labeled as Gen 1, Gen 2, Gen 3, and Gen 4). Each generation consists of three participants. Members of Gen 1 always actively participate. However, the members of subsequent generations only participate depending on the decisions of the previous generations. For each participating generation, there are five decision periods.

For each decision period, each participant receives an endowment of 10 Money Units (i.e., in the total of five periods, each member makes decisions regarding an overall amount of 50 ECU). In each decision period, each participant decides how much of the 10 ECU to keep and how many ECU to assign to a group account. The amount that a participant keeps will be paid to this participant at the end of the experiment. The ECU in the group account will be added up. After each period, each group member will be able to see how many ECU the two other group members have assigned individually to the group account, as well as the sum of the amounts assigned to this account. If after the five decision periods, the sum assigned to the group account by the three group members is 75 ECU or more, the experiment continues, and the next group starts. If the sum is less than 75 ECU, the experiment ends. In this case, all participants of the next groups (please note that we use the terms group and generation interchangeably) do not earn money from the experiment.

In particular, if Gen 1 assigned a total of at least 75 ECU to the group account, then the experiment continues to Gen 2. If Gen 2 assigned at least 75 ECU to the group account, then the experiment continues to Gen 3. If Gen 3 assigned at least 75 ECU to the group account, then the experiment continues to Gen 4. Finally, if the experiment reaches Gen 4, each member of Gen 4 earns 25 ECU, and the experiment ends. If it does not reach Gen 4, the members will earn nothing from the experiment. In each decision period, after finding out how much the other members of the generation have assigned to the group account, participants have the possibility to penalize members of their generation. They can assign to each of the generation members 0, 1, or

at most 2 negative points. Each negative point costs 1 ECU. Each negative point that they assign is multiplied by three and reduces the remaining ECU of the penalized member (i.e., the group member who received the negative points). E.g., if a participant assigns one negative point, the remaining ECU of the penalized member is reduced by 3 ECU. Before each decision period, the participants had the opportunity to communicate via chat messages which represents the basis of this analysis.

## The chat messages

You receive a folder with 14 excel spreadsheets. Please note:

- Each spreadsheet represents one session.
- Each session can contain up to two cohorts.
- Each cohort consists of up to four generations/groups. If a generation fails to reach the threshold, the subsequent generations are not included.
- Each generation/group consists of three participants playing five periods.

You can find all messages of one generation in one sheet within the spreadsheet, please see an example in Figure 5.H.2. Please note that some participants did not use the chat. Therefore, their ID might not appear in the respective column (“nickname”). In each sheet, you find a set of questions (not all of them cover questions on punishment; see Figure 5.H.3). Please use the dropdowns to answer them. Please make sure you answered all questions on all (!) sheets.

session	id	participant	channel	nickname	body	timestamp	Zeit	Gruppe	Generatio
Round 1	4y11wa3	Teilnehmer 2			gt negative Punkte senden? Wusst ihr das?	18.01.23 12:52		1 G1	
	4y11wa3	Teilnehmer 3			esahit haben, sodass man nicht auf 75 Punkte kommt	18.01.23 12:53		1 G1	
	4y11wa3	Teilnehmer 1			ich denken wir?del	18.01.23 12:53		1 G1	
Round 2	4y11wa3	2	6-6-01-38	Teilnehmer 2	Sollte nicht jeder von uns in jeder Runde einfach 5 ECU geben?	18.01.23 12:57		1 G1	
	4y11wa3	2	6-6-01-38	Teilnehmer 2	Sorry wenn das eine double Frage war	18.01.23 12:57		1 G1	
	4y11wa3	1	6-6-01-38	Teilnehmer 1	yes, der meining bin ich auch. hab mich eben vertippt, aber gleich es dann gleich wieder aus :-P	18.01.23 12:58		1 G1	
	4y11wa3			Teilnehmer 2	LOL	18.01.23 12:58		1 G1	
Round 3	4y11wa3			Teilnehmer 3	Wir können das hier beschleunigen, wenn wir einfach den Chat skippen.	18.01.23 13:00		1 G1	
	4y11wa3			Teilnehmer 2	Okay	18.01.23 13:00		1 G1	
Round 4	4y11wa3	2	6-6-01-44	Teilnehmer 2	hallo	18.01.23 13:02		1 G1	
	4y11wa3	2	6-6-01-44	Teilnehmer 2	kurze Frage	18.01.23 13:02		1 G1	
	4y11wa3	2	6-6-01-44	Teilnehmer 2	wie kommt man überhaupt zum 25 C (das Maximal)?	18.01.23 13:03		1 G1	
Round 5									

Figure 5.H.2: Example of one generation's chat

Bitte beantworten Sie die folgenden Fragen: 1. Was war der erste Vorschlag im Chat? 2. Was war der aktuelle Vorschlag am Ende des Chats? 3. Wurde über die Negativpunkte im Allgemeinen gesprochen? Bitte beantworten Sie die folgenden Fragen nur, wenn sie Frage 3 mit "Ja" beantwortet haben. 3a Gab es innerhalb der Diskussion eine Drohung Negativpunkte zu verteilen? 3b Gab es innerhalb der Diskussion eine Vereinbarung keine Negativpunkte zu verteilen?		Session 4y11nwa3	Gruppe 2	Generation G3
---	--	---------------------	-------------	------------------

Figure 5.H.3: Questions

## Questions

**Question 1:** You can choose between “Zu kooperieren,” “Nicht zu kooperieren,” and “Kein Vorschlag.” Please only consider the very first suggestion of any of the participants. Please see an example in Figure 5.H.4. **We define cooperation as the will to reach the threshold of 75 ECU.**

nickname	body
Teilnehmer 1	Hallo
Teilnehmer 3	Ihr packt was aufs Gruppenkonto oder?
Teilnehmer 2	main
Teilnehmer 2	wieso? Verliere ich dadurch nicht nur Geld?
Teilnehmer 3	Mmh okay, also ich w/rd schon was draufpacken, ist gerechter
Teilnehmer 1	wir m/issen ja nicht sozial sein
Teilnehmer 1	aber wenn ihr eh nichts draufmacht, kann ich es auch lassen, dann ist ja eh egal
Teilnehmer 1	hat es nachteile nichts draufpacken?
Teilnehmer 2	Was haben wir davon, wenn es durch alle Gruppen durchgeht?
Teilnehmer 1	ich glaube keine Nachteile
Teilnehmer 1	dann lass nichts draufpacken

First Suggestion

Final Suggestion

Bitte beantworten Sie die folgenden Fragen: 1. Was war der erste Vorschlag im Chat? 2. Was war der aktuelle Vorschlag am Ende des Chats?		Zu kooperieren Nicht zu kooperieren
--	--	--

Figure 5.H.4: Example for Questions 1 and 2

**Question 2:** You can choose between “Zu kooperieren,” “Nicht zu kooperieren,” and “Kein Vorschlag.” Please only consider the very last suggestion of any of the participants. Please see an example in Figure 5.H.4. **We define cooperation as the will to reach the threshold of 75 ECU.**

nickname	body
Teilnehmer 6	Es reicht, wenn jeder je Runde 5 gibt
Teilnehmer 4	Genau, jeder immer 5
Teilnehmer 5	Dachte, gleich alles rein und die letzten halt 0
Teilnehmer 5	aber gut
Teilnehmer 6	kommt aufs Gleiche hinaus :)
Teilnehmer 6	Bloß keine Negativpunkte
Teilnehmer 5	dann gebe ich m/ichste 0 und danach 5
Teilnehmer 6	Easy!

Agreement to not punish

Bitte beantworten Sie die folgenden Fragen: 3. Wurde über die Negativpunkte im Allgemeinen gesprochen? Bitte beantworten Sie die folgenden Fragen nur, wenn sie Frage 3 mit "Ja" beantwortet haben. 3a Gab es innerhalb der Diskussion eine Drohung Negativpunkte zu verteilen? 3b Gab es innerhalb der Diskussion eine Vereinbarung keine Negativpunkte zu verteilen?		Ja Nein Ja
--	--	------------------

Figure 5.H.5: Example for Questions 3, 3a, and 3b

**Question 3, 3a, 3b:** Please note that the questions are not relevant for all sessions; i.e., you will find sheets without those questions. If you answer question 3 with no, you can leave questions 3a and 3b blank. Please see an example in Figure 5.H.5.

Please do not hesitate to reach out to [stenzel@economics.uni-kiel.de](mailto:stenzel@economics.uni-kiel.de) in case of any questions. Thank you very much!



## 5.1 Translated instructions (computer screens)

This section includes the translated instructions for the PUN&COMMS treatment. The PUN (COMMS) treatment instructions are identical, excluding the COMMS (PUN) parts. The baseline instructions are identical, excluding both parts for COMMS and PUN. Each subsection represents a single screen presented to the participants.

### Welcome

Dear participants,

Thank you for participating in our experiment. The experiment is expected to last 60 minutes. The experiment is carried out in strict compliance with the code of conduct for economic experiments of the University of Kiel. This means, among other things, that all information is provided to you truthfully. You can end the experiment at any time without giving a reason. Your data will be treated as strictly confidential and will be processed and published anonymously by the researchers involved. In the following, for reasons of better readability, only the female form will be used, although this explicitly refers to people of any gender.

All participants receive the same instructions. Please read the instructions carefully.

Please mute your microphone. If you have any questions, write to us using the private chat feature in Zoom.

Each participant receives 5€ as a participation fee, regardless of the course of the experiment. Depending on your decisions and the decisions of the other participants during the experiment, you can earn more money.

During the experiment we use Experimental Currency Units (ECU) instead of euro as currency. The amount you get for the decision tasks is calculated and displayed in ECU. At the end of the experiment, the total amount will be converted into euro and (in addition to the 5€ participation fee) transferred to your specified bank account within 2-3 days after the experiment. At the end of the experiment, we will ask you for your account information.

The exchange rate between ECU and euro is 2.50 ECU : 1.00€. This means: For every 2.50 ECU you earn, you receive 1.00€.

### **Group Assignment and Experiment Procedure**

In this experiment, every 12 participants will be randomly and anonymously divided into four groups, each group consists of three participants. The four groups will be labeled during the experiment as Group 1 (with members 1, 2, and 3), Group 2 (with members 4, 5, and 6), Group 3 (with members 7, 8, and 9), and Group 4 (with members 10, 11, and 12). A random mechanism will determine if you will be assigned to Group 1, 2, 3, or 4. Once the computer program starts, you will see on your computer screen if you are assigned to Group 1, 2, 3, or 4.

In this experiment, members of Group 1 always actively participate (i.e., taking decisions). However, the members of subsequent groups only participate depending on the decisions of the previous groups. I.e., Group 2's participation depends on the decisions of Group 1, and similarly, Group 3's participation depends on the decisions of Groups 1 and Group 2. Finally, Group 4's participation depends on the decisions of Groups 1, Group 2, and Group 3.

### **The Order of Decisions**

For each group, there are five decision rounds. You take two decisions per round. All members of the group simultaneously make their decisions. For each decision round, each participant receives an endowment of 10 ECU (i.e., in the total of five rounds, each member makes decisions regarding an overall amount of 50 ECU).

Before each decision round, the 3 group members can chat with each other via a chat window on your computer screen. In the first two rounds, you have 3 minutes, and in the rounds 3-5, you have 1.5 minutes per round to communicate. You can only talk with the two other members of your group. You are not allowed to identify yourself or other group members in this chat.

In each decision round, each participant decides how much of the 10 ECU to keep and how much ECU to assign to a group account. The amount that a participant keeps will be paid to this participant at the end of the experiment.

The ECU in the group account will be added up.

After each round, each group member will be able to see how many ECU the two other group members have assigned individually to the group account, as well as the sum of the amounts assigned to this account.

If after the five decision rounds, the sum assigned to the group account by the three group members is 75 ECU or more, the experiment continues, and the next group starts. If the sum is less than 75 ECU, the experiment ends. In this case, all participants of the next groups do not earn money from the experiment.

In particular, if Group 1 assigned a total of at least 75 ECU to the group account, then the experiment continues to Group 2. If Group 2 assigned at least 75 ECU to the group account, then the experiment continues to Group 3. If Group 3 assigned at least 75 ECU to the group account, then the experiment continues to Group 4. Finally, if the experiment reaches Group 4, each member of Group 4 earns 25 ECU, and the experiment ends. If it does not reach Group 4, the members will earn nothing from the experiment.

Note that besides the earnings from the experiment that depend on your decisions and the decisions of previous groups, every participant is entitled to 5€ as a participation fee, irrespective of the course of the experiment.

### **Negative Points**

After being informed in each round of decisions how much the two other participants in your group have sent to the group account, you make the second decision. You can attribute negative points to the two other participants in your group. You can assign to each of your group members 0, 1, or at most 2 negative points. Each negative point that you assign costs 1 ECU. Each negative point that you assign is multiplied by three and reduces the remaining ECU of the penalized member. You have three possibilities:

- You assign 0 negative points (no costs for you and the remaining ECU of the penalized member is reduced by 0 ECU).
- You assign 1 negative point (costs you 1 ECU and the remaining ECU of the penalized member is reduced by 3 ECU).

- You assign 2 negative points (costs you 2 ECU and the remaining ECU of the penalized member is reduced by 6 ECU).

After your decision, you can see how many negative points you assigned and how many negative points you have received. All decisions are anonymous: You cannot find out which participant has assigned negative points to you, and others cannot find out to which participant you have assigned negative points.

### **Earning per Round and Total Earning**

In case the previous group contributed enough ECUs to the group account, your group was reached, and you took decisions; then, your payoff for each round will be as follows:

- Endowment for this round (10 ECU)
- minus the assigned amount to the group account
- minus assigned negative points
- minus received negative points times 3.

Please note that your payoff for one round can be negative. Your payoff is summed up for the five decision rounds. In cases where your summed payoff is negative, it will be set to 0 (and in any case, you will earn the participation fee of 5€). In case your group was not reached, and you did not take decisions: You will not earn money from the experiment. However, you, as everybody else, are entitled to the 5€ participation fee.

### **Test Questions**

Please consider the following cases and answer the questions. At the end of the page, you will also find a summary of the instructions.

**Case 1:** After the five decision rounds, Participant 1 in Group 1 assigned a total of 30 ECU to the group account. Moreover, after the five decision rounds, the total amount in the group account of Group 1 is 75 ECU.

- What is the total payoff for Participant 1 (in ECU): [20 ECU]
- Does the experiment continue to Group 2? [yes]

**Case 2:** After the five decision rounds, Participant 1 in Group 1 assigned a total of 20 ECU to the group account. Moreover, after the five decision rounds, the total amount in the group account of Group 1 is 70 ECU.

- What is the total payoff for Participant 1 (in ECU): [30 ECU]
- Does the experiment continue to Group 2? [no]
- What is the total payoff for members in Group 2 (in ECU): [0 ECU]

**Case 3:** Participant 2 in Group 2 assigned 0 negative points to Participant 1 and 2 negative points to Participant 3.

- By how many ECU is the payoff of Participant 2 (the sender) reduced? [2]
- By how many ECU is the payoff of Participant 1 reduced? [0]
- By how many ECU is the payoff of Participant 3 reduced? [6]

The experiment will start once you press “continue.”

### Group Assignment

You are Participant [x] and are assigned to Group [y]. {*For participants in Groups 2, 3, and 4:*} Group 1 will start. Please wait until they make their decisions. This can take up to 15 minutes.

### Chat

Remaining time: [Time]

You are Participant [x] in Group [y]. This is Round [z] of 5. You can now chat with your other 2 group members for 3 minutes / 1.5 minutes. You can leave the chat already earlier. You are not allowed to identify yourself or other group members in this chat. In the next round, you can chat with your other 2 group members again. If you want to leave the chat earlier, please click “next.” Please note that you cannot return to the chat in this round.

## Group Account

You are Participant [x] in Group [y]. This is Round [z] of 5.

For this task, you and two additional group members have received an endowment of 10 ECU. On this screen, you can decide how many (0-10) of your ECU you want to assign to the group account. All ECU that would be assigned to this group account will be added up. If after five rounds the sum is equal or larger to 75 ECU, the next group will start to play. If the sum accumulated in the group account after 5 rounds is less than 75 ECU, the experiment will end, and the next group will not make decisions (and will not receive an additional endowment).

Please state how many ECU you want to send to the group account: [...]

## Contributions to the Group Account

You are Participant [x] in Group [y]. This is Round [z] of 5. Here, you are informed about the decisions of the participants in your group.

### This Round

- Your contribution:
- Participant [1]: ...
- Participant [2]: ...

### Up to this round

- Your contribution:
- Participant [1]: ...
- Participant [2]: ...

You and the other members of your group contributed  $[\sum c_i]$  ECU in total up to this point. *{In the last round:}* There are more than 75 ECU in the group account and, therefore, the next group starts. / There are not more than 75 ECU in the group account and, as a consequence, the experiment ends.

### Negative Points

You are Participant [x] in Group [y]. This is Round [z] of 5. On this screen, you can take a decision if you want to assign negative points to another member. A negative point costs you 1 ECU and reduces the earnings of the other participant by 3 ECU. You can assign between 0 and 2 negative points to each participant in each round.

### Round Payoff

You are Participant [x] in Group [y]. This is Round [z] of 5.

- Your endowment: ...
- Your assignment to the group account: ...
- Your assigned negative points: ...
- Negative points received: ...
- **Your round payoff:** ...

### Total Payoff

You are Participant [x] in Group [y]. This is Round [z] of 5. Your payoffs per round:

- Round 1: ...
- Round 2: ...
- Round 3: ...
- Round 4: ...
- Round 5: ...
- **Total:** ...

## Chapter 6

### Conclusions

My dissertation aims to enhance our comprehension of how we can address two market failures present in climate change: information asymmetries and the intergenerational free-rider problem. Chapter 2 discussed information asymmetries with regard to GHG emissions present in firms' supply chains. In Chapter 3 and 4, we evaluated how (self-)verification technologies might enable investors to overcome information asymmetries, e.g., regarding firms' carbon risks. Chapter 5 assessed how decentralized institutions might help to fight the intergenerational free rider problem.

As highlighted in Chapter 2, policy regulations require Scope 3 emissions reporting for a growing number of companies (e.g., the EU's Corporate Sustainability Reporting Directive). Because production processes involve global and complex supplier networks, the reporting entity must obtain data that are beyond its control, facing multiple challenges. Key barriers to data collection include (i) legal and regulatory concerns, (ii) scarcity of data and action interoperability, and (iii) the significant risk of sharing sensitive data, resulting in high data sharing costs for companies. Attempting to overcome these barriers, cross-industry collaborations, technology companies, sector-specific initiatives, standard-setting bodies, reporting institutions, and regulators are establishing overarching standards and technical infrastructures to facilitate the flow of primary data between firms and further exchange with financial markets. Other technological advances, including distributed data storage and processing combined with homomorphic encryption technologies, allow firms to maintain control over their data. Governance frameworks provide a technical environment for data processing and analysis while holding all parties accountable for their actions.

Sharing primary data can also provide additional benefits to individual companies and enable new business models within digitized supply chains.



For example, to account for transportation emissions under Scope 3, a company must collect and process data that can also be used to coordinate collective logistics optimization. Thus, the potential benefits of primary data sharing among companies are even greater than the mere benefits of calculating Scope 3 emissions. In order to leverage primary data sharing for Scope 3 emissions calculation and beyond, different stakeholders need to take a systemic and global perspective to overcome the three barriers to data sharing along supply chains.

The emergence of new data sharing systems, such as Blockchain and privacy-preserving technologies, could further decrease the data sharing costs for firms. By (self-)verifying disclosed data at low cost, investors and consumers might be able to overcome the information asymmetries present in markets and create the necessary incentives for firms to shift to greener-than-average technologies. To model this firm-investor relationship, we developed a novel experimental design in Chapter 3 and 4. The Probabilistic Investment Game builds on sender-receiver games and adds a realistic and arguably important feature of start-up investments, namely their probabilistic nature. This extension is noteworthy because the probabilistic nature of our experiment, as opposed to a deterministic setting, could promote untruthful behavior since investors cannot detect a lie from lottery outcomes (failures can be attributed to bad luck).

First, we conducted a one-shot variation in which firms and investors played the Probabilistic Investment Game only once. We observed that firms were more honest than expected when sending free messages to investors, even after they received a low probability of success (and honesty would seem to be against their interest). Such “overcommunication” is consistent with previous research on “cheap talk” communication (e.g., Dickhaut et al., 1995; Cai et al., 2006; Jin et al., 2021; Fréchette et al., 2022), but our study has extended it to a probabilistic game. In all treatments, this cheap talk communication affected the investors’ beliefs on the true probability of success. Such investors’ trust in cheap talk messages is in line with Forsythe et al. (1999) and Benndorf et al. (2015). Yet, without a verification or self-verification possibility, the true probability of success did not significantly affect investments.

Verification by investors is effective in conditioning investment decisions on the firms’ true (but unknown) probability of success when verification is accurate, but not when it is noisy, albeit the latter is arguably a better representation of the current practice. The literature on cheating and deception

indicates that individuals tend to behave more honestly if they are worried about being exposed as a liar (see, e.g., Kajackaite & Gneezy, 2017). Accordingly, in our experiment, the lowest cheating in cheap talk messages was observed when accurate verification was available to investors. As cheap talk communication affected the investors' beliefs on the true probability of success more than theory predicts, less cheating led to more effective investment decisions.

Next, we tested self-verification by firms. Here, the predicted unraveling did not fully occur. Yet, when a self-verification instrument is sufficiently cheap, it is effective in conditioning investment decisions on the firms' true probability of success. In this case, firms use self-verification more often, allowing investors to make more informed and thus effective decisions. Finally, the data indicated that when (self-)verification instruments are too costly, they may not be efficient from a social viewpoint. Hence, recent technological advances may have the potential for long-term efficiency increase only when they become available at sufficiently low costs.

Following the analysis of the initial findings, we amended our game design and introduced repetition (15 rounds with fixed roles and random matching). Similar to the one-shot version, we observed the "overcommunication" phenomenon in all treatments. As investors trusted the free messages, investments were shown to be significantly greater in "good" firms than in "bad" firms in all treatments. However, this effect is considerably higher for the (self-)verification treatments than in the trust-based treatment(s). With the option of accurate verification, investors made more informed decisions, regardless of whether they were matched with a firm with a low or high probability of success.

In the self-verification treatments, we also found that firms with a high probability of success self-verified less frequently than anticipated. Our findings indicate that firms with a high probability of success gradually increased their self-verification over time. Despite the lack of self-verification, investors, on average, continued to invest a positive amount even when firms did not self-verify and instead sent a free message during the self-verification treatments. However, after receiving a free message indicating a high probability of success (instead of a self-verified one), investors invested less on average compared to the trust-based treatment or when they did not verify the message. Even though investors seemed to become more skeptical, some still believed that a firm with a high probability of success may not necessarily self-verify. In-

terestingly, these beliefs persisted even in the self-verification treatment with very low costs.

Chapter 5 addressed the intergenerational free rider problem utilizing the repeated Intergenerational Goods Game (IGG). In the absence of any institutions, we observed low levels of cooperation in the repeated IGG. In the communication treatment, members of one generation were given the opportunity to communicate freely with each other before making contribution decisions. However, this treatment resulted in low altruistic cooperation (similar to the Baseline), even with the highest incidence of zero contributions. Rather than holding each other accountable, members of a generation seemed to use the chat to agree on the optimal payoff strategy for their generation, which involves not contributing to the group account. In the punishment treatment, members could reduce their peers' earnings at a cost (with a 3:1 fine-to-fee ratio) after learning the others' contribution decisions. While subsequent generations' earnings were higher, the benefits of punishment come with their own drawbacks: More generations now narrowly missed the threshold, resulting in a waste of resources. In addition, the act of punishment itself also led to resource wastage. However, when punishment was combined with communication, the situation changed: Most generations could now meet the threshold with minimal punishment, and those who did not coordinate recognized their limitations early on and avoided wasting resources on a lost cause.

At first glance, this result may not be surprising to those familiar with the literature on classical social dilemmas. However, it is important to note the fundamental difference between the types of cooperation required in the IGG and classical social dilemmas. In the latter, *mutual cooperation* leads to the maximization of the group payoff. The IGG models a situation that requires altruistic cooperation, a form of cooperation that does not yield any material benefits for the decision makers. This is because the current generation benefits the most from non-cooperation, as the losses from non-cooperation are borne entirely by the next generation, leaving each member of the current generation unaffected by the immediate consequences of defection. Therefore, for any given generation, *mutual defection* leads to the maximization of the generational payoff. As with climate change, the current generation benefits at the expense of future generations (intergenerational free riding). The classical social dilemma literature's findings do not directly apply to the altruistic cooperation needed in such cases. In fact, in linear public goods games, both punishment (Fehr & Gächter, 2000, 2002; Nikiforakis & Normann, 2008) and

communication (Isaac & Walker, 1988) alone have each been shown to deter free riding. It is hence noteworthy that neither of these institutions alone is sufficient to support cooperation in the IGG.

In summary, this thesis addressed two market failures present in climate change: information asymmetries (regarding GHG emissions) and the free rider problem (concerning future generations). Our findings regarding information asymmetries have both optimistic and pessimistic interpretations. On the positive side, humans are more truthful and trusting than initially hypothesized. On the negative side, the availability of a cheap self-verification technology does not resolve the information asymmetry. This undermines a potentially important benefit of Blockchain or other (self-)verification technologies (e.g., Gephart et al., 2019; Rajasekaran et al., 2022). Regarding the intergenerational free rider problem, we observed that the interplay between communication and decentralized enforcement catalyzed altruistic cooperation in the repeated IGG. This gives a glimmer of hope that decentralized institutions could solve the intergenerational free rider problem in Climate Change.

## Bibliography

- Barrett, S. (2007). *Why Cooperate? The Incentive to Supply Global Public Goods*. Oxford University Press, USA.
- Battersby, S. (2017). Can humankind escape the tragedy of the commons? *Proceedings of the National Academy of Sciences*, 114(1), 7–10.
- Benndorf, V., Kübler, D., & Normann, H. T. (2015). Privacy concerns, voluntary disclosure of information, and unraveling: An experiment. *European Economic Review*, 75, 43–59.
- Bernardini, E., Di Giampaolo, J., Faiella, I., & Poli, R. (2021). The impact of carbon risk on stock returns: Evidence from the European electric utilities. *Journal of Sustainable Finance & Investment*, 11(1), 1-26.
- Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, 142(2), 517-549.
- Boon-Falleur, M., Grandin, A., Baumard, N., & Chevallier, C. (2022). Leveraging social cognition to promote effective climate change mitigation. *Nature Climate Change*, 12, 332–338.
- Brown, T. C., & Kroll, S. (2017). Avoiding an uncertain catastrophe: Climate change mitigation under risk and wealth heterogeneity. *Climatic Change*, 141(2), 155–166.
- Buchholz, W., & Sandler, T. (2021, June). Global public goods: A survey. *Journal of Economic Literature*, 59(2), 488-545.
- Bui, B., Moses, O., & Houqe, M. N. (2020). Carbon disclosure, emission intensity and cost of equity capital: Multi-country evidence. *Accounting & Finance*, 60(1), 47-71.
- Burton-Chellew, M. N., May, R. M., & West, S. A. (2013). Combined inequality in wealth and risk leads to disaster in the climate change game. *Climatic Change*, 120(4), 815–830.
- Cai, H., Tao, J., & Wang, Y. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior*, 56(1), 7–36.
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods ex-

- periments: A selective survey of the literature. *Experimental economics*, 14, 47–83.
- Dickhaut, J. W., McCabe, K. A., & Mukherji, A. (1995). An experimental study of strategic information transmission. *Economic Theory*, 6(3), 389–403.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980–994.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137–140.
- Forsythe, R., Lundholm, R., & Rietz, T. (1999). Cheap talk, fraud, and adverse selection in financial markets: Some experimental evidence. *The Review of Financial Studies*, 12(3), 481–518.
- Fréchette, G. R., Lizzari, A., & Perego, J. (2022). Rules and commitment in communication: An experimental analysis. *Econometrica*, 90(5), 2283–2318.
- Gephart, J. A., Froehlich, H. E., & Branch, T. A. (2019). Opinion: To create sustainable seafood industries, the United States needs a better accounting of imports and exports. *Proceedings of the National Academy of Sciences*, 116(19), 9142–9146.
- Hauser, O. P., Rand, D. G., Peysakhovich, A., & Nowak, M. A. (2014). Cooperating with the future. *Nature*, 511, 220–223.
- IPCC. (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. (eds Pörtner, H.-O. et al.) (Cambridge University Press).
- Isaac, R. M., & Walker, J. M. (1988). Communication and free-riding behavior: The voluntary contribution mechanism. *Economic Inquiry*, 26(4), 585–608.
- Jacquet, J., Hagel, K., Hauert, C., Marotzke, J., Röhl, T., & Milinski, M. (2013). Intra-and intergenerational discounting in the climate game. *Nature Climate Change*, 3(12), 1025–1028.
- Jin, G. Z., Luca, M., & Martin, D. (2021). Is no news (perceived as) bad news? An experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, 13(2), 141–173.
- Jung, J., Herbohn, K., & Clarkson, P. (2018). Carbon risk, carbon risk awareness and the cost of debt financing. *Journal of Business Ethics*, 150(4), 1151–1171.

- Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, 102, 433–444.
- Lange, A., Löschel, A., Vogt, C., & Ziegler, A. (2010). On the self-interested use of equity in international climate negotiations. *European Economic Review*, 54(3), 359–375.
- Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., & Schellnhuber, H. J. (2008). Tipping elements in the earth’s climate system. *Proceedings of the National Academy of Sciences*, 105(6), 1786–1793.
- Lohse, J., & Waichman, I. (2020). The effects of contemporaneous peer punishment on cooperation with the future. *Nature Communications*, 11(1815).
- Mahmoudian, F., Yu, D., Lu, J., Nazari, J. A., & Herremans, I. M. (2023). Does cost of debt reflect the value of quality greenhouse gas emissions reduction efforts and disclosure? *Journal of International Accounting, Auditing and Taxation*, 52, 100563.
- Matsumura, E. M., Prakash, R., & Vera-Muñoz, S. C. (2014). Firm-value effects of carbon emissions and carbon disclosures. *The Accounting Review*, 89(2), 695–724.
- Milinski, M., Röhl, T., & Marotzke, J. (2011). Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. *Climatic Change*, 109(3), 807–814.
- Naumann, G., Cammalleri, C., Mentaschi, L., & Feyen, L. (2021). Increased economic drought impacts in Europe with anthropogenic warming. *Nature Climate Change*, 11(6), 485–491.
- Nikiforakis, N., & Normann, H.-T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11, 358–369.
- Nordhaus, W. (2019). Climate change: The ultimate challenge for economics. *American Economic Review*, 109(6), 1991–2014.
- Rajasekaran, A. S., Azees, M., & Al-Turjman, F. (2022). A comprehensive survey on blockchain technology. *Sustainable Energy Technologies and Assessments*, 52, 102039.
- Reed, K. A., Wehner, M. F., & Zarzycki, C. M. (2022). Attribution of 2020 hurricane season extreme rainfall to human-induced climate change. *Nature Communications*, 13(1), 1905.
- Stenzel, A., & Waichman, I. (2023). Supply-chain data sharing for scope 3

- emissions. *npj Climate Action*, 2(1), 7.
- Tavoni, A., Dannenberg, A., Kallis, G., & Löschel, A. (2011). Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences*, 108(29), 11825-11829.
- Tebaldi, C., Ranasinghe, R., Vourdoukas, M., Rasmussen, D. J., Vega-Westhoff, B., Kirezci, E., ... Mentaschi, L. (2021). Extreme sea levels at different global warming levels. *Nature Climate Change*, 11(9), 746-751.
- Tosun, J. (2022). Addressing climate change through climate action. *Climate Action*, 1, 1-4.
- Trinks, A., Ibikunle, G., Mulder, M., & Scholtens, B. (2022). Carbon intensity and the cost of equity capital. *The Energy Journal*, 43(2), 1944-9089.
- Waichman, I., Requate, T., Karde, M., & Milinski, M. (2021). Challenging conventional wisdom: Experimental evidence on heterogeneity and coordination in avoiding a collective catastrophic event. *Journal of Environmental Economics and Management*, 109, 102502.
- Wang, Y., Wu, Z., & Zhang, G. (2022). Firms and climate change: A review of carbon risk in corporate finance. *Carbon Neutrality*, 1(1), 6.
- WBCSD. (2021). *Pathfinder Framework: Guidance for the Accounting and Exchange of Product Life Cycle Emissions*. Geneva, Switzerland: World Business Council for Sustainable Development (WBCSD).



## Erklärungen

### Erklärung zum selbständigen Verfassen der Arbeit

Ich erkläre hiermit, dass ich meine Doktorarbeit “Studies and Experiments on Information Asymmetries and the Intergenerational Free Rider Problem in Climate Change” selbstständig und ohne fremde Hilfe angefertigt habe und dass ich als Coautor maßgeblich zu den weiteren Fachartikeln beigetragen habe. Alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen der aufgeführten Beiträge wurden besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert.



München, 18.03.2025

## Erklärung der Co-Autoren

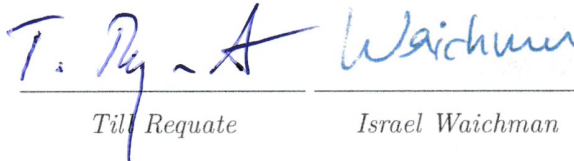
Ich erkläre hiermit, dass ich Mitautor des unten aufgeführten Artikels bin. In dem unter meiner Co-Autorenschaft verfassten Artikel haben alle Co-Autoren wesentlich zur Entwicklung der Forschungsfrage, zur Analyse der Forschungsdaten und zur Erstellung des Manuskripts gemäß den DFG-Richtlinien zur Sicherung guter wissenschaftlicher Praxis beigetragen.

Stenzel, A., & Waichman, I. (2023). Supply-chain data sharing for scope 3 emissions. *npj Climate Action*, 2, 7.



Israel Waichman

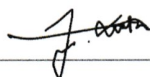
Stenzel, A., Requate, T., & Waichman, I. (2024). The Efficacy of (Self-) Verification Instruments in Risky Investments: An Experiment.



Till Requate

Israel Waichman

Stenzel, A., Lohse, J., Requate, T., & Waichman, I. (2024). Punishment with Communication Sustains Altruistic Cooperation Across Generations.



Johannes Lohse



Till Requate

Israel Waichman