

# **Essays on Social Identity and Moral Behavior**

Inauguraldissertation

zur Erlangung des Grades eines Doktors  
der Wirtschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der  
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

**Jan Luca Henkel**

aus Köln

Bonn

2023

Dekan: Prof. Dr. Jürgen von Hagen  
Erstreferent: Prof. Dr. Armin Falk  
Zweitreferent: Prof. Dr. Florian Zimmermann  
Tag der mündlichen Prüfung: 11. August 2023

# Acknowledgements

Dissertations are not created in isolation but shaped through the continuous efforts of others. This dissertation is no exception. Without the support, insights, and inspiration from my advisors, co-authors, colleagues, fellow doctoral students, friends, and family, this dissertation would not have been possible. I am immensely grateful to them for making it possible. Thank you all.

First and foremost, to my advisors, starting with my first advisor Armin Falk. His talent of not only quickly identifying problems with my projects and ideas but also promptly providing solutions greatly improved my research. Without his at all times honest and constructive feedback, I would not have overcome challenges that appeared to be indomitable. Then, to my second advisor Florian Zimmermann. I could always count on his support no matter how short the notice, be it for something as mundane as an urgently needed signature or for something as important as preparing for a job talk. I am also very grateful to Farzad Saidi, who provided, without hesitation, support and encouragement on research and everything else at a time when I direly needed it. Armin, Florian, and Farzad's guidance has been instrumental not only for this dissertation but for my development as a researcher in general.

Second, to my co-authors with which I had the great pleasure of writing papers together. I am thankful for their willingness to collaborate with me, each providing me with an opportunity to learn and improve. Christian Zimpelmann for the countless hours optimizing every single sentence of our paper. Roland Bénabou for showing me the process of how masterful writing emerges. Jean Tirole for the experience of what mastery of theoretical arguments can achieve. Cornelia Betsch, Robert Böhm, Lars Korn, and Philipp Sprengholz for teaching me the joy and value of interdisciplinary research.

Third, to my colleagues, who I was fortunate to get to know and who together created a great work environment. Among them Suanna Oh, Evan Friedman, and Frederik Schwerter, who always showed unconditional enthusiasm to support my work, Chris Roth, with whom I shared many urgently needed cake breaks, and Arkadev Ghosh and Maximilian Müller, who I could always count on to be ready for lunch precisely at 12:30. Because of them, I always looked forward to coming to the office.

Fourth, to my fellow doctoral students. To Timo Freyer, with whom I shared not only a flat but also endless inspiring discussions about the world, and to Matthias Gnewuch and Martin Dohmen, who tried their best to prevent me from working too much. To Peter Andre, Felix Chopra, and Laurenz Günther, who were always happy to discuss research and everything else with me and to Chui Yee Ho and Ximeng Fang, with whom I shared many fun cooking evenings.

Fifth, to the fantastic support staff in Bonn that often went above and beyond to help me with administrative aspects of conducting research, among them Markus Anthony, Stefanie Sauter, Ilona Krupp, and Simone Jost. A special thanks also to Holger Gerhardt for his dedication to helping me solve technical issues and his provision of public goods such as the LaTeX template this dissertation is written with. I am also grateful for the support provided by the Bonn Graduate School of Economics, Institute for Applied Microeconomics, briq Institute on Behavior & Inequality, BonnEconLab, Collaborative Research Center Transregio 224, and ECONtribute.

Finally, I am thankful to my friends and family, among them Daniela, whose willingness to test and proofread all my experiments prevented countless errors, and Luzie, whose views and insights continue to enrich my view of the world. Most importantly, I am immensely grateful to my mother, uncle, and father for 28 and counting years of unwavering love and support.

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xv</b>
<b>Introduction</b>	<b>1</b>
References	4
<b>1 Proud to Not Own Stocks: How Identity Shapes Financial Decisions</b>	<b>5</b>
1.1 Introduction	5
1.2 Conceptual framework	12
1.3 The prevalence of negative perceptions of stockholders	14
1.3.1 Data	14
1.3.2 Measurement	15
1.3.3 Results	16
1.3.4 Robustness and correlates	17
1.4 The causal effect of perceptions of stockholders on financial decision-making	19
1.4.1 Description experiment	20
1.4.2 Description experiment results	24
1.4.3 Information experiment	26
1.4.4 Information experiment results	28
1.4.5 Follow-up survey for robustness	32
1.5 The association between perceptions of stockholders and stock market participation	34
1.5.1 Data	35
1.5.2 Results	35
1.6 Determinants and implications of perceptions of stockholders	37
1.6.1 Determinants and accuracy of perceptions of stockholders	38
1.6.2 Affective polarization and political attitudes towards stockholders	40

1.6.3	Misreporting in surveys	41
1.7	Conclusion	43
Appendix 1.A	Additional figures	49
Appendix 1.B	Additional tables	55
Appendix 1.C	Selection of character traits	68
Appendix 1.D	Defining the group of stockholders	71
Appendix 1.E	Additional evidence on people’s perceptions of stockholders	72
Appendix 1.F	Validation of the stock description experiment	76
1.F.1	Validating the stock investment decision	76
1.F.2	Validating negative views about stockholders on Prolific	77
Appendix 1.G	Heterogeneous effects of information about stockholders on behavior	80
Appendix 1.H	Follow-up survey for information experiment	83
1.H.1	Survey design	83
1.H.2	Validation	84
1.H.3	Results	84
Appendix 1.I	Correcting stereotypical perceptions	90
1.I.1	Design	90
1.I.2	Results	91
Appendix 1.J	Research transparency	93
1.J.1	Description experiment	93
1.J.2	Information experiment	93
1.J.3	Correcting stereotypes experiment	93
Appendix 1.K	Variables	96
1.K.1	Demographic and financial variables	96
1.K.2	Stock market knowledge variables	97
1.K.3	Behavioral determinants of portfolio choice	97
1.K.4	Determinants of political attitudes	98
1.K.5	Scales	98
Appendix 1.L	Instructions	101
1.L.1	Perception survey view of stockholders elicitation screens	101
1.L.2	Description experiment decision screens	103
1.L.3	Information experiment information provision screens	107
1.L.4	Follow-up survey behavior intention elicitation screen	113
1.L.5	Follow-up survey perceptions elicitation screen	114
References		115
<b>2</b>	<b>The Association Between Vaccination Status Identification and Societal Polarization</b>	<b>121</b>
2.1	Introduction	121

2.2	Measurement and correlates of VSI	123
2.3	Perceptions of public discourse and discrimination	125
2.4	The role of VSI in vaccination policy acceptance	127
2.5	Discussion	129
2.6	Methods	131
2.6.1	First wave (December 2021)	132
2.6.2	Second wave (February 2022)	134
2.6.3	Third wave (July 2022)	135
2.6.4	Additional wave with subsample (March 2022)	135
	Appendix 2.A Supplement	136
	References	147
<b>3</b>	<b>Eliciting Moral Preferences under Image Concerns: Theory and Experiment</b>	<b>151</b>
3.1	Introduction	151
3.1.1	Related literature	153
3.2	Model	154
3.2.1	Preferences	154
3.2.2	Direct elicitation	155
3.2.3	Becker-DeGroot-Marschak or multiple-price list	156
3.2.4	Comparison of DE vs. MPL	158
3.3	Experimental design	159
3.3.1	Saving a Life	159
3.3.2	Treatments	160
3.3.3	Procedure	161
3.4	Hypotheses and results	162
3.4.1	Hypotheses	162
3.4.2	Results	162
3.5	Conclusion	164
	Appendix 3.A Proofs	166
	Appendix 3.B Decision screens	168
	Appendix 3.C Robustness experiment	169
3.C.1	Setup and treatments	170
3.C.2	Procedure	171
3.C.3	Results	171
	Appendix 3.D Instructions	172
3.D.1	Announcement by the experimenter	172
3.D.2	Introduction	173
3.D.3	Treatment <i>DE Low Image</i>	177

3.D.4	Treatment <i>DE High Image</i>	179
3.D.5	Treatment <i>MPL Low Image</i>	181
3.D.6	Treatment <i>MPL High Image</i>	183
3.D.7	Robustness experiment	187
3.D.8	Introduction	187
3.D.9	Treatment <i>DE No-Image</i>	188
3.D.10	Treatment <i>MPL No-Image</i>	189
	References	192
<b>4</b>	<b>Ends versus Means: Kantians, Utilitarians and Moral Decisions</b>	<b>195</b>
4.1	Introduction	195
4.1.1	Related literature	199
4.2	Experimental design	200
4.2.1	Ends-versus-means block	201
4.2.2	Self-versus-other block	208
4.2.3	Questionnaires	210
4.2.4	Procedural details	211
4.3	Results	212
4.3.1	How prevalent are deontological and consequential decision-making?	212
4.3.2	How consistent are deontological and consequential decision-making?	215
4.3.3	How are deontological and consequential decision-making related to prosocial behavior and morality-questionnaire measures?	217
4.3.4	Robustness of correlations between blocks	219
4.4	Conclusion	222
	Appendix 4.A Details on self versus others decision situations	224
	Appendix 4.B Robustness analysis trolley dilemma	225
	4.B.1 Figures	225
	4.B.2 Robustness experiment	228
	Appendix 4.C Robustness analysis ends-versus-means decisions	230
	Appendix 4.D Robustness between blocks	232
	Appendix 4.E Instructions	239
	4.E.1 General information	239
	4.E.2 Ends-versus-means block	241
	4.E.3 Self-versus-others block	254
	References	260



# List of Figures

1.1	People's perceptions of characteristics of stockholders and non-stockholders	18
1.2	Description experiment results: aggregate and individual behavior	24
1.3	Information experiment results: aggregate and individual behavior	30
1.4	The relationship between negative views about stockholders and stock market participation across countries	37
1.A.1	People's perceptions of characteristics of stockholders and non-stockholders around the world part 1	49
1.A.2	People's perceptions of characteristics of stockholders and non-stockholders around the world part 2	50
1.A.3	Combinations of risky option visualization (Draw description)	51
1.A.4	Example of information provided to subjects	52
1.A.5	Distribution of signals generated in the information experiment	53
1.A.6	Distribution of prior beliefs in the information experiment	53
1.A.7	Heterogeneity in the relationship between negative views and stock market participation	54
1.E.1	Replication of Panel A of Figure 1.1 using robustness sample	73
1.E.2	Perceptions of differences in activities between stockholder and non-stockholder	75
1.F.1	Replication of Panel A of Figure 1.1 using Prolific sample	78
1.H.1	Information experiment treatment effect on beliefs in follow-up	88
1.H.2	Information experiment treatment effect on intentions in follow-up	89
1.L.1	Perception survey view of stockholders elicitation Netherlands	101
1.L.2	Perception survey view of stockholders elicitation United States	102
1.L.3	Description experiment stock description intro	103
1.L.4	Description experiment stock description decision	104
1.L.5	Description experiment draw description intro	105
1.L.6	Description experiment draw description decision	106
1.L.7	Information experiment instructions screen 1	107
1.L.8	Information experiment instructions screen 2	108
1.L.9	Information experiment instructions screen 3	109
1.L.10	Information experiment instructions screen 4	110

1.L.11	Information experiment instructions screen 5	111
1.L.12	Information experiment instructions screen 6	112
1.L.13	Follow-up survey investment intention elicitation screen	113
1.L.14	Follow-up survey perceptions elicitation screen	114
2.1	Effects of vaccination status identification (VSI) on perceptions of public discourse and discrimination	126
2.2	Effects of vaccination status identification (VSI) on reactance toward mandatory policies and activism behavior	128
2.A.1	Stability of VSI over time	143
3.1	Equilibrium under direct elicitation (panel A) and multiple-price list (panel B)	156
3.2	Main experimental results	163
3.B.1	Decision screen DE	168
3.B.2	Decision screen MPL	169
3.D.1	Typical appearance of a tuberculosis patient.	175
3.D.2	A worker from Operation ASHA delivers medication to a tuberculosis patient	176
3.D.3	Relationship between the donation and the saving of a life	177
4.1	Decision screen of the trolley problem	203
4.2	Bribe game	205
4.3	Decision screen of the rule-following task	208
4.4	The effect of changing realism and consequences on behavior in the trolley dilemma	212
4.5	Distribution of deontological and consequentialist decision-making in ends-versus-means tradeoff situations	214
4.6	Relationship of trolley behavior with behavior in the other ends-versus-means situations	215
4.7	Consistency of deontological and consequential decision-making	217
4.8	Principal component analyses	221
4.B.1	The effect of changing realism and consequences on behavior in the trolley dilemma using the full sample	225
4.B.2	The effect of changing realism and consequences on behavior in the trolley dilemma among high comprehension subjects	226
4.B.3	Differences in characteristics of subjects between the <i>Hypothetical</i> and <i>Real</i> treatments	227
4.B.4	The effect of changing consequences on behavior in self versus other tradeoff situations	229

4.C.1	Distribution of deontological and consequentialist decision-making in ends-versus-means tradeoff situations among high comprehension subjects	230
4.C.2	Distribution of deontological and consequentialist decision-making in ends-versus-means tradeoff situations using the full sample	231



# List of Tables

1.1	Description experiment illustration of choice options	22
1.2	Information experiment treatment effect on negative perceptions	29
1.3	Information experiment treatment effect on decision-making	31
1.4	Information experiment causal effect of negative perceptions on decision-making	33
1.5	The relationship between negative views about stockholders and stock market participation	45
1.6	Comparison of average actual and predicted differences	46
1.7	Implications of negative perceptions on in-group favoritism and political attitudes	47
1.8	Misreporting of risky financial assets ownership	48
1.B.1	Overview of samples	55
1.B.2	Overview of sample demographics Netherlands	56
1.B.3	Overview of sample demographics cross-country study	56
1.B.4	The relationship between negative views about stockholders and stock market participation across countries	57
1.B.5	Overview of sample demographics Prolific (US)	58
1.B.6	Description experiment balance test	58
1.B.7	Description experiment treatment effect on decision-making	59
1.B.8	Information experiment balance test	60
1.B.9	The relationship between negative views about stockholders and the share of risky financial assets	61
1.B.10	The association of demographic variables with negative views about stockholders	62
1.B.11	OLS-regression on the relationship between being a stockholder and self-assessed character traits	63
1.B.12	Distribution of self-assessments and allocations	64
1.B.13	Linking of administrative data balancing test	65
1.B.14	Determinants sample balancing test	65
1.B.15	The relationship between negative views and stock market participation (Probit)	66
1.B.16	Misreporting of share of risky financial assets	67

1.C.1	List of character traits used for selection	68
1.C.2	Association of traits with stockholder ranking results	69
1.C.3	Identity relevance of traits results	70
1.E.1	List of activities used for robustness	72
1.F.1	Description experiment validation that the stock option is associated with the group of stockholders	77
1.F.2	The relationship between negative views about stockholders and stock market participation on Prolific	79
1.G.1	Heterogeneity in the effect of information about stockholders on decision-making using symmetric cutoffs	81
1.G.2	Heterogeneity in the effect of information about stockholders on decision-making using equal observation cutoffs	82
1.H.1	Follow-up experiment balance test	85
1.H.2	Follow-up experiment effect of binary coded information on investment intention and perceptions	86
1.H.3	Follow-up experiment effect of information on investment intention and perceptions	87
1.I.1	Effect of providing information about actual differences between stockholders and non-stockholders on investment behavior	92
1.J.1	Replication of Table 1.B.7 using preregistered sample	94
1.J.2	Replication of the result of Table 1.3 using preregistered specification	95
1.K.1	Identity scale items	99
1.K.2	Political preferences items	100
2.A.1	Correlations between vaccination status identification (VSI), vaccination intention and the 7C antecedents of vaccination part I	136
2.A.2	Correlations between vaccination status identification (VSI), vaccination intention and the 7C antecedents of vaccination part II	137
2.A.3	Predictors of VSI.	138
2.A.4	Effect of VSI on perception of discriminatory public discourse	139
2.A.5	Effect of VSI on everyday discrimination perception	140
2.A.6	Effect of VSI on ostracism perception	141
2.A.7	Effect of VSI on ingroup preference [EUR]	142
2.A.8	Effect of VSI on reactance toward vaccination mandate	143
2.A.9	Effect of VSI on likelihood of having joined demonstration against mandate	144
2.A.10	Effect of VSI on likelihood of having signed petition against mandate	145
2.A.11	Sociodemographics	146
3.1	Regression analyses of the effect of the elicitation method on prosocial behavior	165

3.C.1	Regression analyses of the effect of the elicitation method on voucher choice	172
4.1	Ends-versus-means decision situations	201
4.2	Prosociality (self-versus-other) decision situations	209
4.3	Correlation matrix for ends-versus-means decision-situation	216
4.4	Correlations of ends-versus-means, self-versus-other, and questionnaire measures	220
4.D.1	Correlations of ends-versus-means, self-versus-other and questionnaire measures comprehension sample	233
4.D.2	Correlations of ends-versus-means, self-versus-other and questionnaire measures memory check sample	234
4.D.3	Correlations of ends-versus-means, self-versus-other and questionnaire measures excluding speeders	235
4.D.4	Correlations between ends-versus-means and self-versus-other blocks using behavior categorization for the later variables	236
4.D.5	Correlations between ends-versus-means and self-versus-other blocks using median splits for the later variables	237
4.D.6	Biserial correlations between ends-versus-means and self-versus-other blocks	238





# Introduction

Understanding individual decision-making is fundamental to understanding economic and societal processes, as individuals play a pivotal role in markets and society. For instance, via direct or indirect investments, individuals move stock markets, steer expenditures and influence corporate decisions. By engaging in or abstaining from social distancing, complying with or undermining mandated protection measures, and deciding whether to be vaccinated or not, individual behavior shapes the course of pandemics. Through the extent to which individuals engage in moral behavior, they determine the response to existential threats that require collective and cooperative action, such as measures against climate change.

A thorough understanding of individual decision-making not only helps to understand economic and societal processes, but also how to address the challenges that come with them. Understanding how individuals make investment decisions helps to assess the influences and dangers of increasing wealth inequality and can motivate changes to governmental systems like pension schemes. Understanding individual's responses towards vaccinations and protection measures help to design more effective and accepted measures to mitigate pandemics. Understanding how individuals make moral decisions and the extent to which such behavior is malleable helps to design effective interventions and regulations that address climate change.

This dissertation contributes to a better understanding of individual decision-making by presenting four independent research papers that study the determinants and consequences of individual decision-making. Each paper conducts a positive analysis of behavior, attempting to enhance our understanding of how individuals *actually make* decisions instead of providing arguments about how they *should make* decisions. To do so, the papers combine theoretical frameworks modeling individual behavior with the rigorous testing of these frameworks using behavioral experiments and surveys. The use of formal theoretical frameworks allows the formulation of behavioral predictions that can be empirically tested while making explicit the assumptions and conditions required to derive the predictions. Experiments are then used to test the predictions because they are one of social science's foremost tools to measure individual behavior and establish causality (Falk and Heckman, 2009).

Specifically, the dissertation investigates decision-making in the domains of financial investment choices, decisions and attitudes around vaccinations in the con-

text of the COVID-19 pandemic, and moral behavior. It draws insights and applies methods from behavioral economics, household finance and social psychology. The first two chapters focus on social identity, i.e., the idea that people's identity – the concept of how they are – depends on the groups they are members of. The chapters show how identification processes help to understand financial decisions and polarization around vaccination decisions. The third and fourth chapter focus on moral behavior, studying the influence of image concerns – individuals' desire to signal to others and maintain to themselves that they are generous, caring, or generally “morally good,” – on moral decision-making and how people resolve moral dilemma situations. Below, I briefly summarize each chapter.

**Chapter 1: Proud to Not Own Stocks: How Identity Shapes Financial Decisions.**

Chapter 1, which is joint work with Christian Zimpelmann, investigates individuals' decision to invest in the stock market. This decision is one of the most important financial decisions of households, as stock investments offer substantially higher expected returns than other asset classes at the expense of increased short-run volatility (Jordà et al., 2019). Interestingly, a large fraction of individuals around the world do not invest in the stock market (Gomes, Haliassos, and Ramadorai, 2021).

In the chapter, we propose an explanation for this widespread aversion behavior: identity concerns induced by people's negative views of stockholders lead them to avoid stock investments. To show the influence of people's negative views on investment behavior, we first measure people's views using surveys we conduct with nearly 8,500 individuals from eleven countries. We find that a large fraction of respondents in all countries views stockholders negatively – they are perceived as greedy, gambler-like, and selfish individuals. In two experiments, we then provide experimental evidence that such perceptions of identity-relevant characteristics causally influence financial decision-making: if people view stockholders more negatively, they are less likely to choose stock-related investments. Moving to field data, by linking survey and administrative data, we show that negative perceptions strongly predict households' stock market participation, more so than leading alternative determinants. Our findings provide a novel explanation for the puzzlingly low stock market participation rates around the world, new perspectives on the malleability of financial decision-making, and evidence for the importance of identity in economic decision-making.

**Chapter 2: The Association Between Vaccination Status Identification and Societal Polarization.**

In Chapter 2, which is joint work with Philipp Sprengholz, Lars Korn, Cornelia Betsch, and Robert Böhm, we investigate polarization behavior around COVID-19 vaccinations. Before the COVID-19 pandemic, whether to vaccinate oneself was largely considered a health decision. However, as the pandemic progressed, public discord between those vaccinated and those unvaccinated for COVID-19 has intensified globally, and the decision to vaccinate became morally loaded.

We propose that identification processes play a key role in the perceptions and behaviors that fuel this intergroup conflict between vaccinated and unvaccinated. Using panel data from 3,267 vaccinated and 2,038 unvaccinated respondents in Germany and Austria, we test whether identification with one's vaccination status is associated with current societal polarization. We find that vaccination status identification (VSI) explains substantial variance in a range of polarizing attitudes and behaviors, such as discrimination experiences and behavior. VSI was also related to higher psychological reactance toward mandatory vaccination policies among the unvaccinated. Similarly, higher levels of VSI reduced the gap between intended and actual counter-behaviors over time by the unvaccinated. Our results highlight the importance of VSI for predicting behavioral responses to vaccination policies and suggest the need for identity-based interventions and de-escalating strategies to increase the acceptance and effectiveness of vaccination campaigns.

**Chapter 3: Eliciting Moral Preferences under Image Concerns: Theory and Experiment.** In Chapter 3, which is joint work with Roland Bénabou, Armin Falk, and Jean Tirole, we study how image concerns interact with features of the decision environment. We focus on two key features of decision environments: single versus multiple simultaneous decisions and certainty versus uncertainty of the consequences. Using direct elicitation versus multiple-price-list or, equivalently, Becker-DeGroot-Marschak schemes as exemplars of different features, we characterize in a theoretical model how image-seeking inflates prosocial giving. We show that the signaling bias induced by image concerns relative to true preferences depends on the interaction between the elicitation method and visibility level: it is greater under direct elicitation for low image concerns and greater under multiple-price-list/Becker-DeGroot-Marschak schemes for high ones. We then test the model's predictions in an experiment with life-saving donations and find the predicted crossing effect. Our results argue for caution in interpreting standard estimates of moral preferences from experiments or contingent-valuation surveys and provide guidance for maximizing public-goods contributions and image or reputational manipulations.

**Chapter 4: Ends versus Means: Kantians, Utilitarians and Moral Decisions.** In Chapter 4, which is joint work with Roland Bénabou and Armin Falk, we study people's moral behavior in situations where the two most influential moral theories, Consequentialism and Deontological ethics, differ in the actions they postulate as morally right. The Consequentialist view postulates that choosing the morally right action should be based on the ends resulting from the action. In contrast, the Deontological view postulates it should be based on the conformity of the means involved with some overarching notion of duty.

Using experiments, we design a series of games that induce ends-versus-means tradeoffs, including a real-stakes version of the classical trolley dilemma. We supplement these games with six classical self-versus-others choice tasks, allowing us to relate consequential/deontological behavior to standard measures of prosociality.

Across the ends-versus-means tradeoff situations, we find a sizeable prevalence (20 to 40%) of non-consequentialist choices by subjects but no evidence of stable individual preference types across situations. In particular, trolley behavior predicts no other ends-versus-means choices. Instead, which moral principle prevails appears to be highly context-dependent. In contrast, we find a substantial level of consistency across self-versus-other decisions, but individuals' degree of prosociality is unrelated to how they choose in ends-versus-means tradeoffs that only affect others. Our findings highlight the relevance of deontological considerations for decision-making but also the challenges of using moral dilemmas to predict moral behavior.

Taken together, the chapters of this thesis produce two key insights that advance our understanding of individual decision-making. First, they show that social identity is a powerful concept to explain behavior. Identification processes can explain a wide array of behaviors related to financial decisions and vaccinations, two domains which previously have not been considered as relevant to social identity. Second, they show that moral behavior is more malleable and to a higher degree shaped by the situation context than previously acknowledged. Not only do image concerns influence the extent of moral behavior, but these concerns interact with the decision environment. Similarly, individuals appear to resolve moral dilemmas depending on the decision context rather than according to some underlying moral principle.

## References

- Falk, Armin, and James J. Heckman.** 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science* 326 (5952): 535–38. <https://doi.org/10.1126/science.1168244>. [1]
- Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai.** 2021. "Household Finance." *Journal of Economic Literature* 59 (3): 919–1000. <https://doi.org/10.1257/jel.20201461>. [2]
- Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor.** 2019. "The Rate of Return on Everything, 1870-2015." *Quarterly Journal of Economics* 134 (3): 1225–98. <https://doi.org/10.1093/qje/qjz012>. [2]

# Chapter 1

## Proud to Not Own Stocks: How Identity Shapes Financial Decisions\*

*Joint with Christian Zimpelmann*

### 1.1 Introduction

The decision to invest in the stock market is a central element of households' financial decision-making. Stock investments offer substantially higher expected returns than other asset classes at the expense of increased short-run volatility (Jordà et al., 2019). They are thus widely recommended by financial experts as a vital aspect of households' long-term saving strategies (e.g., for retirement). Nevertheless, in almost all countries, the majority of households do not invest in stocks (Gomes, Haliassos, and Ramadorai, 2021). Widespread avoidance persists even among wealthy households and remains largely unexplained by classical preference- or constraint-

\* An earlier version of this chapter has appeared as CRC TR 224 Discussion Paper No. 380 and as ECONtribute Discussion Paper No. 206.

**Acknowledgements:** We thank Peter Andre, Kai Barron, Roland Bénabou, Robert Böhm, Felix Chopra, Simon Cordes, Armin Falk, Timo Freyer, Evan Friedman, Nicola Gennaioli, Katrin Gödker, Arkadev Ghosh, Andreas Grunewald, Laurenz Günther, Paul Heidhues, Chui Yee Ho, Zwetelina Iliewa, Michael Kosfeld, Botond Kőszegi, Christine Laudenbach, Maximilian Müller, Axel Niemeyer, Suanna Oh, Christopher Roth, Felix Rusche, Farzad Saidi, Paul Smeets, Frederik Schwerter, Mark Spils, and Florian Zimmermann for helpful comments and discussions. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2126/1- 390838866. Support by the German Research Foundation (DFG) through CRC TR 224 (Project A01 and C01) is gratefully acknowledged.

**Ethics approval:** The study obtained ethics approval from the German Association for Experimental Economic Research (#CS4Wa3W1 and #UXNgX9vj)

**Preregistration:** The study was preregistered at aspredicted.org ([https://aspredicted.org/66D\\_2XD](https://aspredicted.org/66D_2XD), [https://aspredicted.org/1B8\\_MXD](https://aspredicted.org/1B8_MXD) and [https://aspredicted.org/DF3\\_5YR](https://aspredicted.org/DF3_5YR)). See also Appendix 1.J on research transparency.

based explanations.<sup>1</sup> The resulting underdiversification of households' portfolios has major implications for societal challenges such as wealth inequality, financial stability, and the design of retirement pension systems (e.g., Fagereng, Gottlieb, and Guiso, 2017; Kuhn, Schularick, and Steins, 2020).

Despite the potential advantages of stock investments for households and their importance to society, the media frequently characterize stock investors in ways commonly considered to be objectionable or repulsive. For instance, movies such as the 1987 classic "Wall Street" or the more recent "Wolf of Wall Street" provide vivid examples of stock investors and brokers as selfish individuals willing to take extreme risks at the expense of others. Likewise, popular books about stock market investing describe investors as highly motivated by greed (Shefrin, 2002; Nofsinger, 2017).

This paper investigates how people view stockholders and how these views influence financial decision-making. Using large-scale surveys conducted in eleven countries, we document widespread negative perceptions of stockholders. To conceptualize the relationship between perceptions and choice behavior, we develop a theoretical framework in which people care about the characteristics of individuals making similar decisions to themselves. This form of identity concern leads people to experience disutility if they associate themselves through their decisions with a group they perceive to have negative character traits. Testing the framework's predictions in two preregistered experiments, we provide evidence that subjects' perceptions of stockholders causally drive financial decisions. Extending the analysis to field data, we demonstrate that perceptions predict households' stock market participation as identified by administrative data.<sup>2</sup>

In the first step of our empirical analysis, we measure individuals' perceptions of stockholder and non-stockholder characteristics using surveys. Guided by our framework, we focus on identity-relevant characteristics, i.e., character traits that are important to people. To select relevant traits, we provide participants ( $N = 194$ ) with a set of traits, asking them to rate how important each trait is to them and how strongly they associate them with stockholders. We find that the traits greed, being a gambler, and selfishness are rated highest along these two dimensions.

We then measure individuals' perceptions of stockholders and non-stockholders with respect to these three traits using surveys we fielded to 3,272 Dutch respon-

1. As an example, among all households in the Netherlands with a net balance of at least 60,000 € in financial assets (top 20% quantile), 55% do not have stock holdings. Instead, their assets are concentrated in banking and saving accounts. See Guiso and Sodini (2013) for similar evidence in other countries. Explaining such behavior through risk attitudes requires implausibly high degrees of risk aversion (Heaton and Lucas, 2000).

2. Identity concerns induced by negative views operate even in private, which is important because investment decisions are mostly unobserved by others. While presumably amplified when choices are observable, we show the relevance of identity concerns for financial decisions when choices are made anonymously.

dents of the LISS panel and 5,130 respondents living in Australia, Germany, Italy, Japan, Mexico, South Korea, Spain, Sweden, United Kingdom, and the United States ( $N \approx 500$  per country). The LISS panel builds upon a probability sample of the Dutch population, employing special efforts to ensure that the sample is representative of the population. The samples from the other countries are quota-representative with respect to age and gender. In the surveys, respondents consider stockholders and non-stockholders of their respective countries and separately rate how they perceive the individuals of each group with respect to the three character traits greed, being a gambler, and selfishness. Since we empirically validate that these traits are considered negative traits, comparing ratings between groups reveals whether respondents view one group more negatively.

We document that large fractions of respondents view stockholders negatively. In all eleven countries, stockholders are rated significantly more greedy, gambler-like, and selfish than non-stockholders ( $p < 0.001$  in each country). Averaging over the three traits, between 49% and 81% of respondents rate stockholders strictly more negative than non-stockholders. In a series of robustness checks, we replicate widespread negative perceptions using alternative elicitation methods and framing variations. For instance, respondents rate stockholders also significantly more negatively when considering positively framed characteristics. We also verify that negative perceptions do not simply measure a lack of stock market knowledge or financial numeracy.

In the second step of our analysis, we test a key prediction of our framework: since people view stockholders negatively, whether a decision is associated with the stock market should influence their decision-making. An ideal test of this prediction compares the choice behavior of individuals between two identical investments that only differ in their association with the stock market. In reality, however, stock investments differ from other investments in many relevant aspects, such as expected returns, costs, and uncertainty. We thus conduct an experiment (US,  $N = 515$ ) to test whether people dislike stock investments relative to non-stock investments even when all outcome-related features are held constant, a type of behavior we label *stock market aversion*.

In our experiment, we employ a simple incentivized investment choice in which subjects repeatedly choose between a safe and a risky option. Choosing the safe option yields an amount with certainty while choosing the risky option yields a high or low payoff with equal probability. In two treatments, we vary how the options are described to subjects. In the *Stock Description* treatment, the risky option is described as an investment whose outcomes are associated with the stock market, i.e., based on past stock performances. In the *Draw Description* treatment, the risky option is instead described as an investment whose outcomes depend on a random draw. The safe option is described as abstaining from the respective investment. Since the descriptions of options are the only difference between the treatments, the underlying probabilities and payoffs are identical across treatments. By design, both descrip-

tions are of similar length and complexity. We are thus able to identify the effect of varying the association of an investment on behavior, keeping outcome-related features constant.

We find a 27% decrease in subjects' likelihood of choosing the risky option when the option is described as a stock investment instead of an investment in the outcome of a random draw ( $p < 0.001$ ). Notably, the fraction of subjects who refuse to choose the risky option in any decision almost doubles, from 19% in the *Draw Description* treatment to 36% in the *Stock Description* treatment. Using additional within-subject variation in the descriptions reveals that almost 40% of subjects are stock averse. These participants are willing to invest under the *Draw Description* but exhibit a strictly lower willingness to invest under the *Stock Description*. Accordingly, we find support for the prediction of our framework that the mere association of an option with the stock market leads to aversion.

In a third step, we provide causal evidence that subjects' negative perceptions of stockholders are responsible for the documented stock market aversion. To establish causality, we conduct an experiment (US,  $N = 548$ ) in which we exogenously shift subjects' perceptions and measure the shift's impact on their decision-making. For each subject, we randomly drew ten stockholders and ten non-stockholders out of a separate sample of 272 stockholders and non-stockholders that allocated money between themselves and a charity. We then inform subjects about the difference in donation behavior between the stockholders and non-stockholders of their draw, leading to between-subject variation in the direction and magnitude of the generated information. For example, some subjects received the information that in their draw, stockholders donated 10% more to the charity than non-stockholders, while others were informed that in their draw, non-stockholders donated 30% more. After providing the information, we elicit subjects' perceptions and ask them to choose whether to bet on risky options described as a stock investment using the choice paradigm employed in the previous experiment.<sup>3</sup> Since deciding how much to donate instead of taking for oneself is a signal about prosociality, we expect the information to shift subjects' perceptions.

We find that providing subjects with information on the difference in donation behavior between stockholders and non-stockholders significantly influences their perceptions, and crucially, their investment decisions. The stronger the signal that stockholders donated more, the less negatively the subjects view stockholders compared to non-stockholders ( $p < 0.001$ ). This shift in perception translates into behavior: a 10 pp. increase in the donation difference favoring stockholders increases the likelihood of investments in the stock option by 0.9 pp. ( $p = 0.028$ ). This effect

3. Subjects might update about factors influencing investment behavior that are unrelated to perceptions. Using our controlled decision environment allows us to abstract from these effects. For instance, even if the information changes subjects' beliefs regarding the profitability of actual stock investments, the returns of the stock decision are fixed in our experiment.



is sizable since a one standard deviation increase in the signal mitigates the effect of varying descriptions found in the previous experiment by 25%. Employing an instrumental variable approach, we estimate that a one standard deviation decrease in subjects' negative perception of stockholders causally increases investments in the stock option by 14-16 pp. (from an average of 45%). We conclude that negative perceptions are a key contributor to the stock market aversion documented in our experiments.

In a follow-up survey conducted several days later, we show that the treatment variation persistently changes subjects' perceptions of stockholders and influences intentions to invest their own money in the stock market. Subjects who receive the information that stockholders donated more perceive stockholders as less selfish and report a higher intention to invest than those who received the information that non-stockholders donated more ( $p < 0.01$ ). Furthermore, we purposefully framed the follow-up differently to obfuscate the relation to the main experiment. Since we observe that subjects do not perceive a connection between the main experiment and follow-up, experimenter demand effects cannot explain these results (Haaland, Roth, and Wohlfart, 2023).

Extending our analysis to field data, we demonstrate that negative perceptions (i) significantly predict stock market participation revealed through administrative records and (ii) are quantitatively important compared to leading alternative determinants of investment decisions proposed in the literature. Using the LISS panel enables us to link subjects' perceptions of stockholders to their asset allocations based on tax records provided by Statistics Netherlands. To assess the importance of perceptions in predicting investment decisions, we link our survey in the LISS panel with previous questionnaires to obtain measures of subjects' risk aversion, beliefs regarding stock returns, financial numeracy, general trust, political orientation, ambiguity aversion, and likelihood insensitivity.

Our results show that negative perceptions of stockholders strongly predict stock market participation. Controlling for alternative determinants and demographic variables, a one standard deviation increase in negative perceptions is associated with a 4.8 pp. decrease in the likelihood of owning stocks. This is a considerable effect size because only 23% of our sample owns stocks. Moreover, the coefficient is larger than the marginal effects of almost all other determinants; for instance, it is 33% larger than the standardized coefficient of risk aversion. We further show the generalizability of our results in the surveys we fielded to the ten other countries. Using self-reported stock market participation as dependent variable, we find a significant negative relationship in nine out of ten countries. In a regression with country fixed-effects, a one standard deviation increase in negative perceptions is associated with a 6.9 pp. decrease in stock market participation.

In a series of additional analyses, we highlight the importance of people's perceptions of stockholders for attitudes and behavior beyond financial decisions. First, non-stockholders distribute twice as much money to non-stockholders than to stock-

holders in an incentivized allocation decision. The more negatively subjects view stockholders, the more strongly they exhibit this type of in-group favoritism. Second, non-stockholders hold highly polarized opinions regarding the extent to which financial markets should be taxed and regulated. Even after controlling for factors such as subjects' political orientation and redistribution concerns, negative perceptions explain the level of support for increases in taxation and regulation. Third, we examine misreporting behavior in surveys using our linked survey-admin data. Our framework suggests that individuals potentially misreport holding risky financial assets to avoid identity conflicts. Consistent with this hypothesis, we observe that 30% of stock-owning households report not owning stocks in the survey. Importantly, negative perceptions predict this type of misreporting: a one standard deviation increase in negative perceptions is associated with a 4.5 pp. decrease in (correctly) reporting being a stockholder.

Turning to potential determinants of observed negative perceptions of stockholders, we find evidence consistent with the stereotypes model of Bordalo et al. (2016). In the model, people overweight a group's representative types, thereby exaggerating actual group differences. We show that stockholders self-assess as more greedy, gambler-like, and selfish and behave more selfishly than non-stockholders, but these differences are small and concentrated in the tail: stockholders are relatively more likely among very greedy, selfish, and gambler-like individuals. The model then predicts that perceptions are stereotypical – exaggerated representations of reality – and we indeed find that subjects significantly overestimate the actual differences when asked to predict them.

This paper makes several contributions to the literature. First, our paper contributes to research on the determinants of stock market participation. Previous literature has primarily focused on explaining people's stock aversion through preferences and beliefs related to investment outcomes as well as constraints (Gomes, Haliassos, and Ramadorai, 2021).<sup>4</sup> However, stock averse behavior is prevalent in our experiments although outcome- and constraint-based factors are held constant. Hence, we demonstrate that previous explanations neglect an important factor driving stock aversion and provide evidence that people's negative perceptions of stockholders cause this type of behavior.

A number of studies have examined factors predicting investment behavior unrelated to the potential outcomes and constraints of investments. The identity-based

4. Preference-based explanations investigate how different weightings of outcomes induced by risk-, ambiguity- (e.g. Dimmock et al., 2016), loss averse (Barberis, Huang, and Thaler, 2006) or likelihood insensitive preferences (Dimmock, Kouwenberg, and Wakker, 2016) influence stock market participation. Belief-based explanations have focused on factors such as return beliefs (e.g. Giglio et al., 2021) or optimism (Puri and Robinson, 2007). Constraint-based explanations postulate that factors such as fixed costs of participation (Vissing-Jorgensen, 2004), limited stock market literacy (e.g. van Rooij, Lusardi, and Alessie, 2011) and cognitive function (Grinblatt, Keloharju, and Linnainmaa, 2011) prevent people from investing in the stock market.

mechanism proposed in our study provides new ways for interpreting these empirical patterns. Kaustia and Torstila (2011) argue that personal values matter for investment decisions by showing that political orientation is correlated with stock market participation in Finland. Our framework and evidence document a precise mechanism of how values shape investment choices. Another strand of literature finds that cultural background, social interactions and social involvement matter for financial decision-making (Haliassos, Jansson, and Karabulut, 2017; Kuchler and Stroebel, 2021). Our framework suggests a specific channel through which culture and social influences matter: they shape people's perceptions of stockholders, which in turn influence their behavior. Furthermore, a nascent literature investigates how anti-finance sentiments influence investment behavior (Grosfeld, Rodnyansky, and Zhuravskaya, 2013; D'Acunto, Prokopczuk, and Weber, 2019; D'Acunto, 2020; Lenz and Mayer, 2023). While these studies consider individuals' trust in financial markets as a potential mechanism predicting stock market participation (Guiso, Sapienza, and Zingales, 2008), anti-finance sentiments fostering negative images of stockholders provide an alternative rationale for the documented patterns.

Second, we contribute to the emerging literature on the relationship between identity and economic behavior (Akerlof and Kranton, 2000; Shayo, 2020). Identity has been found relevant for workers' productivity in firms (Hjort, 2014; Ghosh, 2022), consumption choices (Atkin, Colson-Sihra, and Shayo, 2021), and labor market supply (Oh, 2021).<sup>5</sup> Bauer and Smeets (2015) find that investors' degree of identification with socially responsible investments is associated with a larger wealth share invested in these assets. We show that identity concerns causally influence investment behavior and contribute conceptually by linking identity concerns to people's perceptions of groups making similar decisions.<sup>6</sup>

Third, we add to the literature studying people's perceptions of others (Bursztyn and Yang, 2022), such as perceptions about income (Cullen and Perez-Truglia, 2022), political opinions (Ahler, 2014) and gender (Bordalo et al., 2019). We study people's perceptions of character traits of others, and show the relevance of these perceptions for an important household decision. By showing that people's perceptions of stockholders are stereotypical, we relate to the economic literature on stereotypes (Bordalo et al., 2016). This literature typically considers group categoriza-

5. More generally, identity influences behavior in settings such as judicial decisions (Shayo and Zussman, 2011), women's income (Bertrand, Kamenica, and Pan, 2015), and conflict (Depetris-Chauvin, Durante, and Campante, 2020).

6. A related idea is that people care about the characteristics of others or themselves and use behavior as signaling devices (Bénabou and Tirole, 2006, 2011). These image or reputational concerns have been shown as influential in both private/anonymous settings (Grossman and van der Weele, 2017; Bursztyn et al., 2020; Falk, 2021; Mechtenberg et al., 2022; Schneider, 2022) and when observed by others (e.g., Ariely, Bracha, and Meier, 2009). In addition, status concerns (Bursztyn et al., 2018) and social pressure (Bursztyn and Jensen, 2017) have been shown to influence economic behavior.

tions wherein changing group membership is impossible (e.g., ethnicity) or takes considerable effort (e.g., nationality). In contrast, we show that in a setting where groups are based on actions, and thus easily changeable, strong stereotypical beliefs are prevalent and causally influence decisions. A recent study combining identity and stereotypes in a theoretical framework of endogenous group choice is Bonomi, Gennaioli, and Tabellini (2021). We empirically document the connection between identity, stereotypes, and financial decision-making.

## 1.2 Conceptual framework

In this section, we formalize the idea that people's perceptions of other people's character traits matter for their decision-making. We use the framework to develop our hypotheses and guide our analyses in the next sections.

An individual  $i$  has to take an action  $a$  out of a set of available actions  $A$ . Each action  $a$  is characterized by a state-contingent prospect  $z_a = (E_1 : x_1, E_2 : x_2 \dots)$  yielding outcome  $x_j$  if event  $E_j$  occurs. Individual  $i$ 's utility of the prospect is a flexible function  $u_i(z_a)$  which captures the various outcome-based forms of belief formation and uncertainty attitudes proposed by the literature (risk- or ambiguity aversion, loss aversion, etc.).<sup>7</sup>

Individuals in our framework not only care about the outcomes of their actions, but also about their identity – their concept of who they are. A large literature in psychology (Burke and Stets, 2009; Stets and Serpe, 2013) shows that identity is group-based, i.e., people think of themselves in terms of groups that they belong to. There exists a set of groups  $G$  to which individuals can belong to and potentially identify with. Group membership in our setting is defined through the actions that members take, meaning that individuals categorized by a group share some unique group-defining feature related to the action set. Let  $g : A \rightarrow G$  denote a mapping indicating which group is associated with each action.

Individuals are characterized by a set of attributes or traits  $(q_1, q_2, \dots, q_H)$ , which influence behavior and attitudes. We focus on identity-relevant characteristics, i.e., individuals care about whether they and others have these characteristics. For example, traits related to morality are central to people (Wojciszke, Bazinska, and Jaworski, 1998; Goodwin, Piazza, and Rozin, 2014; Strohminger and Nichols, 2014). We assume each trait is evaluated on a numerical scale, which higher values indicating a more positive evaluation and individuals care about the average  $\bar{q}$  of the set of identity-relevant traits.<sup>8</sup>

7. In particular,  $u_i(z_a)$  might depend on her weighting function over the events (where subjective probabilities are a special case for subjective expected utility maximizers), as well as her preferences over outcomes and endowments.

8. Our predictions remain unchanged if we instead assume that people care about a unidimensional score based on all traits, with each trait monotonically influencing the score.

Importantly, individuals have identity concerns: they seek to uphold a positive identity. Because identity is based on group membership, identity concerns operate through how individuals evaluate the traits of others. For simplicity, we assume that people care about the average value of the identity-relevant characteristics within a group, however this can be easily expanded to capture more flexible aggregation functions. Accordingly, for a given action  $a$ , individual  $i$  forms a belief over the characteristics of members of group  $g$ :  $\hat{\mathbb{E}}_i[\bar{q} | g(a)]$ . The key element of our framework is then that associating oneself through an action with a group that is perceived as negative (positive) creates negative (positive) utility:

$$U_i(a) = u_i(z_a) + \theta \hat{\mathbb{E}}_i[\bar{q} | g(a)].$$

Because people's identity is based on their group memberships and they desire to uphold a positive identity, joining a negatively perceived group creates an identity conflict and thus disutility.<sup>9</sup> The importance of this type of identity-based utility relative to outcome-based utility  $u_i(z_a)$  is captured by the parameter  $\theta$ . Since it concerns people's identity, action do not have to be visible to others for identity-based utility to matter.<sup>10</sup> Hence, how people view others may influence their decisions even for decisions that are predominantly unobserved by others, such as investment decisions.

In the context of investment decisions, there exists the group of stockholder  $S$  and non-stockholder  $NS$ , defined by whether an individual owns stocks or not. Then, actions based on investing in the stock market, such as buying shares of a mutual fund, putting money in stocks of a specific company, or engaging in options trading, are all associated with the group of stockholders. Suppose individuals can either invest  $a = I$  or not invest  $a = NI$  in the stock market; hence  $g(I) = S$  and  $g(NI) = NS$ . Maximizing  $U_i(a)$ , individual  $i$  chooses  $a = I$  if and only if

$$u_i(z_I) - u_i(z_{NI}) \geq \theta (\hat{\mathbb{E}}_i[\bar{q} | NS] - \hat{\mathbb{E}}_i[\bar{q} | S]).$$

That is, even if the potential material gains from investing are large, people will abstain from investing if the group of stockholders is viewed as sufficiently negative compared to the group of non-stockholders.

Our framework makes two key testable predictions. The first one builds on the idea that the influence of identity concerns depends on which actions are associated with which groups. The attractiveness of an action should, hence, depend on which group the action is associated with.

9. This mechanism is similar to violations of group prescriptions (e.g., gender roles), which can create "anxiety and discomfort in oneself" (Akerlof and Kranton, 2000).

10. It thus differs from status concerns – the desire to signal high income or wealth to others (Heffetz and Frank, 2011).

**Prediction 1.** Assume there are two distinct groups  $G_1$  and  $G_2$ , over which perceptions differ such that  $G_1$  is perceived more positive ( $\hat{\mathbb{E}}_i[\bar{q} | G_1] > \hat{\mathbb{E}}_i[\bar{q} | G_2]$ ). Changing the association of an action  $a$  from group  $G_1$  to  $G_2$ , keeping all other aspects constant, reduces the attractiveness of  $a$ .

Holding the association of actions and groups fixed, our framework furthermore predicts that identity concerns depend on people's views of the respective groups. The second prediction, thus, states that the attractiveness of an action is directly influenced by people's views of the group associated with the action.

**Prediction 2.** Suppose action  $a$  is associated with group  $G$ . If the perception of the characteristics of the individuals belonging to  $G$  becomes more positive (negative), the attractiveness of action  $a$  increases (decreases) relative to other actions not associated with  $G$ .

A necessary condition for identity concerns to be relevant for the decision to invest in stocks is that people's views of stockholders and non-stockholders differ. In the next section, we identify identity-relevant characteristics and measure people's views of stockholders and non-stockholders over these characteristics. Based on the results, we test the predictions of our framework in two experiments in Section 1.4.

### 1.3 The prevalence of negative perceptions of stockholders

In this section, we document people's view of stockholders by measuring their perceptions of identity-relevant characteristics of stockholders and non-stockholders. We start by describing the data sources in Section 1.3.1, introduce our method of measuring perceptions in Section 1.3.2, and provide the results in Section 1.3.3.

#### 1.3.1 Data

We elicit people's views of stockholders using surveys, drawing from two samples with complementary advantages. Our primary sample focuses on a single country, the Netherlands, and contains a broad range of measures for a large and representative population sample. Our secondary sample covers ten countries with different cultural and institutional backgrounds, allowing us to assess the generalizability of our findings.

**Netherlands.** For the Netherlands, we fielded two surveys to the LISS (Longitudinal Internet studies for the Social Sciences) panel. The panel is based on a true probability sample of the Dutch population drawn from the population registers in collaboration with Statistics Netherlands (CBS) and administered by Centerdata (Tilburg University).<sup>11</sup> We focus on participants of the panel who report to be the

11. Special efforts are made to ensure that the panel represents the adult Dutch population. For example, devices are provided to participants who otherwise would lack access.

main financial decision-maker of their respective household. In total 3,272 panelists participated in our first survey in which we elicited their views of stockholders. In addition, we subsequently fielded a second survey to 1,592 non-stockholders to obtain supplementary measures. See Appendix Table 1.B.2 for a summary of demographic characteristics. We chose the LISS panel due to the following three key features that make it ideally suited for our purpose. First, it allows us to measure people's views of stockholders among a broad population sample. Second, we can link participants' responses to Dutch administrative data, which includes information on financial assets of the respective households based on tax records. Third, we can link participants' responses to other surveys fielded to the LISS panel. Through these features we obtain a rich set of individual level measures.

**Cross-country.** To measure people's views more broadly around the world, we fielded a short survey to individuals living in ten countries in cooperation with the survey company Bilendi. In total, we have data from 5,130 respondents, around 500 in each of the following countries: Australia, Germany, Italy, Japan, Mexico, South Korea, Spain, Sweden, United Kingdom, and the United States. Samples are representative for each country with respect to age and gender based on quota sampling. See Appendix Table 1.B.3 for a summary of demographic characteristics. Surveys were translated and back-translated by professional translators. Rare cases of disagreements were resolved by a third translator.

### 1.3.2 Measurement

Section 1.2 formulates two conditions for a character trait  $q$  to influence decisions related to the stock market. First, the trait needs to be identity-relevant. Second, subjects need to perceive differences in the extent to which they associate it with the group of stockholders and non-stockholders (in our framework with respect to trait  $h$ :  $\hat{\mathbb{E}}_i[q_h | S] \neq \hat{\mathbb{E}}_i[q_h | NS]$ ). Based on these conditions, we identified a list of eight potential character traits through a search in media (articles, books and movies). We then conducted a pre-test to select three character traits based on the two mentioned conditions, see Appendix 1.C for details. We selected the following traits which belonged to the top four traits for each criterion, supplemented with accompanying definitions:<sup>12</sup>

**Greed** A strong wish to continuously get more of things like wealth, possessions or social values.

**Gambler** A person that shows the tendency to risk money or other stakes in the hope of being successful.

12. The definitions were based on established psychological formulations, which we slightly altered to increase comprehension. We circulated the definitions among experts and non-experts to ensure that they were both internally valid and easily understood.

**Selfishness** Being willing to accept negative consequences for other people or the environment to gain a personal advantage as a result.

**Elicitation.** To elicit perceptions of stockholders, we asked subjects to consider the entire adult population of their respective country to be divided into two groups: those who hold any risky financial assets and those who do not. We provided subjects with the precise list of assets considered risky financial assets. The categorization is based on the corresponding tax category in the Netherlands, ensuring that Dutch subjects are familiar with the categorization. Since the assets categorized as risky financial assets are typically traded on the stock market, we will use the terms “stockholder” and “has risky financial assets” interchangeably. We justify in more detail our choice of defining stockholders and provide evidence that subjects understand the definition well in Appendix 1.D. Subjects then stated separately for the group of stockholders and non-stockholders their assessment for each trait (‘People who (do not) own risky financial assets are on average ...’) on a scale from 0 ‘totally disagree’ to 10 ‘totally agree’. We chose this measurement because it is easily understood by subjects and straightforward to answer, allowing us to include it even in short surveys.<sup>13</sup>

**Variables.** Through these trait ratings, we obtain a measure of how subjects perceive the characteristics of each group. We define subjects’ *average negative perceptions* about stockholders/non-stockholders as the average rating of each group over the three trait. To obtain a single measure, we define *negative views about stockholders* as the difference between subjects’ average negative perceptions of stockholders and their average negative perceptions of non-stockholders.<sup>14</sup> We label the variable “negative views” because the three selected traits are framed negatively. While, in general, these traits may also have positive aspects, subjects view them predominantly negatively (see Appendix 1.C). Hence, higher values indicate that subjects view stockholders more negatively relative to non-stockholders.

### 1.3.3 Results

**Netherlands.** Figure 1.1 Panel A shows the distribution of subjects’ ratings of stockholders and non-stockholders for the three traits greed, gambler, and selfishness in the Netherlands. Higher values indicate that the respective group is rated more greedy, gambler-like and selfish. For all three traits, the distribution for stockholders lies to the right of the non-stockholder distribution, indicating that stockholders

13. For these reasons, similar trait ratings have been employed in the context of political ideology in political science (e.g., Iyengar, Sood, and Lelkes, 2012; Hobolt, Leeper, and Tilley, 2021)

14. Our results do not rely on averaging over traits. In particular, our analyses in Section 1.5 yield similar results if we consider each trait separately.



are rated more negatively ( $p < 0.001$ , Kolmogorow-Smirnow test)<sup>15</sup>. On the individual level, a large fraction of subjects rates stockholders strictly more negatively than non-stockholders (64% for greed, 81% for gambler, and 47% for selfishness,  $p < 0.001$ , Wilcoxon signed-rank test). Moreover, ratings between traits are highly correlated and show a high degree of internal consistency (Cronbach's  $\alpha = 0.80$  for stockholder and  $\alpha = 0.86$  for non-stockholder ratings). Comparing average negative perceptions, we observe that 81% of respondents rate stockholders more negatively than non-stockholders. In conclusion, subjects in the Netherlands hold substantial negative views about stockholders.

**Cross-country.** Panel B of Figure 1.1 shows that negative views about stockholders are not limited to the Netherlands. For our set of eleven countries, the figure displays subjects average negative perceptions about stockholders and non-stockholders. In every country, stockholders are rated more negatively on average than non-stockholders, often by more than 50%. These differences are significant in every instance ( $p < 0.001$ , Wilcoxon signed-rank test). In the Appendix, Figures 1.A.1 and 1.A.2 replicate the figure of Panel A, showing that the distribution of ratings differ in all countries. On the individual level, the data similarly reveal strong negative views. In total, 64% of subjects rate stockholders strictly more negatively than non-stockholders, with fractions for each country ranging from 49% to 73%. Thus, negative views about stockholders are a general finding in various countries around the world.

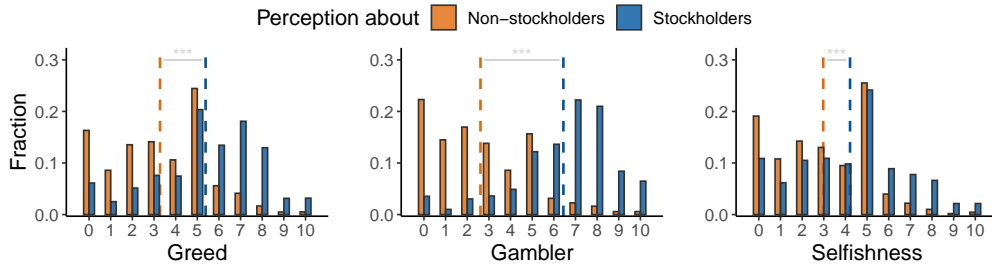
#### 1.3.4 Robustness and correlates

In the following, we show the robustness of our previously documented result. Specifically, we extend our analysis to alternative measures of people's views of stockholders and the relation of views with potential confounds as well as investigate whether views vary with respect to types of investments and socio-demographic background.

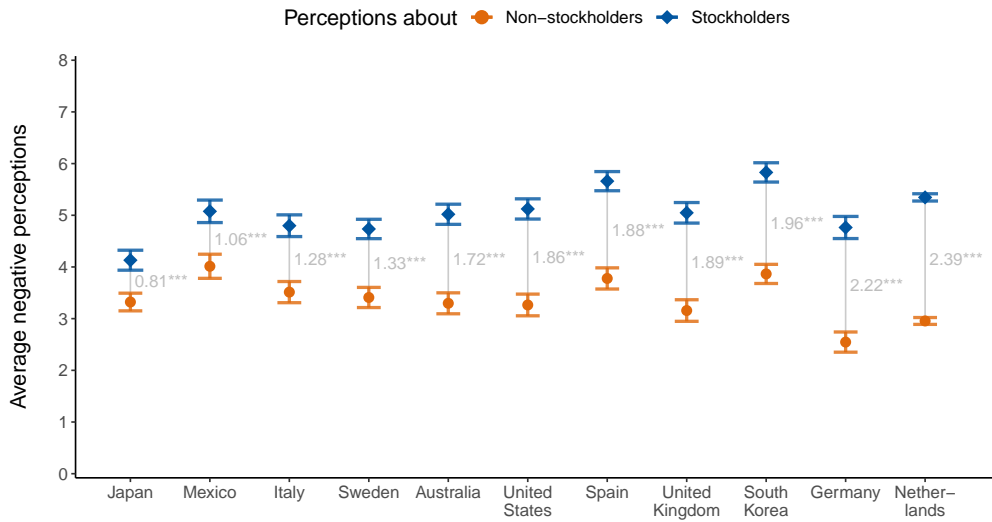
**Robustness to measurement.** Is our finding that people view stockholders negatively an artifact of how we measure views of stockholders? For instance, the use of negative traits and Likert scales might bias subjects' responses. It is important to emphasize, however, that our main measure concerns the difference between subjects' ratings of stockholders and non-stockholders. Thus, biases arising from Likert scales are unlikely to drive the large differences in ratings we find. To further show the robustness of our results, we use a more quantitative elicitation measure and vary whether views are measured over negatively or positively framed items in a separate sample of 1,016 Dutch individuals. In total, we elicit perceptions over nine

15. All tests refer to two-sided tests. We adjusted these p-values and the subsequent ones in this section for multiple hypothesis testing using Bonferroni correction.

**Panel A: Netherlands**



**Panel B: Cross-country sample**



**Figure 1.1.** People’s perceptions of characteristics of stockholders and non-stockholders

Notes: **Panel A** displays the distribution of subjects’ ratings of the group of stockholders and non-stockholders elicited in the LISS panel ( $N = 3,272$ ). Higher values indicate that subjects rate the respective group as more greedy (left graph), gambler-like (middle graph), and selfish (right graph). The dotted lines display mean ratings. **Panel B** displays subjects’ ratings of the group of stockholders and non-stockholders averaged over the three traits greed, gambler, and selfishness for our set of ten countries ( $N = 5,130$ ) and the LISS panel (Netherlands). Higher values indicate that subjects rate the respective group as more negative on average. Bars indicate 95% confidence intervals. Stars indicate significance levels, where \* $p < 0.1$ , \*\* $p < 0.05$  and \*\*\* $p < 0.01$ , obtained from Kolmogorov-Smirnov tests (Panel A) and Wilcoxon signed-rank tests (Panel B).

items that are related to the previously used character traits. For each, we replicate our main finding: subjects view stockholders significantly more negatively if views are elicited over a negatively framed item and less positively if elicited over a positively framed item (in every instance,  $p < 0.001$ , Wilcoxon signed-rank test, Bonferroni corrected). For example, a majority of subjects state that stockholders act less prosocial in an incentivized dictator game. See Appendix 1.E for details on design and results.

**Relation to stock market knowledge and numeracy.** Are views about stockholders merely a proxy for individuals' financial numeracy or stock market knowledge? From a theoretical point of view, perceptions of stockholders' characteristics are distinct from the later concepts. To test whether perceptions are also empirically distinct, we collected self-assessed stock market knowledge, a standard measure of financial numeracy, subjects' self-assessed belief whether they would be successful in the stock market and their belief over stock returns in our second LISS survey (see Appendix 1.K.2 for variable definitions). We find that neither self-assessed stock market knowledge ( $r = -0.13$ ), financial numeracy ( $r = 0.12$ ), success belief ( $r = -0.07$ ), or return beliefs ( $r = 0.06$ ) are meaningfully correlated with negative views about stockholders.

**Perceptions over different investments.** How do people view different types of stock investors? Going beyond eliciting views of the general group of stockholders, we presented participants of our second LISS survey with investors that (i) only invest in socially responsible investments (SRI), (ii) only invest in market index funds, and (iii) explicitly use financial derivatives (options, swaps, and warrants). Participants rated how selfish they perceived each group of investors. We find that participants differentiate perceptually between types of stock investors: compared to stockholders in general, SRI and index fund investors are perceived significantly less negative ( $p < 0.001$ , paired Student's t-test, Bonferroni corrected), while investors that use derivatives are perceived as more negative ( $p = 0.017$ ). Yet, all three groups are still viewed significantly more negative relative to non-stockholders ( $p < 0.001$ ).

**Relation to socio-demographic variables.** Do negative views vary with socio-demographic background? Using the first LISS survey, we regress negative views about stockholders on a set of background variables. We find that women and older subjects have significantly higher negative views and observe no difference with respect to education, income, and wealth. Thus, more wealthy individuals do not hold systemically less negative views of stockholders. For details on the regression results, see Appendix Table 1.B.10.

## 1.4 The causal effect of perceptions of stockholders on financial decision-making

This section investigates whether people's perceptions of stockholders causally influence financial decision-making. Our conceptual framework of Section 1.2 provides an explanation for why perceptions influence behavior: people have identity concerns, i.e., want to uphold a positive identity, and their identity is based on the groups they belong to.

**Motivational evidence.** If group identity is relevant in the stock market context, we would expect non-stockholders to form an identity with the group of non-

stockholders. To test this, we elicit five items taken from established identity scales, adopted to the stock market context, in our non-stockholder sample in the second LISS (see Appendix 1.K.5 for details). We find that non-stockholders express a high degree of identification with their group. For instance, 50% agree that “Not having risky financial investments is an important part of my identity” and 41% even agree that they are “proud to not own risky financial investments”. The extent of this group identification is significantly associated with how negatively subjects view stockholders ( $p < 0.001$ ).<sup>16</sup>

Our framework generates the following hypothesis on how such identification is relevant for behavior: choosing an option associated with a group that is viewed negatively creates an identity conflict, which people seek to avoid. We test this hypothesis in two steps. First, we design an experiment to test whether non-stockholders avoid options associated with stockholders (Prediction 1 of our framework). We then design a second experiment to test whether people’s negative views of stockholders are responsible for this aversion behavior (Prediction 2 of our framework).

**Data.** We use US participants recruited on Prolific for the experiments. We choose Prolific because it is one of the leading market research companies used in social science research and its participants have been shown to provide high-quality responses in terms of attention and comprehension (Eyal et al., 2021; Gupta, Rigotti, and Wilson, 2021). The experiments were preregistered, see Appendix 1.J for details. We used oTree (Chen, Schonger, and Wickens, 2016) for programming.

#### 1.4.1 Description experiment

To test whether non-stockholders avoid choice options associated with stockholders, we present subjects with simple incentivized investment decisions and exogenously vary the association of the decisions’ options with the stock market.

**Decisions.** Subjects in the experiment face two decision parts. Each decision part consists of four investment decisions. In each decision, subjects are endowed with \$30 and choose between a safe and a risky option. By choosing the risky option, subjects pay an amount  $c < \$30$  and receive with 50% probability a high outcome

16. While we focus on the influence of identity when choices are anonymous, we also collected suggestive evidence that visibility to others matters for investment decisions. We elicited non-stockholders intention to invest (7-point Likert scale) under two hypothetical situations. First, everyone they know will find whether they invest in the stock market. Second, no one will find out. Investment intentions are significantly lower in the first situation compared to the second ( $p < 0.001$ , paired Wilcoxon signed-rank test). In total, 24% of subjects indicated a strictly lower intention to invest if others would find out. Hence, on top of generating identity concerns that materialize in private, the negative image of stockholders might also generate social image concerns.

$x_h > c$  and with 50% probability a low outcome  $x_l < c$  as additional payment. Choosing the safe option yields the endowment as additional payment with certainty. The price and outcomes of the risky option vary between the four decisions of each part.

**Descriptions.** To vary the association of the decisions with the stock market, we use two different ways to describe the risky and safe options to subjects. In the treatment *Stock Description*, the risky option is described as a bet on past stock market prices. Subjects are told that they could buy one share of an (actual, existing) ETF<sup>17</sup> at a price that the ETF had traded at a specific point of time in the past (pay amount  $c$ ). The share would then be randomly sold at one of two subsequent past points in time. We selected the points such that the selling price was higher than the buying price at one point and lower at the other. Subjects would then receive the value of the ETF at the selected selling date as payment (receive  $x_h$  or  $x_l$ ). The safe option is described as abstaining from betting on the stock market (receive the endowment with certainty). See the first column of Table 1.1 on how the options were introduced to subjects and an example of the decision described as a stock investment.

In the treatment *Draw Description*, all references to the stock market are replaced by neutral wording. The risky option is described as a bet on the outcome of a random draw. Instead of an ETF share, subjects could buy a ticket for a random draw (pay amount  $c$ ). One out of two outcomes would then be randomly selected. One outcome was higher than the cost of the ticket, while the other was lower. Subjects would then receive the corresponding amount as payment (receive  $x_h$  or  $x_l$ ). The safe option is described as abstaining from betting on the outcome of a random draw (receive the endowment with certainty). See column 2 of Table 1.1 for illustration.

Importantly, the treatments vary only in how the options are described. The buying prices  $c$  and outcomes  $x_l$  and  $x_h$  are identical between treatments. Moreover, we designed the descriptions to be as similar as possible in terms of complexity and length, changing only the labels of the options. To further reduce the scope for misunderstandings between descriptions, we gave subjects the respective expected value of the risky option. Subjects were thus able to easily compare expected payments between risky and safe options.<sup>18</sup>

**Between-subject variation.** For the first decision part, we randomly assigned subjects either to treatment *Draw Description* (where options are described neutrally),

17. We used year-end share prices of the following four ETF's: iShares Nasdaq 100, MSCI World iShares, iShares MSCI EM and Invesco FTSE RAFI US 1000.

18. Nieddu and Pandolfi (2021) also vary descriptions to study the influence of financial literacy. They describe a simple lottery either as a coin toss or using financial terms such as defaults, net return, current and future value. Thus, to understand the payoff structure of the latter, financial literacy is required. We deliberately described the payoff structure without financial terms in both cases and provided the expected value to ensure that subjects understood both descriptions equally well.

**Table 1.1.** Description experiment illustration of choice options

Stock description	Draw description
<p>In this section, you will make 4 decisions. In each decision, you will separately receive \$30 from us. With this money, you can choose between two options:</p> <p><b>Option A:</b> Participate in the <b>stock market</b> by buying a <b>share</b>. The value of the <b>share</b> depends on the <b>movement of the stock market</b>.</p> <p><b>Option B:</b> Do not participate in the <b>stock market</b>.</p> <p>In each decision, you will have the option to buy a different <b>share</b>. Each <b>share</b> has a different price and offers different <b>returns</b>. The decisions are presented independently of each other. That is, your choice in one decision does not affect the other decisions.</p>	<p>In this section, you will make 4 decisions. In each decision, you will separately receive \$30 from us. With this money, you can choose between two options:</p> <p><b>Option A:</b> Participate in a <b>random draw</b> by buying a <b>ticket</b>. The value of the <b>ticket</b> depends on the <b>outcome of the random draw</b>.</p> <p><b>Option B:</b> Do not participate in the <b>random draw</b>.</p> <p>In each decision, you will have the option to buy a different <b>ticket</b>. Each <b>ticket</b> has a different price and offers different <b>prizes</b>. The decisions are presented independently of each other. That is, your choice in one decision does not affect the other decisions.</p>

Example of Option A:

		Share Price MSCI World ETF
2010	(Buying Price)	\$27.19
2011		\$25.06
2016		\$37.21

Example of Option A:

		Prize
Buying Price		\$27.19
Outcome 1		\$25.06
Outcome 2		\$37.21

*Notes:* The table presents an excerpt of the instructions of the description experiment. The experiment features four choices between two options, Option A and Option B. Subjects first receive the displayed text and a more detailed explanation on the next screen. See Appendix 1.L.2 for the full wording. The table furthermore displays an example of Option A. In the *Stock Description* treatment, choosing Option A in this example means buying the share of an MSCI World ETF to a price of \$27.19 in 2010. A computer then sells the share either to a price of \$25.06 in 2011 or \$37.21 in 2016 with equal probability. In the *Draw Description* treatment, choosing Option A means buying a lottery ticket to a price of \$27.19 that offers a prize of \$25.06 or \$37.21 with equal probability. The alternative in both cases is choosing Option B, in which case subjects receive \$30 with certainty.

or to treatment *Stock Description* (where options are described as stock investments). Comparing choices between treatments thus identifies the effect of varying the association of the options with the stock market.

**Within-subject variation.** After the first decision part, subjects answered a couple of general demographic questions intended as filler questions and subsequently

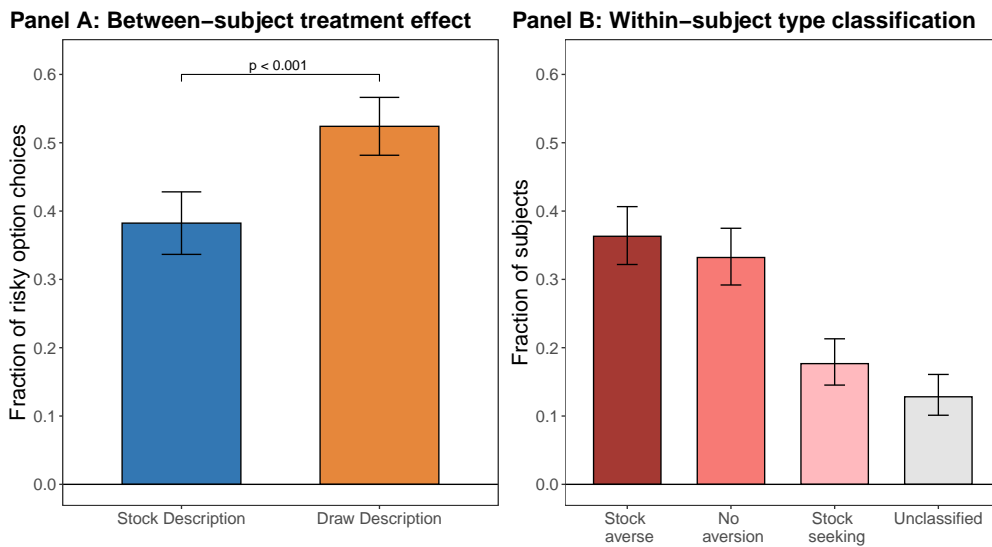
faced the second decision part.<sup>19</sup> Here, subjects again make four decisions. Subjects who received the *Stock Description* before subsequently receive the *Draw Description*, and vice versa. This variation allows us to investigate preferences on the individual level since we observe the same subjects making choices under the different descriptions. In order to avoid consistency effects<sup>20</sup>, we obfuscate that the values of the options are repeated by changing the presentation of the decision's options in two aspects. First, outcomes are visualized in the form of tables in one set of questions and as figures in the other. Second, the currency used is either dollar or the British pound. Since payments on Prolific are always made in pounds and subjects are located in the US, they are familiar with both currencies. Figure 1.A.3 in the appendix displays the four possible visualizations. The visualization and currency used for each decision part is randomized on the subject level independently of the description.

**Validation.** Since the outcome-related features are kept constant between treatments, outcome-based theories of decision-making predict no difference in choice behavior. However, if the risky option in the *Stock Description* treatment is, to a higher degree, associated with stockholders compared to the *Draw Description* treatment and subjects view stockholders negatively, our framework predicts a difference (Prediction 1). We argue that the risky option in the *Stock Description* treatment features a higher association because instead of betting on a random draw, subjects bet on the value of an index traded on the stock market. This option is thus more similar to real-world stock investments compared to the *Draw Description* option. Indeed, we validate that the *Stock Description* risky option is empirically more strongly associated with stockholders: even controlling for risk attitudes, stockholders are significantly more likely to choose the option than non-stockholders. We further show that participants on Prolific hold substantial negative views about stockholders, just as the general US population. For details, see Appendix 1.F.

**Procedure.** At the beginning of the experiment, subjects made two choices between a certain amount and a binary lottery. We use these responses to calibrate the payout values of the risky options in the subsequent decision parts (by picking different selling years for the stock option). Subjects who reveal a high degree of risk averse behavior during the initial two questions receive more favorable risky options later, which reduces the fraction of subjects that either always pick the safe option or the risky option in all decisions. This procedure thus increases the power

19. In the first decision part, subjects were not aware that a second decision part would follow.

20. Subjects could be motivated to choose the same options across decision parts because they want to appear congruous in their choice behavior (Falk and Zimmermann, 2017). This behavior would mitigate the within-subject influence of the descriptions. Naturally, it cannot play a role in the between-subject comparison.



**Figure 1.2.** Description experiment results: aggregate and individual behavior

Notes: **Panel A** displays the between-subject treatment effect of the description experiment. The outcome variable denotes the average fraction where the risky instead of the safe option is chosen. *Stock Description* denotes the treatment in which the risky option is described as a stock investment and *Draw Description* the treatment in which the option is described as a random draw. **Panel B** displays the distribution of types identified by the within-subject analysis. Subjects are *stock averse* if they weakly prefer the risky option described as a random draw to the risky option described as a stock investment, with at least one preference being strict. A preference being strict means choosing the risky option under the draw description and the safe option under the stock description. Subjects are *stock seeking* if they weakly prefer the risky option described as stock to the draw one, with at least one strict preference. Subjects show *no aversion* if they consistently choose the same option between descriptions in all decisions. Remaining subjects that do not show a consistent choice pattern are labeled *unclassified*. Error bars indicate 95% confidence intervals.

of our experimental comparison, but is independent of treatments, ensuring that the outcomes of the draw and stock options were identical as previously described.

**Sample and incentives.** 651 subjects on Prolific with residence in the US completed the experiment. The median completion time was slightly above 7 minutes, and subjects received £0.9 ( $\approx$  \$1.13) for completion. Additionally, ten randomly selected subjects had one of their decisions implemented with real consequences. In total, the average payment was £1.38 per subject ( $\approx$  \$14.80 per hour), which is well above the US federal minimum wage level. In accordance with the preregistration, we exclude 136 subjects who indicated that they are stockholders. Our sample thus consists of 515 subjects. See Appendix Table 1.B.5 for sample demographics.

#### 1.4.2 Description experiment results

**Between-subject analysis.** In total, 260 subjects were part of the *Draw Description* treatment and 255 were part of the *Stock Description* treatment. Appendix Table



1.B.6 shows that treatments were balanced across demographic variables. Panel A of Figure 1.2 displays the treatment effect on decision-making. We find that subjects in the *Draw Description* treatment choose the risky option in, on average, 52% of their decisions (i.e., in 2.10 out of the 4 possible decisions). In comparison, subjects in the *Stock Description* treatment choose the risky option in only 38% of decisions. Hence, once the risky option is described as a bet on the stock market instead of a bet on a random draw, subjects are 27% less likely ( $p < 0.001$ , Wilcoxon rank-sum test) to choose the option, even though the values underlying the option are identical.<sup>21</sup> Panel A of Appendix Table 1.B.7 provides complimentary regression results. In the table, we pool decisions and regress a dummy indicating if the risky option was chosen on a dummy indicating the *Stock Description* treatment. The regression confirms a large and significant treatment effect. Notably, we find that while only 19% of subjects never choose the risky option within the *Draw Description* treatment, this fraction almost doubles to 36% in the *Stock Description* treatment.

**Within-subject analysis.** For the within-subject analysis, we use the choices of both decision parts. We find that the within-subject effect is very similar to the between-subject effect. While subjects choose the risky option in 50% of decisions when described as a bet on a random draw, this fraction decreases to 38% when described as a bet on stock market movements. Panel B of Table 1.B.7 in the Appendix shows the corresponding regression results. We again find a significant effect that is robust to the inclusion of demographic controls.

Utilizing the fact that we vary descriptions within-subject, we can categorize subjects into distinct behavioral types. Since subjects face the same four decisions between risky and safe options under both descriptions, we have four choice pairs that reveal people's preferences. For a given choice pair, we define a strict preference for the random draw if a subject chooses the risky option when it is described as random draw and the safe option when the risky option is described as stock investment. We define a strict preference for the stock investment if the reverse happens: a subject chooses the safe option under the draw description but the risky option under the stock description. Subjects are *stock averse* if they, across the four choice pairs, show a strict preference for the random draw for at least one pair and for none of the pairs a strict preference for the stock investment. Accordingly, subjects are *stock seeking* if they show at least once a strict preference for the stock investment and no strict preference for the random draw. Subjects display *no aversion* if they have no strict preference, meaning they consistently choose the same option within each choice pair. The remaining subjects are labeled *unclassifiedunclassified*.<sup>22</sup> Figure 1.2

21. Reassuringly, neither the form of visualization ( $p = 0.83$ , Wilcoxon rank-sum test) nor the currency ( $p = 0.85$ , Wilcoxon rank-sum test) had any effect on how often the risky choice is chosen.

22. These subjects show at least one strict preference for the stock investment and at least one for the random draw across the four choice pairs.

displays the distribution of types. In total, 36% of subjects are *stock averse*, 18% are *stock seeking*, 33% display *no aversion*, and the remaining 13% show no consistent preferences across descriptions (*unclassified*).

In summary, the description experiment reveals a strong aversion against a choice option if it is associated with the stock market – even if we hold any other aspect of the choice option constant. Hence, we find evidence supporting Prediction 1 of our model.

### 1.4.3 Information experiment

Next, we examine whether subjects' negative perception of stockholders is the mechanism driving the stock market aversion documented in the last section. Specifically, we test Prediction 2 of our framework: a change in people's views of stockholders changes their attitude towards stock investments. To do so, we require an exogenous shift to people's perceptions of stockholders. We thus conduct a second experiment in which we provide subjects with information about the difference in donation behavior between stockholders and non-stockholders. We choose information about donation behavior because it is closely linked to people's perception of traits such as greed and selfishness, and people consider it relevant and informative (e.g., Ariely, Bracha, and Meier, 2009; Grossman and van der Weele, 2017; Bénabou et al., 2022; Exley, Pezzuto, and Serra-Garcia, 2023). Therefore, we expect this information to shift subjects' perceptions of the difference in the traits of selfishness and greed between stockholders and non-stockholders.

**Information generation.** In order to generate data for the information, we ran a separate experiment on Prolific with 272 participants, which we label allocators. Allocators consisted of both stockholders and non-stockholders. They were given \$100, which they could freely divide between themselves and a charity that supports children with critical illnesses. For a randomly selected subset, this donation decision was implemented with real consequences.

**Experimental variation.** The key feature of the design is that we exogenously vary the information that subjects receive. For each subject, we individually and randomly selected ten allocators who indicated that they hold stocks and ten who indicated that they do not hold any stocks. To control for the fact that stockholders often have higher income, which could interfere with subjects' interpretation of the signal, we drew these 20 allocators from the same income bin (using three bins). The entire procedure was transparently described to subjects.<sup>23</sup> As informa-

23. They were also informed about the relationship between behavior in the donation decision and other relevant real-life behavior. For example, we explained that previous research has shown that people who donate more in such a decision are also more likely to do voluntary work (e.g. Falk et al., 2018).

tion, subjects received the percentage difference in donation behavior between the randomly drawn stockholders and non-stockholders. Specifically, subjects were presented the following sentence: “For the randomly selected participants, we found that [GROUP 1] donate [X]% more than [GROUP 2].” See Appendix Figure 1.A.4 for a screenshot. Consequently, some subjects receive the information that stockholders donated more than non-stockholders, while others learn that in their draw non-stockholders donated more with different percentage differences. To ensure that subjects paid attention, they had to repeat the information on the subsequent page.

**Treatment variables.** For the analyses, we use the signal about differences in donation behavior as independent variables in two different ways. First, a dummy variable equal to one if the subject received the signal that stockholders donated more than non-stockholders, and equal to zero if they donated a lesser or an equal amount.<sup>24</sup> Second, the full signal as continuous variable where higher values indicate that the selected stockholders donated more relative to the selected non-stockholders.

**Outcome variable.** To cleanly identify the causal effect of varying perceptions on financial decision-making, we make use of the controlled decision environment of the previous experiment (Section 1.4.1). After receiving the information about donation behavior, subjects made four incentivised decisions between a safe option and a risky option described as a stock investments (stock option). Our main outcome variable is thus subjects’ willingness to invest in the stock option.

We chose the same choice paradigm as in the last experiment to fix the investments’ outcomes and probabilities. Thus, even if subjects update about secondary factors unrelated to their perceptions of stockholders such as their beliefs of the profitability or riskiness of stock investments, this secondary updating should not influence decision-making in our experiment. Furthermore, since all subjects receive information, we hold constant factors such as priming or attention that could influence decision-making independent of the information’s content. We are, hence, confident in assuming that outcome-based utility  $u_i(z_I)$  is unaffected by our treatment variation, allowing us to directly observe the causal effect of varying perceptions about characteristics of stockholders compared to non-stockholders  $\hat{\mathbb{E}}_i[\bar{q} | S] - \hat{\mathbb{E}}_i[\bar{q} | NS]$ .

**Procedure.** At the start of the experiment, subjects made two calibration choices and then four incentivized choices between a safe option and a risky option described as a random draw using the same procedure as in Section 1.4.1. This allows

24. In total, less than 5% (25 subjects) received the signal that both groups donated an equal amount. We chose to pool them to maintain a dichotomous variable. Our results are quantitatively very similar if we pool these subjects with those that received the signal that stockholders donated more or if we exclude them altogether from the analysis.

us to use subjects' decisions in a non-stock related setting as a control when analyzing the effect of the information, increasing statistical power. Subjects subsequently state their prior belief over the difference in donation behavior between stockholders and non-stockholders, and then receive the signal about the difference in donation behavior. After receiving this information, subjects made four incentivised decisions between a safe option and a risky option described as a stock investments, as described previously. Finally, we elicited subjects' perceptions of stockholders and non-stockholders (posterior belief) using the module developed in Section 1.3.

**Sample and incentives.** 652 subjects on Prolific with residence in the US completed the experiment. Median completion time was 9.5 minutes and subjects received £1.2 for completion. Additionally, ten randomly selected subjects had one of their decisions implemented with real consequences. In total, average payment was £1.71 per subject ( $\approx$  \$13.50 per hour). In accordance with the preregistration, we exclude 104 subjects who indicated that they are stockholders. Our sample thus consists of 548 subjects. See Appendix Table 1.B.5 for sample demographics.

#### 1.4.4 Information experiment results

**Signal distribution and prior beliefs.** In total, 61% of subjects received the information that non-stockholders donated more than stockholders or that they donated the same amount, while 39% of subjects received the information that stockholders donated more. The mean of the full signal variable is -10% (median -6%), but with substantial variation, see Appendix Figure 1.A.5 for the whole distribution. Replicating the results of Section 1.3, we find that prior to receiving the actual difference, subjects believe the randomly selected group of stockholders donate 15 pp. less than the group of non-stockholders. Appendix Figure 1.A.6 displays the distribution. Reassuringly, prior beliefs were not correlated with the randomly generated signal ( $r = 0.01, p = 0.89$ ). Table 1.B.8 in the appendix further shows that signals were not correlated with demographic variables either.

**Effect on posterior beliefs.** We start by investigating the impact of the signal on posterior beliefs. In Table 1.2, we regress differences in subjects' rating of the selfishness (column (1)) and greed (column (2)) of stockholders compared to non-stockholders on the signal about the difference in donation behavior between both groups. As intended, the information significantly shifts perceptions: the higher the signal, the lower the subject's negative perception of stockholders. At the same time, we observe only a limited and insignificant impact of the signal on differences in ratings of the trait "gambler" (column (3)), which is reassuring as the information provided was not related to this trait.

**Effect on behavior.** Our main variable of interest is subjects' likelihood of choosing the stock option. Panel A of Figure 1.3 displays the results for the binarized signal

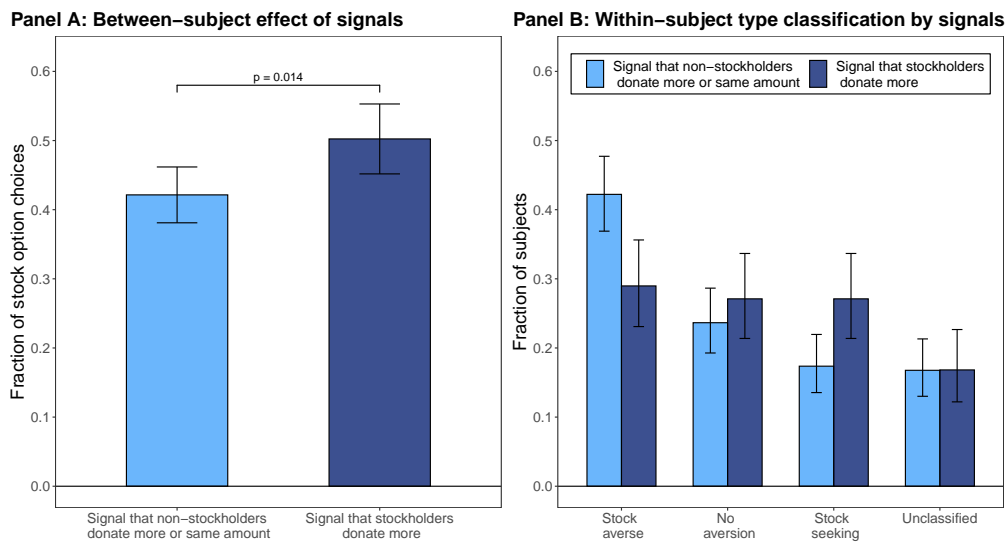
**Table 1.2.** Information experiment treatment effect on negative perceptions

	<i>Dependent variable:</i>		
	Perceptions about stockholders		
	Selfishness	Greed	Gambler
	(1)	(2)	(3)
Signal over donation behavior	-0.190*** (0.034)	-0.150*** (0.033)	-0.049 (0.031)
Constant	1.516*** (0.132)	1.963*** (0.130)	5.010*** (0.152)
Prior beliefs	X	X	X
Observations	548	548	548
R <sup>2</sup>	0.071	0.053	0.028

Notes: The table displays OLS-estimates. The dependent variable is the difference in perceptions between stockholders and non-stockholders, with higher values indicating that stockholders are perceived to be more selfish in column (1), greedy in column (2) and gambler in column (3). Signal over donation behavior denotes the signal received over the difference in donation behavior. Higher values indicate a higher signal in the direction that stockholders donate more relative to non-stockholders, with the unit being 10% differences. That is, a one unit increase means a signal that stockholders donate 10% more than non-stockholders. Prior beliefs is subjects belief over the differences in donation behavior between stockholders and non-stockholders. Robust standard errors in parentheses. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

variable. Subjects who receive the information that non-stockholders donated more choose the stock option in 42% of cases. This number increases to 50% for those subjects who learn that stockholders donated more ( $p = 0.014$ , Wilcoxon rank-sum test). In Panel B of Figure 1.3, we repeat our classification of subjects into the four behavioral types (*stock averse*, *stock seeking*, *no aversion*, *unclassified*). When subjects receive the information that stockholders donate more than non-stockholders, they are significantly less likely to be *stock averse* ( $p = 0.002$ , two sample test of proportions), and instead are more likely to show *no aversion* or *stock seeking* behavior.<sup>25</sup> At the same time, the fraction of unclassified subjects is not affected by the information. Therefore, we obtain causal evidence that information about differences in donation behavior influences stock investment behavior.

25. Similarly, we see a significant decrease in the fraction of subjects who refuse to choose the stock option in any decision ( $p = 0.039$ , two sample test of proportions).



**Figure 1.3.** Information experiment results: aggregate and individual behavior

*Notes:* **Panel A** displays the effect of varying the signal in the information experiment. The outcome variable denotes the average fraction that the risky option described as a stock investment instead of the safe option is chosen. The left bar displays the choice behavior of subjects who received a signal that non-stockholders donated more or the same amount compared to stockholders. The right bar displays behavior for subjects who received a signal that stockholders donated more than non-stockholders. **Panel B** displays the distribution of types identified by the within-subject analysis. Subjects are *stock averse* if they weakly prefer the risky option described as a random draw to the risky option described as a stock investment, with at least one preference being strict. A preference being strict means choosing the risky option under the draw description and the safe under the stock description. Subjects are *stock seeking* if they weakly prefer the risky option described as stock to the draw one, with at least one strict preference. Subjects show *no aversion* if they consistently choose the same option between descriptions in all decisions. Remaining subjects that do not show a consistent choice pattern are labeled *unclassified*. Error bars indicate 95% confidence intervals.

In Table 1.3, we include the full distribution of signals, controlling for subjects' prior beliefs and their respective choice when options are described as a random draw. The dependent variable is the choice of the risky option when described as a stock investment. The independent variable is the signal that subjects receive over the difference in donation behavior between stockholders and non-stockholders as continuous variable, with higher values indicating that stockholders donate more relative to non-stockholders. We find a significant effect of the signal on behavior: the more positive the information that subjects receive regarding the donation behavior of stockholders, the higher the likelihood that they choose the stock option. More specifically, a 10% increase in the signal increases the likelihood by 0.9 pp. This effect is sizable, because a one standard deviation (40%) increase in the signal helps to decrease the description effect found in the previous experiment by 25% (3.6 out of 14.2 pp.).

**Table 1.3.** Information experiment treatment effect on decision-making

	<i>Dependent variable:</i>	
	Choice of stock option	
	(1)	(2)
Signal over donation behavior	0.009** (0.004)	0.009** (0.004)
Choice of draw option	0.239*** (0.025)	0.240*** (0.025)
Mean dep. variable	0.45	0.45
Demographic controls		X
Prior beliefs	X	X
Subjects	548	541
Observations	2,192	2,164

*Notes:* The table displays OLS-estimates. All four binary choices between the risky option described as a stock investment and the safe option of a subject enter as separate observations. The dependent variable is an indicator whether the risky option described as a stock investment is chosen. “Signal over donation behavior” denotes the signal subjects receive regarding the difference in donation behavior between non-stockholders and stockholders. Higher values indicate that subjects receive the signal that stockholders donate more relative to non-stockholders, with the unit being 10% differences. “Choice of draw option” is an indicator of whether the risky option is chosen when described as a random draw investment. “Prior beliefs” refers to subjects’ belief over the difference in donation behavior between stockholders and non-stockholders. “Choice of draw option” and “Prior beliefs” are both elicited before subjects receive the signal. Demographic controls include age, gender, education, income, total financial assets and involvement in financial decision-making. Standard errors (in parentheses) are clustered at the subject level. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

**Heterogeneity.** In Appendix 1.G, we investigate heterogeneity in the effect of the information on behavior. Since we elicited people’s prior beliefs about the difference in donation behavior, we can compare subjects that received positive or negative information relative to their priors with subjects that received information confirming their priors. We find, as expected, that receiving positive information about stockholders increases stock option choices, while receiving negative information decreases it. However, the effect of information is not symmetric: positive information has a stronger effect on behavior than negative information of the same magnitude.

**Instrumental variable analysis.** Complementary to our regression analysis, we use the exogenously assigned signal as an instrumental variable to estimate the causal effect of perceptions on financial decision-making. The analysis rests on the assumption that the documented effect of our treatment on financial decision-making operates solely through changes in perceptions of stockholders. This assumption seems plausible, as by design, the identifying variation comes from differing information about the difference in donation behavior between stockholders and non-stockholders. Empirically, as indicated in Table 1.2, we have a strong first stage. For the two treated traits of greed and selfishness, the respective F-statistic are 21.01 and 32.24. Table 1.4 displays the results of the 2SLS-regressions. Columns (1) and (2) focus on perceptions of greed. A one standard deviation increase in subjects' perceptions of stockholders, induced by the signal, decreases the likelihood of investing in the stock option by about 16 pp. Columns (3) and (4) reveal that when focusing on perceptions of stockholders' selfishness, the effect is also statistically significant and very similar. We further find that both effects are stronger than the reduced form effects obtained by regressing the likelihood of investing in the stock option on negative perceptions of greed and selfishness. In these specifications capturing the correlational effect, a one standard deviation increase in negative perceptions is associated with a 5 pp. decrease in subjects' likelihood to choose the stock option.

#### 1.4.5 Follow-up survey for robustness

We conducted a follow-up survey after the information experiment to (1) address the concern that experimenter demand effects drive our results, (2) assess the persistence of our effects on perceptions, and (3) investigate whether the information changed subjects' intentions to invest their own money in the stock market. Experimenter demand effects occur when participants try to guess the experimenter's objective from the instructions and alter their behavior accordingly. However, since our primary outcome is incentivized, previous literature suggests that demand effects are unlikely to drive our results.<sup>26</sup> Nevertheless, to address this concern in the context of our study, we obfuscated the follow-up survey. The idea is to run a separate study where the same subjects are invited, but they are unaware that the studies are connected. If this obfuscation is successful, demand effects cannot drive effects found in the follow-up study.

**Design.** Two days after conducting the information experiment, we fielded another study via Prolific. We invited only subjects of the main experiment, utilizing that subjects on Prolific regularly receive survey invitations. In the survey, we

26. Experimental evidence shows that demand effects often have little impact on responses (De Quidt, Haushofer, and Roth, 2018; Danz et al., 2023) even for hypothetical questions where it is presumably less costly for subjects to alter their answers relative to their "natural" choice (Mummolo and Peterson, 2019).



**Table 1.4.** Information experiment causal effect of negative perceptions on decision-making

	<i>Dependent variable:</i>			
	Choice of stock option			
	(1)	(2)	(3)	(4)
Perceptions about stockholders' greed	-0.161** (0.075)	-0.165** (0.075)		
Perceptions about stockholders' selfishness			-0.135** (0.061)	-0.140** (0.062)
Choice of draw option	0.233*** (0.026)	0.233*** (0.026)	0.235*** (0.025)	0.234*** (0.026)
Mean dep. variable	0.45	0.45	0.45	0.45
F-statistic first stage	21.01	21.26	32.24	32.16
Demographic controls		X		X
Prior beliefs	X	X	X	X
Subjects	548	541	548	541
Observations	2,192	2,164	2,192	2,164

*Notes:* The table displays 2SLS-estimates. All four binary choices between the risky option described as a stock investment and the safe option of a subject enter as separate observations. The dependent variable is an indicator whether the risky option described as a stock investment is chosen. The instrument is the signal received regarding the difference in donation behavior between stockholders and non-stockholders (see Table 1.2 for the first stage results). The instrumented variable is in columns (1-2) "Perceptions about stockholders' greed" and in (3-4) "Perceptions about stockholders' selfishness". Both denote difference in ratings between stockholders and non-stockholders, with higher values indicating that stockholders are rated to more selfish (1-2) and more greedy (3-4). "Choice of draw option" is an indicator whether the risky option is chosen when described as a random draw investment. "Prior beliefs" is subjects' belief over the difference in donation behavior between stockholders and non-stockholders. Demographic controls include age, gender, education, income, total financial assets and involvement in financial decision-making. Standard errors (in parentheses) are clustered at the subject level. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

elicited subjects intention to invest into the stock market in the future using a 7-point Likert scale. We elicited perceptions by asking subjects how selfish they view non-stockholders compared to stockholders (9-point Likert scale).

**Obfuscation.** To obfuscate the follow-up survey, we altered the survey's description, visual style, responsible researcher and institution. We also embedded our variables of interest within a battery of questions on other topics. In particular, we elicited subjects intention to engage in a variety of behaviors such as the intention to behave more environmentally friendly or buy disability insurance. Similarly, our perception

question was embedded in a larger battery of questions asking subjects about different groups and personality traits. Hence, in both cases did our variable of interest appear to subjects as one of many variables. Appendix 1.H provides further details.

**Obfuscation validation.** By offering high incentives (an extrapolated hourly wage of \$18.65), we were able to recruit 428 subjects (78%) of the main experiment for the follow-up survey.<sup>27</sup> Between the main experiment and the follow-up survey, subjects completed, an average of 15 other studies. At the end of our survey, we asked subjects to indicate the number of similar studies they had completed in the past two weeks. In total, 82% of subjects answered with “none”, indicating that we successfully created a survey that was perceived being distinct from the main study.<sup>28</sup> Furthermore, not a single subject of the follow-up referenced our main experiment when asked “If you had to guess, what would you say was the purpose of this study?”. It thus appears that our obfuscation measures were successful in creating a survey that subjects perceive as distinct from the main experiment.

**Results.** We find that subjects who received information that stockholders donated more for a good cause are significantly more willing to consider investing in the future ( $p = 0.020$ , Wilcoxon rank-sum test) compared to those that received the information that non-stockholders donated more or the same amount. They also view non-stockholders significantly more selfish compared to stockholders ( $p = 0.005$ , Wilcoxon rank-sum test). We find slightly weaker but still persistent and significant effects when we look at the full continuous signal variable. As placebo check, we also assess the influence of the information on the other non-stock market related variables of the survey. Reassuringly, we generally find no systematic effect of the information on intentions and perceptions of other traits and groups. We show the details of these results in Appendix 1.H.

In summary, our data confirm Prediction 2: perceptions about characteristics of stockholders causally change the attractiveness of choices associated with the stock market and influence subjects’ investment intentions.

## 1.5 The association between perceptions of stockholders and stock market participation

Having established that negative views about stockholders causally influence financial decisions in experiments, we next turn to field data for external validity. In this

27. We see no evidence of selection effects. Whether subjects participated in the follow-up or not was not correlated with whether they received positive or negative information about stockholders ( $r = -0.02, p = 0.65$ ), the magnitude of the signal ( $r = 0.01, p = 0.78$ ) or prior beliefs ( $r = -0.06, p = 0.17$ ). Demographics were similarly balanced.

28. Our results are similar if we only consider these subjects in the analysis.

section, we show that negative views are (i) predictive of households' actual stock market participation and (ii) a quantitatively important determinant of stock market participation relative to leading alternative determinants proposed by the literature.

### 1.5.1 Data

To investigate the relationship between people's views about stockholders and their stock market participation, we require individual-level asset data that can be related to our survey-based perception measure. The LISS panel is ideal for this purpose, because it allows us to link survey responses with tax record data provided by Statistics Netherlands. Thus, for each subject that consented to the linkage (89% of our sample), we observe their respective households' financial asset allocation.<sup>29</sup> To assess how well our explanation predicts stock market participation relative to the previous literature, we utilize the broad scope of the LISS panel. By linking our data to previous surveys, we obtain several preference, attitude and belief measures. We focus on variables that the literature has identified as the most important predictors of households' portfolio choices, namely risk aversion, beliefs about stock returns, financial numeracy, general trust, political orientation, ambiguity aversion, and likelihood insensitivity (full set available for 46% of our sample). Each measure was elicited using state-of-the-art elicitation methods; see Appendix 1.K.3 for details. In our cross-country sample, we naturally do not have access to these measures and rely on self-reported stock market participation. Nevertheless, this sample allows us to generalize our findings to different countries.

### 1.5.2 Results

**Netherlands.** We regress a dummy variable indicating whether the household holds any risky financial assets on their standardized views of stockholders using OLS.<sup>30</sup> Table 1.5 displays the results. Column (1) reveals that the more negatively subjects view stockholders compared to non-stockholders, the less likely they possess risky financial assets themselves: an increase in negative views about stockholders by one standard deviation is associated with decreased likelihood of possessing risky financial assets by almost 6 percentage points. This is a substantial effect as the baseline likelihood is 23 percentage points. Column (2) reveals that the relation is only slightly smaller and remains significant when we control for the demographic variables age, gender, education, income, and wealth. In columns (3) and (4), we focus on the smaller sample of subjects for which we have measures of behavioral vari-

29. Whether we can link a subject to administrative data is unrelated to their views about stockholders and the demographic variables we observe in the LISS: age, gender, and education (more details in Table 1.B.13).

30. Probit regressions yield similar results, see Table 1.B.15.

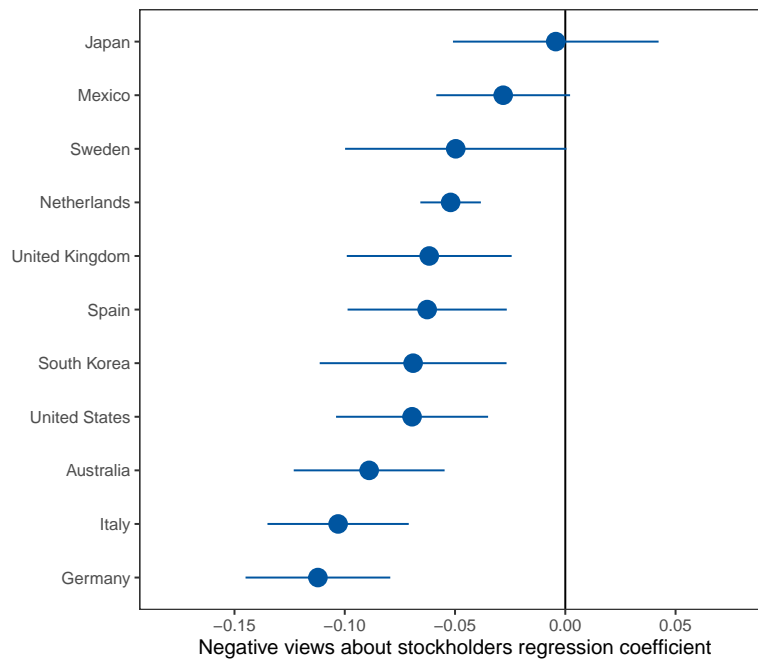
ables that possibly influence portfolio choice.<sup>31</sup> We replicate the relationship with stock market participation documented by the literature for all other behavioral predictors with the exception of financial numeracy and political orientation.<sup>32</sup> The effect of negative perceptions is still substantial and significant. We find that a one standard deviation increase in negative views is associated with a 4.8 percentage point decrease in the likelihood of owning risky financial assets. The marginal effect is larger than the marginal effect of, for instance, risk or ambiguity aversion. Views about stockholders are, therefore, among the strongest behavioral predictors.<sup>33</sup> In Appendix Table 1.B.9, we focus on the set of households that hold a non-zero amount of risky financial assets and use as the dependent variable the share of risky financial assets of total financial assets. Since decisions over the share of one's portfolio in stocks conditionally on owning stocks are associated with the group of stockholders, our framework predicts that negative perceptions of stockholders should play no, or only a limited role. Indeed, we find no significant association between the share invested in risky financial assets and negative perceptions of stockholders in the full sample and only a small association in the determinants sample.

**Cross-country.** Do negative perceptions of stockholders predict stock market participation in other countries as well? Figure 1.4 displays for each country the result of an OLS regression with negative views of stockholders as independent and a dummy variable indicating whether a subject holds any risky financial assets as dependent variable. We find that negative views of stockholders predict participation in all countries except Japan, with seven out of ten countries having a coefficient significantly different from zero at the 5% level. Running a regression on the full set of countries with country fixed-effects, we find that a one standard deviation increase in negative perceptions is associated with a 6.9 percentage point decrease in stock market participation (see Appendix Table 1.B.4). Hence, our findings from the Netherlands generalize to our larger set of surveyed countries.

31. The determinants sample differs from the full sample along some demographic variables (e.g. age). This is expected, as most of the variables are elicited in earlier questionnaires. Importantly, negative views about stockholders and stock market participation rates do not vary significantly between the samples (more details in Appendix Table 1.B.14).

32. Without including other variables, financial numeracy and political orientation significantly predict whether households own risky financial assets. However, once we include the other behavioral and demographic variables in the regression, financial numeracy and political orientation are no longer significant predictors.

33. A complementary approach to comparing the coefficients of the variables is to compare how well each variable explains variation in stock market participation. To do so, we apply the Shapley value method (Lipovetsky and Conklin, 2001). This method provides the marginal contribution of each independent variable in explaining variance in the dependent variable by aggregating the  $R^2$  of each possible combination of independent variables. The results further support the relevance of perceptions: with the exception of the return belief variable, perceptions explain the highest share of variance among the behavioral variables presented in Table 1.5.



**Figure 1.4.** The relationship between negative views about stockholders and stock market participation across countries

*Notes:* This figure shows OLS coefficients from regressing a dummy variable indicating whether subjects' report to hold any risky financial assets on their (standardized) negative views about stockholders, controlling for age and gender. Negative views about stockholders is defined as the mean of the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. For the Netherlands the coefficient of the specification of column (2) in Table 1.5 is used. Bars indicate 95% confidence intervals.

**Heterogeneity.** Among which demographic groups is the effect of negative perceptions on stockownership concentrated? We use the rich demographic data from the LISS panel and interact them with our negative perception measure. We find that the effect of negative perceptions is more strongly associated with stock-ownership among wealthy, male, educated and older subjects (see Figure 1.A.7 in the Appendix for details). Our identity-based explanation thus can descriptively account for the fact that even wealthy and educated households abstain from participating in the stock market.

Overall, we conclude that the extent to which individuals hold negative perceptions of stockholders robustly predict their stock market participation. The estimated effect size is large, also relative to the effect size of other predictors brought forward by the literature, and present among a diverse set of countries.

## 1.6 Determinants and implications of perceptions of stockholders

This section first investigates the origins of the documented negative perceptions of stockholders. In Section 1.6.1, we test the idea that negative perceptions emerge from stereotypical beliefs about stockholders, applying insights from the stereotypes model of Bordalo et al. (2016). We then explore important implications of negative perceptions of stockholders beyond investment decisions. They are related to subjects (i) favoring non-stockholders over stockholders, (ii) supporting policies that are less favorable towards stockholders (both Section 1.6.2) and (iii) misreporting financial assets in surveys (Section 1.6.3).

### 1.6.1 Determinants and accuracy of perceptions of stockholders

As starting point in examining the origin of negative perceptions of stockholders, we investigate whether they are based on accurate or stereotypical beliefs, i.e., correct or overly negative representations of reality. To assess their accuracy, we need to compare actual differences between stockholders and non-stockholders with subjects' predictions over these differences.

**Actual differences.** We measure actual differences by asking subjects in the LISS panel to rate themselves with respect to the selected traits greedy, gambler, and selfish (agreement to "I'm kind of greedy/a gambler/selfish" 0-10 Likert scale). Comparing the answers of stockholders and non-stockholders gives us a direct measure of the differences in self-assessments. We indeed find differences between the two groups, as displayed in Panel A of Table 1.6. Stockholders rate themselves on average as being significantly more greedy, selfish, and as gamblers compared to non-stockholders ( $p < 0.001$ , Wilcoxon rank-sum test).

**Predicted differences.** We measure predicted differences by asking subjects to guess how stockholders and non-stockholders rate themselves on average for each of the three character traits. By comparing actual with predicted differences, we can thus test whether subjects have biased perceptions.<sup>34</sup> Panel B of Table 1.6 shows the results. While subjects are, on average, quite close to the true average for non-stockholders, they systematically overestimate the answers of stockholders. Actual differences are significantly exaggerated, ranging from 9% for selfishness to nearly 120% for gambler-like (Panel C). Thus, subjects in our representative sample

34. Such a clean comparison would not be possible with our negative views about stockholders measure, as it concerns subjects' subjective opinions of stockholders and non-stockholders, for which no objective truth exists. Empirically, negative views about stockholders and predictions about response behavior are highly correlated ( $r = 0.42$ ).

of Dutch financial deciders incorrectly believe that stockholders assess themselves more negatively than they actually do.

**Formation of stereotypes.** How do the documented stereotypical representations of reality emerge? The stereotypes model of Bordalo et al. (2016) provides a potential explanation. In the model, the formation of stereotypes is linked to the representativeness heuristic by Tversky and Kahneman (1983). When assessing groups, people base their judgment on attributes that are diagnostic and thus representative of a group. Applied to our context, negative perceptions of stockholders emerge because stockholders are *relatively* more likely among very greedy (and selfish and gambler, respectively) individuals. Subjects then perceive individuals holding stocks to be highly representative of greedy individuals, leading them to (vastly) overestimate the existing group differences.<sup>35</sup>

We indeed find evidence for the explanation proposed by the stereotypes model. While average differences are quite small, we find them substantially more pronounced in the tails of the distribution. Stockholders are twice as likely to rate themselves as very greedy, gambler-like, and selfish (7-10 on the 10-point scale) compared to non-stockholders (see Panel A of Appendix Table 1.B.12). At the same time, there is almost no difference between the groups among the moderately greedy, gambler, and selfish (1-3 on the 10-point scale) subjects.

**Validation.** In order to show that the previous results extend beyond self-assessments, we employ a behavioral measure with US participants on Prolific. We ask stockholders and non-stockholders to allocate 100€ between themselves and a charity ( $N = 272$ , sample used to generate information for the information experiment of Section 1.4.3) and a separate sample to predict the resulting difference in donation behavior between the two groups ( $N = 652$ , information experiment full sample). We find that stockholders donate, on average, 6 percentage points less compared to non-stockholders. These differences are again concentrated in the tails. While both groups do not differ in their likelihood of taking between 51€ and 99€ for themselves, stockholders are over 75% more likely to allocate the entire endowment to themselves (see Panel B of Appendix Table 1.B.12). In line with the stereotypes model, we find a large exaggeration of the true difference. On average, subjects believed stockholders donate 15 percentage points less, more than twice the actual difference.

35. We believe this explanation to be intuitively compelling in our context because a small subclass of stockholders, namely traders and brokers, is the main focus of popular media's display of stockholders. These individuals are usually painted as extremely greedy, selfish, and gambler-like. Consequently, people might view them as being representative of stockholders in general and form negative perceptions of the entire group of stockholders.

**Correcting stereotypical perceptions.** Since perceptions of stockholders are overly negative, an immediate policy question is whether a debiasing intervention could influence people’s investment behavior. We thus tested in our second survey fielded to the LISS panel the effectiveness of providing subjects with information on the actual differences between stockholders and non-stockholders. A randomly selected half of the participants receive information on the actual differences in self-assessments for the traits greed, gambler, and selfishness, while the other half received no new information. Afterward, subjects faced a (probabilistically incentivized) investment choice, in which they could allocate 100 € between a safe option and an ETF that tracks the Amsterdam Exchange index.

We find that this information successfully mitigates biased beliefs. Subjects with biased prior perceptions hold less bias perceptions after receiving information and subjects with approximately correct prior perceptions do not change their perceptions. This pattern translates to behavior. Subjects in the former group are now significantly more likely to invest money into the ETF compared to subjects in the control group with similar prior perceptions, while the investment behavior of the latter group does not significantly differ. These findings suggest that interventions that provide information on actual differences between stockholders and non-stockholders have the potential to effectively increase investment behavior without producing backlash effects. See Appendix 1.I for details on the experiment and results.

### 1.6.2 Affective polarization and political attitudes towards stockholders

In Sections 1.4 and 1.5, we focused on the influence of negative views about stockholders on investment behavior. However, research on social identity has shown that identification processes affect not only behavior, but also attitudes. We would therefore expect that negative views about stockholders predict attitudes towards the stock market more generally. Consequently, we investigate the relationship of negative views about stockholders with affective polarization, i.e., animosities towards stockholders and with political attitudes towards stockholders and the stock market in our second survey fielded to the LISS panel.

**Affective polarization.** Does identification with the group of non-stockholders translate into differential behavior towards stockholders, resulting in affective polarization due to in-group favoritism? Building on Enke, Rodríguez-Padilla, and Zimmermann (2022) and related work, we measure in-group favoritism using an allocation decision. We endow subjects with 100€, which they can freely distribute between two other participants of the LISS panel. One participant is a stockholder, the other a non-stockholder, and we inform subjects that both have a similar amount of income and wealth. Since stockholders are, on average, wealthier than non-stockholders, this feature allows us to abstract from animosities towards individuals of higher status. This decision was incentivized, as for one randomly selected subject, the allocation decision was implemented with real consequences. To bench-



mark subjects in-group favoritism towards non-stockholders and control for general in-group tendencies, we subsequently employ the Moral Universalism short-scale (Enke, Rodríguez-Padilla, and Zimmermann, 2022) in the survey.

We observe substantial in-group favoritism among non-stockholder, resulting in strong discriminatory behavior against stockholders. Non-stockholders allocate, on average, 67.23€ out of the 100€ endowment to the recipient not holding stocks. This degree of favoritism is even stronger than subjects' favoritism towards individuals of their own nationality (compared to individuals living anywhere in the world), to whom they distribute, on average, 62.08€. Column (1) in Table 1.7 reveals that subjects' in-group favoritism towards non-stockholders is significantly influenced by their views about stockholders. An increase in negative views by one standard deviation is associated with an 0.1 standard deviations increase in in-group favoritism towards non-stockholders. Furthermore, column (2) shows that when controlling for several demographic variables and, in particular, their in-group favoritism along the other dimensions of the universalism scale (higher values indicate more universalism, i.e., less in-group favoritism), the relation remains largely unchanged. That is, even comparing subjects that show similar degrees of general in-group favoritism, we find a strong relationship between perceptions and allocations to non-stockholders.

**Political attitudes towards stock market.** To measure the relation of negative views with political attitudes towards the stock market, we elicit subjects' support for five policy proposals concerning the taxation and regulation of stockholders and the stock market (using 7-point Likert scales). For example, we ask subjects whether they support the introduction of a financial transaction tax and whether the existing wealth tax in the Netherlands should be higher for investments in risky assets than for safe assets (Table 1.K.2 in the Appendix reports the full list). We aggregate the items to obtain a measure for which higher levels indicate political preferences that are less favorable for stockholders in terms of regulation and taxation.

We find substantial variation in non-stockholders' attitudes towards all five policy proposals. For instance, 26% support introducing a financial transaction tax, while 44% oppose it, and 43% support a higher wealth tax for risky assets, to which 32% oppose. Importantly, the extent of negative views about stockholders moderate these polarized opinions. As Column (3) in Table 1.7 shows, the more negatively subjects view stockholders, the more strongly they support policies that are less favorable for stockholders. In column (4), we control for several measures capturing subjects' redistributive concerns. In particular, we elicited subjects' support for an increase in wealth taxation (independent of asset classes), their self-reported political orientation and opposition to income inequality. Negative views about stockholders remains a strong and significant predictor of support for policies favoring non-stockholders.

### 1.6.3 Misreporting in surveys

As administrative data on wealth is not available in most countries, researchers often need to rely on self-reported asset data. Negative views about could lead individuals to misreport having risky financial assets to avoid identity conflicts. Suppose, for instance, somebody has received stocks not by choice but through inheritance, a gift, or some company participation program. If this person perceives stockholders negatively, merely thinking about their stock-related assets might create disutility (Bénabou and Tirole, 2011). Moreover, reporting stock ownership in a survey makes the affiliation to this group explicit and “official” to the individual, which could create an identity conflict. Some might resolve this conflict by simply denying holding any risky financial assets.

**Measurement.** Since we observe both administrative and self-reported survey data, we can test whether negative perceptions predict directional misreporting. As described before, for each subject we observe their households’ asset holdings identified through tax records. At the same time, the LISS panel surveys every adult member of the respective subject’s household, and asks for their asset holdings. We aggregate reports on the household level to match administrative records.<sup>36</sup> Since the survey uses the same asset categories and subjects are asked to state the balance at exactly the same time point for which we have the administrative data balance, deviations are likely to be caused by misreporting.

**Aggregate results.** We first note that for 7% of the households, the dummy variable indicating if the household reports holding any risky financial assets in the LISS contradicts the respective variable in the administrative data. This result is in line with previous findings, which find substantial response errors in self-reported income and asset data (e.g. Hill, 2006; Meyer, Mok, and Sullivan, 2015; Bollinger et al., 2019). Importantly, the differences are asymmetric: 30% of households with risky assets do not report their holdings, while only 2% of the households without risky assets report having them. This leads to a severe under-reporting of risky assets. Based on the self-reported data only 16% of all households report holding any risky financial assets even though 21% actually own them based on administrative data.<sup>37</sup> At the same time, for those households that correctly report their stock-ownership status, there is no asymmetric misreporting in the share of assets invested in risky financial assets. While households on average invest 39% of their total financial assets in stock market related assets, their reported share is 40%.

36. Only the financial decider is asked to report assets jointly owned by the household which ensures that assets are not counted multiple times when we aggregate the individual data.

37. The numbers in this paragraph are based on the sample of households (65 %) for which we observe self-reported asset data for all adult household members. In Table 1.8, we use the full set of households and add a dummy indicating whether we observe all adult household members as control variable.

**Predicting misreporting.** In Table 1.8, we show that negative perceptions of stockholders predict the documented under-reporting of risky financial assets. We focus on the subset of households that hold risky financial assets based on official data and use as the dependent variable an indicator of whether they (correctly) state that they hold them in the survey. As column (1) reveals, the higher a subject's negative perceptions, the more likely the subject self-reports that they do to hold any risky financial assets. This effect is economically important: a one standard deviation increase in negative perceptions is associated with a more than five percentage point decrease in reporting ownership of risky assets. When we include demographic controls and financial numeracy in column (2), the coefficient is still on a similar level and statistically significant at the 10% level. We furthermore examine whether negative perceptions are related to survey response error per se. Negative perceptions neither predict misreporting on the intensive margin of risky financial assets (Appendix Table 1.B.16) nor misreporting in house ownership. Both findings are in line with the identity concerns mechanism. Thus, our results suggest that negative perceptions are related to misreporting due to identity concerns.

## 1.7 Conclusion

This paper proposes that people's views of stockholders matter for their investment decisions. Using a series of surveys, we document that a large fraction of individuals perceive stockholders negatively. We then show experimentally that these negative perceptions of stockholders causally influence people's decision-making. Furthermore, they are an important predictor of actual stock market participation, which we show using linked survey-administrative data. Moreover, perceptions significantly contribute to affective polarization, political attitudes toward stockholders, and misreporting behavior in surveys.

Our results highlight limitations to the persuasive power of outcome-based strategies to influence households' financial decisions and offer perspectives on the design of alternative, potentially more effective strategies. The strong aversion to any form of stock-related options we document in our experiments suggests that classic strategies, such as highlighting potential gains from investing or improving knowledge might have limited effects on the substantial share of highly identity-minded households. Indeed, a meta-analysis covering 201 studies indicates that interventions to improve financial literacy generally have very limited effects on behavior (Fernandes, Lynch, and Netemeyer, 2014). Similarly, interventions or advertisements that appeal to the skills and knowledge of financial experts might not convince households whose decisions are highly motivated by identity concerns to invest in stocks. Such strategies could even backfire if they reinforce the belief that the population of stockholders is fundamentally different. Being confronted with

individuals perceived as representative of selfish or greedy people could further support stereotypical views of the entire population of stockholders.

Instead, our results suggest that interventions aimed at decreasing negative perceptions of stockholders have the potential to substantially influence behavior. We show that a relatively light-touch intervention – providing subjects with information about differences between stockholders and non-stockholders – significantly affect decision-making in our experiment. Broader interventions, such as providing detailed information about differences over a longer duration of time, could thus induce changes in households' actual stock market participation. In particular, combining the insights from our results and framework with concepts developed from research on inter-group relations (Böhm, Rusch, and Baron, 2020) may be very effective in reaching the group of identity-motivated households. For instance, “de-categorization” is a strategy whose goal is to alleviate the belief that opposing groups form homogeneous units. Applied to the current context, it means highlighting that very different members of society invest in stocks. Another example is the concept of “re-categorization”, which proposes the communication of similarities and common goals between stockholders and non-stockholders.

Since we have documented that a substantial fraction of people hold stereotypical and thus biased beliefs regarding stockholders, normative arguments favoring such interventions can be made. This is an important factor distinguishing our explanation of limited participation from preference-based explanations. Not only are deeply held preferences hard or even impossible to change, but it is also normatively questionable as to whether an attempt to change them through interventions should be made at all. In contrast, our results indicate that perceptions of stockholders are malleable and frequently incorrect, providing a much larger normative scope for behavioral change. However, the presence of identity concerns complicates a straightforward welfare assessment. We view disentangling the welfare effects of changing perceptions when one's identity depends on perceptions as an exiting avenue for future research.

We conclude by pointing out that our framework and methodology are not tied to financial decision-making. It can readily be applied more generally to other economic contexts in which strong aversion behavior appears to be at odds with outcome- or constraint-based explanations. For instance, our mechanism could help to explain the widespread non-take-up of social transfers (Currie, 2006). Similarly, significant fractions of people abstain from holding debt and avoid insurance markets (Guiso and Sodini, 2013). Since people naturally form perceptions over groups defined by economic decisions, our channel has the potential to significantly influence decision-making and particularly aversion behavior in these domains.

**Table 1.5.** The relationship between negative views about stockholders and stock market participation

	<i>Dependent variable: Has risky financial assets</i>			
	Full sample		Determinants sample	
	(1)	(2)	(3)	(4)
Negative views about stockholders	-0.058*** (0.007)	-0.052*** (0.007)	-0.060*** (0.010)	-0.048*** (0.010)
<b>Behavioral variables</b>				
General trust				0.024** (0.010)
Ambiguity aversion				-0.021** (0.010)
Likelihood insensitivity				-0.021* (0.011)
Belief over positive stock returns				0.063*** (0.011)
Risk aversion				-0.036*** (0.010)
Financial numeracy				-0.000 (0.011)
Right-wing political orientation				0.011 (0.010)
<b>Socio-demographic variables</b>				
Female		-0.075*** (0.015)	-0.050** (0.021)	-0.017 (0.021)
Age		0.000 (0.000)	0.000 (0.001)	0.001 (0.001)
Education: upper secondary		0.021 (0.018)	0.015 (0.026)	0.007 (0.026)
Education: tertiary		0.123*** (0.019)	0.121*** (0.028)	0.102*** (0.030)
Income 2nd tercile		-0.009 (0.017)	-0.013 (0.026)	-0.011 (0.026)
Income 3rd tercile		0.083*** (0.021)	0.061* (0.032)	0.048 (0.032)
Wealth 2nd tercile		0.070*** (0.017)	0.090*** (0.024)	0.074*** (0.025)
Wealth 3rd tercile		0.267*** (0.020)	0.315*** (0.029)	0.289*** (0.030)
Mean dep. variable	0.228	0.226	0.238	0.238
Observations	2915	2903	1410	1410

*Notes:* The table shows OLS regression coefficients. The dependent variable is an indicator variable equal to one if the subject owns stocks and zero otherwise. "Negative views about stockholders" is defined as the mean of the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. See Appendix 1.K.3 for details on the other independent variables. All displayed behavioral variables have been standardized. Robust standard errors in parentheses. Significance levels: \*p<0.1, \*\*p<0.05 and \*\*\*p<0.01.

**Table 1.6.** Comparison of average actual and predicted differences

	Greed	Gambler	Selfishness
<b>Panel A: Actual average self-assessment</b>			
Stockholders	3.71	3.38	2.91
Non-stockholders	2.54	2.10	2.05
Difference	1.17	1.29	0.86
<b>Panel B: Predicted average self-assessment</b>			
Stockholders	4.27	5.22	3.47
Non-stockholders	2.70	2.40	2.53
Difference	1.57	2.82	0.94
<b>Panel C: Exaggeration (Predicted / Actual )</b>			
Exaggeration of difference (%)	34.8%	119.3%	9.4%

*Notes:* Panel A displays the average response of subjects in the LISS panel when asked to self-assess themselves regarding the characteristics greedy, gambler and selfish (agreement to “I’m kind of [...]”, 0-10 Likert scale). Panel B displays the average prediction of subjects over these self-assessments. Panel C displays the ratio of predicted differences between stockholders and non-stockholders and actual differences.

**Table 1.7.** Implications of negative perceptions on in-group favoritism and political attitudes

	<i>Dependent variable:</i>			
	Money allocated to non-stockholder		Support for higher taxation and regulation of stock market	
	(1)	(2)	(3)	(4)
Negative views about stockholders	0.103*** (0.027)	0.086*** (0.026)	0.175*** (0.026)	0.152*** (0.024)
Moral universalism		-0.132*** (0.033)		
Support for wealth tax				0.416*** (0.027)
Right-wing political orientation				-0.028** (0.012)
Opposition to income inequality				0.118*** (0.027)
Demographic controls		X		X
Observations	1,594	1,594	1,594	1,397
R <sup>2</sup>	0.011	0.063	0.031	0.307

*Notes:* The table displays OLS-estimates. The dependent variable in columns (1) and (2) is subjects' behavior in an allocation game, with higher values indicating more money is allocated to non-stockholders and thus less to stockholders. In columns (3) and (4), the dependent variable is subjects' mean answers on the five item scale capturing political attitudes towards stock market, with higher values indicating a higher support for taxation and regulation of the stock market and stockholders. "Negative views about stockholders" is defined as the mean over the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. See main text and Appendix 1.K.4 for details on the other independent variables. All displayed independent variables have been standardized. Robust standard errors in parentheses. Demographic controls include age, gender, education, income and total financial assets. Significance levels are \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

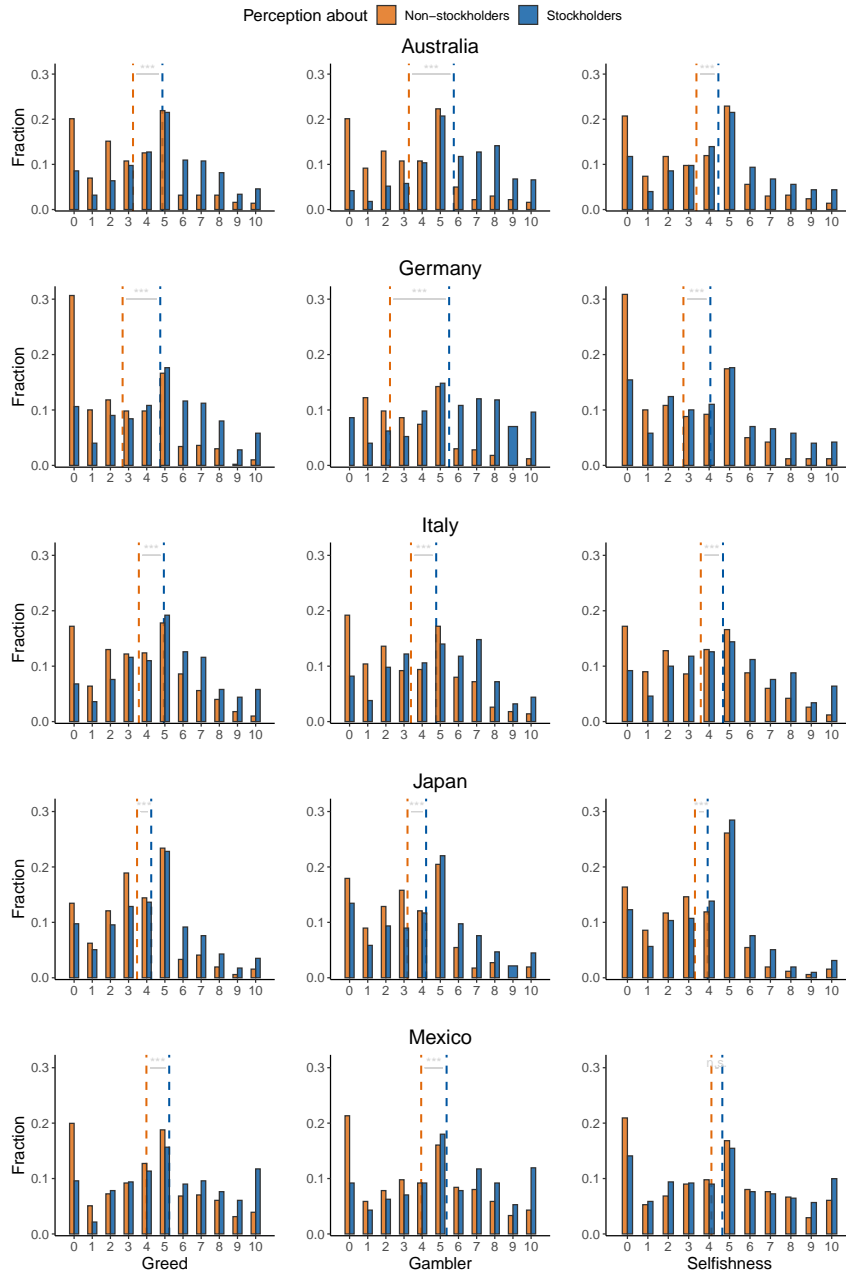
**Table 1.8.** Misreporting of risky financial assets ownership

	<i>Dependent variable:</i> Reports having risky financial assets	
	(1)	(2)
Constant	0.519*** (0.035)	
Negative views about stockholders	-0.051** (0.022)	-0.045* (0.025)
Financial numeracy		0.139*** (0.033)
Demographic controls		X
Observations	593	354

*Notes:* The table shows OLS regression coefficients. The dependent variable is a dummy variable equal to one if anybody in the household reports that they have any risky financial assets in the LISS panel. The sample is restricted to all households which hold any risky financial assets based on Dutch administrative data. “Negative views about stockholders” is defined as the mean over the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. “Financial numeracy” is defined as the number of correct answers to the four numeracy questions of van Rooij, Lusardi, and Alessie (2011). Both independent variables have been standardized. Robust standard errors in parentheses. Controls include age, gender, education, income, wealth, and a dummy whether we observe all adult household members. Significance levels: \* $p < 0.1$ , \*\* $p < 0.05$  and \*\*\* $p < 0.01$ .

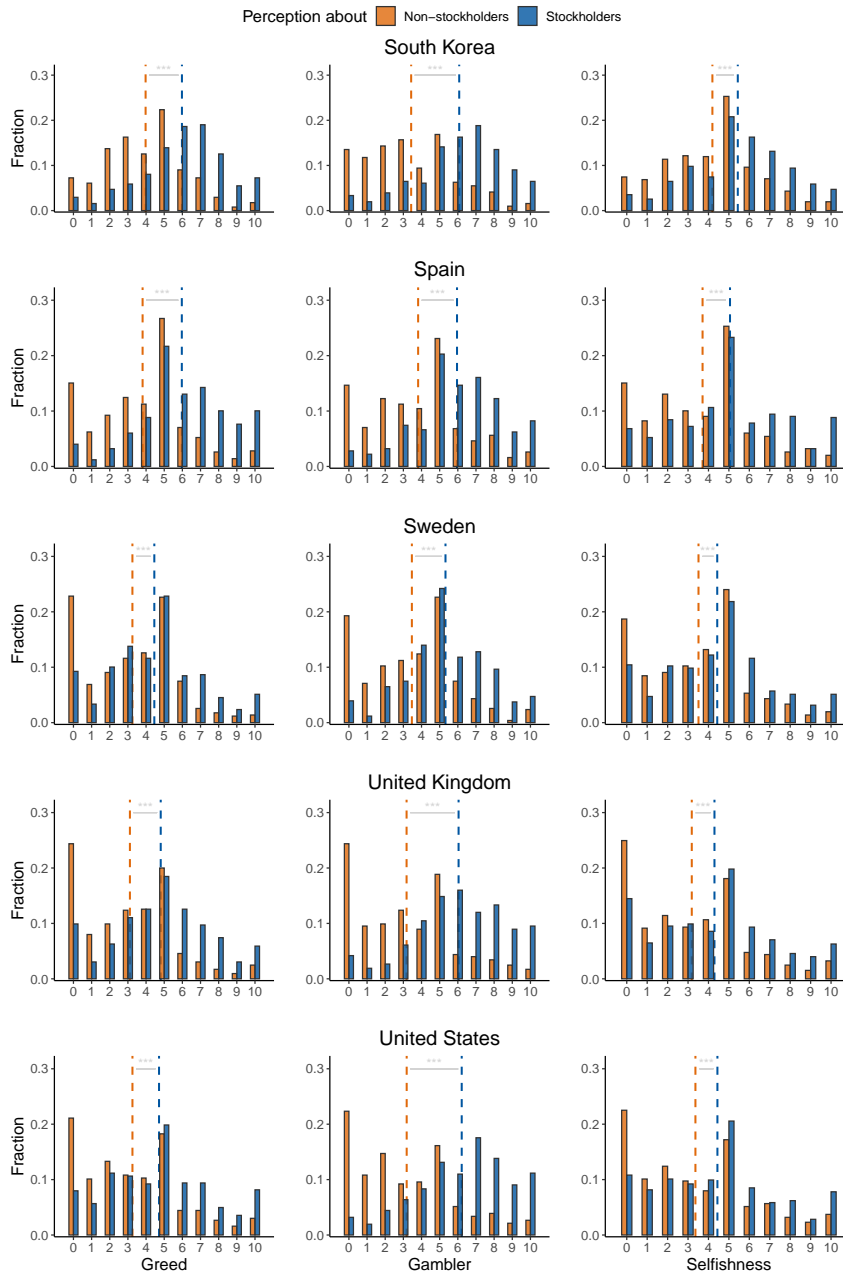


## Appendix 1.A Additional figures



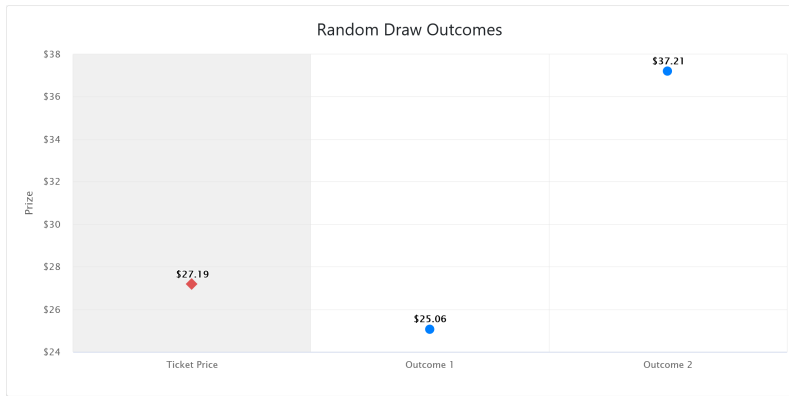
**Figure 1.A.1.** People’s perceptions of characteristics of stockholders and non-stockholders around the world part 1

Notes: The figure displays for each country the distribution of subjects’ ratings of the group of stockholders and non-stockholders. Higher values indicate that subjects rate the respective group as more greedy (left column), gambler-like (middle column), and selfish (right column). The dotted lines display mean ratings. Stars indicate significance levels, where \*p<0.1, \*\*p<0.05 and \*\*\*p<0.01, obtained from Kolmogorov-Smirnov tests.

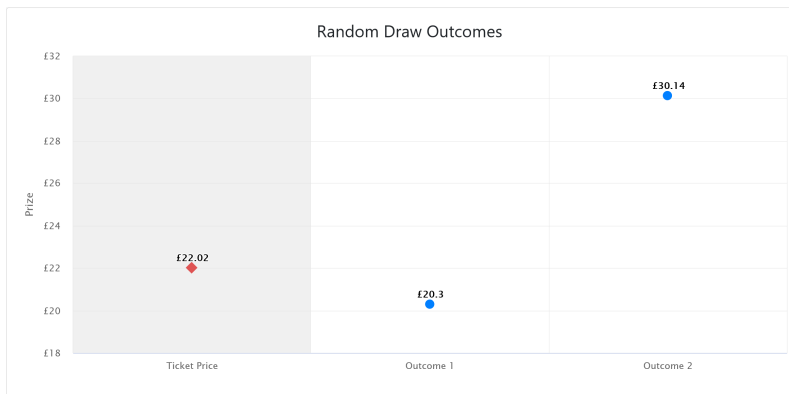


**Figure 1.A.2.** People’s perceptions of characteristics of stockholders and non-stockholders around the world part 2

*Notes:* The figure displays for each country the distribution of subjects’ ratings of the group of stockholders and non-stockholders. Higher values indicate that subjects rate the respective group as more greedy (left column), gambler-like (middle column), and selfish (right column). The dotted lines display mean ratings. Stars indicate significance levels, where \*p<0.1, \*\*p<0.05 and \*\*\*p<0.01, obtained from Kolmogorov-Smirnov tests.



(a) Figure visualization and dollar currency



(b) Figure visualization and pound currency

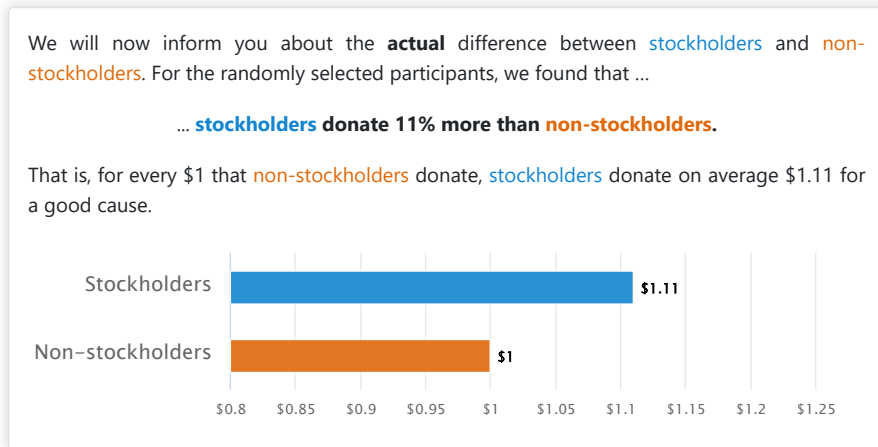
Prize in Dollar	
Buying Price	\$27.19
Outcome 1	\$25.06
Outcome 2	\$37.21

(c) Table visualization and dollar currency

Prize in Pound	
Buying Price	£22.02
Outcome 1	£20.3
Outcome 2	£30.14

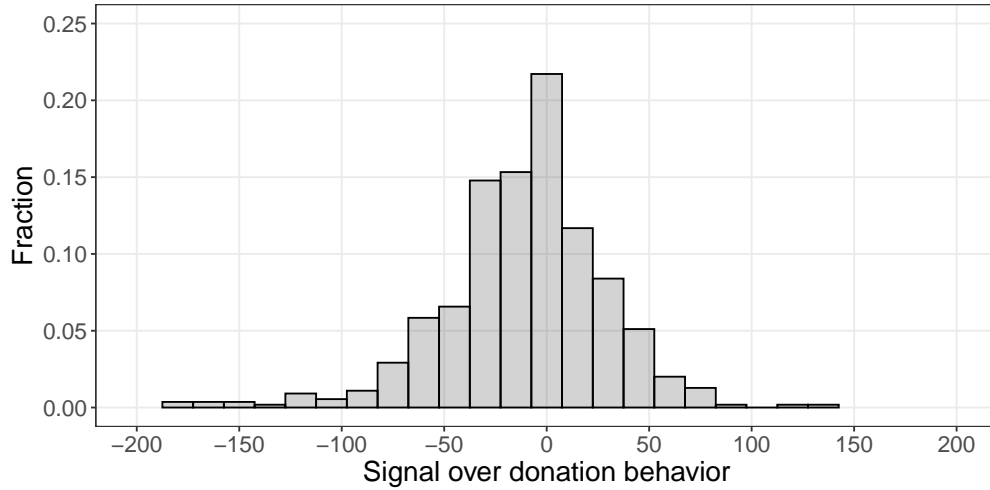
(d) Table visualization and pound currency

**Figure 1.A.3.** Combinations of risky option visualization (Draw description)



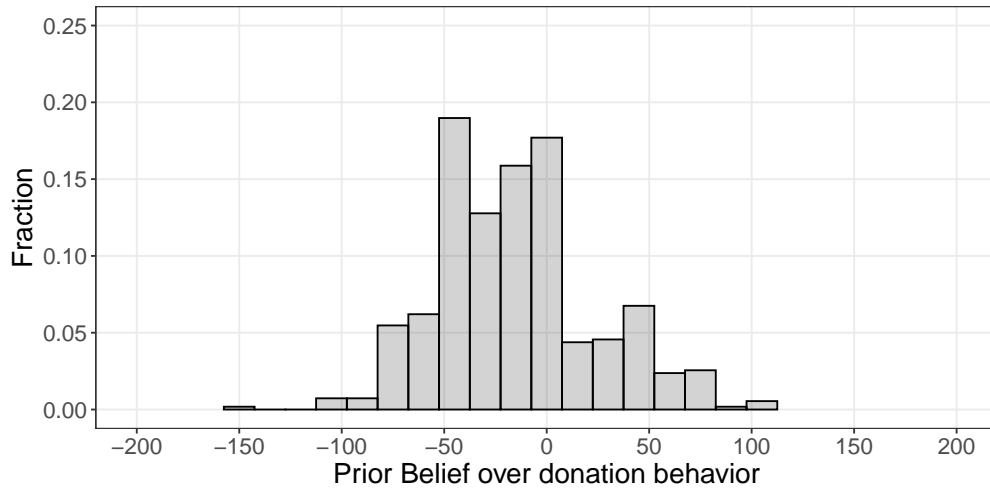
Next

Figure 1.A.4. Example of information provided to subjects



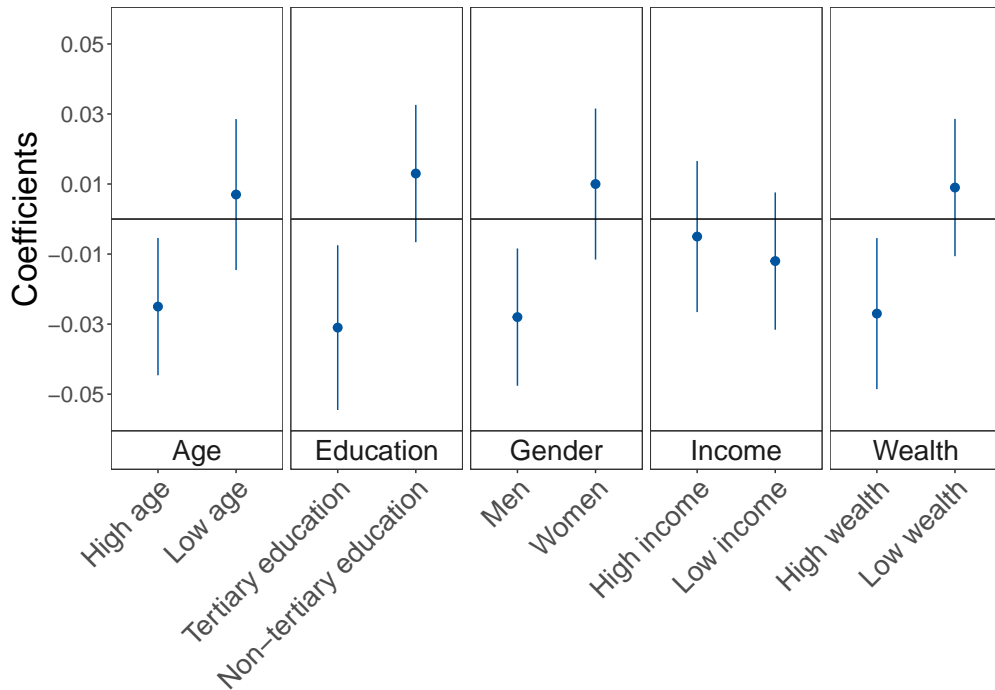
**Figure 1.A.5.** Distribution of signals generated in the information experiment

Notes: This figure displays in a histogram the distribution of signals subjects received in the information experiment over the difference in donation behavior between stockholder and non-stockholder. Higher values indicate signals that stockholder donated more relative to non-stockholder.



**Figure 1.A.6.** Distribution of prior beliefs in the information experiment

Notes: This figure displays in a histogram the distribution of subjects prior beliefs over the difference in donation behavior between stockholder and non-stockholder in the information experiment. Higher values indicate subject's beliefs that stockholders donate more relative to non-stockholders.



**Figure 1.A.7.** Heterogeneity in the relationship between negative views and stock market participation

*Notes:* This figure displays interactions term coefficients obtained from an OLS regression. The specification is the same as in Column (4) of Table 1.5, with the addition of interactions of negative views about stockholders with the displayed demographic variables. The more negative the coefficient, the stronger is the association of increases in negative views with decreases in stock market participation among the specified subgroup.

## Appendix 1.B Additional tables

**Table 1.B.1.** Overview of samples

Label	Provider	N	Covered in	Description
Perception pre-test	Pureprofile	194	Section 1.3, Appendix 1.C	Survey to select character traits based on identity-relevance and association with stockholders.
First LISS panel survey	Centerdata	3,272	Sections 1.3, 1.5, 1.6.1 and 1.6.3	Survey to measure people's perceptions of stockholders and non-stockholders in the Netherlands.
Second LISS panel survey	Centerdata	1,594	Sections 1.3, 1.1 and 1.6.2	Follow-up to the first survey to collect additional measures, investigate implications and the effect of correcting stereotypes.
Cross-country survey	Bilendi	5,130	Sections 1.3 and 1.5	Survey to measure people's perceptions of stockholders and non-stockholders in Australia, Germany, Italy, Japan, Mexico, South Korea, Spain, Sweden, United Kingdom, and United States.
Perception robustness survey	Panel Inzicht	1,016	Section 1.3, Appendix 1.E	Additional surveys to measure perceptions using different elicitation methods
Information generation survey	Prolific	272	Section 1.4.3, 1.6, Appendix 1.F	Survey to generate the information about the donation behavior of stockholders and non-stockholders for the information experiment.
Description experiment	Prolific	515	Section 1.4.1	Experiment to test prediction 1 of the conceptual framework.
Information experiment	Prolific	548	Section 1.4.3	Experiment to test prediction 2 of the conceptual framework.
Information experiment follow-up survey	Prolific	428	Section 1.4.3	Survey to investigate experimenter demand effects as well as persistence and validity of the treatment effects of the information experiment.

**Table 1.B.2.** Overview of sample demographics Netherlands

Variable	CBS (2020)	Full LISS panel	First survey LISS panel	Second survey LISS panel
<b>Gender</b>				
Women	51%	54%	51%	53%
<b>Age</b>				
16 - 34 years	29%	25%	13%	9%
35 - 64 years	48%	46%	49%	46%
65 years and older	23%	28%	38%	45%
<b>Education</b>				
University degree	33%	40%	29%	27%
<b>Net income</b>				
Below 20,000 euros/dollars	20%	20%	17%	20%
20,000 - 50,000 euros/dollars	51%	51%	59%	61%
Above 50,000 euros/dollars	30%	29%	24%	19%
Sample size		6,462	3,272	1,594

Notes: CBS stands for Centraal Bureau voor de Statistiek, the statistical bureau of the Netherlands.

**Table 1.B.3.** Overview of sample demographics cross-country study

Variable	Australia	Germany	Italy	Japan	Mexico	South Korea	Spain	Sweden	United Kingdom	United States
<b>Gender</b>										
Women	50%	50%	50%	50%	52%	50%	51%	51%	48%	50%
<b>Age</b>										
16 - 34 years	30%	27%	24%	25%	44%	27%	23%	31%	29%	26%
35 - 64 years	52%	58%	59%	55%	51%	62%	62%	54%	61%	51%
65 years and older	18%	15%	16%	20%	5%	12%	15%	15%	11%	23%
Sample size	502	499	500	513	511	510	498	508	525	564



**Table 1.B.4.** The relationship between negative views about stockholders and stock market participation across countries

<i>Dependent variable: Participates in the Stock Market</i>						
	All countries	Australia	Germany	Italy	Japan	Mexico
	(1)	(2)	(3)	(4)	(5)	(6)
Negative views about stockholders	-0.069*** (0.006)	-0.089*** (0.017)	-0.112*** (0.017)	-0.103*** (0.016)	-0.004 (0.024)	-0.028* (0.015)
Observations	5,130	502	499	500	513	511

<i>Dependent variable: Participates in the Stock Market</i>					
	South Korea	Spain	Sweden	United Kingdom	United States
	(7)	(8)	(9)	(10)	(11)
Negative views about stockholders	-0.069*** (0.022)	-0.063*** (0.018)	-0.050* (0.026)	-0.062*** (0.019)	-0.069*** (0.018)
Observations	510	498	508	525	564

*Notes:* The table shows OLS regression coefficients. The dependent variable is an indicator variable equal to one if the subject reports to own stocks and zero otherwise. "Negative views about stockholders" is defined as the mean of the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. The variable is standardized. Robust standard errors in parentheses. The regression further includes as control variables subjects' age and gender as well as country fixed effects in column (1). Significance levels: \*p<0.1, \*\*p<0.05 and \*\*\*p<0.01.

**Table 1.B.5.** Overview of sample demographics Prolific (US)

Variable	ACS (2020)	Information generation survey	Description experiment	Information experiment
<b>Gender</b>				
Women	51%	79%	69%	78%
<b>Age</b>				
16 - 34 years	46%	72%	58%	70%
35 - 64 years	38%	27%	38%	28%
65 years and older	16%	1%	4%	2%
<b>Education</b>				
University degree	28%		36%	34%
<b>Net income</b>				
Below 20,000 euros/dollars	29%	20%	24%	28%
20,000 - 50,000 euros/dollars	34%	37%	36%	32%
Above 50,000 euros/dollars	37%	43%	40%	40%
Sample size		272	515	548

Notes: ACS stands for American Community Survey, administered by the U.S. Census Bureau. Education was not elicited in the survey used to generate the donation information (third column).

**Table 1.B.6.** Description experiment balance test

Variable	Treatment stock description (1)	Treatment draw description (2)	$H_0 : (1) = (2)$ p-value (3)
Age	35.52	35.03	0.89
Women	0.70	0.68	0.58
High income	0.53	0.48	0.23
High wealth	0.65	0.62	0.40
Is financial decider	0.65	0.62	0.45
Has college degree	0.35	0.38	0.51

Notes: Column (1) and (2) displays mean values of variables for both treatments of the description experiment. The last column shows p-values obtained using Wilcoxon rank-sum tests comparing the variables of both treatments.

**Table 1.B.7.** Description experiment treatment effect on decision-making

<b>Panel A: Between-subject effect</b>		
	<i>Dependent variable: Choice of risky option</i>	
	(1)	(2)
Constant	0.524*** (0.021)	0.600*** (0.084)
Stock description	-0.142*** (0.032)	-0.132*** (0.032)
Demographic controls		X
Subjects	515	509
Observations	2,060	2,036
<b>Panel B: Within-subject effect</b>		
	<i>Dependent variable: Choice of risky option</i>	
	(1)	(2)
Constant	0.498*** (0.016)	0.583*** (0.069)
Stock description	-0.117*** (0.016)	-0.117*** (0.016)
Demographic controls		X
Subjects	515	509
Observations	4,120	4,072

*Notes:* The table displays OLS-estimates. The dependent variable is an indicator variable equal to one if the risky option is chosen and zero if the safe option is chosen. Stock description is an indicator variable equal to one if the risky option is described as a stock investment and zero if described as a random draw investment. Standard errors (in parentheses) are clustered at the subject level. Demographic controls include age, gender, education, income, total financial assets and involvement in financial decision-making. Significance levels are \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

**Table 1.B.8.** Information experiment balance test

Variable	Correlation coefficient $r$ (1)	$H_0 : r = 0$ p-value (2)
Prior Belief	0.01	0.89
Number of random draw option choices	-0.05	0.24
Age	0.03	0.46
Women	0.02	0.57
High income	-0.01	0.84
High wealth	-0.03	0.51
Is financial decider	-0.05	0.20
Has college degree	-0.02	0.70

Notes: The table displays in column (1) are pairwise Pearson correlation coefficients between the displayed variable and the variable indicating the signal over donation behavior that subjects received. Higher values for this variable indicate signals that stockholder donated more relative to non-stockholder. Column (2) show p-values obtained using Pearson correlation tests.

**Table 1.B.9.** The relationship between negative views about stockholders and the share of risky financial assets

	<i>Dependent variable: Share of risky financial assets</i>			
	Full sample		Determinants sample	
	(1)	(2)	(3)	(4)
Negative views about stockholders	-0.015 (0.013)	-0.013 (0.013)	-0.038** (0.017)	-0.030* (0.017)
<b>Behavioral variables</b>				
General trust				0.011 (0.021)
Ambiguity aversion				-0.004 (0.019)
Likelihood insensitivity				0.012 (0.019)
Belief over positive stock returns				0.030* (0.016)
Risk aversion				-0.040* (0.021)
Financial numeracy				0.011 (0.023)
Right-wing political orientation				-0.007 (0.018)
<b>Socio-demographic variables</b>				
Female		-0.006 (0.023)	-0.012 (0.033)	0.023 (0.035)
Age		0.004*** (0.001)	0.002* (0.001)	0.003** (0.001)
Education: upper secondary		0.023 (0.040)	0.076 (0.047)	0.080 (0.049)
Education: tertiary		0.072* (0.039)	0.116** (0.048)	0.116** (0.049)
Income 2nd tercile		-0.023 (0.035)	-0.036 (0.045)	-0.032 (0.045)
Income 3rd tercile		-0.001 (0.034)	-0.022 (0.047)	-0.030 (0.047)
Wealth 2nd tercile		-0.015 (0.041)	0.088 (0.061)	0.084 (0.062)
Wealth 3rd tercile		0.030 (0.045)	0.160*** (0.060)	0.158*** (0.060)
Mean dep. variable	0.359	0.352	0.345	0.345
Observations	665	657	335	335

*Notes:* The table shows OLS regression coefficients. The dependent variable is the share of risky financial assets as a percentage of all financial assets. The sample is restricted to households with any risky financial assets. "Negative views about stockholders" is defined as the mean over the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. See Appendix 1.K.3 for details on the other independent variables. All displayed behavioral variables have been standardized. Robust standard errors in parentheses. Significance levels: \* $p < 0.1$ , \*\* $p < 0.05$  and \*\*\* $p < 0.01$ .

**Table 1.B.10.** The association of demographic variables with negative views about stockholders

	<i>Dependent variable: Negative views about stockholders</i>
	(1)
Female	0.083** (0.038)
Age	0.004*** (0.001)
Education: upper secondary	-0.006 (0.058)
Education: tertiary	-0.055 (0.058)
Income 2nd tercile	0.071 (0.051)
Income 3rd tercile	0.004 (0.052)
Wealth 2nd tercile	0.045 (0.049)
Wealth 3rd tercile	-0.054 (0.055)
Observations	2903
$R^2$	0.008

*Notes:* The table show coefficients of an OLS regression. The dependent variable is “Negative views about stockholders” defined as the mean over the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. Robust standard errors in parentheses. Significance levels: \* $p < 0.1$ , \*\* $p < 0.05$  and \*\*\* $p < 0.01$ .

**Table 1.B.11.** OLS-regression on the relationship between being a stockholder and self-assessed character traits

	<i>Dependent variable:</i>					
	Self-assessed greed		Self-assessed gambler		Self-assessed selfishness	
	(1)	(2)	(3)	(4)	(5)	(6)
Is stockholder	1.166*** (0.111)	0.733*** (0.114)	1.287*** (0.105)	1.136*** (0.112)	0.860*** (0.098)	0.461*** (0.102)
Constant	2.541*** (0.050)	4.056*** (0.234)	2.097*** (0.046)	2.734*** (0.218)	2.049*** (0.043)	3.303*** (0.202)
Demographic controls		X		X		X
Observations	3,271	3,267	3,271	3,267	3,271	3,267
R <sup>2</sup>	0.034	0.126	0.047	0.091	0.025	0.101

*Notes:* The table displays OLS-estimates. The dependent variable in columns (1) - (6) is subjects self-assessment over the character traits greed (1-2), gambler (3-4), and selfishness (5-6) on a scale from 0 to 10 (LISS data, Netherlands). Higher values indicate higher degrees of greed, gambler, and selfishness, respectively. Is stockholder is a dummy variable equal to one if the subject self-reports to have risky financial assets and zero otherwise. Robust standard errors in parentheses. Controls include age, gender, education, income and total financial assets. Significance levels are \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

**Table 1.B.12.** Distribution of self-assessments and allocations

<b>Panel A: Netherlands</b>				
	Self-assessed greed (0-10 point scale)			
	0	1-3	4-6	7-10
Stockholder	14.7%	35.7%	31.6%	17.9%
Non-stockholder	32.7%	35.6%	22.8%	9%
	Self-assessed gambler (0-10 point scale)			
	0	1-3	4-6	7-10
Stockholder	14.7%	41.5%	30.3%	13.4%
Non-stockholder	37%	38.7%	18.3%	6.1%
	Self-assessed selfishness (0-10 point scale)			
	0	1-3	4-6	7-10
Stockholder	16.6%	48.1%	26.8%	8.5%
Non-stockholder	35.6%	40.6%	19.7%	4.1%
<b>Panel B: United States</b>				
	Amount allocated to self (0€ - 100€)			
	0-49	50	51-99	100
Stockholder	17.0%	24.5%	35.8%	22.6%
Non-stockholder	19.6%	30.6%	37%	12.8%

*Notes:* This table shows in Panel A the distribution of subjects self-assessment over the character traits greed, gambler, and selfishness. Elicited on a scale from 0 to 10, displayed are the proportions of subjects rating themselves as 0, 1-3, 4-6, or 7-10 for the respective traits separately for the group of stockholders and non-stockholders. Panel B shows in a similar fashion the distribution of money allocated by subjects to themselves instead to a charity in an allocation game. Displayed are the proportions of subjects allocation 0-49, exactly 50, 52-99 and exactly 100 € to themselves separately for the group of stockholders and non-stockholders.



**Table 1.B.13.** Linking of administrative data balancing test

	Not linked to admin data	Linked to admin data	$H_0 : (1) = (2)$
	(1)	(2)	(3)
Age	56.15	56.15	0.86
Female	0.56	0.51	0.12
Education: upper secondary	0.33	0.34	0.88
Education: tertiary	0.44	0.44	0.96
Negative views about stockholders	2.31	2.40	0.29
Observations	365	2903	

Notes: Column (1) displays mean values of variables of subjects we could not link to administrative income and wealth data. Column (2) does the same for subjects we could link. The last column shows p-values obtained using Wilcoxon rank-sum tests comparing the variables of the two samples.

**Table 1.B.14.** Determinants sample balancing test

	Any determinant missing	Determinants sample	$H_0 : (1) = (2)$
	(1)	(2)	(3)
Age	51.18	61.41	0.00
Female	0.56	0.45	0.00
Education: upper secondary	0.34	0.33	0.59
Education: tertiary	0.46	0.42	0.06
Negative views about stockholders	2.37	2.43	0.39
Income 2nd tercile	0.32	0.34	0.39
Income 3rd tercile	0.34	0.33	0.58
Wealth 2nd tercile	0.33	0.33	0.97
Wealth 3rd tercile	0.28	0.39	0.00
Has risky financial assets	0.22	0.24	0.31
Observations	1493	1410	

Notes: Column (1) displays mean values of variables of subjects for which any of the determinants displayed in Table 1.5 is missing. Column (2) does the same for the determinants sample. The last column shows p-values obtained using Wilcoxon rank-sum tests comparing the variables of the two samples.

**Table 1.B.15.** The relationship between negative views and stock market participation (Probit)

	<i>Dependent variable: Has risky financial assets</i>			
	Full sample		Determinants sample	
	(1)	(2)	(3)	(4)
Negative views about stockholders	-0.205*** (0.027)	-0.208*** (0.030)	-0.249*** (0.043)	-0.212*** (0.045)
<b>Behavioral variables</b>				
General trust				0.127*** (0.047)
Ambiguity aversion				-0.102** (0.044)
Likelihood insensitivity				-0.076* (0.045)
Belief over positive stock returns				0.221*** (0.040)
Risk aversion				-0.157*** (0.041)
Financial numeracy				0.004 (0.052)
Right-wing political orientation				0.058 (0.044)
<b>Socio-demographic variables</b>				
Female		-0.283*** (0.057)	-0.186** (0.085)	-0.040 (0.090)
Age		-0.001 (0.002)	-0.001 (0.003)	0.003 (0.003)
Education: upper secondary		0.099 (0.088)	0.053 (0.120)	0.019 (0.127)
Education: tertiary		0.459*** (0.084)	0.442*** (0.115)	0.392*** (0.125)
Income 2nd tercile		0.012 (0.080)	-0.003 (0.115)	-0.013 (0.119)
Income 3rd tercile		0.296*** (0.081)	0.216* (0.120)	0.160 (0.125)
Wealth 2nd tercile		0.422*** (0.082)	0.607*** (0.135)	0.582*** (0.138)
Wealth 3rd tercile		1.039*** (0.083)	1.332*** (0.133)	1.299*** (0.138)
Mean dep. variable	0.228	0.226	0.238	0.238
Observations	2915	2903	1410	1410

Notes: The table shows Probit regression coefficients. The dependent variable is an indicator variable equal to one if the subject owns stocks and zero if not. "Negative views about stockholders" is defined as the mean over the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. See Appendix 1.K.3 for details on the other independent variables. All displayed behavioral variables have been standardized. Robust standard errors in parentheses. Significance levels: \*p<0.1, \*\*p<0.05 and \*\*\*p<0.01.

**Table 1.B.16.** Misreporting of share of risky financial assets

	<i>Dependent variable:</i> Reported share of risky financial assets	
	(1)	(2)
Constant	0.154*** (0.022)	
Negative views about stockholders	-0.009 (0.010)	-0.009 (0.013)
Financial numeracy		0.011 (0.022)
Share of risky assets	0.735*** (0.036)	0.742*** (0.049)
Demographic controls		X
Observations	372	228
$R^2$	0.584	0.624

*Notes:*

The table shows OLS regression coefficients. The dependent variable is the share of risky financial assets as a percentage of all financial assets. The sample is restricted to all households which hold any risky financial assets based on Dutch administrative data and also report any risky financial assets. "Negative views about stockholders" is defined as the mean over the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. "Financial numeracy" is defined as the number of correct answers to the four numeracy questions of van Rooij, Lusardi, and Alessie (2011). Both independent variables have been standardized. Robust standard errors in parentheses. Controls include age, gender, education, income, wealth, and a dummy whether we observe all adult household members. Significance levels: \* $p < 0.1$ , \*\* $p < 0.05$  and \*\*\* $p < 0.01$ .

## Appendix 1.C Selection of character traits

This section describes the pilot experiment that we used to select three character traits for our measure capturing perceptions of stockholders. As explained in Section 1.3.2, we used two criteria to select the traits: the traits need to be associated with stockholders and identity-relevant to subjects. Based on these conditions, we searched media outlets (articles, books, and movies) and gathered a list of eight candidates. For each candidate, we adapted a description from established psychological definitions. We further added two additional characteristics (non-religiousness and non-athleticness) as a validity check. We predicted no association with stockholders for these two characteristics and thus would expect subjects to rate them accordingly. Table 1.C.1 displays the ten characteristics together with the respective definitions.

**Table 1.C.1.** List of character traits used for selection

Variable	Definition
Aggressiveness	The tendency toward social dominance, threatening behavior, and hostility.
Arrogance	The tendency to show an attitude of overbearing superiority or to make presumptuous claims or assumptions.
Dishonesty	The tendency to lack truthfulness, uprightness, and integrity.
Gambler	A person that shows the tendency to risk money or other stakes in the hope of being successful.
Greed	The tendency to continuously want more of things like wealth, possessions or social values.
Impatience	The tendency to be restless or short of temper, especially under irritation, delay, or opposition
Impulsiveness	The tendency to act hastily and without adequate reflection on the possible consequences.
Selfishness	The tendency to accept negative consequences for other people or the environment to gain a personal advantage as a result.
Non-athleticness	A person that lacks agility, muscular strength, or broad-shouldered physique.
Non-religiousness	The tendency to not have a religious character or not relate to or believe in a religion.

We presented this list (in randomized order) and the accompanying definitions to 194 subjects in a Dutch Online Panel supplied by the provider Pureprofile. In order to check the first condition, subjects were asked to rank the traits according to how strongly they associate the traits with stockholders. We used the same definition for stockholders as employed throughout the paper. Using subjects rankings, we computed for each trait the average rank, with one being ranked highest (most strongly associated) and ten the lowest rank. Table 1.C.2 displays the results. We

obverse that people perceive the trait greed to be most strongly associated with stockholders, followed by gambler and selfishness. All three traits differ significantly from the random benchmark ( $p < 0.001$ , Wilcoxon rank-sum test). Reassuringly, we find that the characteristics non-athletic and non-religious are ranked last, indicating that subjects answered deliberately.

**Table 1.C.2.** Association of traits with stockholder ranking results

Characteristic	Average rank
Greed	3.85
Gambler	4.31
Selfishness	4.83
Arrogance	4.84
Impulsiveness	5.02
Impatience	5.51
Aggressiveness	5.89
Dishonesty	6.48
Non-athleticness	6.70
Non-religiousness	7.58

*Notes:* The table show the ordered average rank of the ten selected characteristics. Lower values indicate higher rank, i.e., a stronger association of the trait with stockholders.

To test the second condition, we asked subjects to state for each trait how “important it is for you that you do not appear to have this characteristic and that others do not see you as such a person”. Subjects could rate the traits using a scale from one (“not at all important”) to ten (“very important”). As displayed in Table 1.C.3, the previously highest ranked traits greed, gambler and selfishness are also among the four highest-rated traits with respect to their identity relevance. Based on these results, we chose the three character traits greed, gambler, and selfishness as our leading variables used to measure perceptions of stockholders. Because subjects indicate that it is important to them that they do not appear to have the three selected traits, these traits are consider negatively by subjects.

**Table 1.C.3.** Identity relevance of traits results

Characteristic	Identity importance
Gambler	6.34
Greedy	5.90
Impulsiveness	5.68
Selfishness	5.65
Aggressiveness	5.53
Arrogance	5.30
Impatience	4.77
Dishonesty	4.49
Non-athleticness	4.47
Non-religiousness	3.35

*Notes:* The table show the ordered average identity relevance score of the ten selected characteristics. Higher values indicate higher self-assessed importance.

## Appendix 1.D Defining the group of stockholders

The definition of stockholders we use in our surveys is based on the official asset categorization of Statistics Netherlands. Throughout our surveys, we define stockholders as individuals “who possess any risky financial investments”, where “risky financial investments include growth funds, share funds, bonds, debentures, stocks, options, and warrants. They do not include banking accounts, saving accounts, bank savings schemes, insurance policies, or real estate.” We chose this definition because the assets categorized as risky financial investments form a specific tax category in the Netherlands and this definition is used in the asset questionnaire of the LISS panel which panelist answer every other year. For these reasons, our Dutch participants are familiar with the definition we provided. Furthermore, this categorization allows us to compare our survey data with administrative data because the latter is based on tax records. To guarantee comparability between countries, we used the same definition also in our surveys fielded to other countries.

However, an important question is whether the objective definition we use coincides with subjects’ subjective definition of stockholders. In particular, subjects could consider other types of investments relevant for their definition of what defines a stockholder. Notably, our definition does not include “indirect” stock market investments through retirement saving plans. The distinction between direct and indirect investments is particularly important in the Netherlands because the vast majority of employed Dutch individuals are automatically enrolled in retirement saving plans that contain stock market investments to varying degrees.<sup>38</sup>

Results based on the linking of survey data from the LISS panel with administrative records indicate that individuals correctly disregard indirect investments in their subjective definition of stockholders. In the linked data, we can directly compare subjects’ own categorization with the categorization of our definition. If subjects consider other types of investments outside of our definition as an important part of being a stockholder, we would expect many individuals self-reporting to be stockholders even if they are not based on our definition. Reassuringly, we see no evidence for such over-reporting of being a stockholder. Only 2% of subjects that are non-stockholders based on our definition report being a stockholder. If anything, we find that some stockholders report being non-stockholders, a phenomenon we explore in more detail in Section 1.6.3. The evidence thus supports the notion that our definition of stockholders coincides with subjects’ subjective distinction between stockholders and non-stockholders.

38. In our framework, only active decisions matter for identity. Thus, participating indirectly in the stock market through automatic enrollment should not influence group identification. We indeed find substantial group identification in our Dutch sample despite most subjects being enrolled in retirement saving plans that contain stock market investments (see Section 1.4).

## Appendix 1.E Additional evidence on people's perceptions of stockholders

This section describes the additional surveys that we used as robustness exercise for the main result presented in Section 1.3.

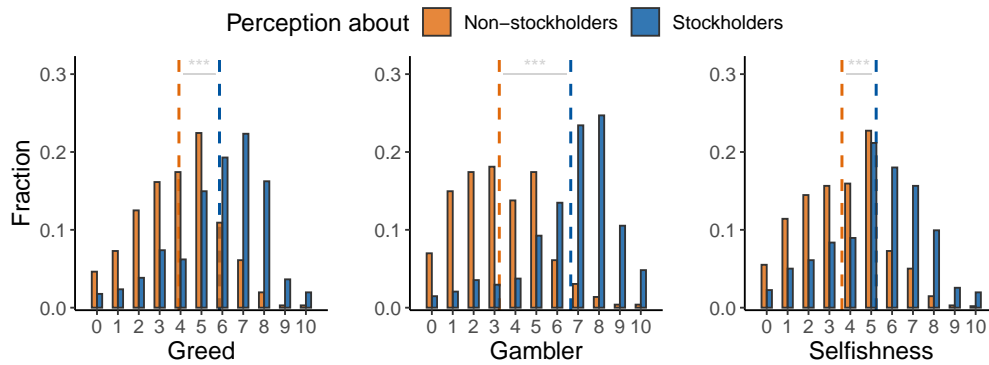
**Table 1.E.1.** List of activities used for robustness

Variable	Related trait	Observations	Survey question Out of [GROUP], how many stated that they...
Voluntary work	Selfishness	1,016	...currently do voluntary work or did so in the past two years?
Donating money	Selfishness	670	...donated money to a good cause in the past two years?
Helping strangers	Selfishness	346	...helped a stranger in need at some point in the past two years?
Casino loss	Gambler	1,016	...lost money in a casino at least once?
Excessive risk	Gambler	346	...at least once been in a situation where they regretted that they took too much risk?
Unnecessary buying	Greed	1,016	...at least once bought a product which they didn't really need just for the sake of having it?
Dissatisfaction	Greed	1,016	...felt dissatisfied because they wanted to possess more things at some point during the past two years?

**Design.** In total, we collected data from 1,016 subjects who are broadly representative of the Dutch population in terms of age and gender provided by Panel Inzicht. After answering a couple of demographic questions, subjects were presented with a list of activities. The activities were selected to be related to the three character traits selfishness, gambler and greed. For each activity, subjects were asked to estimate the proportion of people engaging in the activity. Subjects could enter any percentage number from 0 to 100 as answer and were asked separately about stockholders and non-stockholders. In order to not overload subjects with too many activities, we varied between subjects some of the activities. Table 1.E.1 displays the set of activities used together with the respective number of observations and formulations of the survey questions.

Additionally, we described a dictator game to a subset of 346 subjects. We told them that survey participants are provided an endowment of five euros. The partici-





**Figure 1.E.1.** Replication of Panel A of Figure 1.1 using robustness sample

*Notes:* The figure displays the distribution of subjects' ratings of the group of stockholders and non-stockholders elicited using the Panel Inzicht sample. Higher values indicate that subjects rate the respective group as more greedy (left graph), gambler-like (middle graph), and selfish (right graph). The dotted lines display mean ratings. Stars indicate significance levels, where \* $p < 0.1$ , \*\* $p < 0.05$  and \*\*\* $p < 0.01$ , obtained from Kolmogorov-Smirnov tests.

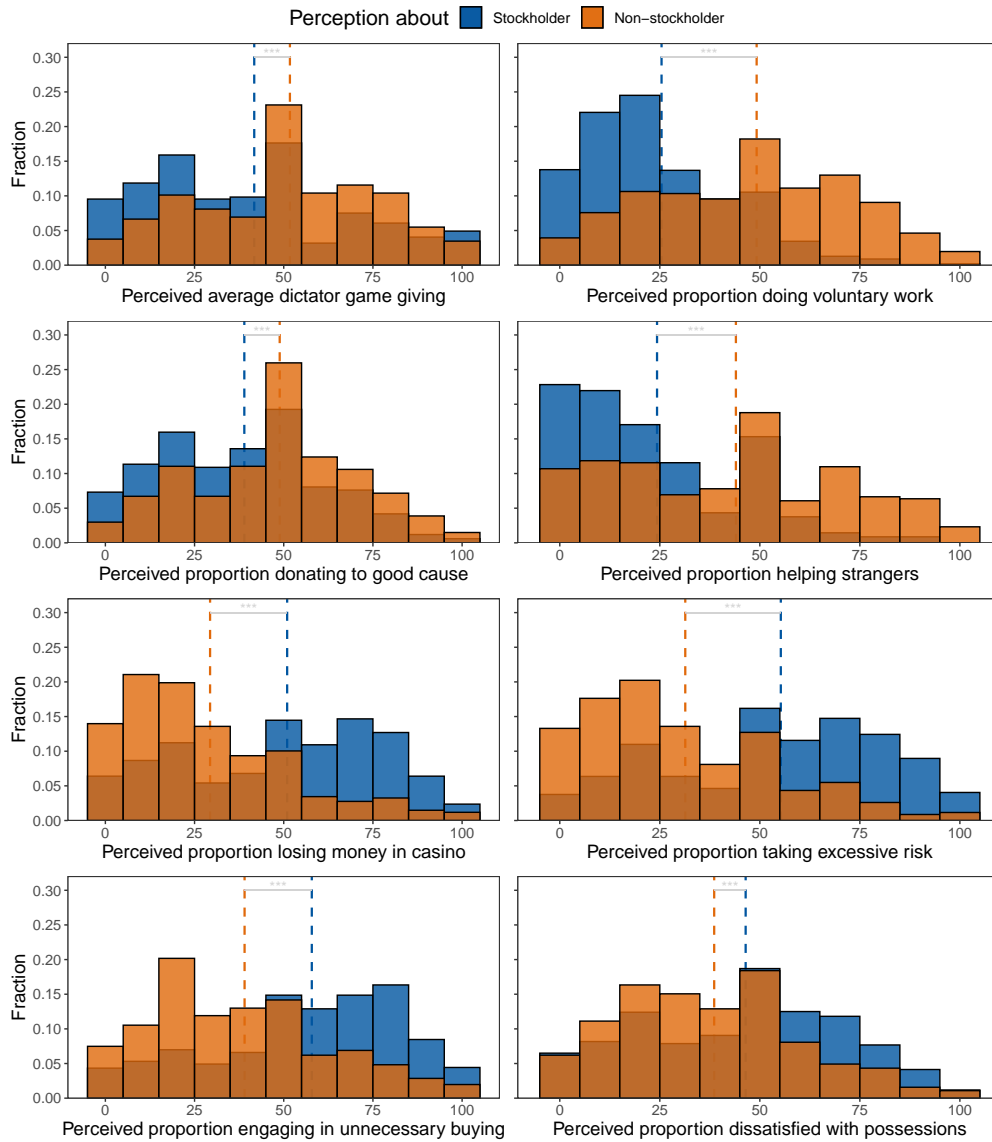
pants can allocate the endowment between themselves and the charitable organization Artsen zonder Grenzen (Doctors Without Borders). A randomly selected subset gets their decision implemented with real consequences. Similar to the elicitation of the activities, we ask subjects to estimate how these participants allocate the money. Specifically, they were asked to estimate the average amount donated to the charity as percentage of the endowment (0-100%). Subjects make their guess separately for the group of stockholders and non-stockholders.

Lastly, we wanted to replicate our finding of negative perceptions over the three selected character traits presented in the main text for the LISS panel also for this sample. Accordingly, after eliciting subject's estimates regarding the activities and dictator game, we included the standard module that elicits perceptions of the character traits greed, gambler, and selfishness, as described in Section 1.3.

**Results.** Figure 1.E.1 shows that we can replicate the finding of large negative perceptions of stockholders with respect to character traits in this sample. The distribution of ratings for stockholders again lie to the right of the non-stockholder distribution in all three cases ( $p < 0.001$ , Kolmogorow-Smirnow test). On the individual level, we again observe that a majority views stockholders as strictly more greedy, gambler and selfish, with 80% holding strictly negative views about stockholders.

Figure 1.E.2 shows the results for the dictator game and activities. In all nine cases, the distributions of estimates differs significantly ( $p < 0.001$ , Kolmogorow-Smirnow test). If the activities are positively described (dictator game, voluntary work, donating money, helping strangers), subjects predict stockholder to be sig-

nificantly less likely to be engaged in the activity. If the activities are negatively described (casino loss, excessive risk, unnecessary buying, and dissatisfaction), subjects predict stockholder to be significantly more likely to be engaged in the activity. In almost all cases, the average perceived difference is larger than 10, in some cases even larger than 20 percentage points. Furthermore, in all cases does a majority of subjects view stockholder as more negative or less positive, respectively. For instance, 58% of subjects predicting stockholder to give strictly less to charity compared to non-stockholder in the dictator game and 51% predict the fraction of stockholders that have donated money to a good cause in the past two years to be strictly lower than the fraction of non-stockholders.



**Figure 1.E.2.** Perceptions of differences in activities between stockholder and non-stockholder

Notes: The figure displays the distribution of subjects' estimates of the fraction of stockholders and non-stockholders engaging in the described activities. Higher values indicate that subjects estimate a higher fraction of individuals engaging in the activity. The dotted lines display mean values. Stars indicate significance levels, where \* $p < 0.1$ , \*\* $p < 0.05$  and \*\*\* $p < 0.01$ , obtained from Kolmogorov-Smirnov tests.

## Appendix 1.F Validation of the stock description experiment

Our framework predicts a difference (Prediction 1) between the *Stock Description* and *Draw Description* treatment if (i) the risky option in the former is, to a higher degree, associated with the stock market compared to the risky option in the latter and (ii) subjects view stockholders negatively. In the following, we provide evidence supporting these two necessary conditions underlying the prediction.

### 1.F.1 Validating the stock investment decision

To vary the association of the risky option with the stock market, we describe the option either as stock investments or as investment into random draws. Our experimental stock investments differs in two important aspects from real stock investments. First, we use stock prices that have already materialized, allowing us to control the uncertainty process generating stock price movements. This process is ambiguous in reality, i.e., generally not quantifiable by exact probabilities. Since subjects in our experiment still bet on stock price movements, we argue that the *Stock Description* investments are more strongly associated with actual stockholders than respective *Draw Description* investments. Second, subjects in our experiment do not receive fractional ownership of a company through their investment as they would in reality. Since households use stock investments mainly for wealth accumulation and not to exercise voting rights over a company, we argue that this difference also does not equalize associations.

**Sample.** To show empirically that the investment decision in *Stock Description* is to a higher degree associated with the group of stockholders, we use the full sample of 651 subjects of the description experiment. While the sample used for the main analysis consists solely of non-stockholders (as preregistered), the full sample also contains 136 stockholders. We thus use the full sample to compare the likelihood that the stock option is chosen among stockholders compared to non-stockholders.

**Results.** As hypothesized, these stockholders are significantly more likely to choose the stock option, i.e., the risky option described as stock investment. While non-stockholders choose this option in 38% out of all decisions, stockholders choose the option in 57% of decisions. Table 1.F.1 shows that this result remains significant when controlling for demographic variables and the respective choice in the *Draw Description* treatment. That is, controlling for subjects' choice behavior towards an investment with identical risk and returns outside the stock market context, stockholders are still significantly more likely to choose the stock option than non-stockholders. This result provides evidence that the stock option in our experiment is, to a higher degree, associated with the group of stockholders than the random draw option.

**Table 1.F.1.** Description experiment validation that the stock option is associated with the group of stockholders

	<i>Dependent variable:</i>			
	Choice of stock option			
	(1)	(2)	(3)	(4)
Constant	0.381*** (0.017)	0.444*** (0.078)	0.222*** (0.016)	0.269*** (0.075)
Is stockholder	0.193*** (0.037)	0.184*** (0.038)	0.152*** (0.035)	0.148*** (0.036)
Choice of draw option			0.318*** (0.024)	0.312*** (0.024)
Demographic Controls		X		X
Subjects	651	643	651	643
Observations	2,604	2,572	2,604	2,572

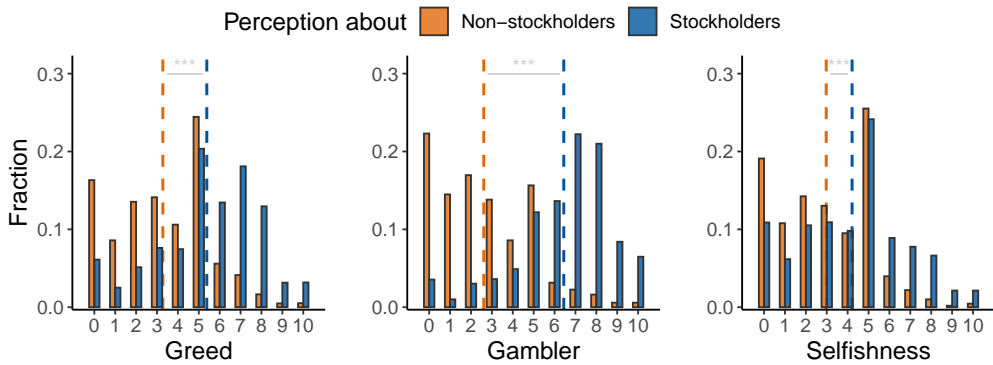
*Notes:* The table displays OLS-estimates. The dependent variable is an indicator variable equal to one if the risky option described as stock investment is chosen and zero if the safe option is chosen. Is stockholder is a indicator variable equal to one if the subject self-reports to have risky financial assets and zero otherwise. Choice of draw option is an indicator variable equal to one if the risky option described as random draw investment is chosen and zero if the safe option is chosen. Standard errors (in parentheses) are clustered at the subject level. Demographic controls include age, gender, education, income, total financial assets and involvement in financial decision-making. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

## 1.F.2 Validating negative views about stockholders on Prolific

The goal is to replicate the main finding of Section 1.3 - people view stockholders negatively - for our experimental population on Prolific. To do so, we ran a separate survey on Prolific to measure participants' views about stockholders.

**Sample.** 272 subjects on Prolific with residence in the US completed the survey. See Appendix Table 1.B.5 for sample demographics. The median completion time was less than 5 minutes, and subjects received £0.5 ( $\approx$  \$0.63) for completion. We also used this sample to generate the donation behavior used for the information experiment of Section 1.4.3. Hence, we presented subjects with an allocation decision and afterward elicited their perceptions of stockholders as described in 1.3.2 together with demographic variables.

**Results.** Figure 1.F.1 displays for all three traits the distribution of ratings separately for each group. For all three traits, we can reject the null hypothesis that both distributions are drawn from the same distribution (all  $p < 0.01$ , Kolmogorov-Smirnov test, Bonferroni corrected). On the individual level, 55% of subjects view stockholders as strictly more greedy, 88% view them as strictly more as gamblers, and 46% strictly more selfish. On average, 83% of subjects view stockholders strictly more negative than non-stockholders. In Table 1.F.2, we further show that these views predict subjects' stock market participation. A one standard deviation increase in negative views is associated with a 5.7 pp. decrease in (self-reported) stock market participation. Hence, we replicate the existence of negative views about stockholders among respondents on Prolific and additionally show their predictive power for stock market participation.



**Figure 1.F.1.** Replication of Panel A of Figure 1.1 using Prolific sample

*Notes:* The figure displays the distribution of subjects' ratings of the group of stockholders and non-stockholders elicited using the Prolific sample. Higher values indicate that subjects rate the respective group as more greedy (left graph), gambler-like (middle graph), and selfish (right graph). The dotted lines display mean ratings. Stars indicate significance levels, where \* $p < 0.1$ , \*\* $p < 0.05$  and \*\*\* $p < 0.01$ , obtained from Kolmogorov-Smirnov tests.

**Table 1.F.2.** The relationship between negative views about stockholders and stock market participation on Prolific

	<i>Dependent variable:</i>	
	Participates in the stock market	
	(1)	(2)
Negative views about stockholders	-0.057** (0.023)	-0.058** (0.023)
Constant	0.195*** (0.024)	
Demographic Controls		X
Subjects	272	269

*Notes:* The table shows OLS regression coefficients. The dependent variable is an indicator variable equal to one if the subject owns stocks and zero otherwise. "Negative views about stockholders" is defined as the mean of the negative perceptions over the traits greed, gambler and selfishness, with higher values indicating that stockholders are perceived more negatively. The variable has been standardized. Robust standard errors in parentheses. Controls include age, gender, education, income and wealth. Significance levels: \* $p < 0.1$ , \*\* $p < 0.05$  and \*\*\* $p < 0.01$ .

## Appendix 1.G Heterogeneous effects of information about stockholders on behavior

In Section 1.4.3, subjects received information the difference in donation behavior between stockholders and non-stockholders. This section analyzes heterogeneity in the effect of the information on behavior. By design, some subjects in our experiment receive a signal that the difference in donation behavior is more in favor of stockholders than they thought. Other subjects receive a more favorable signal towards non-stockholders relative to their prior beliefs. Lastly, there is a fraction of subjects that get confirming information because their signal is close to their prior beliefs. Therefore, we can use this variation to investigate whether the information has asymmetric effects on behavior: do subjects react stronger if they receive positive or negative news about stockholders?

**Variable.** We define the variable *signal relative to prior* as the signal over the difference in donation behavior subjects receive minus their belief over this difference, which was elicited prior to the information provision. Higher values indicate that subjects receive, relative to their prior belief, a more positive signal of stockholders donation behavior compared to non-stockholders. Based on this variable, we create three bins representing the three previously described groups, with -20 and 20 as the cutoffs. We chose this cutoffs to have symmetric cutoffs around zero while balancing the number of subjects in each bin. In total, there are 162 subjects whose signal is 20 percentage points or more below their prior beliefs (category *Negative signal relative to prior*). 163 subjects receive a signal that is within 20 percentage points of their prior belief, and 223 subjects receive a signal that is 20 percentage points or more above their prior beliefs (category *Positive signal relative to prior*). The larger number of subjects in the last bin is caused by subjects' beliefs being skewed in favor of non-stockholders compared to the actual difference (see Section 1.6.1).

**Results.** In Table 1.G.1, we regress the categorical variable's three groups on subjects' likelihood to choose the stock option. Subjects that receive confirming information act as reference group in the regression, and we control for the level of subjects' prior beliefs and their respective choice when options are described as a random draw. We observe that subjects receiving a signal favoring non-stockholders (relative to their prior) choose the stock option less, while subjects receiving a positive signal over stockholders donation behavior are more likely to choose the stock option. Crucially, only the later is statistically significant, and also about 50% larger in magnitude. Hence, subjects receiving positive information of stockholders react more strongly than subjects receiving negative information of the same magnitude relative to subjects priors. In Table 1.G.2, we replicate this effect using a different binning approach: instead of using -20 and 20 as cutoffs, we choose the cutoffs such



that each bin has the same number of subjects (183 subjects each, which results in cutoffs of -13 and 31).

**Table 1.G.1.** Heterogeneity in the effect of information about stockholders on decision-making using symmetric cutoffs

	<i>Dependent variable:</i>	
	Choice of stock option	
	(1)	(2)
Negative signal relative to prior	-0.051 (0.045)	-0.052 (0.045)
Positive signal relative to prior	0.078** (0.040)	0.083** (0.040)
Choice of draw option	0.241*** (0.025)	0.241*** (0.025)
Demographic controls		X
Prior beliefs	X	X
Subjects	548	541
Observations	2,192	2,164

*Notes:* The table displays OLS-estimates. All four binary choices between the risky option described as a stock investment and the safe option of a subject enter as separate observations. The dependent variable is an indicator whether the risky option described as a stock investment is chosen. “Negative signal relative to prior” and “Positive signal relative to prior” are categories indicating whether subjects received negative or positive news over the donation behavior of stockholders relative to their priors. The reference category captures receiving news aligned with subjects’ prior beliefs. Cutoffs defining the categories were chosen to achieve symmetry around 0. “Choice of draw option” is an indicator whether the risky option is chosen when described as a random draw investment. “Prior beliefs” is subjects’ belief about the difference in donation behavior between stockholders and non-stockholders. Both, “Choice of draw option” and “Prior beliefs”, are elicited before subjects receive the signal. Demographic controls include age, gender, education, income, total financial assets and involvement in financial decision-making. Standard errors (in parentheses) are clustered at the subject level. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

**Table 1.G.2.** Heterogeneity in the effect of information about stockholders on decision-making using equal observation cutoffs

	<i>Dependent variable:</i>	
	Choice of stock option	
	(1)	(2)
Negative signal relative to prior	-0.035 (0.043)	-0.032 (0.043)
Positive signal relative to prior	0.082** (0.039)	0.099** (0.039)
Choice of draw option	0.241*** (0.025)	0.242*** (0.025)
Demographic controls		X
Prior beliefs	X	X
Subjects	548	541
Observations	2,192	2,164

*Notes:* The table displays OLS-estimates. All four binary choices between the risky option described as a stock investment and the safe option of a subject enter as separate observations. The dependent variable is an indicator whether the risky option described as a stock investment is chosen. “Negative signal relative to prior” and “Positive signal relative to prior” are categories indicating whether subjects received negative or positive news over the donation behavior of stockholders relative to their priors. The reference category captures receiving news aligned with subjects’ prior beliefs. Cutoffs defining the categories were chosen such that every category has the same number of subjects. “Choice of draw option” is an indicator whether the risky option is chosen when described as a random draw investment. “Prior beliefs” is subjects’ belief about the difference in donation behavior between stockholders and non-stockholders. Both, “Choice of draw option” and “Prior beliefs”, are elicited before subjects receive the signal. Demographic controls include age, gender, education, income, total financial assets and involvement in financial decision-making. Standard errors (in parentheses) are clustered at the subject level. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

## Appendix 1.H Follow-up survey for information experiment

This section provides a detailed description of the follow-up survey we conducted after the information experiment. As explained in the main text, one of the main goals of the survey was to address experimenter demand concerns. To do so, we designed the survey to appear to subjects as a completely distinct study that is unrelated to the information experiment. If subjects indeed perceive no connection between follow-up and information experiment, but we still find treatment effects in the survey, then experiment demand effects cannot confound them.

### 1.H.1 Survey design

We invited participants of the information experiment to our follow-up survey two days after the experiment finished. This invitation appeared on participants' messaging boards on Prolific alongside invitations to studies from other researchers. Invitations include a title, short description, and name of the responsible researcher, which we thus all varied between information experiment and follow-up survey. We described the former as a decision-making experiment with the first author as researcher and the University of Bonn as institution, and the latter as an opinion survey about general topics with the second author and IZA as researcher and institution. Once subjects accepted the invitation, they were directed to the survey, which used a different welcome page and visual style compared to the information experiment. The survey itself consisted of two parts, one eliciting intentions and one eliciting perceptions.

**Intentions.** In the first part (see Appendix 1.L.4 for a screenshot), we elicit intentions to engage in specific behaviors for five items using 7-point Likert scales ranging from strongly disagree to strongly agree. The five items were the following:

*In the next five years, I will under no circumstances or situations whatsoever ...*

- *donate more money to a good cause than I currently do.*
- *invest any of my own money into financial assets traded on the stock market such as funds or shares.*
- *do more sports than I currently do.*
- *make more environmentally friendly decisions.*
- *buy disability insurance or increase coverage.*

Our main variable of interest is the second item on investment intention. We choose the formulation “under no circumstances or situation whatsoever” to retain variation in answers given the strong opposition to the stock market in our sample. The other items function as filler items to obfuscate our interest in investment intentions.

**Perceptions.** In the second part, we elicit subjects' perceptions about group differences for a list of three groups and four personality traits using 9-point Likert scales. We ask about the difference in personality traits between:

- *Those who regularly make donations for a good cause ('Regular donors') and those who do not regularly make donations for a good cause ('No regular donors')*
- *Those who invest any money into financial assets traded on the stock market such as funds or shares ('Stockholder') and those without financial assets traded on the stock market ('Non-stockholder')*
- *Those who have a disability insurance ('Insured') and those who do not have a disability insurance ('Non-insured').*

For each of these groups, we elicit perceptions of the traits intelligence, selfishness, generosity, and close-mindedness. Our main variable of interest is how different subjects perceive stockholders and non-stockholders with respect to selfishness (see Appendix 1.L.5 for a screenshot). Again, we added the other groups and traits to obfuscate the connection to the information experiment.

### 1.H.2 Validation

To attract as many subjects of the main experiment as possible, we offered more than twice the minimum pay required by Prolific (an extrapolated hourly wage of \$18.65). With these incentives, we were able to recruit 428 subjects (78%) of the main experiment for the follow-up survey. As displayed in Table 1.H.1, we find no evidence of selection effects. Which information subjects received or which decisions they made in the information experiment does not differ between the group of subjects participating in the follow-up survey and those not participating.

Between the main experiment and the follow-up survey, subjects completed an average of 15 other studies. At the end of the follow-up survey, we asked subjects to indicate the number of similar studies they had completed in the past two weeks. In total, 82% of subjects answered with "none", indicating that we successfully created a survey perceived as distinct from the main study. This fraction is a lower bound on the fraction of individuals who do not perceive a connection between follow-up and main experiment, as our survey might be similar to one of the other studies that subjects participated in. Furthermore, not a single subject of the follow-up referenced our main experiment when asked "If you had to guess, what would you say was the purpose of this study?". It thus appears that our obfuscation measures were successful in creating a survey that subjects perceive as distinct from the main experiment.

### 1.H.3 Results

We find that subjects who received information that stockholders donated more for a good cause are significantly more willing to consider investing in the future

**Table 1.H.1.** Follow-up experiment balance test

Variable	Participated in follow-up survey (1)	Did not participate in follow-up survey (2)	$H_0 : (1) = (2)$ p-value (3)
Prior belief	-16.52	-10.99	0.34
Signal over donation behavior	-10.02	-11.16	0.98
Signal over donation behavior (dummy)	0.39	0.41	0.65
Number of stock option choices	1.85	1.69	0.36
Number of random draw option choices	2.25	2.17	0.51
Age	31.53	29.62	0.19
Women	0.76	0.85	0.04
High income	0.49	0.57	0.10
High wealth	0.58	0.61	0.57
Is financial decider	0.64	0.57	0.23
Has college degree	0.35	0.34	0.93
Observations	428	120	

*Notes:* Column (1) displays mean values of variables of subjects that participated in the obfuscated follow-up survey conducted after the information experiment. Column (2) does the same for subjects that did not participate. The last column shows p-values obtained using Wilcoxon rank-sum tests comparing the variables of the info experiment with the follow-up survey.

( $p = 0.020$ , Wilcoxon rank-sum test) compared to those that received the information that non-stockholders donated more or the same amount. They also view non-stockholders as significantly more selfish compared to stockholders ( $p = 0.005$ , Wilcoxon rank-sum test). Table 1.H.2 shows these effects using an OLS regression, controlling for prior beliefs. We find slightly weaker but still persistent and significant effects when we look at the full continuous signal variable (see Table 1.H.3). A weaker effect compared to the binary classification of information seems plausible because information on which group donated more is easier to memorize than the exact percentage difference.

As a placebo check, we also assess the influence of the information on the other variables of the survey. Reassuringly, we find neither an effect of the information on perceptions of other traits, nor on other groups (Figures 1.H.1). For intentions, only the intention to donate has a coefficient that is significantly different from zero (Figures 1.H.2). Given the large number of placebo checks, it is not surprising that one of them also turns out to be significant. The general pattern clearly shows that there is no systematic relation between the information and the elicited non-stock market related perceptions and intentions.

**Table 1.H.2.** Follow-up experiment effect of binary coded information on investment intention and perceptions

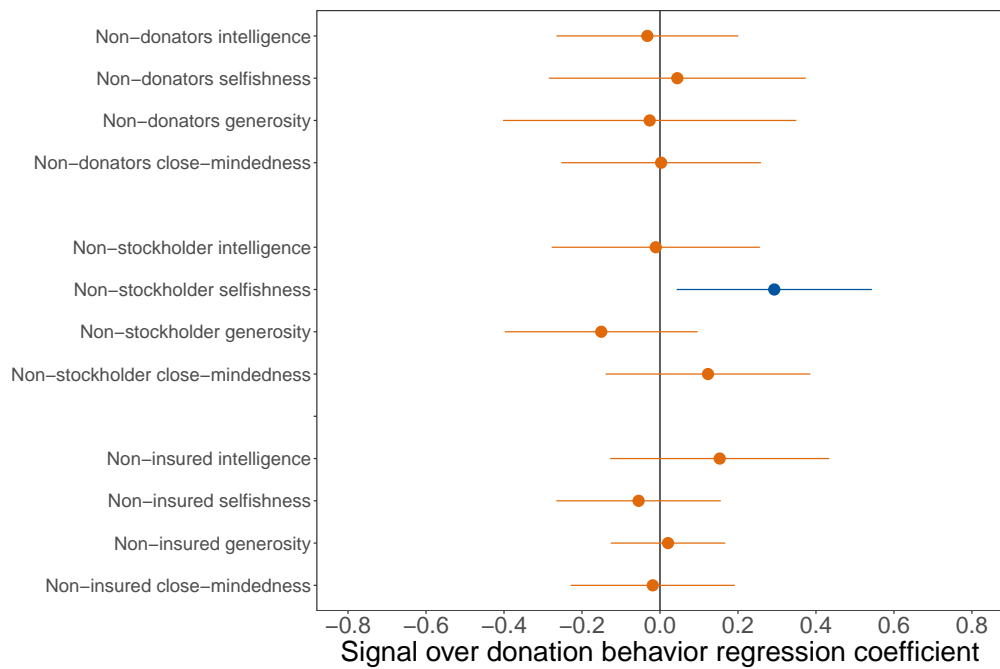
	<i>Dependent variable:</i>			
	Investment intention		Perception over non-stockholder selfishness	
	(1)	(2)	(3)	(4)
Signal over donation behavior (dummy)	0.469*** (0.174)	0.433** (0.176)	0.293** (0.127)	0.342*** (0.125)
Prior belief	0.074*** (0.023)	0.084*** (0.024)	0.041** (0.017)	0.052*** (0.016)
Constant	4.722*** (0.118)	4.982*** (0.522)	4.354*** (0.085)	4.022*** (0.348)
Demographic controls		X		X
Observations	428	422	428	422
R <sup>2</sup>	0.039	0.107	0.025	0.100

*Notes:* The table displays OLS-estimates. The dependent variable in columns (1) and (2) is subjects' intention to invest in the stock market at some point in the next five years on a scale from 1 to 7. In columns (3) and (4), the dependent variable is the extent to which subjects perceive non-stockholder to be more selfish than stockholders on a scale from 1 to 9. Signal over donation behavior is a dummy variable equal to one if subjects received the signal that stockholders donated more than non-stockholders and zero if non-stockholders donated more or the same amount. Prior beliefs is subjects belief over the differences in donation behavior between stockholders and non-stockholders. Higher values indicate that subjects believe stockholders to donate more relative to non-stockholders, with the unit being 10% differences. That is, a one unit increase means that a subject believes stockholders to donate 10% more than non-stockholders. Robust standard errors in parentheses. Demographic controls include age, gender, education, income, total financial assets and involvement in financial decision-making. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

**Table 1.H.3.** Follow-up experiment effect of information on investment intention and perceptions

	<i>Dependent variable:</i>			
	Investment intention		Perception over non-stockholder selfishness	
	(1)	(2)	(3)	(4)
Signal over donation behavior (full signal)	0.051** (0.025)	0.044* (0.025)	0.030 (0.019)	0.036** (0.017)
Prior belief	0.076*** (0.023)	0.086*** (0.024)	0.042** (0.017)	0.053*** (0.016)
Constant	4.956*** (0.089)	5.271*** (0.522)	4.499*** (0.065)	4.253*** (0.340)
Demographic controls		X		X
Observations	428	422	428	422
R <sup>2</sup>	0.036	0.103	0.022	0.096

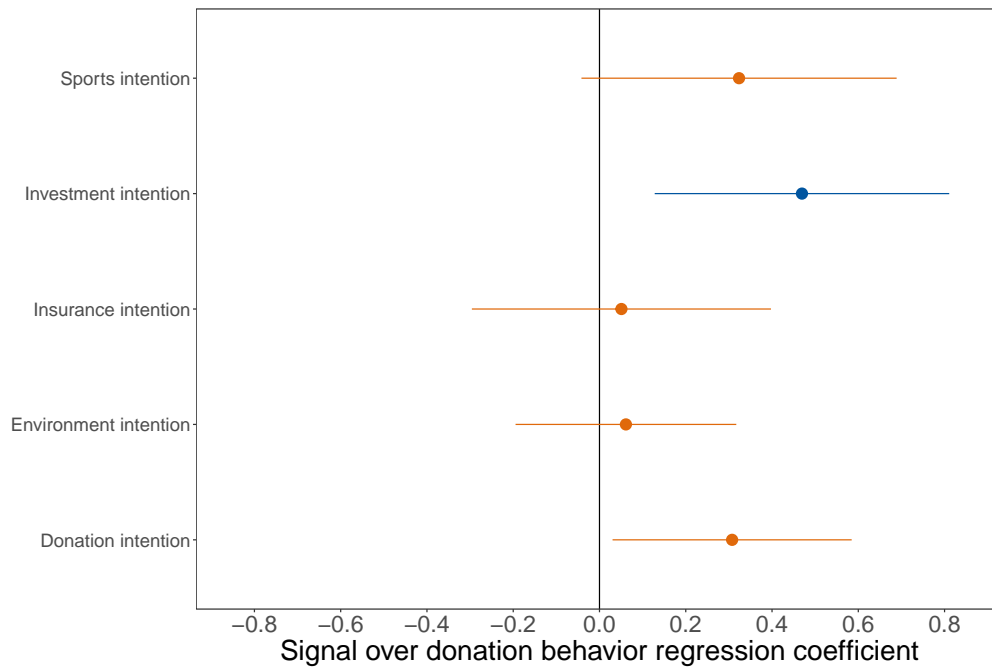
*Notes:* The table displays OLS-estimates. The dependent variable in columns (1) and (2) is subjects intention to invest in the stock market at some point in the next five years on a scale from 1 to 7. In columns (3) and (4), the dependent variable is the extent to which subjects perceive non-stockholder to be more selfish than stockholders on a scale from 1 to 9. Signal over donation behavior denotes the signal received over the difference in donation behavior. Higher values indicate a higher signal in the direction that stockholders donate more relative to non-stockholders, with the unit being 10% differences. That is, a one unit increase means a signal that stockholders donate 10% more than non-stockholders. Prior beliefs is subjects belief over the differences in donation behavior between stockholders and non-stockholders, with the same coding and unit as the signal over donation behavior variable. Robust standard errors in parentheses. Demographic controls include age, gender, education, income, total financial assets and involvement in financial decision-making. Significance levels are \*p<0.1; \*\*p<0.05; \*\*\* p<0.01.



**Figure 1.H.1.** Information experiment treatment effect on beliefs in follow-up

*Notes:* This figure shows OLS estimates of the dummy variable equal to one if subjects received the signal that stockholders donated more than non-stockholders and zero if non-stockholders donated more on different belief variables elicited in the follow-up to the information experiment. Beliefs are elicited for different characteristics and groups using scales from 1 to 9. Higher values indicate that the displayed group is rated higher with the respective trait than the complementary group. Bars indicate 95% confidence intervals.





**Figure 1.H.2.** Information experiment treatment effect on intentions in follow-up

*Notes:* This figure shows OLS estimates of the dummy variable equal to one if subjects received the signal that stockholders donated more than non-stockholders and zero if non-stockholders donated more or the same amount on different intention variables elicited in the follow-up to the information experiment. The intention variables measure subjects' intention to engage in the described activity. Their intention is elicited as agreement to the statement "In the next five years, I will under no circumstances or situations whatsoever..." on 7-point Likert scales, which we reverse code for ease of exposition. Higher values thus indicate a higher intention to perform the displayed activity. Bars indicate 95% confidence intervals.

## Appendix 1.1 Correcting stereotypical perceptions

This section describes the design and results of the information experiment conducted in the LISS panel mentioned in Section 1.6.1. The idea of the experiment is to investigate the effect of providing subjects with information on the actual differences between stockholders and non-stockholders. Since subjects' perceptions are overly negative, this type of information could be a natural policy intervention. That such a debiasing intervention could be effective appears plausible in light of the experiment described in Section 1.4.3, which was designed to identify the causal effect of negative perceptions on investment behavior. However, based on the previous literature it is not obvious that such a debiasing intervention is successful in changing beliefs and behavior in the general population. While providing information on actual differences has generally been found to be successful in reducing misperceptions and changing behavior (Bursztyn and Yang, 2022), notable exceptions exist (e.g., Alesina, Miano, and Stantcheva, 2023). Moreover, direct attempts to correct people's beliefs could even backfire, instead increasing stereotypes and animosities (Nyhan and Reifler, 2010; Fouka, 2020). After all, there are differences between stockholders and non-stockholders we show in Section 1.6.1, which, albeit small, could strengthen subjects' confidence of holding stereotypical beliefs or trigger resistance.

### 1.1.1 Design

To investigate the effect of correcting people's stereotypes about stockholders, we conducted an experiment in the second survey we fielded to the LISS panel. In total, 1,596 non-stockholders participated in the survey. These subjects had participated in our first survey, where we elicited their perceptions (see Section 1.3) and predicted differences in response behavior (see Section 1.6.1). We randomly selected half of the subjects and provided them with information on the actual differences between stockholders and non-stockholders. Specifically, they learned the difference in self-assessments for the traits greed, gambler, and selfishness (1.17, 1.29, and 0.86 points, respectively, see Section 1.6.1). They also received complementary information on differences in activities related to these self-assessments. As a reference, we reminded subjects of their predictions of these differences. The other half received no information on the actual difference, only the reminder of their predictions. Afterwards, we elicited subjects' posterior negative perceptions using the module described in Section 1.3. We subsequently gave them an incentivized investment choice: subjects were endowed with 100 €, which they could allocate between a savings account that pays no interest (safe option) and an ETF that tracks the Amsterdam Exchange index (stock option). A randomly selected subset of 16 subjects had their decisions implemented with real consequences. For them, the money was invested in the savings account and/or the ETF for a year and paid out afterward. By

comparing the treatment group with the control group, we are thus able to identify the effect of attempting to debias subjects on beliefs and investment behavior in a representative sample of non-stockholders.

### 1.1.2 Results

**Effect on posterior perceptions.** We find that providing information on the actual differences successfully reduces people's negative perception of stockholders. In the control group, subjects rated stockholders on average 2.83 Likert scale points more negatively than non-stockholders. In the treatment group, stockholders are only rated 2.19 points more negatively ( $p < 0.001$ , Wilcoxon rank-sum test). An important question, however, is how subjects that hold beliefs close to the actual differences react. To investigate their behavior, we split our sample along the median of subjects' predictions about the response behavior of subjects.<sup>39</sup> We find that subjects with median or below stereotypes in the treatment group do not show a significant change in their negative perceptions ( $p = 0.91$ , paired Wilcoxon signed-rank test). At the same time, subjects in the treatment group with above median stereotypes strongly decrease them when comparing their prior with posterior negative perceptions ( $p < 0.001$ , paired Wilcoxon signed-rank test).

**Effect on behavior.** What is the impact of the information on investment behavior? In the full sample, we observe a 6% increase in the average amount invested in the ETF, an insignificant positive effect ( $p = 0.55$ , Wilcoxon rank-sum test). However, this average effect once again masks substantial heterogeneity. Since subjects with priors close to the actual differences are given information that confirms their beliefs, we would not predict changes in their investment behavior. Instead, the effect should be concentrated on subjects who receive information correcting their stereotypes. Using the same median split as before, Table 1.1.1 displays the treatment effect of providing information about the true differences on the amount of money invested in the stock option. We find in Column (1) that subjects with ex-ante above median stereotypes significantly increase the amount invested in the stock option by about 20%. At the same time, as displayed in Column (3), subjects with median or below median stereotypes reduce their investment only by an insignificant amount. Adding controls in Columns (2) and (4) does not change the results. Hence, we observe a significant effect on subjects who receive information correcting their stereotypes,

39. The median prediction is 2 (on a scale of 0 to 10). Within the group of subjects on or below the median, average predictions are 0.55 for greed, 1.46 for gambler, and 0.33 for selfish. Thus, this group has, on average, beliefs close to the actual differences. Within the group of subjects with above median predictions, averages are 3.91, 5.70, and 2.86, revealing substantial stereotypes. Using instead the negative views of stockholders measure defined in Section 1.3 or other splits yield similar results.

**Table 1.1.1.** Effect of providing information about actual differences between stockholders and non-stockholders on investment behavior

	<i>Dependent variable: money invested in exchange-traded fund</i>			
	Above median stereotypes		Median or below stereotypes	
	(1)	(2)	(3)	(4)
Information treatment	6.083** (2.721)	5.415** (2.620)	-1.774 (2.608)	-1.006 (2.541)
Constant	29.743*** (1.846)		33.214*** (1.830)	
Demographic controls		X		X
Observations	736	736	858	858

*Notes:* The table displays OLS-estimates. The dependent variable denotes the amount of money (out of 100€) that subjects allocate to the stock option instead of the safe option. Information treatment is an indicator variable equal to one if the subject receives information about the actual differences between stockholders and non-stockholders in self-assessed greed, gambler, and selfishness and related activities and zero otherwise. Demographic controls include age, gender, education, income and total financial assets. Robust standard errors in parentheses. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

and no backlash effects by subjects who receive information more closely aligned with their prior beliefs.<sup>40</sup>

40. We also elicited investment intentions both in the short run (six month) and long run (five years). In line with the strong stock market aversion documented previously, we find intentions to be very low. For instance, 68% agree with the statement that they would under no circumstance whatsoever invest in the stock market in the next five years. Although the treatment increases intentions in the expected direction, the effects are not significant.

## Appendix 1.J Research transparency

All surveys that include experimental variation were preregistered at [aspredicted.org](https://aspredicted.org). The preregistrations include details on the experimental design, the sampling process and planned sample size, exclusion criteria, hypotheses, and the main analyses. This section documents deviations from the preregistration.

### 1.J.1 Description experiment

The description experiment (Section 1.4.1) was preregistered at [https://aspredicted.org/66D\\_2XD](https://aspredicted.org/66D_2XD). In the preregistration, we specified that we would exclude subjects that chose the safe option each time in the calibration part. To maximize sample size and comparability with the information experiment, we report the results in the main text without excluding these subjects. Table 1.J.1 replicates the results using the preregistered exclusion restrictions. The resulting treatment effect is highly significant and slightly larger than the one reported in the main text.

### 1.J.2 Information experiment

The information experiment (Section 1.4.3) was preregistered at [https://aspredicted.org/1B8\\_MXD](https://aspredicted.org/1B8_MXD). In the preregistration, instead of the OLS analysis shown in Table 1.3 with the choice of the stock option as dependent variable, we specified a slightly different specification in which choices of stock option and draw options enter as separate observations. Table 1.J.2 replicates the results using the preregistered specification. As in the specification reported in the main text, the preregistered specification documents that the information about donation behavior significantly influences the likelihood that the stock option is chosen. The follow-up survey reported at the end of Section 1.4.4 was not preregistered.

### 1.J.3 Correcting stereotypes experiment

The experiment that corrected stereotypes (Section 1.6.1, with details in Appendix 1.I) was preregistered at [https://aspredicted.org/blind.php?x=DF3\\_5YR](https://aspredicted.org/blind.php?x=DF3_5YR). The preregistration specified four exclusion restrictions. Our analysis only employs the first two restrictions. The latter two restrictions concern the time respondents spend on the survey. Due to errors in the time spent variable, we cannot employ these restrictions. However, judging from previous experience with LISS surveys, these restrictions would have affected only a very small set of respondents anyway (less than 5%). Furthermore, the preregistration included three outcomes related to financial decision-making. To keep the analysis aligned with the previous experiments reported in Section 1.4, we mainly focus on the first, the incentivized investment decision. In Footnote 40, we discuss the other two. The preregistration also specified a

**Table 1.J.1.** Replication of Table 1.B.7 using preregistered sample

<b>Panel A: Between-subject effect</b>		
	<i>Dependent variable: Choice of risky option</i>	
	(1)	(2)
Constant	0.579*** (0.022)	0.648*** (0.085)
Stock description	-0.154*** (0.034)	-0.147*** (0.034)
Demographic controls		X
Subjects	418	412
Observations	1,672	1,648
<b>Panel B: Within-subject effect</b>		
	<i>Dependent variable: Choice of risky option</i>	
	(1)	(2)
Constant	0.560*** (0.017)	0.614*** (0.071)
Stock description	-0.139*** (0.018)	-0.140*** (0.019)
Demographic controls		X
Subjects	418	412
Observations	3,344	3,296

*Notes:* The table displays OLS-estimates. The dependent variable is an indicator variable equal to one if the risky option is chosen and zero if the safe option is chosen. Stock description is an indicator variable equal to one if the risky option is described as a stock investment and zero if described as a random draw investment. Standard errors (in parentheses) are clustered at the subject level. Controls include age, gender, education, income, wealth and involvement in financial decision-making. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

series of additional heterogeneity and exploratory analyses. These are not included in the paper. The analyses using the median split were not preregistered.

**Table 1.J.2.** Replication of the result of Table 1.3 using preregistered specification

	<i>Dependent variable:</i>	
	Choice of risky option	
	(1)	(2)
Constant	0.558*** (0.016)	
Stock description	-0.080*** (0.019)	-0.082*** (0.019)
Prior belief	0.002 (0.004)	0.001 (0.004)
Prior belief × stock description	0.009** (0.004)	0.009** (0.004)
Signal over donation behavior	-0.004 (0.004)	-0.006 (0.004)
Signal over donation behavior × stock description	0.012*** (0.005)	0.012*** (0.005)
Demographic controls		X
Subjects	548	541
Observations	4,384	4,328

*Notes:* The table displays OLS-estimates. All eight binary choices between the risky option described as a stock investment and the safe option of a subject enter as separate observations. The dependent variable is an indicator variable equal to one if the risky option is chosen and zero if the safe option is chosen. “Stock description” is an indicator variable equal to one if the risky option is described as a stock investment and zero if described as a random draw investment. “Prior beliefs” is subjects belief over the difference in donation behavior between stockholders and non-stockholders. Higher values indicate that subjects believe stockholders to donate more relative to non-stockholders, with the unit being 10% differences. That is, a one unit increase means that a subject beliefs stockholders to donate 10% more than non-stockholders. “Signal over donation behavior” denotes the signal received over the difference in donation behavior, with the same coding and unit as “Prior beliefs”. Standard errors (in parentheses) are clustered at the subject level. Controls include age, gender, education, income, total financial assets and involvement in financial decision-making. Significance levels are \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

## Appendix 1.K Variables

### 1.K.1 Demographic and financial variables

This section provides further information about the calculation of background variables for the Dutch sample. For the US samples on Prolific we elicited all demographic variables during our own data collection.

**Age** Refers to the financial decider who participated in the survey. Obtained from the Dutch administrative data (or the LISS background questionnaire for regressions not involving admin data).

**Gender** Refers to the financial decider who participated in the survey. Obtained from the Dutch administrative data (or the LISS background questionnaire for regressions not involving admin data).

**Education** Based on achieved educational level. Obtained from the LISS background questionnaire. The Dutch educational levels are categorized as follows:

**Lower secondary and below:** primary school, vmbo

**Upper secondary:** mbo, havo, vwo

**Tertiary:** hbo, wo

In the US sample, we use the following categories: “Less than high school degree”, “Graduated High school or equivalent”, “Associate degree”, “Professional degree (JD, MD)”, “Some college but no degree”, “Bachelor’s degree”, “Post-graduate degree”.

**Income** Monthly net income of the household. The measure is equalized by dividing through the square root of the number of household members (square root scale). In regressions, we add terciles of this variable as categorical variable. Obtained from Dutch admin data (or the LISS background questionnaire for analyses not involving admin data). In the US sample, we use household net income as elicited in our own questionnaire and do not equalize the measure as we do not have information about the number of household members.

**Wealth** Wealth of the household including financial assets (safe and risky financial assets) and non-financial assets like real estate. Debts are subtracted. The measure is equalized by dividing through the square root of the number of household members (square root scale). In regressions, we add terciles of this variable as categorical variable. Obtained from Dutch admin data for the beginning of the year 2020.

**Financial assets** Total financial assets of the household including safe and risky financial assets. The measure is equalized by dividing through the square root of the number of household members (square root scale). In regressions, we add terciles of this variable as categorical variable. We use this measure for analyses



not involving admin data. Elicited in the LISS asset questionnaire. In the US sample, we use personal total financial assets as elicited in our own questionnaire.

**Owns any risky financial assets** Dummy variable if risky financial assets in the household are larger than 0. Obtained from Dutch admin data for the beginning of the year 2020.

**Share of risky financial assets** Risky financial assets of the household divided by total financial assets of the household. Set to missing if total financial assets do not exceed 0. Obtained from Dutch admin data for the beginning of the year 2020.

### 1.K.2 Stock market knowledge variables

**Self-assessed stock market knowledge** Based on the agreement on a 7-point likert scale to the statement “At the moment, I am confident in my knowledge about how the stock market works”. Elicited in our second survey filed to the LISS panel.

**Financial numeracy** We employ the four questions of van Rooij, Lusardi, and Alessie (2011) measuring financial numeracy. We count the number of correct answers and standard normalize the measure. Included in the LISS panel by Gaudecker, Wogrolly, and Zimpelmann (2022) in May 2021.

**Self-assessed ability to be successful in the stock market** Based on the agreement on a 7-point likert scale to the statement “I currently do not have what it takes to be successful in trading on the stock market”. Variable is reverse coded for ease of interpretation. Elicited in our second survey filed to the LISS panel.

**Belief over stock returns** Following Hurd, Rooij, and Winter (2011), subjects are asked “Suppose someone invests 1000 euros in the AEX today and in six months looks at what the AEX has done. How likely is it that this investment will be worth more than 1000 euros?”. Elicited in the LISS panel in May 2020.

### 1.K.3 Behavioral determinants of portfolio choice

**Risk aversion index** We employ a quantitative lottery choice task and a qualitative risk questions for general decisions under risk based on Falk et al. (2022). We use the experimentally validated weights by Falk et al. (2022) to calculate the index such that the qualitative risk component is weighted slightly higher at 53% (after standard normalizing both components). Included in the LISS panel by Gaudecker, Wogrolly, and Zimpelmann (2022) in May 2021.

**Financial numeracy** We employ the four questions of van Rooij, Lusardi, and Alessie (2011) measuring financial numeracy. We count the number of correct answers and standard normalize the measure. Included in the LISS panel by Gaudecker, Wogrolly, and Zimpelmann (2022) in May 2021.

**Ambiguity aversion** We calculate the ambiguity aversion index as defined in Baillon et al. (2018). The measure is based on up to 28 binary choices between an ambiguous option related to a bet on the stock market and a risky option with known winning probability. Included in the LISS panel by Gaudecker, Wogrolly, and Zimpelmann (2022) in May 2021.

**Likelihood insensitivity** We calculate the index of ambiguity induced likelihood insensitive as defined in Baillon et al. (2018). The measure is based on up to 28 binary choices between an ambiguous option related to a bet on the stock market and a risky option with known winning probability. Included in the LISS panel by Gaudecker, Wogrolly, and Zimpelmann (2022) in May 2021.

**Belief over stock returns** Following Hurd, Rooij, and Winter (2011), subjects are asked “Suppose someone invests 1000 euros in the AEX today and in six months looks at what the AEX has done. How likely is it that this investment will be worth more than 1000 euros?”. Elicited in the LISS panel in May 2020.

**General trust** Following Guiso, Sapienza, and Zingales (2008), we use the response to the question “Generally speaking, would you say that most people can be trusted, or that you can’t be too careful in dealing with people? Please indicate a score of 0 to 10.” Obtained from the 2020 personality questionnaire of the LISS panel.

**Right-wing political orientation** Following Kaustia and Torstila (2011), we use the response to the question “In politics, a distinction is often made between “the left” and “the right”. Where would you place yourself on the scale below, where 0 means left and 10 means right?”. Obtained from the 2020 politics and values questionnaire of the LISS panel.

#### 1.K.4 Determinants of political attitudes

**Opposition to income inequality** Based on the question “Some people believe that differences in income should increase in our country. Others feel that they should decrease. Still others hold an opinion that lies somewhere in between. Where would you place yourself on a scale from 1 to 5, where 1 means that differences in income should increase and 5 means that these should decrease?” Obtained from the 2020 politics and values questionnaire of the LISS panel.

#### 1.K.5 Scales

This section describes the identity and political attitudes towards stock market scales mentioned in Sections 1.4 and 1.6.2 in more detail.

**Identity scale.** Table 1.K.1 provides the wording of the items used for the identity scale. As mentioned in the main text, we selected five items from established group

identification scales (Doosje, Ellemers, and Spears, 1995; Klor and Shayo, 2010)<sup>41</sup> and applied them to the stockholder/non-stockholder context. We use the same asset categorization to define the group of stockholders and non-stockholders as in our elicitation of people’s views of stockholders (Section 1.3). For each item, subjects were asked whether they agree or disagree with the respective statement on a scale from one (“fully disagree”) to seven (“fully agree”). Following standard practice, we code a subject as agreeing to a statement if they select at least five on the 7-point scale.

**Table 1.K.1.** Identity scale items

Item
1 Not having risky financial investments is an important part of my identity.
2 I am proud that I have no risky financial investments.
3 When a person or the media criticizes people for not having risky financial investments it feels like a personal insult.
4 I have very little in common with people who have risky financial investments.
5 If I would learn that a person has risky financial investments, I would immediately feel less connected to that person.

**Political attitudes towards stock market scale.** Table 1.K.2 shows the items we used to elicit political preferences toward stockholders and the stock market. As it is common in the literature, we provided subjects with additional explanations for some of the more technical items. These are marked in *italic* in the Table. As before, subjects were asked whether they agree or disagree with the respective statement of each item on a scale from one (“fully disagree”) to seven (“fully agree”). Following standard practice, we code a subject as agreeing to a statement if they select at least five on the 7-point scale, disagreeing if they select at most 3, and undecided if they select four. The first item was purposefully designed to be unrelated to the stock market and instead designed to measure subjects’ general distributional preferences in the context of wealth taxation.<sup>42</sup> Therefore, it is not part of the political preferences measure and instead used as a control variable in Table 1.7. We define our measure capturing subject’s political attitudes toward the stock market as the average over the remaining five items. In doing so, we reverse code the responses to item three so that higher values on the scale indicate political attitudes in favor of non-stockholder compared to stockholder in terms of regularization and taxation.

41. Henkel et al. (2023) show that this scale, used in the context of vaccination identification, is strongly related to the broader identification scale of Leach et al. (2008).

42. Note that such a tax currently exists in the Netherlands. In the instructions, we use the commonly used Dutch term for the tax.

**Table 1.K.2.** Political preferences items

Item
<p>1 The tax on personal investments and savings should be increased. <i>Currently, the effective tax rate is 0.59% each year for individuals with a net asset value (after deducting tax free amount) below EUR 100,001, 1.39% for a value between EUR 100,001 and EUR 1,000,000, and 1.60% above EUR 1,000,000.</i></p>
<p>2 Investments in risky financial assets like stocks or funds should be taxed with a higher tax rate than savings on a banking account. <i>Currently, they are taxed with the same tax rate.</i></p>
<p>3 Investments into stocks and funds for private retirement should be incentivized more strongly by the government, for example, through subsidies or tax exemptions. [Reverse coded]</p>
<p>4 There should be a financial transaction tax introduced to the Netherlands. <i>The financial transaction tax is a small tax applied every time a financial asset (stocks and derivative) is sold. Currently, there is no such tax in the Netherlands.</i></p>
<p>5 Investment products that enable the betting on prices of staple food should be prohibited.</p>
<p>6 There should be a tax on gains from trading risky financial assets in the Netherlands. <i>Currently, gains from trading risky financial assets are not taxed extra in the Netherlands (except when a substantial interest exists, which is the case when at least 5% of the shares, options or profit-sharing certificates in a company are owned).</i></p>

## Appendix 1.L Instructions

This appendix shows the central experimental decisions screens and instructions. The complete instructions are available at the following link:  
[https://osf.io/qz8ab/?view\\_only=8a6ac9a3c6894138ab77721a79bd35f7](https://osf.io/qz8ab/?view_only=8a6ac9a3c6894138ab77721a79bd35f7).

### 1.L.1 Perception survey view of stockholders elicitation screens

We now ask you about **other people of the general Dutch population** who take care of financial matters in their household. Consider the population is divided into two groups of people: First, individuals who **possess** any risky financial investments (e.g., stocks or funds [details]). Second, individuals who **do not possess** any risky financial investments. We would like to know **what you think** about the individuals in those two groups. Please indicate to what extent you agree or disagree with each statement.

[Details: Risky financial investments include growth funds, share funds, bonds, debentures, stocks, options and warrants. They do not include banking accounts, saving accounts, bank savings schemes, insurance policies, or real estate.]

Individuals who **possess** any risky financial investments such as stocks are, on average, ...

kind of greedy [Likert scale]

[info-box: greedy - a strong wish to continuously get more of things like wealth, possessions or social values.]

kind of gamblers [Likert scale]

[info-box: being a gambler - a person that shows the tendency to risk money or other stakes in the hope of being successful.]

kind of selfish [Likert scale]

[info-box: selfish - being willing to accept negative consequences for other people or the environment to gain a personal advantage as a result.]

Individuals who **do not possess** any risky financial investments such as stocks are, on average, ...

kind of greedy [Likert scale]

[info-box: greedy - a strong wish to continuously get more of things like wealth, possessions or social values.]

kind of gamblers [Likert scale]

[info-box: being a gambler - a person that shows the tendency to risk money or other stakes in the hope of being successful.]

kind of selfish [Likert scale]

[info-box: selfish - being willing to accept negative consequences for other people or the environment to gain a personal advantage as a result.]

[Likert scales from 0 (strongly disagree) to 10 (strongly agree)]

**Figure 1.L.1.** Perception survey view of stockholders elicitation Netherlands

### Questions

Please consider that the adult U.S. population is divided into two groups of people:

First, individuals who **do** possess any risky financial investments (e.g., stocks or funds [\(click for details\)](#)).

Second, individuals who **do not** possess any risky financial investments.

We would like to know what you personally think about the individuals in those two groups. Please indicate to what extent you agree or disagree with each statement below. The statements concern the following three characteristics:

**Greedy:** a strong wish to continuously get more of things like wealth, possessions or social values.

**Being a gambler:** a person that shows the tendency to risk money or other stakes in the hope of being successful.

**Selfish:** being willing to accept negative consequences for other people or the environment to gain a personal advantage as a result.

Individuals who **do** possess any risky financial investments such as stocks are, on average, ...

**... kind of greedy.**

Totally disagree  0  1  2  3  4  5  6  7  8  9  10 Totally agree

**... kind of gamblers.**

Totally disagree  0  1  2  3  4  5  6  7  8  9  10 Totally agree

**... kind of selfish.**

Totally disagree  0  1  2  3  4  5  6  7  8  9  10 Totally agree

Individuals who **do not** possess any risky financial investments such as stocks are, on average, ...

**... kind of greedy.**

Totally disagree  0  1  2  3  4  5  6  7  8  9  10 Totally agree

**... kind of gamblers.**

Totally disagree  0  1  2  3  4  5  6  7  8  9  10 Totally agree

**... kind of selfish.**

Totally disagree  0  1  2  3  4  5  6  7  8  9  10 Totally agree

Next

Figure 1.L.2. Perception survey view of stockholders elicitation United States

## 1.L.2 Description experiment decision screens



### Instructions

---

In this section, you will make 4 decisions. In each decision, you will separately receive \$30 from us. With this money, you can choose between two options:

**Option A:** Participate in the stock market by buying a share. The value of the share depends on the movement of the stock market.

**Option B:** Do not participate in the stock market.

In each decision, you will have the option to buy a different share. Each share has a different price and offers different returns. The decisions are presented independently of each other. That is, your choice in one decision does not affect the other decisions.

Note that each decision is going to have the bonus sign. Your choices thus have a chance of being implemented.

Next

**Figure 1.L.3.** Description experiment stock description intro



### Decision 2

You have \$30. This decision features a different share of an exchange-traded fund with different prices than before. Accordingly, you can choose whether to participate in the stock market (Option A) or not (Option B). We start by explaining Option A.

**Option A:** Participate in the stock market by buying a share of an exchange-traded fund (ETF) that replicates the MSCI World index.

Below you can see how the price of a share of the ETF has developed in the years 2010, 2011 and 2016. Each row represents the [actual](#) price in a particular year. The first row shows the price to which you can buy the share.

Share Price of the MSCI World ETF in Dollar	
2010 (Buying Price)	\$27.19
2011	\$25.06
2016	\$37.21

**Consequences of Option A:** If you choose Option A, you buy in the year 2010 the share of the ETF that replicates the MSCI World index to the corresponding price of **\$27.19**. You receive the money not invested in the stock market,  $\$30 - \$27.19 = \$2.81$ , as additional payment with certainty. Additionally, you receive further money depending on the movement of the stock market.

This works as follows: A computer randomly selects one of the two selling years (2011 or 2016). Each year has the same chance of being selected. At the selected selling year, the share you bought will be sold. You then additionally receive the price of the share at the selected year as payment on top of the money that was not invested.

Accordingly, this investment has an expected value of \$31.13. Together with the money not invested (\$2.81), in [expectation](#) you receive **\$33.94** from choosing Option A.

**Option B:** Do not participate in the stock market.

**Consequences of Option B:** If you choose Option B, you receive the \$30 as additional payment. Your payment does not depend on the movement of the stock market in this case.

#### Your Choice

Which option do you choose?

- Option A: I want to participate in the stock market
- Option B: I do not want to participate in the stock market

Confirm decision

**Figure 1.L.4.** Description experiment stock description decision





## Instructions

---

In this section, you will make 4 decisions. In each decision, you will separately receive \$30 from us. With this money, you can choose between two options:

**Option A:** Participate in a random draw by buying a ticket. The value of the ticket depends on the outcome of the random draw.

**Option B:** Do not participate in the random draw.

In each decision, you will have the option to buy a different ticket. Each ticket has a different price and offers different prizes. The decisions are presented independently of each other. That is, your choice in one decision does not affect the other decisions.

Note that each decision is going to have the bonus sign. Your choices thus have a chance of being implemented.

Next

**Figure 1.L.5.** Description experiment draw description intro



### Decision 2

You have \$30. This decision features a different random draw with a different price and outcomes than before. Accordingly, you can choose whether to participate in a random draw (Option A) or not (Option B). We start by explaining Option A.

**Option A:** Participate in a random draw by buying a ticket. The random draw yields different outcomes. Each outcome pays a different prize.

Below you can see the prizes that come with the potential outcomes of the random draw. Each row represents one prize for a potential outcome. The first row shows the price to which you can buy the ticket.

	Prize in Dollar
<b>Buying Price</b>	<b>\$27.19</b>
Outcome 1	\$25.06
Outcome 2	\$37.21

**Consequences of Option A:** If you choose Option A, you buy the ticket at a price of **\$27.19**. You receive the money not used for the ticket,  $\$30 - \$27.19 = \mathbf{\$2.81}$ , as additional payment with certainty. Additionally, you receive further money depending on the outcomes of a random draw.

This works as follows: A computer randomly selects an outcome (1 or 2). Each outcome has the same chance of being selected. You then additionally receive the prize of the selected outcome as payment on top of the money that was not used to buy the ticket.

Accordingly, this ticket has an expected value of \$31.13. Together with the money not used (\$2.81), in [expectation](#) you receive **\$33.94** from choosing Option A.

**Option B:** Do not participate in the random draw.

**Consequences of Option B:** If you choose Option B, you receive the \$30 as additional payment. Your payment does not depend on the outcome of the random draw in this case.

### Your Choice

Which option do you choose?

- Option A: I want to participate in the random draw
- Option B: I do not want to participate in the random draw

Confirm decision

**Figure 1.L.6.** Description experiment draw description decision

### 1.L.3 Information experiment information provision screens

The following part is about a decision recently presented to a large number of people in the US who, just like you, participate in online surveys on Prolific.

These participants were given \$100 from us. They could divide this money freely between themselves and a charitable organization that helps and supports children with critical illnesses. Each dollar that they distributed to themselves would be given to them as additional payment. Each dollar that they distributed to the charity would be donated.

**This decision was not hypothetical:** All participants knew that after making their decision, a randomly selected subset of participants would get their decision implemented with real consequences. Since the decision thus led to actual payments and donations, the participants took it very seriously.

Next

**Figure 1.L.7.** Information experiment instructions screen 1

The decision of how much to donate to a charity in such a situation has been found to be strongly associated with people's behavior in relevant real-life situations and their general personality.

For example, the people that donate less, also tend to be less supportive of their friends and relatives and less willing to help people in need. They are also less likely to engage in voluntary work for a good cause. Conversely, the more people donate, the more they tend to be supportive, helpful, and willing to work voluntarily.

It is also well established that the donation decision is related to how egoistic, self-centered, and greedy a person is. The less people donate, the more egoistic and greedy they are. That is, the decision is indicative of people's pattern of thinking and feelings alongside these characteristics.

To summarize, the result of many research studies is the following: The more people donate money to a good cause when given an amount to freely distribute, the more they care about the well-being of other people in general and act to improve it. Note that this finding is not driven by differences in people's financial situation. For example, comparing people with the same income, it is well documented that people who donate more also generally care more about others.

Next

**Figure 1.L.8.** Information experiment instructions screen 2

Out of all the people who made the previously described decision, a computer randomly selected 10 **stockholders**. Being a stockholder means these people indicated that they participate in the stock market by holding risky financial assets such as stocks or funds ([click for details](#)). Further, a computer randomly selected 10 **non-stockholders** with a similar level of annual income as the **stockholders**. Accordingly, **non-stockholders** do not participate in the stock market.

We are now interested in what you believe about the donation behavior of these two groups. Do you think the group of **non-stockholders** donated on average more to the charity than the group of **stockholders** or was it the other way around? Or do you think they donated the same amount? Keep in mind that both groups have a similar level of income.

- I think **non-stockholders** donated more than **stockholders**
- I think **stockholders** donated more than **non-stockholders**
- I think **non-stockholders** and **stockholders** donated the same amount

Next

**Figure 1.L.9.** Information experiment instructions screen 3

You have indicated that you think non-stockholders donated more than stockholders. How many percent do you think did non-stockholders donated more compared to stockholders?

For example, if you type 100, you think non-stockholders donated 100% more than stockholders, i.e., twice as much. A number of 30 means you think non-stockholders donated 30% more and so on.

<input type="text"/>	%
----------------------	---

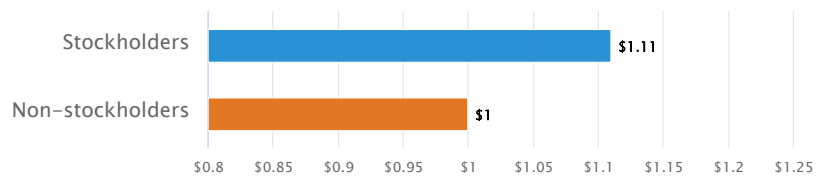
Next

Figure 1.L.10. Information experiment instructions screen 4

We will now inform you about the **actual** difference between **stockholders** and **non-stockholders**. For the randomly selected participants, we found that ...

... **stockholders donate 11% more than non-stockholders**.

That is, for every \$1 that **non-stockholders** donate, **stockholders** donate on average \$1.11 for a good cause.



Next

Figure 1.L.11. Information experiment instructions screen 5

Please repeat the information you just received on the previous page:

Which group donated more on average?

- Non-stockholders donated more
- Stockholders donated more
- Non-stockholders and stockholders donated the same amount
- I don't recall

Next

**Figure 1.L.12.** Information experiment instructions screen 6



### 1.L.4 Follow-up survey behavior intention elicitation screen

How much do you agree or disagree with the statements below?

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
In the next five years, I will under <b>no</b> circumstances or situations whatsoever <b>donate more money to a good cause</b> than I currently do.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In the next five years, I will under <b>no</b> circumstances or situations whatsoever <b>invest any of my own money into financial assets</b> traded on the stock market such as funds or shares.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In the next five years, I will under <b>no</b> circumstances or situations whatsoever <b>do more sports</b> than I currently do.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In the next five years, I will under <b>no</b> circumstances or situations whatsoever <b>make more environmentally friendly decisions</b> .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In the next five years, I will under <b>no</b> circumstances or situations whatsoever <b>buy disability insurance or increase coverage</b> .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 1.L.13. Follow-up survey investment intention elicitation screen

### 1.L.5 Follow-up survey perceptions elicitation screen

Please think about the personality traits of the following two groups: Those who **invest any money into financial assets traded on the stock market such as funds or shares** (labeled "Stockholder" below) and those **without financial assets traded on the stock market** (labeled "Non-stockholder" below).

For each character trait, please indicate what you think about people belonging to the two groups on the following scale from 1 to 9.

For example, if you think stockholder are much more intelligent, select 1 using the slider. If you think non-stockholder are much more intelligent, selected 9. If you see no difference between the two groups, select 5 by clicking on the respective position.

Stockholder 1 2 3 4 No difference 5 6 7 Non-Stockholder 8 9

Intelligent



Selfish



Generous



Close-minded



Figure 1.L.14. Follow-up survey perceptions elicitation screen

## References

- Ahler, Douglas J.** 2014. "Self-Fulfilling Misperceptions of Public Polarization." *Journal of Politics* 76 (3): 607–20. [11]
- Akerlof, George A., and Rachel E. Kranton.** 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–53. [11, 13]
- Alesina, Alberto, Armando Miano, and Stefanie Stantcheva.** 2023. "Immigration and Redistribution." *Review of Economic Studies* 90 (1): 1–39. [90]
- Ariely, Dan, Anat Bracha, and Stephan Meier.** 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review* 99 (1): 544–55. [11, 26]
- Atkin, David, Eve Colson-Sihra, and Moses Shayo.** 2021. "How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption." *Journal of Political Economy* 129 (4): 1193–251. [11]
- Baillon, Aurélien, Zhenxing Huang, Asli Selim, and Peter P. Wakker.** 2018. "Measuring Ambiguity Attitudes for All (Natural) Events." *Econometrica* 86 (5): 1–15. [98]
- Barberis, Nicholas, Ming Huang, and Richard H. Thaler.** 2006. "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing." *American Economic Review* 96 (4): 1069–90. [10]
- Bauer, Rob, and Paul Smeets.** 2015. "Social Identification and Investment Decisions." *Journal of Economic Behavior and Organization* 117: 121–34. [11]
- Bénabou, Roland, Armin Falk, Luca Henkel, and Jean Tirole.** 2022. "Eliciting Moral Preferences: Theory and Experiment." *Working Paper*. [26]
- Bénabou, Roland, and Jean Tirole.** 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78. [11]
- Bénabou, Roland, and Jean Tirole.** 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2): 805–55. [11, 42]
- Bertrand, Marianne, Emir Kamenica, and Jessica Pan.** 2015. "Gender Identity and Relative Income within Households." *Quarterly Journal of Economics* 130 (2): 571–614. [11]
- Böhm, Robert, Hannes Rusch, and Jonathan Baron.** 2020. "The Psychology of Intergroup Conflict: A Review of Theories and Measures." *Journal of Economic Behavior and Organization* 178: 947–62. [44]
- Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak.** 2019. "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch." *Journal of Political Economy* 127 (5): 2143–85. [42]
- Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini.** 2021. "Identity, Beliefs, and Political Conflict." *Quarterly Journal of Economics* 136 (4): 2371–411. [12]
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer.** 2016. "Stereotypes." *Quarterly Journal of Economics* 131 (4): 1753–94. [10, 11, 38]
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer.** 2019. "Beliefs about Gender." *American Economic Review* 109 (3): 739–73. [11]
- Burke, Peter J., and Jan E. Stets.** 2009. *Identity Theory*. New York: Oxford University Press. [12]
- Bursztyn, Leonardo, Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao.** 2018. "Status Goods: Experimental Evidence from Platinum Credit Cards." *Quarterly Journal of Economics* 133 (3): 1561–95. [11]

- Bursztyn, Leonardo, Bruno Ferman, Ali Hasanain, Michael Callen, Saad Gulzar, and Noam Yuchtman.** 2020. "Political Identity: Experimental Evidence on Anti-Americanism in Pakistan." *Journal of the European Economic Association* 18 (5): 2532–60. [11]
- Bursztyn, Leonardo, and Robert Jensen.** 2017. "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure." *Annual Review of Economics* 9: 131–53. [11]
- Bursztyn, Leonardo, and David Y. Yang.** 2022. "Misperceptions about Others." *Annual Review of Economics* 14: 425–52. [11, 90]
- Chen, Daniel L., Martin Schonger, and Chris Wickens.** 2016. "oTree-An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9: 88–97. [20]
- Cullen, Zoë, and Ricardo Perez-Truglia.** 2022. "How Much Does Your Boss Make? The Effects of Salary Comparisons." *Journal of Political Economy* 130 (3): 766–822. [11]
- Currie, Janet.** 2006. "The Take-up of Social Benefits." In *Poverty, the Distribution of Income, and Public Policy*, edited by Alan Auerbach, David Card, and John Quigley, 80–148. New York: Russell Sage. [44]
- D'Acunto, Francesco.** 2020. "Tear Down This Wall Street: Anti-finance Rhetoric, Subjective Beliefs, and Investment." *Working Paper*. [11]
- D'Acunto, Francesco, Marcel Prokopczuk, and Michael Weber.** 2019. "Historical Antisemitism, Ethnic Specialization, and Financial Development." *Review of Economic Studies* 86 (3): 1170–206. [11]
- Danz, David, Marissa Lepper, Guillermo Lezama, Priyoma Mustafi, Lise Vesterlund, Alistair Wilson, and K. Pun Winichakul.** 2023. "Simon Doesn't Say: Minimal Qualitative Distortions from Experimenter Demand." *Working Paper*. [32]
- De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth.** 2018. "Measuring and Bounding Experimenter Demand." *American Economic Review* 108 (11): 3266–302. [32]
- Depetris-Chauvin, Emilio, Ruben Durante, and Filipe Campante.** 2020. "Building Nations through Shared Experiences: Evidence from African Football." *American Economic Review* 110 (5): 1572–602. [11]
- Dimmock, Stephen G., Roy Kouwenberg, Olivia S. Mitchell, and Kim Peijnenburg.** 2016. "Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence." *Journal of Financial Economics* 119 (3): 559–77. [10]
- Dimmock, Stephen G., Roy Kouwenberg, and Peter P. Wakker.** 2016. "Ambiguity Attitudes in a Large Representative Sample." *Management Science* 62 (5): 1363–80. [10]
- Doosje, Bertjan, Naomi Ellemers, and Russell Spears.** 1995. "Perceived Intragroup Variability as a Function of Group Status and Identification." *Journal of Experimental Social Psychology* 31 (5): 410–36. [99]
- Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann.** 2022. "Moral Universalism: Measurement and Economic Relevance." *Management Science* 68 (5): 3590–603. [40]
- Exley, Christine L, John-Henry Pezzuto, and Marta Serra-Garcia.** 2023. "Beliefs About Giving Across Contexts." *Working Paper*. [26]
- Eyal, Peer, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina.** 2021. "Data Quality of Platforms and Panels for Online Behavioral Research." *Behavior Research Methods* 54: 1643–62. [20]
- Fagereng, Andreas, Charles Gottlieb, and Luigi Guiso.** 2017. "Asset Market Participation and Portfolio Choice over the Life-Cycle." *Journal of Finance* 72 (2): 705–50. [6]

- Falk, Armin.** 2021. "Facing Yourself - A Note on Self-Image." *Journal of Economic Behavior and Organization* 186: 724–34. [11]
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde.** 2018. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics* 133 (4): 1645–92. [26]
- Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde.** 2022. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Management Science* 69 (4): 1935–50. [97]
- Falk, Armin, and Florian Zimmermann.** 2017. "Consistency as a Signal of Skills." *Management Science* 63 (7): 2197–210. [23]
- Fernandes, Daniel, John G. Lynch, and Richard G. Netemeyer.** 2014. "Financial Literacy, Financial Education, and Downstream Financial Behaviors." *Management Science* 60 (8): 1861–83. [43]
- Fouka, Vasiliki.** 2020. "Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I." *Review of Economic Studies* 87 (1): 204–39. [90]
- Gaudecker, Hans-martin Von, Axel Wogroly, and Christian Zimpelmann.** 2022. "The Distribution of Ambiguity Attitudes." *Working Paper*. [97, 98]
- Ghosh, Arkadev.** 2022. "Religious Divisions and Production Technology: Experimental Evidence from India." *SSRN Electronic Journal*. [11]
- Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus.** 2021. "Five Facts about Beliefs and Portfolios." *American Economic Review* 111 (5): 1481–522. [10]
- Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai.** 2021. "Household Finance." *Journal of Economic Literature* 59 (3): 919–1000. <https://doi.org/10.1257/jel.20201461>. [5, 10]
- Goodwin, Geoffrey P., Jared Piazza, and Paul Rozin.** 2014. "Moral Character Predominates in Person Perception and Evaluation." *Journal of Personality and Social Psychology* 106 (1): 148–68. [12]
- Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa.** 2011. "IQ and Stock Market Participation." *Journal of Finance* 66 (6): 2121–64. [10]
- Grosfeld, Irena, Alexander Rodnyansky, and Ekaterina Zhuravskaya.** 2013. "Persistent Antimarket Culture: A Legacy of the Pale of Settlement after the Holocaust." *American Economic Journal: Economic Policy* 5 (3): 189–226. [11]
- Grossman, Zachary, and Joël J. van der Weele.** 2017. "Self-Image and Willful Ignorance in Social Decisions." *Journal of the European Economic Association* 15 (1): 173–217. [11, 26]
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales.** 2008. "Trusting the Stock Market." *Journal of Finance* 63 (6): 2557–600. [11, 98]
- Guiso, Luigi, and Paolo Sodini.** 2013. "Household Finance: An Emerging Field." In *Handbook of the Economics of Finance*, edited by George M. Constantinides, Milton Harris, and Rene M. Stulz, 2: 1397–532. Elsevier. [6, 44]
- Gupta, Neeraja, Luca Rigotti, and Alistair Wilson.** 2021. "The Experimenters' Dilemma: Inferential Preferences over Populations." *Working Paper*. [20]
- Haaland, Ingar, Christopher Roth, and Johannes Wohlfart.** 2023. "Designing Information Provision Experiments." *Journal of Economic Literature* 61 (1): 3–40. [9]
- Haliassos, Michael, Thomas Jansson, and Yigitcan Karabulut.** 2017. "Incompatible European Partners? Cultural Predispositions and Household Financial Behavior." *Management Science* 63 (11): 3780–808. [11]
- Heaton, John, and Deborah Lucas.** 2000. "Portfolio Choice in the Presence of Background Risk." *Economic Journal* 110 (460): 1–26. [6]

- Heffetz, Ori, and Robert H. Frank.** 2011. "Preferences for Status: Evidence and Economic Implications." In *Handbook of Social Economics*, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 1: 69–91. North-Holland: Elsevier B.V. [13]
- Henkel, Luca, Philipp Sprengholz, Lars Korn, Cornelia Betsch, and Robert Böhm.** 2023. "The Association between Vaccination Status Identification and Societal Polarization." *Nature Human Behaviour* 7 (2): 231–39. [99]
- Hill, Daniel H.** 2006. "Wealth Dynamics: Reducing Noise in Panel Data." *Journal of Applied Econometrics* 21 (6): 845–60. [42]
- Hjort, Jonas.** 2014. "Ethnic Divisions and Production in Firms." *Quarterly Journal of Economics* 129 (4): 1899–946. [11]
- Hobolt, Sara B., Thomas J. Leeper, and James Tilley.** 2021. "Divided by the Vote: Affective Polarization in the Wake of the Brexit Referendum." *British Journal of Political Science* 51 (4): 1476–93. [16]
- Hurd, Michael, Maarten Van Rooij, and Joachim Winter.** 2011. "Stock Market Expectations of Dutch Households." *Journal of Applied Econometrics* 26 (3): 416–36. [97, 98]
- Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes.** 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (3): 405–31. [16]
- Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor.** 2019. "The Rate of Return on Everything, 1870-2015." *Quarterly Journal of Economics* 134 (3): 1225–98. <https://doi.org/10.1093/qje/qjz012>. [5]
- Kaustia, Markku, and Sami Torstila.** 2011. "Stock Market Aversion? Political Preferences and Stock Market Participation." *Journal of Financial Economics* 100 (1): 98–112. [11, 98]
- Klor, Esteban F., and Moses Shayo.** 2010. "Social Identity and Preferences over Redistribution." *Journal of Public Economics* 94 (3-4): 269–78. [99]
- Kuchler, Theresa, and Johannes Stroebel.** 2021. "Social Finance." *Annual Review of Financial Economics* 13: 37–55. [11]
- Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins.** 2020. "Income and Wealth Inequality in America, 1949-2016." *Journal of Political Economy* 128 (9): 3469–519. [6]
- Leach, Colin Wayne, Martijn van Zomeren, Sven Zebel, Michael L.W. Vliek, Sjoerd F. Pennekamp, Bertjan Doosje, Jaap W. Ouwerkerk, and Russell Spears.** 2008. "Group-Level Self-Definition and Self-Investment: A Hierarchical (Multicomponent) Model of In-Group Identification." *Journal of Personality and Social Psychology* 95 (1): 144–65. [99]
- Lenz, Guido, and Maximilian Mayer.** 2023. "Hollywood, Wall Street, and Mistrusting Individual Investors." *Journal of Economic Behavior and Organization* 210: 117–38. [11]
- Lipovetsky, Stan, and Michael Conklin.** 2001. "Analysis of Regression in Game Theory Approach." *Applied Stochastic Models in Business and Industry* 17 (4): 319–30. [36]
- Mechtenberg, Lydia, Grischa Perino, Nicolas Treich, Jean-Robert Tyran, and Stephanie Wang.** 2022. "Self-Signaling in Voting." *Working Paper*. [11]
- Meyer, Bruce D., Wallace K.C. Mok, and James X. Sullivan.** 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29 (4): 199–226. [42]
- Mummolo, Jonathan, and Erik Peterson.** 2019. "Demand Effects in Survey Experiments: An Empirical Assessment." *American Political Science Review* 113 (2): 517–29. [32]
- Nieddu, Marco, and Lorenzo Pandolfi.** 2021. "Cutting Through the Fog: Financial Literacy and Financial Investment Choices." *Journal of the European Economic Association* 19 (1): 237–74. [21]
- Nofsinger, John R.** 2017. *The Psychology of Investing*. Routledge. [6]

- Nyhan, Brendan, and Jason Reifler.** 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32 (2): 303–30. [90]
- Oh, Suanna.** 2021. "Does Identity Affect Labor Supply?" *Working Paper*. [11]
- Puri, Manju, and David T. Robinson.** 2007. "Optimism and Economic Choice." *Journal of Financial Economics* 86 (1): 71–99. [10]
- Schneider, Florian H.** 2022. "Signaling Ideology through Consumption." *Working Paper*. [11]
- Shayo, Moses.** 2020. "Social Identity and Economic Policy." *Annual Review of Economics* 12: 355–89. [11]
- Shayo, Moses, and Asaf Zussman.** 2011. "Judicial Ingroup Bias in the Shadow of Terrorism." *Quarterly Journal of Economics* 126 (3): 1447–84. [11]
- Shefrin, Hersh.** 2002. *Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing*. New York: Oxford Academic. [6]
- Stets, Jan E., and Richard T. Serpe.** 2013. "Identity Theory." In *Handbook of Social Psychology*, edited by John DeLamater and Amanda Ward, 31–60. Springer. [12]
- Strohinger, Nina, and Shaun Nichols.** 2014. "The Essential Moral Self." *Cognition* 131 (1): 159–71. [12]
- Tversky, Amos, and Daniel Kahneman.** 1983. "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90 (4): 293–315. [38]
- van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie.** 2011. "Financial Literacy and Stock Market Participation." *Journal of Financial Economics* 101 (2): 449–72. [10, 48, 67, 97]
- Vissing-Jorgensen, Annette.** 2004. "Perspectives on Behavioral Finance: Does "Irrationality" Disappear with Wealth? Evidence from Expectations and Actions." In *NBER Macroeconomics Annual 2003*, edited by Mark Gertler and Kenneth Rogoff, 139–94. Cambridge, MA: MIT Press. [10]
- Wojciszke, Bogdan, Roza Bazinska, and Marcin Jaworski.** 1998. "On the Dominance of Moral Categories in Impression Formation." *Personality and Social Psychology Bulletin* 24 (12): 1251–63. [12]





## Chapter 2

# The Association Between Vaccination Status Identification and Societal Polarization\*

*Joint with Philipp Sprengholz, Lars Korn, Cornelia Betsch, and Robert Böhm*

### 2.1 Introduction

While COVID-19 vaccines have been widely available in many countries for some time, a significant proportion of people remain unvaccinated (Mallapaty, 2022). As uptake slowed down despite the availability of vaccines, public discourse around the issue intensified, prompting calls for mandatory vaccination in many countries. Thousands of people have participated in public demonstrations both for and against vaccination and, more specifically, for and against associated mandates (DW, 2022). As this situation is unfamiliar, little is known about what causes people to join one of these polarized camps.

How can we explain this shift from relating to one's own vaccination status as an individual and private identity to a publicly shared group identity involving col-

\* This chapter has appeared as Preprint on PsyArXiv (<https://doi.org/10.31234/osf.io/mgqk5>). A modified version of this chapter has been published in *Nature Human Behavior* (Henkel et al., 2023).

**Acknowledgments:** Financial support for this study was provided in part by grants from University of Erfurt and Thüringer Staatskanzlei (awarded to Cornelia Betsch) and the University of Vienna (awarded to Robert Böhm) as well as the German Research Foundation (DFG, BE BE3970/11-1). Luca Henkel: Funding by the German Research Foundation under Germany's Excellence Strategy (EXC 2126/1-390838866) is gratefully acknowledged. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

**Ethics approval:** The study obtained ethics approval from the University of Erfurt's institutional review board (#20211215).

**Preregistration:** The study was preregistered at [aspredicted.org](https://aspredicted.org) (<https://aspredicted.org/nn2as.pdf>).

lective action intentions? Some researchers have argued that individual vaccination status can become an important part of people's identity (Attwell and Smith, 2017), and identification with these social groups has likely increased for some during the long-lasting pandemic, which has had (and continues to have) a great impact on people's lives (Motta et al., 2021; Wakefield and Khauser, 2021). Moreover, belonging to one group or the other has had real-world impacts in many countries, e.g., due to access restrictions based on vaccination status.

Research on opinion-based groups shows that social groups can form around shared attitudes (Bliuc et al., 2007; McGarty et al., 2009). While in the real world groups often cluster based on socio-demographic characteristics, opinion-based groups can profit especially from online interaction (Garcia, Galaz, and Daume, 2019). According to social identity theory (Tajfel and Turner, 1979; Turner and Reynolds, 2008b), individuals use their group memberships to define who they are (i.e., their self-concept) and to determine what to think and how to behave. The related self-categorization theory (Turner and Reynolds, 2008a) proposes that social categories are represented as a set of attributes that capture similarities within and differences between groups, including the groups' values, norms, and behaviors. In the context of vaccination and related policies, this could include, for instance, values related to belief in science, freedom to make one's own decisions, or trust in certain media. According to these prominent theoretical perspectives, people tend to view their own social groups (ingroups) as distinctive and superior to other groups (outgroups) and engage in behaviors that confirm this belief. Thus, strong group identification can fuel intergroup conflict. Such conflict may unfold in terms of how people perceive ingroup and outgroup members and how they actually behave in their interactions with ingroup and outgroup members (Brewer, 2001). Furthermore, individuals' group identification could determine how they respond to different situations that threaten their ingroup's status (Morrison, Fast, and Ybarra, 2009). For example, previous research on the polarization of the American electorate has shown that partisans discriminate against opposing partisans (threatening their status) to a degree that exceeds discrimination based on race (Iyengar and Westwood, 2015).

Using this strong theoretical basis, the present studies tested the fundamental idea that identification with one's vaccination status is a crucial factor in the polarization of related attitudes and behaviors. Using correlational analysis, we provide evidence on the extend of people's identification and its association with measures of societal polarization. We collected panel data from Germany and Austria ( $N = 5,305$ ) in three waves (December 2021, February 2022, and July 2022; plus an additional data collection with a subsample in March 2022) to investigate the correlates and consequences of what we label "vaccination status identification" (VSI). In the total sample, 62% indicated to be vaccinated in December 2021. Compared to the adult German (GER) and Austrian (AUT) population (Bundesanstalt Statistik Österreich, 2022a, 2022b, 2022c; Statistisches Bundesamt, 2022a, 2022b, 2022c),

our sample was slightly younger ( $M_{\text{Sample AUT}} = 44.5$ ,  $M_{\text{Population AUT}} = 49.6$ ,  $M_{\text{Sample GER}} = 47.6$ ,  $M_{\text{Population GER}} = 51.3$ ), roughly balanced across genders (Sample AUT: 53.3% female, Population AUT: 51.3%; Sample GER: 53.0%, Population GER: 51.1%), more educated (Sample AUT: 48.4% have university entrance qualification, Population AUT: 44.1%; Sample GER: 53.1%, Population GER: 37.0%) and roughly similar in their likelihood to be employed (Sample AUT: 60.2%, Population AUT: 60.7%; Sample GER: 61.9%, Population GER: 58.2%). The timing of the studies is unique, as in both countries data were collected during heated public debates about the value of vaccination and different vaccination policies, such as vaccine mandates.

## 2.2 Measurement and correlates of VSI

To measure VSI, we adapted five items from established group identification scales (e.g., “I am proud (not) to be vaccinated against COVID-19”) (Doosje, Ellemers, and Spears, 1995; Roth and Mazziotta, 2015). The items were chosen to capture different dimensions of group identification without needing to assess a large number of items as required by established scales. Data from March 2022 showed that VSI was strongly related to other (broader or more specific) social identification measures developed by Leach and al. (2008) ( $r = .79$ ,  $CI = [.75, .82]$ ) and Doosje, Ellemers, and Spears (1995) ( $r = .72$ ,  $CI = [.68, .76]$ ). As further indicators of the validity of VSI, we assessed typical intergroup phenomena that should be stronger with greater identification with one’s own group. Indeed, when rating the distinctiveness of the groups of vaccinated and unvaccinated people, higher levels of VSI were positively correlated with greater perceived intergroup differences for both vaccinated ( $r_{\text{vaccinated}} = .41$ ,  $CI = [.30, .51]$ ) and unvaccinated individuals ( $r_{\text{unvaccinated}} = .29$ ,  $CI = [.17, .40]$ ). In contrast (and as preregistered), both groups perceived members of their respective ingroup to be more similar to each other the higher VSI was ( $r_{\text{vaccinated}} = .59$ ,  $CI = [.50, .66]$ ;  $r_{\text{unvaccinated}} = .52$ ,  $CI = [.43, .61]$ ). Outgroup members were also perceived to be more similar to each other as respondent’s VSI increased, although this result was more noticeable among unvaccinated ( $r_{\text{unvaccinated}} = .23$ ,  $CI = [.11, .35]$ ) than among vaccinated respondents ( $r_{\text{vaccinated}} = .08$ ,  $CI = [-.05, .20]$ ).

Additionally, VSI proved sufficiently distinct from vaccination intention and vaccine-related feelings and beliefs. The latter were measured using the 7C scale of vaccination readiness (Geiger and al., 2021), which includes confidence in vaccines, complacency, calculation, constraints, collective responsibility, compliance, and conspiracy thinking. The following correlations refer to December 2021 but did not change qualitatively in later data collection timepoints (see Appendix Tables 2.A.1 and 2.A.2). Among unvaccinated individuals, correlations between VSI and the 7C ranged between  $-.28$  (for confidence in vaccines) and  $.39$  (for conspiracy thinking);

the correlation with vaccination intention was  $-.26$ . Among vaccinated individuals, correlations between VSI and the 7C ranged between  $-.64$  (for the perception of constraints) and  $.58$  (for collective responsibility); the correlation with their intention to receive a booster shot was  $.50$ . Taken together, the validation results indicate that (1) VSI is indeed well described as a group identity among the vaccinated and the unvaccinated in the context of the COVID-19 pandemic, and (2) VSI is empirically related but conceptually different from other vaccination-related perceptions and behavioral intentions.

In December 2021, mean VSI was medium to high and varied considerably between individuals, with somewhat higher overall levels among the vaccinated ( $M = 4.74$ ,  $SD = 1.36$ ) than among the unvaccinated ( $M = 4.36$ ,  $SD = 1.25$ ;  $t(4582.60) = 10.42$ ,  $p < .001$ ;  $d = 0.29$ ). Similar levels were found in February and July 2022 (see Appendix Figure 2.A.1). The (small) difference between vaccinated and unvaccinated people appears plausible given that group membership is more likely to change for the unvaccinated as they are able to change their vaccination status and, thus, their group identity. Indeed, for participants whose vaccination status did not change over time, VSI was remarkably stable (correlation between December 2021 and July 2022:  $r_{\text{vaccinated}} = .67$ ,  $CI = [.65, .69]$ ;  $r_{\text{unvaccinated}} = .57$ ,  $CI = [.53, .61]$ ). Participants who decided to get vaccinated between December 2021 and July 2022 ( $n = 144$ ) showed lower levels of VSI in December ( $M = 3.81$ ,  $SD = 1.33$ ) compared to unvaccinated participants who did not get vaccinated and thus did not change their group membership ( $M = 4.51$ ,  $SD = 1.25$ ;  $t(173.90) = 5.79$ ,  $p < .001$ ;  $d = 0.52$ ). Unvaccinated individuals' VSI in December 2021 was thus predictive of their subsequent likelihood of vaccine uptake – the lower their previous VSI, the higher the likelihood of getting vaccinated.

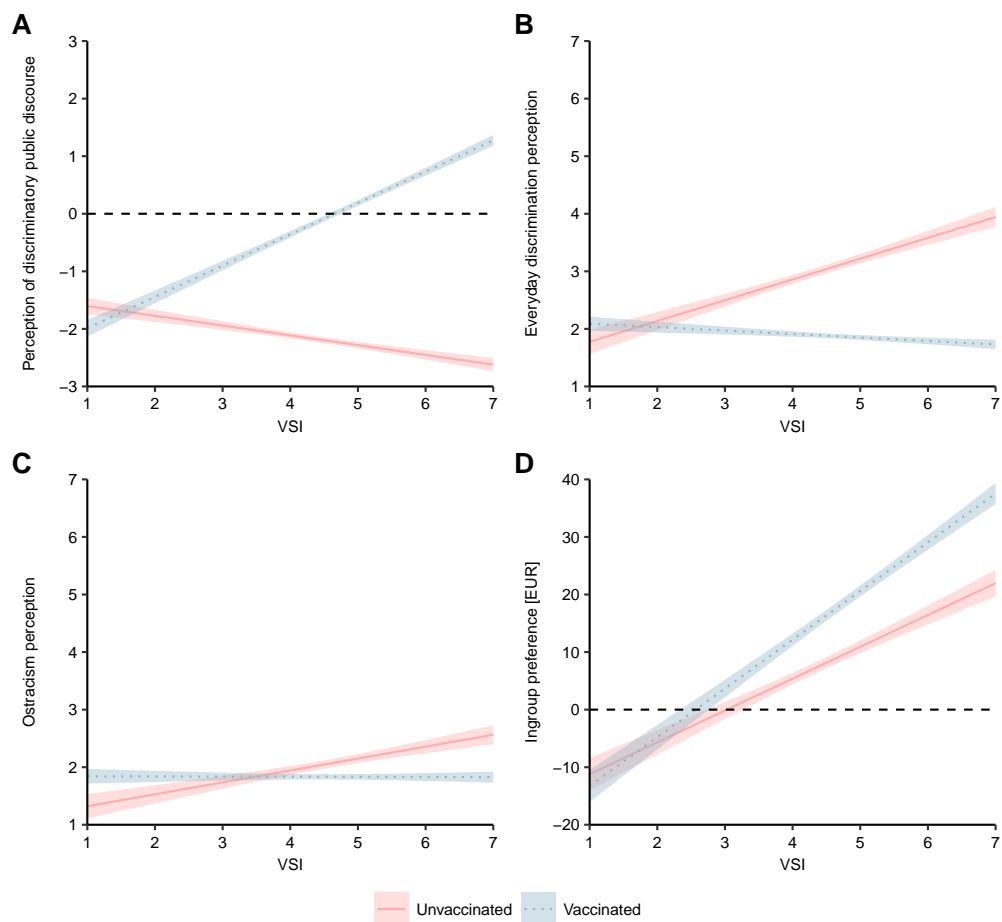
Potential predictors of VSI were explored separately for vaccinated and unvaccinated participants in December 2021 by using regression analyses including socio-demographic variables, news consumption behaviors, trust in the government, and perceptions of social norms around vaccination. Among the vaccinated, individuals were found to identify more strongly with their vaccination status when they were older ( $\beta = 0.19$ ,  $b = 0.01$ ,  $CI = [0.01, 0.01]$ ), trusted the government more ( $\beta = 0.29$ ,  $b = 0.17$ ,  $CI = [0.14, 0.19]$ ), when they reported that people important to them were vaccinated (descriptive norm to be vaccinated;  $\beta = 0.10$ ,  $b = 0.09$ ,  $CI = [0.04, 0.14]$ ) and, most notably, when they assumed that people important to them thought that one should be vaccinated (injunctive norm to be vaccinated;  $\beta = 0.34$ ,  $b = 0.25$ ,  $CI = [0.20, 0.30]$ ). VSI was also stronger among those who reported searching more frequently for COVID-19-related information ( $\beta = 0.21$ ,  $b = 0.14$ ,  $CI = [0.11, 0.17]$ ). Left-wing voters identified less with being vaccinated than centrists ( $b = -0.10$ ,  $CI = [-0.21, 0.00]$ ), and the same was true for right-wing voters ( $b = -0.30$ ,  $CI = [-0.42, -0.12]$ ) and nonvoters ( $b = -0.33$ ,  $CI = [-0.43, -0.23]$ ).

The unvaccinated identified more strongly with being unvaccinated when they did not perceive a descriptive norm to be vaccinated ( $\beta = -0.14$ ,  $b = -0.08$ ,  $CI = [-0.13, -0.04]$ ), when they trusted the government less ( $\beta = -0.18$ ,  $b = -0.15$ ,  $CI = [-0.20, -0.09]$ ), when they claimed to vote for right-wing parties rather than centrist parties ( $b = 0.36$ ,  $CI = [0.17, 0.55]$ ), and when they consumed less information from traditional news sources, such as TV, radio, newspapers, or government websites ( $\beta = -0.13$ ,  $b = -0.46$ ,  $CI = [-0.67, -0.25]$ ), and instead consumed more information from alternative sources, such as social media and messaging services ( $\beta = 0.16$ ,  $b = 0.62$ ,  $CI = [0.40, 0.84]$ ). Importantly, the results of both regression analyses remained qualitatively unchanged when controlling for vaccination intention and the 7C scale (see Appendix Table 2.A.3).

## 2.3 Perceptions of public discourse and discrimination

According to previous research (Korn et al., 2020), vaccination is perceived as a social contract. Because it has positive consequences for others, those who get vaccinated (and comply with the contract) tend to treat others who also comply more favorably than unvaccinated (non-compliant) others. Importantly, the vaccinated also tend to punish unvaccinated others (Korn et al., 2020; Weisel, 2021; Bor, Jørgensen, and Petersen, 2023), which can be interpreted as a manifestation of conflict and polarization. In line with this tendency, in December 2021, 82% of the unvaccinated perceived public discourse around vaccination as unfair, moralistic, and patronizing, while only 23% of the vaccinated reported feeling this way. Importantly, this perception was moderated by VSI (Figure 2.1 Panel A); higher levels of VSI were associated with perceiving the public discourse as more positive among the vaccinated but as more negative among the unvaccinated. The results were similar for general perceptions of being discriminated against, as measured by a short five-item version of the Everyday Discrimination Scale (e.g., “Other people act as if I am not intelligent”) (Sternthal, Slopen, and Williams, 2011). Among vaccinated individuals, average perceived discrimination was low (December 2021:  $M = 1.87$ ,  $SD = 1.22$ ); among the unvaccinated, it was higher ( $M = 2.99$ ,  $SD = 1.76$ ;  $t(3259.70) = 25.23$ ,  $p < .001$ ;  $d = 0.74$ ) and increased with VSI (Figure 2.1 Panel B). The same pattern was found for severe forms of discrimination measured with the Ostracism Short Scale (Rudert et al., 2020) in February 2022 (e.g., “Others exclude me from conversations”). While perceived ostracism was low in both groups, unvaccinated individuals ( $M = 2.03$ ,  $SD = 1.45$ ) had experienced slightly more social exclusion than vaccinated participants ( $M = 1.83$ ,  $SD = 1.27$ ;  $t(2705.90) = 4.59$ ,  $p < .001$ ,  $d = 0.15$ ). Importantly, ostracism was not related to VSI among the vaccinated but was found to be positively related among the unvaccinated (Figure 2.1 Panel C).

To investigate whether perceived discrimination had any factual basis, participants were asked to play two dictator games. They were asked to distribute



**Figure 2.1.** Effects of vaccination status identification (VSI) on perceptions of public discourse and discrimination

*Notes:* Linear regression analyses of VSI, vaccination status, and their interaction, predicting (**Panel A**) perceptions of public discourse (average of three 7-point scales ranging from unfair, moralistic, and patronizing to fair, objective, and respectful; data from December 2021),  $R^2 = .46$ ; (**Panel B**) perceptions of everyday discrimination (mean of five items measured on 7-point scales; higher values indicate more discrimination, data from December 2021),  $R^2 = .16$ ; (**Panel C**) perceptions of being ostracized (mean of four items measured on 7-point scales; higher values indicate being more ostracized, data from February 2022),  $R^2 = .02$ ; (**Panel D**) ingroup preference in two dictator games (positive values indicate greater ingroup preference, i.e. discrimination of the outgroup, data from December 2021),  $R^2 = .16$ . Lines represent the linear fit, with ribbons visualizing 95% confidence intervals. The pattern of results did not change qualitatively when controlling for sociodemographic variables and the 7C scale (see Appendix Tables 2.A.4, 2.A.5, 2.A.6 and 2.A.7).

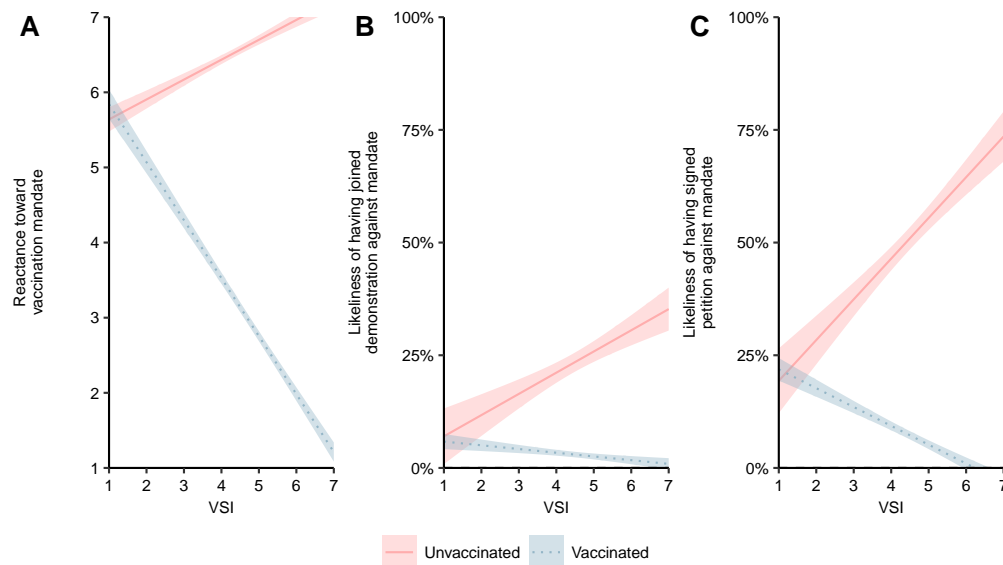
100 EUR between themselves and a vaccinated person (game 1) or an unvaccinated person (game 2, randomized order). Ingroup preference was measured as the difference between the distributed amounts and indicated the strength of discrimination. The games were conducted in December 2021 and repeated in February and July 2022. At all three timepoints, the average ingroup pref-

erence of vaccinated individuals (December 2021:  $M = 18.40$  EUR,  $SD = 29.80$  EUR) was higher compared to unvaccinated participants ( $M = 7.37$  EUR,  $SD = 23.90$  EUR;  $t(4981.00) = 14.86$ ,  $p < .001$ ,  $d = 0.41$ ). While vaccinated and unvaccinated individuals granted comparable amounts to ingroup members (December 2021:  $M_{\text{vaccinated}} = 48.06$  EUR,  $SD_{\text{vaccinated}} = 23.83$  EUR,  $M_{\text{unvaccinated}} = 45.93$  EUR,  $SD_{\text{unvaccinated}} = 25.11$  EUR;  $t(4151.40) = 3.07$ ,  $p = .002$ ,  $d = .08$ ), vaccinated participants gave considerably less money to outgroup members than unvaccinated individuals ( $M_{\text{vaccinated}} = 29.66$  EUR,  $SD_{\text{vaccinated}} = 26.55$  EUR,  $M_{\text{unvaccinated}} = 38.56$  EUR,  $SD_{\text{unvaccinated}} = 25.30$  EUR;  $t(4478.10) = 12.23$ ,  $p < .001$ ,  $d = .34$ ). Furthermore, ingroup preference among the unvaccinated increased with VSI, and this effect was even stronger among vaccinated individuals (Figure 2.1 Panel D). Thus, the more vaccinated people identified with being vaccinated, the more they discriminated against unvaccinated players. The unvaccinated also did this but to a smaller extent. The stronger discrimination behavior of vaccinated individuals matches the finding that perceptions of being discriminated against were reported more frequently by unvaccinated people. This suggests that reports of discrimination are not fiction but fact.

## 2.4 The role of VSI in vaccination policy acceptance

Previous research suggests that low vaccination intentions predict psychological reactance to mandatory vaccination policies, eliciting behaviors that oppose such regulations (Sprengholz, Betsch, and Böhm, 2021; Schmelz and Bowles, 2022). However, there are also some vaccinated people who oppose such mandates (Sprengholz and al., 2022). To better understand the relationship between vaccination status and reactance to mandatory vaccination, we investigated the potential moderating role of VSI. In December 2021, we conducted a between-participants experiment in which participants were asked to imagine that a vaccination mandate would be enforced in the near future. As public discussions loomed around various policy drafts in both Germany and Austria at that time, the experimental conditions took up these discussions and accordingly varied the affected age groups (mandating vaccination for people aged 12 and older vs. 18 and older) and sanctions in case of non-vaccination (fine vs. fine and work bans). Reactance was measured with a single item asking the participants how angry they were about the described mandate.

In line with previous research (Sprengholz, Betsch, and Böhm, 2021), average reactance toward mandatory vaccination was stronger for unvaccinated ( $M = 6.53$ ,  $SD = 1.34$ ) than vaccinated individuals ( $M = 2.96$ ,  $SD = 2.19$ ;  $t(5301.30) = 73.75$ ,  $p < .001$ ,  $d = 1.97$ ). Linear regression analysis revealed that the effect of vaccination status was moderated by VSI (Figure 2.2 Panel A); when VSI was low (i.e., people did not identify with their vaccination status), both vaccinated and unvaccinated individuals reported high levels of reactance. When VSI was high, reactance was



**Figure 2.2.** Effects of vaccination status identification (VSI) on reactance toward mandatory policies and activism behavior

Notes: Linear regression analyses of VSI, vaccination status, and their interaction, predicting (**Panel A**) psychological reactance to a hypothetical vaccination mandate (measured by anger assessed on a 7-point scale, data from December 2021),  $R^2 = .56$ ; (**Panel B**) whether participants reported demonstrating against a vaccine mandate since January 2022 (binary variable, data from February 2022),  $R^2 = .12$ ; (**Panel C**) whether participants reported signing a petition against a vaccine mandate since January 2022 (binary variable, data from February 2022),  $R^2 = .28$ . The pattern of results did not change qualitatively when controlling for sociodemographic variables, the 7C, and, in the case of Panel A, further experimental manipulations (see Appendix Tables 2.A.8, 2.A.9 and 2.A.10). Lines represent the linear fit, with ribbons visualizing 95% confidence intervals.

even greater among the unvaccinated but smaller among the vaccinated. This interaction effect was independent of the manipulated factors (mandated age group and sanctions). Reactance correlated strongly with intended behaviors opposing the mandate, including signing a petition ( $r = .81$ ,  $CI = [.80, .82]$ ), joining a demonstration ( $r = .66$ ,  $CI = [.65, .68]$ ), and mobilizing others to fight the mandate ( $r = .67$ ,  $CI = [.65, .68]$ ). This supports the view that VSI plays an important role in societal polarization as a response to vaccination policies. Higher levels of reactance were also strongly related to the intention to avoid vaccination if it became mandatory ( $r = .76$ ,  $CI = [.74, .77]$ ).

While we only assessed intentions to engage in activism against the mandate in December 2021, we explored the link between VSI and actual behavior reported in February 2022. Specifically, we asked participants whether they had participated in a demonstration or signed a petition against the introduction of vaccination mandates since the beginning of 2022. For low levels of VSI, we found that similar fractions of vaccinated and unvaccinated were involved in demonstrations (Figure 2.2 Panel B)



or signing petitions (Figure 2.2 Panel C). However, for high levels of VSI, we found polarized behavior: the fractions strongly increased for the unvaccinated, while they decreased for vaccinated individuals. Utilizing the panel structure, we further investigated whether those individuals who indicated that they would demonstrate against mandatory vaccinations or sign a petition in December 2021 actually reported having done so in February 2022 and whether this link was influenced by VSI (as measured in December 2021). Indeed, intention was found to predict behavior ( $r_{\text{demonstration}} = .46$ ,  $CI = [.43; .48]$ ;  $r_{\text{petition}} = .55$ ,  $CI = [.53, .57]$ ). A regression analysis investigating the effects of intention, VSI, and their interaction on behaviors of unvaccinated individuals revealed that behavior was more likely when intention was high (demonstration:  $\beta = 0.40$ ,  $b = 0.04$ ,  $CI = [0.01, 0.06]$ ; petition:  $\beta = 0.33$ ,  $b = 0.05$ ,  $CI = [0.01, 0.08]$ ). The effect was moderated by VSI, with higher identification increasing the effect of intention on behavior (demonstration:  $\beta = 0.05$ ,  $b = 0.01$ ,  $CI = [0.01, 0.01]$ ; petition:  $\beta = 0.05$ ,  $b = 0.01$ ,  $CI = [0.02, 0.02]$ ).

## 2.5 Discussion

The findings indicate that the strength of identifying with one's vaccination status is associated with several measures of polarization of the current debate on COVID-19 vaccination. Vaccination status identification (VSI) accounts for much of the variance between vaccinated and unvaccinated individuals' perceptions of public discourse, factual and perceived discrimination, as well as the quality and strength of their responses to mandatory vaccination policies. While our results do not allow for causal interpretation, stronger VSI was associated with greater discrimination against people whose vaccination status differed. Interestingly, the tendency to discriminate was especially pronounced among vaccinated participants. This may be explained by other findings, indicating that vaccination is perceived as a social contract among vaccinated people where violating this social contract by not getting vaccinated is more harshly punished by vaccinated people than conforming to it is punished by unvaccinated people (Korn et al., 2020; Weisel, 2021; Bor, Jørgensen, and Petersen, 2023). Stronger VSI was also found to be related to higher psychological reactance to mandatory vaccination policies among the unvaccinated and to their intentions and actions to resist and evade such regulations. As VSI also related to patterns of traditional and social media use, political preferences, and differences in perceived social norms, it seems plausible that the unvaccinated and vaccinated constitute coherent and distinct social clusters ("bubbles") (Schmidt et al., 2018), which can be seen as a further catalyst of group conflict.

The results presented in this article have some limitations. First, the sample is not representative of the German or Austrian population. Collecting data from thousands of unvaccinated individuals meant to lift sampling requirements. While the demographics show considerable variance (see Appendix Table 2.A.11), generaliza-

tion may still be limited. For instance, while similar to the German and Austrian adult population with respect to gender and employment, our sample is younger and more educated. Second, our design and analyses are correlational and causal interpretations are not possible. For instance, we cannot conclude that VSI drives discrimination; instead, the relationship could also work in the other direction or be bidirectional. Future experimental research should investigate these relationships by manipulating VSI in experimental settings. Third, all variables were self-reported and may have differed from the respondents' actual behaviors. For instance, unvaccinated participants may have exaggerated their intentions to avoid vaccination in the case of mandatory policies.

Despite these limitations, the results suggest that VSI may be important to understanding when a private and personal vaccination decision can become an important aspect of group membership that defines people's self-concept. While group processes have been assumed to be relevant when discussing anti-vaccination attitudes among anthroposophical parents (Sobo, 2015), this study reveals strong correlates of identification with being (not) vaccinated in a general population convenience sample. The current data extend the theoretical perspective on vaccination behavior and its societal consequences and can help predict people's behaviors both within and outside of the health sector. Three key examples demonstrate this. First, unvaccinated individuals with lower levels of VSI in December 2021 were more likely to be vaccinated in July 2022. Second, unvaccinated individuals with higher levels of VSI were more likely to translate their intentions of performing counter-behaviors related to vaccination policies (e.g., attending demonstrations) into actual behaviors. Third, in unrelated money games, unvaccinated people were discriminated against by vaccinated people (and vice versa). Given the importance of these behaviors and what they represent for society in the context of the current pandemic (i.e., vaccine uptake, counter-behaviors in response to vaccination policies, and polarization), VSI appears to be a useful concept that should be considered in future research.

The findings also have practical implications. While a shared social identity is known to act as a buffer against stress from COVID-19-related threats (Rudert and al., 2021), our findings highlight the potential negative consequences of strongly identifying with one's vaccination status. Besides being associated with polarization and potential conflicts between vaccinated and unvaccinated groups, VSI may also impede the success of vaccination campaigns. Appeals to identify with vaccination or a vaccinated society may thus backfire. For instance, such campaigns might increase VSI of vaccinated individuals, who potentially increase their discriminatory behavior toward unvaccinated individuals. According to previous research on rejection identification processes (Branscombe, Schmitt, and Harvey, 1999; Ramos et al., 2012), discrimination against unvaccinated people, in turn, might increase their identification with non-vaccination, lowering vaccination intentions even more and further fueling societal polarization. Talking about vaccination as a simple health intervention may be more successful. These possibilities should be explored in fur-

ther research once a causal role for VSI has been established. Previous research indicates that maintaining procedural fairness, e.g., by government officials treating vaccinated and unvaccinated individuals in a fair, respectful and neutral way, can emphasize the inappropriateness of aggressive interactions and decrease discrimination and ostracism behaviors (Böhm, Rusch, and Baron, 2020). While such measures may not increase vaccination rates directly, they could help de-escalate the situation and provide a new basis for discussing and implementing effective and acceptable vaccination policies in the future. In this vein, the rationale behind vaccination mandates could be revisited in public discussions. While sanctions can increase vaccine uptake (Lee and Robinson, 2016), mandatory regulations may also curb polarization. When individuals become vaccinated to avoid penalties, they cease to be part of the unvaccinated and their identification with this group should subsequently decrease. Similarly, and somewhat counterintuitively, we speculate that mandating vaccinations could help reduce the identification of those who have been vaccinated for a long time. Vaccination will then not be something that expresses individual preferences. And once almost all people have been vaccinated due to mandatory regulations, one will be unable to separate oneself from others with regard to vaccination status, and being vaccinated will consequently become a less important part of one's self-concept. In this way, effective (i.e., enforced) mandates could not only help to end the pandemic but also mitigate conflicts between vaccinated and unvaccinated groups, thereby fostering social cohesion.

## 2.6 Methods

The present data were collected in three longitudinal waves: (1) December 15–27, 2021, (2) February 11–23, 2022, and (3) June 30–July 12, 2022. There was an additional data collection with a subsample conducted on March 14–18, 2022. During these times, infection numbers were high in Germany (with a peak of 297,845 new infections on March 18, 2022) and Austria (with a peak of 51,951 new infections at the same date), and discussions about the introduction of mandatory vaccinations loomed.

**Participants.** In the first wave (December 2021), the panel included  $N = 5,305$  participants from Germany (2,003 vaccinated and 1,230 unvaccinated) and Austria (1,264 vaccinated and 808 unvaccinated). To recruit as many unvaccinated individuals as possible, sociodemographic sampling requirements were lifted. Participants were 18 to 99 years old ( $M = 46.42$ ,  $SD = 16.78$ , with a negligible difference between the vaccinated and unvaccinated,  $d = 0.06$ ). 53% were female, and most participants (51%) indicated high education status (i.e., having university entrance qualifications). In the second wave (February 2022), 4,406 of the original participants (83%) participated again (2,906 vaccinated and 1,500 unvaccinated). Between the first and second wave, 126 participants were vaccinated for the first

time. In the third wave (July 2022), 3,660 of the original participants (69%) participated again (2,442 vaccinated and 1,218 unvaccinated). Between the second and third wave, further 18 participants were vaccinated for the first time. In an additional data collection (March 2022), a randomly selected subset of 600 participants from the second wave was invited to participate in order to validate the VSI measure. In total, 498 participated, with 249 being vaccinated and 249 unvaccinated at the time of data collection.

**Ethical compliance.** The study was conducted in accordance with German Psychological Association guidelines. Ethical clearance was obtained from the University of Erfurt's institutional review board (#20211215) and all participants provided informed consent to use and share their data for scientific purposes without disclosure of their identities. Participants were compensated for their participation by the panel provider.

### 2.6.1 First wave (December 2021)

All measures were assessed in the order of appearance. For all materials and survey questions, participants were told that being vaccinated referred to having received at least one dose of an approved COVID-19 vaccine.

**Voting preferences.** Participants were asked which political party best represents them. Depending on the country, participants could choose a party from a list of German or Austrian parties or indicate that no party represents them. Based on their selection, participants were classified as left-, center-, right- or non-voters. Note that when voting preferences are considered as predictors in regression analyses, no standardized estimates are presented in this article.

**Information behavior.** Participants were asked how often they are searching for information about the coronavirus and COVID-19. Answers were recorded on a 7-point scale ranging from “never” to “very often”. Participants were also asked if they used various types of media to stay informed, including TV, radio, newspapers, news websites, governmental websites, social media (e.g., Facebook and Instagram), and messengers (e.g., WhatsApp and Telegram). Answers were recorded on binary scales (“used” or “not used”).

**Trust in the government.** Participants were asked how confident they were that the federal government could handle the coronavirus properly. Answers were recorded on a 7-point scale, ranging from not at all to very much.

**7C antecedents of vaccination.** Antecedents of vaccination were measured using the short version of the 7C scale (Geiger and al., 2021). It included seven statements about confidence (“I am convinced the appropriate authorities do only allow effective and safe vaccines”), complacency (“I get vaccinated because it is too risky to get infected”), constraints (“Vaccinations are so important to me that I prioritize

getting vaccinated over other things”), calculation (“I only get vaccinated when the benefits clearly outweigh the risks”), collective responsibility (“I see vaccination as a collective task against the spread of diseases”), compliance (“It should be possible to sanction people who do not follow the vaccination recommendations by health authorities”), and conspiracy beliefs (“Vaccinations cause diseases and allergies that are more serious than the diseases they ought to protect us from”). Participants were asked about their agreement on a 7-point scale ranging from “very strongly disagree” to “very strongly agree”.

**Vaccination intention.** If the participants were unvaccinated, they were asked how likely they would be to get vaccinated if they had the chance to do so next week. If the participants were already vaccinated, they were asked how likely they would be to get a booster shot if it was available and recommended to them. Intentions were recorded on a 7-point scale, ranging from “not getting vaccinated at all” to “definitively getting vaccinated”.

**Social norms.** Participants were asked about descriptive and injunctive vaccination norms using two items: “People who are important to me are vaccinated, and People who are important to me think one should be vaccinated”. Answers were recorded on a 7-point scale ranging from “nobody” to “everybody”.

**Perception of the public discourse.** Participants were asked how they perceived the public discourse around vaccination using three 7-point items (1 to 7) with the poles unfair–fair, moralistic–objective, and patronizing–respectful. Answers were mean-averaged, and scores below 3 were considered negative perceptions, while scores above 5 were considered positive perceptions.

**Discrimination perceptions.** General discrimination perceptions were assessed using the Everyday Discrimination Scale (Cronbach’s  $\alpha = .90$ ) (Sternthal, Slopen, and Williams, 2011). Participants were asked to indicate their agreement with five statements (e.g., “Other people act as if I am not intelligent”). Answers were recorded on a 7-point scale ranging from “very strongly disagree” to “very strongly agree”.

**VSI.** Vaccination status identification was measured with five items adapted from established group identification scales (Doosje, Ellemers, and Spears, 1995; Leach and al., 2008; Roth and Mazziotta, 2015): (1) “I am proud (not) to be vaccinated against COVID-19”; (2) “When people are criticized for (not) being vaccinated against COVID-19, it feels like a personal insult to me”; (3) “I have little in common with people who have (not) been vaccinated against COVID-19” (reverse-coded); (4) “I have no problem telling others that I have (not) been vaccinated against COVID-19”; and (5) “If I learned that another person had (not) been vaccinated against COVID-19, I would directly feel more connected to that person”. Answers were recorded on a 7-point scale ranging from “very strongly disagree” to “very

strongly agree”. We explicitly decided for creating a new measure to capture different dimensions of group identification without the need to assess a large number of items as required by established scales like the social identification scale by Leach and al. (2008). While internal consistency was acceptable (December 2021: Cronbach’s  $\alpha = .68$ ), the scale was also successfully validated against other identification measures in an additional wave in March 2022 (see below).

**Ingroup preference in dictator games.** Participants were asked to distribute 100 EUR between themselves and a vaccinated person (game 1) or an unvaccinated person (game 2, randomized order). The games were incentivized by the random selection of one decision by one participant for payout (in case the participant had assigned money to another person, this person was also selected randomly). Ingroup preference was measured as the difference between the amounts distributed to in-group members and outgroup members.

**Experiment on vaccination mandates.** Participants were asked to imagine that a vaccination mandate would soon be enforced. Depending on the experimental condition, the mandate referred to different age groups (12 and older vs. 18 and older) and entailed different sanctions (fine vs. fine and work ban), resulting in a  $2 \times 2$  between-subjects design. Participants were allocated randomly (n12 and older + fine = 1,320, n18 and older + fine = 1,346, n12 and older + fine and work ban = 1,294, n18 and older + fine and work ban = 1,345). After receiving the information, participants were asked how much they supported the presented regulation and how angry they were about it (on a 7-point scale ranging from “not at all” to “very much”). The latter item was adapted from the Salzburg State Reactance Scale (Sittenthaler et al., 2015). Furthermore, activism and avoidance intentions were assessed by asking participants if they would join a demonstration against the mandate, if they would sign a petition against it, if they would mobilize others to fight the regulation, and if they would try to search for ways around the mandate. Answers to these four items were collected on a 7-point scale ranging from “not at all” to “definitively”. Finally, unvaccinated participants were asked if they would get vaccinated if the presented regulation came into force. Vaccination intentions were recorded on a 7-point scale ranging from “not getting vaccinated at all” to “definitively getting vaccinated”.

### 2.6.2 Second wave (February 2022)

In the second wave, the measures from the first wave were assessed again, except for specific media usage and the experiment on vaccination mandates. Furthermore, ostracism experiences and activism behaviors were surveyed.

**Ostracism experiences.** Using the Ostracism Short Scale (Rudert et al., 2020), participants were asked how often they had experienced different forms of ostracism during the last two months (Cronbach’s  $\alpha = 0.93$ ; e.g., “Others exclude me from

conversations”). Answers were recorded on a 7-point scale ranging from “never” to “always”.

**Activism behavior.** Participants were asked if they had participated in a demonstration and if they had signed a petition against vaccination mandates since the beginning of 2022. The answers to both items were recorded on a binary scale (“yes” or “no”).

### 2.6.3 Third wave (July 2022)

In the third wave, the measures from the first wave were assessed again, except for the experiment on vaccination mandates.

### 2.6.4 Additional wave with subsample (March 2022)

In this additional wave, a subsample of vaccinated and unvaccinated participants from the second wave was surveyed to validate the VSI measure (assessed as in the first and second waves) against the following identification scales.

**Social identification scales.** Participants were asked to indicate their agreement with 15 items of a social identification scale adapted from Leach and al. (2008); example items: “(Not) being vaccinated against COVID-19 is an important part of how I see myself”; “I am a typical person who is (not) vaccinated against COVID-19”. Answers were recorded on a 7-point scale ranging from “very strongly disagree” to “very strongly agree” (Cronbach’s  $\alpha = .95$ ). Participants were also asked to indicate their agreement with four items adapted from the social identification scale developed by Doosje, Ellemers, and Spears (1995); example items: “I consider myself a member of the group that has (not) been vaccinated against COVID-19”; “I identify with the group that has (not) been vaccinated against COVID-19”. Answers were recorded on a 7-point scale ranging from “very strongly disagree” to “very strongly agree” (Cronbach’s  $\alpha = .96$ ).

**Perceived intergroup similarity.** Participants were asked how similar they perceived the groups of vaccinated and unvaccinated people by selecting one of five figures showing two circles (representing the two groups) that had an overlap of 0–100%. Answers were coded from 1 to 5, with higher values indicating more intergroup similarity.

**Perceived intragroup similarity.** Using two items, participants were asked “How similar to each other are individuals that are (not) vaccinated against COVID-19”? Answers were recorded on a 7-point scale ranging from “not at all” to “very much”.

## Appendix 2.A Supplement

**Table 2.A.1.** Correlations between vaccination status identification (VSI), vaccination intention and the 7C antecedents of vaccination part I

<b>December 2021, vaccinated participants</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Confidence	1.00	-0.51	-0.59	-0.13	0.68	0.48	-0.47	0.59	0.53
(2) Complacency	-0.51	1.00	0.61	0.14	-0.61	-0.44	0.32	-0.56	-0.52
(3) Constraints	-0.59	0.61	1.00	0.17	-0.64	-0.53	0.33	-0.57	-0.64
(4) Calculation	-0.13	0.14	0.17	1.00	-0.18	-0.15	0.22	-0.20	-0.16
(5) Collective responsibility	0.68	-0.61	-0.64	-0.18	1.00	0.57	-0.49	0.66	0.58
(6) Compliance	0.48	-0.44	-0.53	-0.15	0.57	1.00	-0.27	0.44	0.56
(7) Conspiracy thinking	-0.47	0.32	0.33	0.22	-0.49	-0.27	1.00	-0.51	-0.30
(8) Booster intention	0.59	-0.56	-0.57	-0.20	0.66	0.44	-0.51	1.00	0.50
(9) VSI	0.53	-0.52	-0.64	-0.16	0.58	0.56	-0.30	0.50	1.00
<b>December 2021, unvaccinated participants</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Confidence	1.00	-0.47	-0.56	0.13	0.60	0.46	-0.39	0.51	-0.28
(2) Complacency	-0.47	1.00	0.61	-0.15	-0.57	-0.47	0.23	-0.58	0.24
(3) Constraints	-0.56	0.61	1.00	-0.06	-0.64	-0.66	0.23	-0.63	0.21
(4) Calculation	0.13	-0.15	-0.06	1.00	0.14	0.04	-0.04	0.13	-0.12
(5) Collective responsibility	0.60	-0.57	-0.64	0.14	1.00	0.52	-0.31	0.59	-0.26
(6) Compliance	0.46	-0.47	-0.66	0.04	0.52	1.00	-0.14	0.45	-0.14
(7) Conspiracy thinking	-0.39	0.23	0.23	-0.04	-0.31	-0.14	1.00	-0.31	0.39
(8) Vaccination intention	0.51	-0.58	-0.63	0.13	0.59	0.45	-0.31	1.00	-0.26
(9) VSI	-0.28	0.24	0.21	-0.12	-0.26	-0.14	0.39	-0.26	1.00
<b>February 2022, vaccinated participants</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Confidence	1.00	-0.55	-0.61	-0.12	0.71	0.51	-0.51	0.63	0.54
(2) Complacency	-0.55	1.00	0.68	0.14	-0.64	-0.46	0.39	-0.60	-0.55
(3) Constraints	-0.61	0.68	1.00	0.19	-0.69	-0.56	0.38	-0.61	-0.63
(4) Calculation	-0.12	0.14	0.19	1.00	-0.16	-0.16	0.20	-0.18	-0.15
(5) Collective responsibility	0.71	-0.64	-0.69	-0.16	1.00	0.57	-0.51	0.65	0.61
(6) Compliance	0.51	-0.46	-0.56	-0.16	0.57	1.00	-0.24	0.48	0.57
(7) Conspiracy thinking	-0.51	0.39	0.38	0.20	-0.51	-0.24	1.00	-0.49	-0.34
(8) Booster intention	0.63	-0.60	-0.61	-0.18	0.65	0.48	-0.49	1.00	0.52
(9) VSI	0.54	-0.55	-0.63	-0.15	0.61	0.57	-0.34	0.52	1.00



**Table 2.A.2.** Correlations between vaccination status identification (VSI), vaccination intention and the 7C antecedents of vaccination part II

<b>February 2022, unvaccinated participants</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Confidence	1.00	-0.43	-0.45	0.14	0.55	0.34	-0.40	0.42	-0.30
(2) Complacency	-0.43	1.00	0.64	-0.17	-0.58	-0.46	0.22	-0.52	0.19
(3) Constraints	-0.45	0.64	1.00	-0.09	-0.62	-0.67	0.21	-0.52	0.18
(4) Calculation	0.14	-0.17	-0.09	1.00	0.16	0.03	-0.07	0.13	-0.11
(5) Collective responsibility	0.55	-0.58	-0.62	0.16	1.00	0.47	-0.31	0.53	-0.27
(6) Compliance	0.34	-0.46	-0.67	0.03	0.47	1.00	-0.13	0.35	-0.13
(7) Conspiracy thinking	-0.40	0.22	0.21	-0.07	-0.31	-0.13	1.00	-0.28	0.43
(8) Vaccination intention	0.42	-0.52	-0.52	0.13	0.53	0.35	-0.28	1.00	-0.24
(9) VSI	-0.30	0.19	0.18	-0.11	-0.27	-0.13	0.43	-0.24	1.00
<b>July 2022, vaccinated participants</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Confidence	1.00	-0.59	-0.64	-0.16	0.72	0.47	-0.48	0.66	0.56
(2) Complacency	-0.59	1.00	0.68	0.17	-0.69	-0.47	0.36	-0.66	-0.58
(3) Constraints	-0.64	0.68	1.00	0.23	-0.71	-0.57	0.35	-0.67	-0.65
(4) Calculation	-0.16	0.17	0.23	1.00	-0.21	-0.13	0.30	-0.24	-0.16
(5) Collective responsibility	0.72	-0.69	-0.71	-0.21	1.00	0.55	-0.49	0.73	0.63
(6) Compliance	0.47	-0.47	-0.57	-0.13	0.55	1.00	-0.21	0.48	0.59
(7) Conspiracy thinking	-0.48	0.36	0.35	0.30	-0.49	-0.21	1.00	-0.50	-0.31
(8) Booster intention	0.66	-0.66	-0.67	-0.24	0.73	0.48	-0.50	1.00	0.57
(9) VSI	0.56	-0.58	-0.65	-0.16	0.63	0.59	-0.31	0.57	1.00
<b>July 2022, unvaccinated participants</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Confidence	1.00	-0.49	-0.52	0.19	0.58	0.43	-0.37	0.44	-0.31
(2) Complacency	-0.49	1.00	0.65	-0.19	-0.58	-0.53	0.17	-0.49	0.20
(3) Constraints	-0.52	0.65	1.00	-0.11	-0.61	-0.73	0.17	-0.50	0.17
(4) Calculation	0.19	-0.19	-0.11	1.00	0.20	0.06	-0.07	0.19	-0.15
(5) Collective responsibility	0.58	-0.58	-0.61	0.20	1.00	0.53	-0.23	0.55	-0.23
(6) Compliance	0.43	-0.53	-0.73	0.06	0.53	1.00	-0.13	0.38	-0.10
(7) Conspiracy thinking	-0.37	0.17	0.17	-0.07	-0.23	-0.13	1.00	-0.19	0.45
(8) Vaccination intention	0.44	-0.49	-0.50	0.19	0.55	0.38	-0.19	1.00	-0.23
(9) VSI	-0.31	0.20	0.17	-0.15	-0.23	-0.10	0.45	-0.23	1.00

**Table 2.A.3.** Predictors of VSI.

	Dependent variable:			
	Vaccinated (1)	Vaccinated (2)	Non-Vaccinated (3)	Non-Vaccinated (4)
Descriptive norm	0.085*** (0.035, 0.136)	0.040* (-0.001, 0.082)	-0.082*** (-0.126, -0.038)	-0.065*** (-0.109, -0.022)
Injunctive norm	0.247*** (0.198, 0.296)	0.065*** (0.024, 0.107)	-0.009 (-0.052, 0.034)	0.034 (-0.009, 0.077)
Trust in Government	0.166*** (0.141, 0.191)	0.058*** (0.035, 0.082)	-0.145*** (-0.198, -0.092)	-0.036 (-0.092, 0.020)
Frequency news	0.137*** (0.107, 0.168)	0.050*** (0.023, 0.078)	0.003 (-0.032, 0.038)	0.001 (-0.033, 0.034)
News classic sources	0.098 (-0.078, 0.273)	-0.009 (-0.161, 0.144)	-0.458*** (-0.671, -0.245)	-0.260** (-0.466, -0.053)
News alternative sources	0.221*** (0.007, 0.435)	0.342*** (0.159, 0.524)	0.617*** (0.398, 0.836)	0.328*** (0.116, 0.540)
Age	0.011*** (0.008, 0.013)	0.006*** (0.004, 0.008)	0.000 (-0.004, 0.004)	-0.002 (-0.006, 0.002)
Male	-0.088** (-0.164, -0.012)	-0.017 (-0.082, 0.048)	0.082 (-0.030, 0.195)	-0.002 (-0.110, 0.105)
Diverse	-0.061 (-0.668, 0.546)	0.209 (-0.578, 0.997)	-0.140 (-0.706, 0.427)	-0.070 (-0.679, 0.539)
Edu 10+ years w/o uni qual	-0.049 (-0.172, 0.074)	-0.013 (-0.120, 0.093)	-0.066 (-0.231, 0.098)	-0.072 (-0.231, 0.088)
Edu 10+ years with uni qual	0.012 (-0.110, 0.134)	0.031 (-0.073, 0.135)	-0.208** (-0.377, -0.040)	-0.215*** (-0.377, -0.052)
No Migr. background	0.003 (-0.099, 0.106)	0.009 (-0.080, 0.097)	0.047 (-0.097, 0.191)	0.069 (-0.068, 0.205)
Migr. not known	0.038 (-0.349, 0.424)	-0.179 (-0.502, 0.144)	-0.313 (-0.973, 0.346)	-0.267 (-0.851, 0.317)
Income middle	0.021 (-0.086, 0.128)	0.034 (-0.060, 0.127)	-0.141** (-0.280, -0.003)	-0.137** (-0.268, -0.006)
Income high	-0.063 (-0.174, 0.047)	-0.046 (-0.142, 0.049)	-0.061 (-0.210, 0.087)	-0.075 (-0.215, 0.066)
Income no-response	0.016 (-0.128, 0.160)	-0.011 (-0.131, 0.109)	-0.104 (-0.275, 0.068)	-0.108 (-0.271, 0.054)
Non-religious	-0.001 (-0.077, 0.074)	-0.009 (-0.074, 0.056)	-0.187*** (-0.294, -0.080)	-0.177*** (-0.280, -0.075)
Unemployed	0.203*** (0.106, 0.301)	0.178*** (0.096, 0.260)	-0.072 (-0.199, 0.055)	-0.013 (-0.135, 0.109)
Healthjob	-0.074 (-0.208, 0.061)	-0.097* (-0.211, 0.017)	0.059 (-0.136, 0.253)	0.110 (-0.078, 0.298)
System-relevant job	-0.010 (-0.103, 0.083)	-0.036 (-0.117, 0.045)	0.024 (-0.101, 0.148)	-0.007 (-0.125, 0.110)
Left-voter	-0.104* (-0.210, 0.001)	-0.078* (-0.169, 0.014)	0.045 (-0.202, 0.291)	0.079 (-0.154, 0.312)
Nonvoter	-0.331*** (-0.433, -0.230)	-0.271*** (-0.359, -0.183)	-0.025 (-0.208, 0.158)	-0.070 (-0.242, 0.101)
Other-voter	-0.293** (-0.520, -0.066)	-0.073 (-0.245, 0.099)	0.050 (-0.222, 0.322)	0.040 (-0.218, 0.297)
Right-voter	-0.296*** (-0.415, -0.177)	-0.162*** (-0.266, -0.058)	0.360*** (0.169, 0.552)	0.239*** (0.058, 0.420)
7C confidence		0.062*** (0.030, 0.094)		-0.019 (-0.065, 0.028)
7C complacency		-0.039*** (-0.065, -0.013)		0.061** (0.014, 0.108)
7C constraints		-0.175*** (-0.201, -0.149)		0.000 (-0.079, 0.080)
7C calculation		-0.009 (-0.025, 0.007)		-0.030** (-0.055, -0.006)
7C coll. responsibility		0.062*** (0.027, 0.097)		-0.023 (-0.070, 0.024)
7C compliance		0.138*** (0.117, 0.159)		0.004 (-0.071, 0.080)
7C consp. thinking		0.009 (-0.017, 0.036)		0.199*** (0.159, 0.239)
Vaccination Intention				-0.040 (-0.091, 0.011)
Constant	1.054*** (0.778, 1.329)	2.767*** (2.348, 3.186)	4.976*** (4.526, 5.427)	3.652*** (2.866, 4.438)
Observations	3,267	3,267	2,038	2,038
R <sup>2</sup>	0.381	0.544	0.143	0.233

Note: Data from December 2021. 95% confidence intervals in parentheses. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 2.A.4.** Effect of VSI on perception of discriminatory public discourse

	<i>Dependent variable: discriminatory public discourse</i>		
	(1)	(2)	(3)
Identification	-0.170*** (-0.215, -0.124)	-0.145*** (-0.190, -0.100)	-0.066*** (-0.104, -0.028)
Is vaccinated	-1.098*** (-1.376, -0.821)	-0.990*** (-1.264, -0.715)	-0.537*** (-0.773, -0.301)
Is vaccinated * Identification	0.713*** (0.654, 0.772)	0.653*** (0.593, 0.713)	0.223*** (0.158, 0.288)
Constant	-1.434*** (-1.646, -1.221)	-0.765*** (-1.082, -0.448)	-1.388*** (-1.829, -0.946)
Sociodemographics		X	X
7C			X
Observations	5,305	5,305	5,305
R <sup>2</sup>	0.458	0.477	0.556

Notes: Data from December 2021. 95% confidence intervals in parentheses. Sociodemographic variables include age, gender, education, migration background, household income, religion, political preference, employment status, and if participants work in healthcare or a system-relevant job. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

**Table 2.A.5.** Effect of VSI on everyday discrimination perception

	<i>Dependent variable: discrimination perception</i>		
	(1)	(2)	(3)
Identification	0.361*** (0.300, 0.422)	0.364*** (0.304, 0.425)	0.264*** (0.202, 0.326)
Is vaccinated	0.736*** (0.436, 1.037)	0.327** (0.025, 0.629)	0.244 (-0.056, 0.544)
Is vaccinated * Identification	-0.421*** (-0.489, -0.353)	-0.322*** (-0.392, -0.253)	-0.178** (-0.254, -0.102)
Constant	1.414*** (1.152, 1.676)	3.036*** (2.678, 3.394)	3.068*** (2.605, 3.532)
Sociodemographics		X	X
7C			X
Observations	5,305	5,305	5,305
R <sup>2</sup>	0.158	0.222	0.283

Notes: Data from December 2021. 95% confidence intervals in parentheses. Sociodemographic variables include age, gender, education, migration background, household income, religion, political preference, employment status, and if participants work in healthcare or a system-relevant job. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

**Table 2.A.6.** Effect of VSI on ostracism perception

	<i>Dependent variable: ostracism perception</i>		
	(1)	(2)	(3)
Identification	0.207*** (0.150, 0.264)	0.223*** (0.165, 0.281)	0.120*** (0.061, 0.179)
Is vaccinated	0.730*** (0.438, 1.022)	0.531*** (0.236, 0.826)	0.428*** (0.133, 0.723)
Is vaccinated * Identification	-0.210*** (-0.276, -0.143)	-0.172*** (-0.240, -0.103)	-0.067* (-0.145, 0.012)
Constant	1.115*** (0.866, 1.364)	2.655*** (2.292, 3.018)	2.811*** (2.331, 3.291)
Sociodemographics		X	X
7C			X
Observations	4,406	4,406	4,406
R <sup>2</sup>	0.018	0.084	0.159

Notes: Data from February 2022. 95% confidence intervals in parentheses. Sociodemographic variables include age, gender, education, migration background, household income, religion, political preference, employment status, and if participants work in healthcare or a system-relevant job. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

**Table 2.A.7.** Effect of VSI on ingroup preference [EUR]

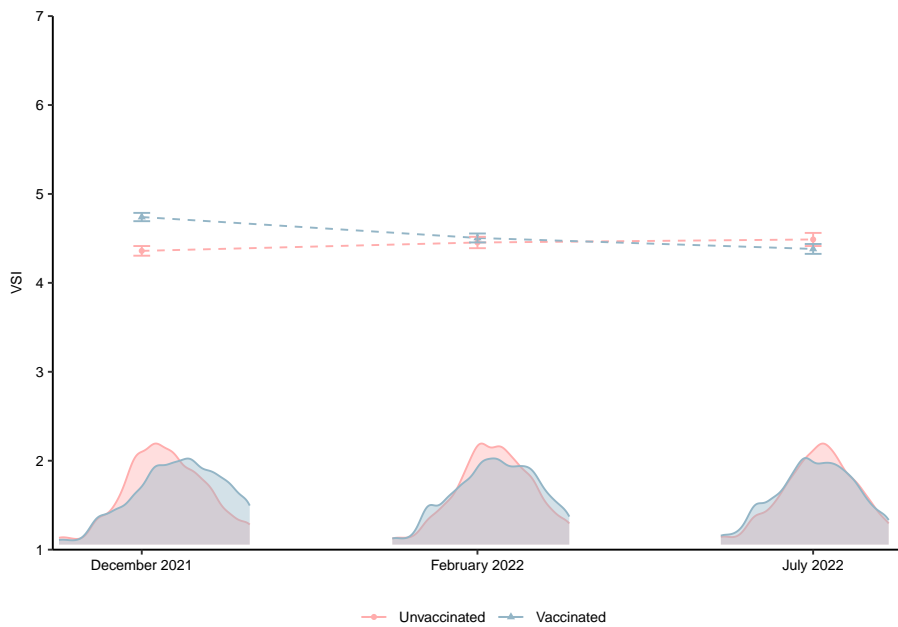
	<i>Dependent variable: ingroup preference</i>		
	(1)	(2)	(3)
Identification	5.530*** (4.540, 6.521)	5.352*** (4.347, 6.357)	5.600*** (4.549, 6.651)
Is vaccinated	-4.987* (-10.190, 0.215)	-3.943 (-9.308, 1.422)	-1.653 (-7.151, 3.845)
Is vaccinated * Identification	2.937*** (1.725, 4.149)	2.723*** (1.450, 3.997)	1.082 (-0.397, 2.561)
Constant	-16.744*** (-20.848, -12.640)	-19.961*** (-26.167, -13.755)	-14.607*** (-23.829, -5.385)
Sociodemographics		X	X
7C			X
Observations	5,305	5,305	5,305
R <sup>2</sup>	0.162	0.171	0.182

*Notes:* Data from December 2021. 95% confidence intervals in parentheses. Sociodemographic variables include age, gender, education, migration background, household income, religion, political preference, employment status, and if participants work in healthcare or a system-relevant job. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

**Table 2.A.8.** Effect of VSI on reactance toward vaccination mandate

	<i>Dependent variable: ingroup preference</i>			
	(1)	(2)	(3)	(4)
VSI	0.256*** (0.207, 0.323)	0.250** (0.188, 0.311)	0.248*** (0.187, 0.309)	0.071** (0.015, 0.127)
Status: vaccinated (Baseline: unvaccinated)	1.238*** (0.853, 1.623)	1.080*** (0.682, 1.478)	1.066*** (0.667, 1.464)	0.174 (-0.169, 0.518)
Interaction VSI x status	-1.036*** (-1.113, -0.960)	-0.983*** (-1.065, -0.902)	-0.981*** (-1.062, -0.899)	-0.233*** (-0.320, -0.145)
Constant	5.375*** (5.091, 5.659)	5.773*** (5.332, 6.215)	5.666*** (5.217, 6.114)	6.037*** (5.432, 6.642)
Sociodemographics		X	X	X
7C			X	X
Experimental manipulations				X
Observations	5,305	5,305	5,305	5,305
R <sup>2</sup>	0.561	0.568	0.569	0.679

Notes: Data from December 2021. 95% confidence intervals in parentheses. Sociodemographic variables include age, gender, education, migration background, household income, religion, political preference, employment status, and if participants work in healthcare or a system-relevant job. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.



**Figure 2.A.1.** Stability of VSI over time

**Table 2.A.9.** Effect of VSI on likelihood of having joined demonstration against mandate

	<i>Dependent variable: joined demonstration against mandate</i>		
	(1)	(2)	(3)
Identification	0.047*** (0.031, 0.063)	0.042*** (0.026, 0.058)	0.027*** (0.010, 0.043)
Is vaccinated	0.044 (-0.033, 0.120)	0.024 (-0.052, 0.100)	-0.002 (-0.078, 0.074)
Is vaccinated * Identification	-0.055*** (-0.072, -0.038)	-0.047*** (-0.065, -0.030)	-0.018* (-0.036, 0.001)
Constant	0.023 (-0.049, 0.095)	0.115** (0.021, 0.208)	0.160*** (0.044, 0.276)
Sociodemographics		X	X
7C			X
Observations	4,406	4,406	4,406
R <sup>2</sup>	0.119	0.146	0.170

Notes: Data from February 2022. 95% confidence intervals in parentheses. Sociodemographic variables include age, gender, education, migration background, household income, religion, political preference, employment status, and if participants work in healthcare or a system-relevant job. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.



**Table 2.A.10.** Effect of VSI on likelihood of having signed petition against mandate

	<i>Dependent variable: signed petition against mandate</i>		
	(1)	(2)	(3)
Identification	0.090*** (0.071, 0.109)	0.076*** (0.058, 0.095)	0.050*** (0.031, 0.069)
Is vaccinated	0.157*** (0.060, 0.255)	0.114** (0.019, 0.209)	0.054 (-0.039, 0.147)
Is vaccinated * Identification	-0.132*** (-0.153, -0.111)	-0.111*** (-0.132, -0.091)	-0.045*** (-0.067, -0.023)
Constant	0.103** (0.016, 0.190)	0.140** (0.031, 0.249)	0.132* (-0.006, 0.271)
Sociodemographics		X	X
7 C			X
Observations	4,406	4,406	4,406
R <sup>2</sup>	0.284	0.317	0.356

Notes: Data from February 2022. 95% confidence intervals in parentheses. Sociodemographic variables include age, gender, education, migration background, household income, religion, political preference, employment status, and if participants work in healthcare or a system-relevant job. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

**Table 2.A.11.** Sociodemographics

	<b>Unvaccinated</b>	<b>Vaccinated</b>
Age	45.9 (14.7)	46.8 (17.9)
<b>Gender</b>		
Male	829 (40.7%)	1639 (50.2%)
Female	1198 (58.8%)	1619 (49.6%)
Other	11 (0.5%)	9 (0.3%)
<b>Education</b>		
Up to 9 years	304 (14.9%)	456 (14.0%)
10+ years without university entrance qualification	776 (38.1%)	1050 (32.1%)
10+ years with university entrance qualification	958 (47.0%)	1761 (53.9%)
<b>Employment</b>		
Yes	1305 (64.0%)	1944 (59.5%)
No	733 (36.0%)	1323 (40.5%)
Observations	2,038	3,267

## References

- Attwell, Katie, and Daniel T. Smith.** 2017. "Parenting as Politics: Social Identity Theory and Vaccine Hesitant Communities." *International Journal of Health Governance* 22: 183–98. [122]
- Bliuc, Ana-Maria, Craig McGarty, Katherine Reynolds, and Daniel Muntele.** 2007. "Opinion-Based Group Membership as a Predictor of Commitment to Political Action." *European Journal of Social Psychology* 37: 19–32. [122]
- Böhm, Robert, Hannes Rusch, and Jonathan Baron.** 2020. "The Psychology of Intergroup Conflict: A Review of Theories and Measures." *Journal of Economic Behavior & Organization* 178: 947–62. [131]
- Bor, Alexander, Frederik Jørgensen, and Michael Bang Petersen.** 2023. "Discriminatory Attitudes against Unvaccinated People during the Pandemic." *Nature* 613 (7945): 704–11. [125, 129]
- Branscombe, Nyla R., Michael T. Schmitt, and Richard D. Harvey.** 1999. "Perceiving Pervasive Discrimination among African Americans: Implications for Group Identification and Well-Being." *Journal of Personality and Social Psychology* 77: 135–49. [130]
- Brewer, Marilynn B.** 2001. "Ingroup Identification and Intergroup Conflict: When Does Ingroup Love Become Outgroup Hate?" In *Social Identity, Intergroup Conflict, and Conflict Reduction*, edited by Richard D. Ashmore, Lee Jussim, and David Wilder, 17–41. Oxford University Press. [122]
- Bundesanstalt Statistik Österreich.** 2022a. *Tabelle: Bevölkerung Zu Jahresbeginn Nach Bundesland, Alter, Geschlecht Sowie Österreichischer/Ausländischer Staatsangehörigkeit 2002 Bis 2022.* [122]
- Bundesanstalt Statistik Österreich.** 2022b. *Tabelle: Bildungsstand Der Bevölkerung 2020 Ab 15 Jahren.* [122]
- Bundesanstalt Statistik Österreich.** 2022c. *Tabelle: Erwerbstätigkeit.* [122]
- Doosje, Bertjan, Naomi Ellemers, and Russell Spears.** 1995. "Perceived Intragroup Variability as a Function of Group Status and Identification." *Journal of Experimental Social Psychology* 31: 410–36. [123, 133, 135]
- DW.** 2022. *Thousands Protest COVID Curbs in Europe amid Omicron Surge.* [121]
- Garcia, Dan, Victor Galaz, and Stephanie Daume.** 2019. "EATLancet vs Yes2meat: The Digital Backlash to the Planetary Health Diet." *Lancet* 394: 2153–54. [122]
- Geiger, Marc, and et al.** 2021. "Measuring the 7Cs of Vaccination Readiness." *European Journal of Psychological Assessment*, 1–9. [123, 132]
- Henkel, Luca, Philipp Sprengholz, Lars Korn, Cornelia Betsch, and Robert Böhm.** 2023. "The Association between Vaccination Status Identification and Societal Polarization." *Nature Human Behaviour* 7 (2): 231–39. [121]
- Iyengar, Shanto, and Sean J. Westwood.** 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59: 690–707. [122]
- Korn, Lars, Robert Böhm, Nils W. Meier, and Cornelia Betsch.** 2020. "Vaccination as a Social Contract." *Proceedings of the National Academy of Sciences* 117: 14890–99. [125, 129]
- Leach, Colin W., and et al.** 2008. "Group-Level Self-Definition and Self-Investment: A Hierarchical (Multicomponent) Model of in-Group Identification." *Journal of Personality and Social Psychology* 95: 144–65. [123, 133–135]
- Lee, Charlotte, and James L. Robinson.** 2016. "Systematic Review of the Effect of Immunization Mandates on Uptake of Routine Childhood Immunizations." *Journal of Infection* 72: 659–66. [131]

- Mallapaty, Smriti.** 2022. "Researchers Fear Growing COVID Vaccine Hesitancy in Developing Nations." *Nature* 601: 174–75. [121]
- McGarty, Craig, Ana-Maria Bliuc, Emma F. Thomas, and Renata Bongiorno.** 2009. "Collective Action as the Material Expression of Opinion-Based Group Membership." *Journal of Social Issues* 65: 839–57. [122]
- Morrison, Kimberly R., Nathanael J. Fast, and Oscar Ybarra.** 2009. "Group Status, Perceptions of Threat, and Support for Social Inequality." *Journal of Experimental Social Psychology* 45: 204–10. [122]
- Motta, Matt, Timothy Callaghan, Steven Sylvester, and Kendra Lunz-Trujillo.** 2021. "Identifying the Prevalence, Correlates, and Policy Consequences of Anti-Vaccine Social Identity." *Politics, Groups, and Identities*, 1–15. [122]
- Ramos, Miguel R., Clare Cassidy, Stephen Reicher, and S. Alexander Haslam.** 2012. "A Longitudinal Investigation of the Rejection-Identification Hypothesis." *British Journal of Social Psychology* 51: 642–60. [130]
- Roth, Jenny, and Agnes Mazziotta.** 2015. "Adaptation and Validation of a German Multidimensional and Multicomponent Measure of Social Identification." *Social Psychology (Gott)* 46: 277–90. [123, 133]
- Rudert, Selma C., and et al.** 2021. "Us and the Virus." *European Psychologist* 26: 259–71. [130]
- Rudert, Selma C., Matthias D. Keller, Andy H. Hales, Mirella Walker, and Rainer Greifeneder.** 2020. "Who Gets Ostracized? A Personality Perspective on Risk and Protective Factors of Ostracism." *Journal of Personality and Social Psychology* 118: 1247–68. [125, 134]
- Schmelz, Konstantin, and Samuel Bowles.** 2022. "Opposition to Voluntary and Mandated COVID-19 Vaccination as a Dynamic Process: Evidence and Policy Implications of Changing Beliefs." *Proceedings of the National Academy of Sciences* 119. [127]
- Schmidt, Ana Lucia, Fabiana Zollo, Antonio Scala, Cornelia Betsch, and Walter Quattrociochi.** 2018. "Polarization of the Vaccination Debate on Facebook." *Vaccine* 36: 3606–12. [129]
- Sittenthaler, Sandra, Eva Traut-Mattausch, Christina Steindl, and Eva Jonas.** 2015. "Salzburger State Reactance Scale (SSR Scale)." *Zeitschrift für Psychologie* 223: 257–66. [134]
- Sobo, Elisa J.** 2015. "Social Cultivation of Vaccine Refusal and Delay among Waldorf (Steiner) School Parents." *Medical Anthropology Quarterly* 29: 381–99. [130]
- Sprengholz, Philipp, and et al.** 2022. "Attitude toward a Mandatory COVID-19 Vaccination Policy and Its Determinants: Evidence from Serial Cross-Sectional Surveys Conducted throughout the Pandemic in Germany." *Vaccine*. [127]
- Sprengholz, Philipp, Cornelia Betsch, and Robert Böhm.** 2021. "Reactance Revisited: Consequences of Mandatory and Scarce Vaccination in the Case of COVID-19." *Applied Psychology: Health and Well-Being*. [127]
- Statistisches Bundesamt.** 2022a. *Tabelle 12211-0001: Bevölkerung, Erwerbstätige, Erwerbslose, Erwerbspersonen, Nichterwerbspersonen Aus Hauptwohnsitzhaushalten: Deutschland, Jahre, Geschlecht, Altersgruppen (2022)*. [122]
- Statistisches Bundesamt.** 2022b. *Tabelle 12211-0100: Bevölkerung Ab 15 Jahren in Hauptwohnsitzhaushalten: Deutschland, Jahre, Geschlecht, Altersgruppen, Allgemeine Achulausbildung (2022)*. [122]
- Statistisches Bundesamt.** 2022c. *Tabelle 12411-0013: Bevölkerung: Bundesländer, Stichtag, Geschlecht, Altersjahre (2022)*. [122]
- Sternthal, Michelle J., Natalie Slopen, and David R. Williams.** 2011. "Racial Disparities in Health." *Du Bois Review: Social Science Research on Race* 8: 95–113. [125, 133]

- Tajfel, Henri, and John C. Turner.** 1979. "An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, edited by Stephen Worchel and William G. Austin, 33–47. Brooks/Cole. [122]
- Turner, John C., and Katherine J. Reynolds.** 2008a. "Self-Categorization Theory." In *Handbook of Theories of Social Psychology*, 399–417. SAGE Publications Ltd. [122]
- Turner, John C., and Katherine J. Reynolds.** 2008b. "The Social Identity Perspective in Intergroup Relations: Theories, Themes and Controversies." In *Blackwell Handbook of Social Psychology: Intergroup Processes*, edited by Rupert Brown and Samuel L. Gaertner. Wiley-Blackwell. [122]
- Wakefield, John R. H., and Ayesha Khauser.** 2021. "Doing It for Us: Community Identification Predicts Willingness to Receive a COVID-19 Vaccination via Perceived Sense of Duty to the Community." *Journal of Community & Applied Social Psychology* 31: 603–14. [122]
- Weisel, Ori.** 2021. "Vaccination as a Social Contract: The Case of COVID-19 and US Political Partisanship." *Proceedings of the National Academy of Sciences* 118. [125, 129]



## Chapter 3

# Eliciting Moral Preferences under Image Concerns: Theory and Experiment\*

*Joint with Roland Bénabou, Armin Falk and Jean Tirole*

### 3.1 Introduction

Individuals' desire to signal to others and maintain to themselves that they are generous, caring, or generally "morally good," is a powerful driver of behavior. People act more responsibly when knowing their choices will be observed and less so when given the opportunity to remain ignorant of potential harms they might cause.

The previous literature on image motives (see, e.g. Bursztyn and Jensen (2017) for an overview) has extensively documented this *level* effect on the prosociality of choices. We explore here a new channel, namely the *interaction* of image with different choice mechanisms. We focus on two key features of the latter: single versus multiple simultaneous decisions, and certainty versus uncertainty of the consequences. Both vary across charitable-contribution schemes, and they critically distinguish the two methods most commonly used to elicit preferences: direct elicitation (*DE*) and

\* A modified version of this chapter has appeared as CRC TR 224 Discussion Paper No. 441.

**Acknowledgments:** We are grateful to Ingela Alger, Jean-François Bonnefon, Gary Charness, Franz Ostrizek, Pëllumb Reshidi, Marie-Claire Villeval and Joël van der Weele for valuable comments. Ana Luisa Dutra, Juliette Fournier, Pierre-Luc Vautrey, Ben S. Young, Youpeng Zhang and Egshiglen Batbayar provided superb research assistance. Bénabou gratefully acknowledges financial support from the Canadian Institute for Advanced Study, Tirole and Falk from the European Research Council (European Community's Seventh Framework Program Grant Agreement no. 249429 and no. 340950, the European Union's Horizon 2020 research and innovation program, Grant Agreement no. 669217) as well as the German Research Foundation (DFG) through CRC TR 224 (Project A01).

**Ethics approval:** The study was approved by the ethical committees of the University of Bonn (no. 2019-01), Toulouse School of Economics and Princeton University (no. 11818).

Becker-DeGroot-Marschak (*BDM*), for instance in its multiple-price list (*MPL*) format.<sup>1</sup> The former features a single choice implemented with certainty, the latter multiple decisions (at different prices), of which one is randomly chosen and implemented.

Taking *DE* and *MPL* (or *BDM*) as exemplars of choice sets' interactions with signaling, we present a simple model and experiment in which agents incur a cost to do good, or forfeit a "bribe" for causing harm. The model identifies three effects that make the mechanisms *differentially* image sensitive and, when combined, generate a specific "crossing" pattern: when image concerns are low (but positive) *DE* will yield more contributions than *MPL*, and when they are high the ordering reverses. A Kantian-resembling unwillingness to compromise on moral values "at any price" is then also more likely to be displayed under *MPL*.

To understand the effects at work, consider first a (*DE*-type) situation in which individuals may contribute to a cause (generate an externality  $e > 0$ ) at some opportunity cost  $c$ , in time or money. In the relevant population there are two types, represented by Alex and Bob, who intrinsically value the cause at  $v_H e$  and  $v_L e < v_H e$ . When social or self image concerns are present but not very strong, there is a range of prices  $c > v_L e$  for which Bob will contribute in order to look as good as Alex, whereas for  $c'$  closer to  $v_H e$  he will decline. Suppose now that the relevant audience will learn of decisions made under both circumstances (e.g., time costs, tax deductions). The fact that Alex contributes at  $c'$  now makes it futile for Bob to contribute even at  $c$ : the aggregation of information hinders pooling, reducing total contributions. In an *MPL/BDM* format, similarly, the rich choice set makes pooling more difficult, as Bob would have to state a willingness to pay of at least  $v_H e$ ; this is too high for him, so he will decline to contribute at *any* list price  $c > v_L e$ . This *discouragement effect* underlies the result that *MPL/BDM* yields less giving than *DE* when image concerns are positive but relatively weak.

Working in the other direction are two effects arising from the contingent nature of *MPL/BDM* bids, which effectively lower the purchase price of image. First, the randomly drawn list price could exceed one's bid, making the latter partly *cheap talk*. This is related to random implementation, but more closely to the ability of participants in a public auction to "posture" with a high bid, while hoping that someone else will outbid them. Second is what we term the *cheap-act* effect: conditional on a bid  $c$  being binding ex-post, the average price paid is only  $E[\tilde{c} | \tilde{c} \leq c]$ . As image concerns intensify, Bob's desire to pool and Alex's desire to separate lead to increasingly high bids, so the cheap-talk effect weakens (implementation becomes more certain). In contrast, the cheap-act effect strengthens (for standard distributions the "discount"  $c - E[c | \tilde{c} \leq c]$  grows), causing *MPL* contributions to rise above those under *DE*.

1. As explained below, *MPL* and *BDM* are formally equivalent.



We test the model's predictions using an experiment in which about 700 participants face a choice between: (i) directing a 350€ donation to a charity in India that will use the money to treat five tuberculosis patients, resulting statistically in the expected saving of one human life; or (ii) taking money for themselves, where the amount is either a fixed 100€ under *DE*, or determined by the subjects' cutoff on an *MPL* where prices range from 0 to 200€. These two elicitation conditions are crossed with low and high moral-image treatments. Comparing the fractions of subjects choosing the "saving a life" contribution over taking 100€, we find a sizeable reversal between *DE* and *MPL* as image concerns go from weak to strong, as predicted by the theory. In the *Low Image* treatments, the fraction opting to save a life is 48% under *MPL* versus 59% under *DE*, while in the *High Image* condition it is 63% under *DE* versus 72% under *MPL*.<sup>2</sup> On the cautionary side, statistical significance is only at the 6-7 percent level, so our simple experiment should be seen as proof-of-concept for the mechanisms brought to light by the model, opening them up to more systematic exploration.

### 3.1.1 Related literature

Previous research on social and self image has primarily focused on how they spur prosocial behaviors, and how this signaling incentive is affected by the presence of rewards (Bénabou and Tirole, 2006; Ariely, Bracha, and Meier, 2009; Bénabou and Tirole, 2011a, 2011b; Ashraf, Bandiera, and Jack, 2014; Grossman and van der Weele, 2017) or excuses (Dana, Weber, and Kuang, 2007; DellaVigna, List, and Malmendier, 2012; Exley, 2016). Our analysis highlights instead their interaction with the mechanism through which choices are made. Not only are schemes such as *DE* vs *MPL/BDM*) differentially sensitive to image concerns, but their effectiveness at measuring intrinsic preferences, or on the contrary spurring higher contributions, can even reverse as reputational motives intensify.

Another strand of work focuses on decision makers' probability of being pivotal (Feddersen, Gailmard, and Sandroni, 2009; Grossman, 2015; Falk, Neuber, and Szech, 2020; Bartling et al., 2022). We show how, in mechanisms such as *MPL*, the probability of having one's choice implemented varies systematically with the intensity of image concerns, as does the expected cost at which the choice will be implemented, and we analyze how both effects shape equilibrium behavior. This relates the paper to work on auctions with signaling, in which bidders seek to demonstrate goodness, wealth, or a strong aftermarket position (Goeree, 2003; Giovannoni and Makris, 2014; Bos and Pollrich, 2020; Bos and Truys, 2022). In our setting, an agents' distribution of potential outcomes depends only on his own choices, and

2. We also conduct a placebo experiment with 366 additional subjects, keeping all aspects unchanged except that choices are now over a non-moral good, for which no image concerns arise. As expected, we find no significant difference between the two elicitation methods.

this lower strategic complexity allows us to identify intuitive effects and testable predictions.

With respect to experimental methodology, we contribute to the study of alternative elicitation mechanisms. Substantial research has compared how *DE*, *BDM*, *MPL* or random implementation (Selten, 1967) affects behavior in one-shot, anonymous games such as dictator or public-goods (Brandts and Charness, 2011; Chen and Schonger, 2016).<sup>3</sup> There is also a large body of research on elicitation methods for risk, time and ambiguity preferences (Charness, Gneezy, and Imas, 2013; Cox, Sadiraj, and Schmidt, 2015; Cohen et al., 2020; Baillon, Halevy, and Li, 2022). To our knowledge, no such study has explored reputationally sensitive decisions like those analyzed here. For choices in the moral domain, self-image (at least) is almost inevitably at play, and can create differences between elicitation methods.<sup>4</sup>

Finally, the paper relates to the debate between consequentialist and deontological principles. The evidence on how people behave in practice is mixed: the literature on public-goods contributions and charitable giving finds that choices are generally sensitive to the implied consequences (Ledyard, 1995; Goeree, Holt, and Laury, 2002), including the risk of having no impact (Brock, Lange, and Ozbay, 2013) and overhead costs (Gneezy, Keenan, and Gneezy, 2014). At the same time, there is evidence of “warm glow” altruism, in which utility is derived from the act as such (Andreoni, 1989, 1990). Experiments that directly focus on consequentialist versus deontological or expressive choices (Falk, Neuber, and Szech, 2020; Van Leeuwen and Alger, 2021; Bénabou, Falk, and Henkel, 2022; Chen and Schonger, 2022) also suggest a mix of preferences. Our paper shows that, when image concerns are important, a mechanism like *MPL* or *BDM* can easily lead consequentialist agents to adopt deontological-looking behaviors.

## 3.2 Model

### 3.2.1 Preferences

Agents are risk-neutral, with a two-period horizon,  $t = 1, 2$ . At date 1, an individual can engage in prosocial behavior ( $a = 1$ ) or act selfishly ( $a = 0$ ). Choosing  $a = 1$  involves a personal cost  $c > 0$  but generates a public good or externality  $e \geq 0$ . Agents differ in their intrinsic motivation to act morally: given  $e$ , it is either  $v_H e$  (high

3. Concerning *DE* with deterministic versus random implementation (an intermediate case relative to *MPL*), the overview by Charness, Gneezy, and Halladay (2016) reports generally ambiguous effects. As the model will make clear, it is only in the presence of sufficient signaling concerns that probabilistic implementation will matter. In contrast, risk attitudes play no role in the effects that we identify, which directly affect expected returns.

4. In the non-moral domain, in contrast, the literature tends to find no difference between *DE* and *BDM* (Miller et al., 2011; Berry, Fischer, and Guiteras, 2020; Cole et al., 2020).

type) or  $v_L e$  (low type), with probabilities  $\rho$  and  $1 - \rho$ ,  $v_H > v_L \geq 0$ , and average  $\bar{v} = \rho v_H + (1 - \rho)v_L$ .

Besides the externality, the second feature of action  $a = 1$  tying it to the moral domain is that it can be reputationally valuable, conferring a social or self-image benefit at date 2. In the social context, the agent knows his type but the audience (peer group, firms, potential partners) does not. In the self-signaling context, he has an immediate, “intuitive” sense of his deep preferences at the moment of action – for instance, how much empathy or spite he experiences – but later on the intensity of that feeling is imperfectly accessible (“forgotten”), and only the deed itself,  $a = 0$  or  $1$ , can be reliably recalled to assess his own moral identity.

Under either interpretation, an agent of type  $v = v_H, v_L$  has expected utility

$$(ve - c)a + \mu \hat{v}(a), \quad (3.1)$$

where  $\hat{v}(a)$  is the expected type conditional on the action  $a \in \{0, 1\}$  and the circumstances under which it took place (deterministic cost, random draw from a list, etc.), while  $\mu$  is the strength of self or social-image concerns, common to all agents. This utility may be additively augmented by any externalities generated by others, but since that term is independent of the agent’s action we omit it here. Note that these preferences are consequentialist: an agent’s desire to behave prosocially trades off the externality he expects his actions to have, the personal costs involved, and the reputational consequences.

As common in signaling models, multiple equilibria may coexist: when

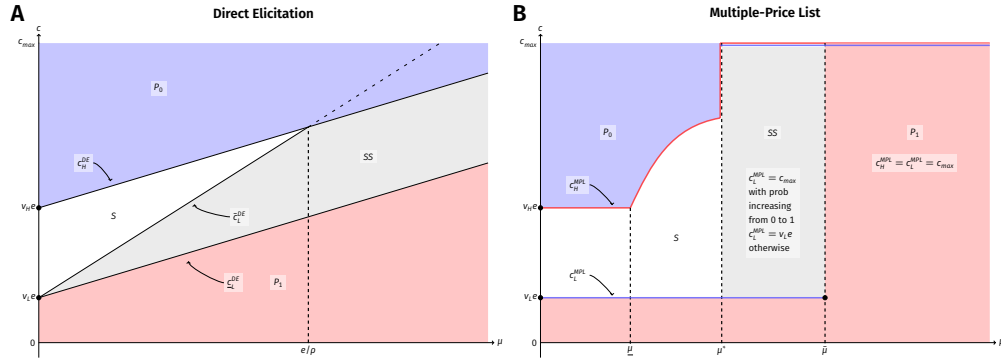
$$\max \{v_L e - c + \mu(v_H - v_L), v_H e - c + \mu(v_H - \bar{v})\} \leq 0 \leq v_H e - c + \mu(v_H - v_L),$$

there is both a pooling equilibrium at  $a = 0$  and a separating one in which the  $v_H$  type contributes, with a mixed-strategy one in-between. In case of multiplicity (see the Appendix), we choose the equilibrium that is best for both types, namely the no-contribution pooling equilibrium. Indeed, separation yields lower payoffs for both, since  $\mu v_L < \mu \bar{v}$  and  $v_H e - c + \mu v_H \leq \mu \bar{v}$ .

This simple framework readily implies that an agent is more likely to act morally the higher the externality  $e$ , his preference  $v \in \{v_H, v_L\}$ , and/or his image concern  $\mu$ . We extend use it to study how people’s (un)willingness to accept different tradeoffs between personal gain and harm to others varies across two “canonical” preference-elicitation mechanisms. All proofs are gathered in the Appendix.

### 3.2.2 Direct elicitation

Under *DE*, the individual faces a take-it-or-leave-it opportunity to incur a given cost (or forfeit a given prize)  $c$  to create an external benefit  $e$ . This decision can be made under varying levels of reputational concern  $\mu$ , corresponding to the public visibility or private memorability of choices.



**Figure 3.1.** Equilibrium under direct elicitation (panel A) and multiple-price list (panel B)

Notes:  $P_0$ : pooling at  $a_H = a_L = 0$ ;  $S$ : separation,  $a_H = 1, a_L = 0$ ;  $SS$ : semi-separation:  $a_H = 1, a_L \in (0, 1)$ ;  $P_1$ : pooling at  $a_H = a_L = 1$ .

As illustrated in Panel A of Figure 3.1 (for  $\rho < 1/2$ ), equilibrium behavior is characterized by three cost thresholds, increasing in  $\mu$ , that delineate regions of separation, semi-separation, and pooling:

$$v_H e - c_H^{DE}(\mu) + \mu(v_H - \bar{v}) \equiv 0, \quad (3.2)$$

$$v_L e - \bar{c}_L^{DE}(\mu) + \mu(v_H - v_L) \equiv 0, \quad (3.3)$$

$$v_L e - \underline{c}_L^{DE}(\mu) + \mu(\bar{v} - v_L) \equiv 0. \quad (3.4)$$

Denoting  $a_H^{DE}(c, \mu)$  and  $a_L^{DE}(c, \mu)$ , or  $a_H$  and  $a_L$  for short, the two types' probabilities of choosing  $a = 1$ , we show

**Proposition 3.1.** *The outcome of direct elicitation is as follows:*

- (1) For low costs,  $c < \min\{\underline{c}_L^{DE}, c_H^{DE}\}$ , everyone behaves morally,  $a_H = a_L = 1$ .
- (2) For intermediate costs,  $c \in (\underline{c}_L^{DE}, c_H^{DE})$ , the high type behaves morally ( $a_H = 1$ ), but the low type's probability  $a_L(c)$  of doing so decreases with  $c$ , and then equals 0 for  $c \geq \min\{\bar{c}_L^{DE}, c_H^{DE}\}$ .
- (3) For high costs,  $c \geq c_H^{DE}$ , both types behave immorally,  $a_H = a_L = 0$ .

Relative to “pure” (intrinsic) moral preferences  $v e$ , decision thresholds are inflated due to reputational concerns; see (3.2)-(3.4). In particular, the range of costs  $[\bar{c}_L^{DE}, c_H^{DE}]$  where full separation occurs shrinks with  $\mu$ , becoming empty for  $\mu > e/\rho$ .

### 3.2.3 Becker-DeGroot-Marschak or multiple-price list

Under *BDM*, the individual “names his price” by stating what maximum cost  $c \in [0, c_{\max}]$  he is willing to incur for taking action  $a = 1$ , where  $0 \leq v_L e < v_H e < c_{\max}$ . Equivalently,  $c$  represents his willingness to accept a “bribe” to make the immoral choice,  $a = 0$ . This elicitation is made incentive-compatible by drawing some  $\tilde{c} \in$

$[0, c_{\max}]$  according to a preannounced distribution  $G(\tilde{c})$ , and implementing  $a = 1$  at cost  $\tilde{c}$  only when  $\tilde{c} \leq c$ . With *MPL*, the price range is discretized and subjects state contingent choices at each level. Both schemes generate identical incentives, so we gather them under the label of *MPL*, since that is the format we implement experimentally.

In experiments,  $G$  is typically uniform, but we allow any other case, including  $c_{\max} = +\infty$ . Let  $L(c)$  denote the low type's net loss from selecting the cutoff  $c$ :

$$L(c) \equiv \int_{v_L e}^c (\tilde{c} - v_L e) dG(\tilde{c}) = \underbrace{\mathbb{P}(\tilde{c} \in [v_L e, c])}_{\text{cheap-talk effect}} \underbrace{(\mathbb{E}[\tilde{c} | \tilde{c} \in [v_L e, c]] - v_L e)}_{\text{cheap-act effect}}$$

and assume  $L(c_{\max}) < \infty$ , for which it suffices that  $E_G[\tilde{c}] < \infty$ . We will say that a subject is *observationally deontological* if he turns down all prices on the proposed list (with distribution  $G$ ): given the available data, he behaves as someone who would not act immorally “at any price.”

We now solve for both types' willingness to accept (WTA) under the multiple-price list, denoted  $c_H^{MPL}$  and  $c_L^{MPL}$  respectively. Note first that, *absent* reputation concerns ( $\mu = 0$ ), *MPL* and *DE* are equivalent, and reveal true preferences:  $c_H^{DE} = c_H^{MPL} = v_H e$ ,  $c_L^{DE} = \tilde{c}_L^{DE} = c_L^{MPL} = v_L e$ . For  $\mu > 0$ , comparing  $L(c)$  to the reputational stakes  $\mu(v_H - v_L)$  and  $\mu(v_H - \bar{v})$  yields both types' equilibrium strategies, illustrated in Panel B of Figure 3.1 and characterized again by critical thresholds between separating, semi-separating and pooling regions:

$$\underline{\mu} \equiv \frac{L(v_H e)}{v_H - v_L} < \mu^* \equiv \frac{L(c_{\max})}{v_H - v_L} < \frac{L(c_{\max})}{\rho(v_H - v_L)} \equiv \bar{\mu}. \quad (3.5)$$

**Proposition 3.2.** *The outcome of the MPL mechanism is as follows:*

(1) *When the (self) reputational concern  $\mu$  is low,  $\mu < \mu^*$ , the high type's WTA for behaving immorally is  $c_H^{MPL} = \max\{v_H e, L^{-1}(\mu(v_H - v_L))\}$ , while the low type finds it too costly to pool and accepts  $c_L^{MPL} = v_L e$ .*

*Initially, for  $\mu \leq \underline{\mu}$ , separation is costless for the high type, then as  $\mu$  rises he has to raise his reservation price to separate from the low type.*

(2) *When  $\mu$  is intermediate,  $\mu \in [\mu^*, \bar{\mu}]$ , the high type can no longer separate and becomes observationally deontological,  $c_H^{MPL} = c_{\max}$ . The low type randomizes, with probability  $a_L(\mu)$  increasing in  $\mu$ , between that same “virtuousness” ( $c_L^{MPL} = c_{\max}$ ) and revealing himself (accepting  $c_L^{MPL} = v_L e$ ).*

(3) *When  $\mu > \bar{\mu}$ , (self) image concerns are strong enough that both types' behavior is observationally deontological:  $c_H^{MPL} = c_L^{MPL} = c_{\max}$ .*

### 3.2.4 Comparison of DE vs. MPL

Under both elicitation schemes, image concerns naturally raise contributions, as seen in Figure 3.1. More novel and complex are the following questions:

1. Is one elicitation scheme *more image-sensitive* than the other?
2. Which one yields *more expected contributions*?

Formally, at a given cost  $c \in [0, c_{\max}]$ , what fraction of people  $\bar{a}^{DE}(c, \mu)$  accept forfeiting  $c$  to implement  $a = 1$  under *DE*, versus what fraction  $\bar{a}^{MPL}(c, \mu)$  state a willingness to pay of at least  $c$  under *MPL*? And how does  $\bar{a}^{DE}(c, \mu) - \bar{a}^{MPL}(c, \mu)$  depend on  $\mu$ ?

While the answers generally depend on the specific value of  $c$ , the cases of sufficiently low and high image concerns yield clear predictions. We will denote as  $\mu^{**}$  the solution to  $\underline{c}_L^{DE}(\mu) = c_{\max}$ , or

$$\mu^{**} \equiv \frac{c_{\max} - v_L e}{\bar{v} - v_L} > \frac{L(c_{\max})}{\bar{v} - v_L} = \bar{\mu}. \quad (3.6)$$

Putting together Propositions 3.1 and 3.2, we have:

**Proposition 3.3.** *For each type  $\tau = H, L$ ,*

- (1) *Visibility raises contributions: for any  $c \in [0, c_{\max}]$ ,  $a_\tau^{DE}(c, \mu)$  and  $a_\tau^{MPL}(c, \mu)$  coincide at  $\mu = 0$ , then both increase (weakly) as  $\mu$  rises, reaching 1 for  $\mu$  large enough.*
- (2) *Under low image concerns, DE yields more contributions: for all  $\mu \in (0, \underline{\mu})$ ,  $a_\tau^{DE}(c, \mu) \geq a_\tau^{MPL}(c, \mu)$ , with strict inequality for  $c \in (v_L e, \bar{c}_L^{DE}(\mu))$  and  $\bar{c} \in (v_H e, \bar{c}_H^{DE}(\mu))$ , both nonempty.*
- (3) *Under high image concerns, MPL yields more contributions: for all  $\mu \geq \bar{\mu}$ ,  $a_\tau^{DE}(c, \mu) \leq a_\tau^{MPL}(c, \mu) = 1$ , with strict inequality for  $c \in (\underline{c}_L^{DE}(\mu), c_{\max})$ , which is nonempty whenever  $\mu \in (\bar{\mu}, \mu^{**})$ .*
- (4) *The average behavior over types,  $\bar{a}^m(c, \mu) \equiv \rho a_H^m(c, \mu) + (1 - \rho) a_L^m(c, \mu)$ ,  $m = DE, MPL$ , inherits these same properties.*

The first result is standard, while the others stem from the interplay of three effects.

*Weak image concerns: discouragement effect dominates.* When  $\mu > 0$  is low enough that separation under *MPL* is costless, we have  $\bar{c}_H^{MPL}(\mu) = v_H e < \bar{c}_H^{DE}(\mu)$  and  $\underline{c}_L^{MPL}(\mu) = v_L e < \underline{c}_L^{DE}(\mu)$ , hence the second result. Intuitively, *MPL* raises the cost to the low type of mimicking the high one, since to do so he must forego up to  $v_H e$ , and for low reputational gain such a discrete cost is not worth it. Under *DE*, in contrast, he pays only in proportion to the gain. This intuition is reflected in the fact that the lower boundary of the separating region is linear in Panel A of Figure 3.1, whereas it is initially flat in Panel B.

*Strong image concerns: cheap-act effect dominates.* At high values of  $\mu$ , reputational concerns become paramount, and the cost of signaling is lower under *MPL*

than under  $DE$ , since high values of  $c$  must only be paid with a probability less than 1: the effective cost of stating a cutoff  $c$  is only  $E[\tilde{c}|\tilde{c} \leq c] < c$ . It is even bounded by  $L(c_{\max}) + v_L e < \infty$ , which limits the extent to which the high type can separate, so that for  $\mu > \bar{\mu}$  full pooling occurs:  $c_H^{MPL} = c_L^{MPL} = c_{\max}$ , so  $a^{MPL}(c, \mu) = 1$ , whereas  $\bar{a}_L^{DE}(c, \mu) < 1$  as long as  $\mu < \mu^{**}$ . Most importantly:

**Property 1.** For any distribution satisfying the monotone hazard rate property ( $g/(1-G)$  increasing), the “discount”  $c - E[\tilde{c}|\tilde{c} \leq c]$  is increasing in  $c$ . Therefore, as  $\mu$  rises and with it each type’s cutoff, the *cheap-act* effect becomes stronger, which increases MPL contributions relative to  $DE$ .

*Intermediate image concerns.* Inside  $(\underline{\mu}, \bar{\mu})$ , a third “*cheap-talk*” effect is also important. Under  $MPL$ , an agent who states a cutoff  $c < c_{\max}$  has only a probability  $G(c) < 1$  of being called upon to actually “deliver”: if  $\tilde{c} > c$  is drawn, he neither incurs a cost *nor* generates the externality  $e$ . This makes it safer to state high cutoffs, thus adding to the *cheap-act* effect. The latter is not as strong in this range as for high values of  $\mu$ , and conversely the *cheap-talk* effect weakens as  $\mu$  rises, pushing  $G(c^{MPL})$  closer to 1. The net balance of the three effects is generally ambiguous in this intermediate range, and consequently so is the sign of  $a^{DE} - a^{MPL}$ .

*Implications.* Three main predictions emerge from the model. First, as usual, greater visibility increases contributions. Second, at low but positive levels of visibility,  $DE$  leads to more prosocial outcomes, as the *discouragement effect* dominates. Third, at high levels (but not so high as to push everyone to  $a = 1$  under  $DE$ ), this ordering reverses:  $MPL$  induces more moral decisions, due to the now dominating *cheap-act* effect.

The inequalities in Proposition 3.3 can be weak or strong, depending on the region of the parameter space. This is a standard feature of models with discrete types and action spaces, which typically disappears when there is sufficient heterogeneity to span all cases. For this reason, when confronting the model with data, we will tighten the predicted inequalities to be strict ones.<sup>5</sup>

## 3.3 Experimental design

### 3.3.1 Saving a Life

We adopt the *Saving a Life* paradigm from Falk and Graeber (2020), in which subjects can either take money for themselves or implement a fixed, life-saving donation to a charity dedicated to the treatment of tuberculosis in India. According to the

5. Our tests will vary  $(\tau, \mu) \in \{DE, MPL\} \times \{\mu_L, \mu_H\}$  while maintaining the same realized cost  $c$ . By Proposition 3.3, the set of parameters such that  $0 < \mu_L < \underline{\mu} < \bar{\mu} < \mu_H < \mu^{**}$  and  $\bar{a}^{DE}(c, \mu_L) - \bar{a}^{MPL}(c, \mu_L) > 0 > \bar{a}^{DE}(c, \mu_H) - \bar{a}^{MPL}(c, \mu_H)$  is nonempty provided that  $c_L^{DE}(\bar{\mu}) < c_H^{DE}(\underline{\mu})$ , which reduces to  $L(c_{\max}) - (1 - \rho)L(v_H e) < (v_H - v_L)e$ . Sufficient conditions are easily found; with a uniform  $G$ , for example,  $c_{\max}/e \in (v_H - (v_H^2 - v_L^2)^{1/2}, v_H + (v_H^2 - v_L^2)^{1/2})$  suffices.

World Health Organization, tuberculosis is one of the ten leading causes of death worldwide, even though there are highly effective antibiotic treatments available. Together with the Indian non-profit organization *Operation ASHA*, we calculated a specific monetary amount sufficient to identify, treat, and cure a number of patients such that – in expectation – one patient will be saved from death by tuberculosis due to the donation. Combining public information on the charity’s operations with estimates from peer-reviewed studies on mortality due to tuberculosis and treatment effectiveness for the specific location considered (Kolappan et al., 2008; Straetmans et al., 2011; Tiemersma et al., 2011), we determined that level to be 350€: by allowing for the treatment of five patients, such a donation allows the (expected) saving of one human life.

This paradigm contrasts the option of saving a life (major positive externality  $e$ ) by triggering a donation of 350€ versus that of taking money for oneself (opportunity cost  $c$ ), inducing a clear tradeoff between morality and self-interest.

### 3.3.2 Treatments

We use a  $2 \times 2$  between-subjects design, varying the elicitation method (DE vs. MPL) as well as the visibility and moral salience of choices (*Low Image* vs. *High Image*) at the payment stage.

Under DE, subjects faced the binary choice between receiving  $c = 100\text{€}$  ( $\approx \$110$ ) as payment, or saving a human life in expectation. As part of the experimental design, we predetermined this single value of  $c = 100\text{€}$  as a compromise between two practical concerns: (i)  $c$  must be high enough to generate choices of both types; (ii) in contrast to *MPL*, each implemented decision has a sure cost to the experimental budget of either  $c$  or the full 350€ donation, which quickly adds up.

For the *MPL* conditions, we used a price-list design: starting with  $c = 0\text{€}$  and proceeding in 10€ increments up to  $c = 200\text{€}$ , subjects could indicate in each of the 21 contingent choices whether they wanted to save a life or take  $c$  for themselves. Each price was then equally likely to be drawn for implementation (uniform  $G$ ).<sup>6</sup> Figures 3.B.1 and 3.B.2 in Appendix 3.B display the corresponding decision screens.

Turning to visibility, recall that the two key forces underlying Proposition 3.3, namely the *discouragement* and the *cheap-act* effects, both require a non-zero level of image concerns. To ensure a minimal level of image concern in both treatments, we notified subjects at the start that: (i) they were anonymously paired with another participant in the same session; (ii) they would see, at the end of the experiment, their own and their partner’s choices displayed alongside on their screens, as would their partner. Apart from observing the partner’s choices, subjects received no information about them, so that no other aspect of the dyad would influence decisions.

6. To avoid inconsistent answers, we enforced a single-switching rule.



To keep image concerns minimal in the *Low Image* treatment ( $\mu = \mu_L$ ), we made the payment procedure double-blind, so that not even the experimenter could link subjects' decisions to their identity. Following Barmettler, Fehr, and Zehnder (2012), at the start of each session one subject was randomly designated to carry out all payments: they did not participate in the regular experiment, and thus had no knowledge about the choices. At the end, payments were stuffed into envelopes and the selected subject handed them out, in an adjacent room, to those who had actively participated.

The *High Image* treatment ( $\mu = \mu_H$ ), in contrast, was designed to induce strong image concerns. Subjects were informed that upon receiving payment: (i) their choice would be compared to that of their matched partner by a committee of three persons, sitting in the room where payments would take place; (ii) both partners' choices would be projected onto a wall, and they would have to read both aloud, using two predetermined sentences.

### 3.3.3 Procedure

697 subjects (405 female, mean age = 24.01, SD = 6.21) participated in 36 sessions at the BonnEconLab of the University of Bonn: 178 in the *MPL-Low Image* treatment, 178 in *MPL-High Image*, 165 in *DE-Low Image*, and 176 in *DE-High Image*. Subjects were recruited using Hroot (Bock, Baetge, and Nicklisch, 2014), and the experiment was conducted using oTree (Chen, Schonger, and Wickens, 2016). Sessions lasted about 60 minutes, with a show-up fee of 12€. For each session, one matched pair of subjects was randomly drawn, and their choices implemented. Thus, in the DE treatments, each of the two either received 100€, or triggered a life-saving 350€ donation. In the MPL treatments, one price from the list was randomly drawn (uniformly), and the pre-stated choices of both partners for this price were implemented. Therefore, each one either triggered the donation or received up to 200€. <sup>7</sup>

At the beginning of each session, subjects received a verbal introduction to the experiment. In the *Low Image* treatments, the procedure ensuring anonymity was explained and demonstrated. In the *High Image* treatments, the committee setup was shown. Subsequently, all subjects received detailed information about tuberculosis, its effects, and treatment. The instructions also linked to a website where they were invited to confirm the validity of the information. We then introduced the charity and its working procedure, and explained our calculations regarding the life-saving effect of the 350€ donation. Subjects then learned about their choice options and, after answering a couple of comprehension questions, made their decisions. Finally, they completed a short questionnaire and were paid in a separate room, with pay-

7. This random implementation adds another layer of the cheap-talk effect, but one that affects DE and MPL in exactly the same way (formally equivalent to dividing  $\mu$  by the probability of implementation), and thus leaves all comparisons between the two unaffected.

ment procedures depending on treatment status, as explained above. For further details on the procedure and instruction, see Appendix 3.D.

### 3.4 Hypotheses and results

Our outcome variable is the fraction  $\bar{a}^m(c, \mu)$  of subjects who choose to save a life over receiving  $c$ , given an elicitation method  $m \in \{DE, MPL\}$  and a level of visibility  $\mu \in \{\mu_L, \mu_H\}$ . For brevity, we will refer to  $\bar{a}^m(c, \mu)$  as “total contributions”.

#### 3.4.1 Hypotheses

Based on Proposition 3.3, we state:

**Hypothesis 3.1.** *For both DE and MPL, total contributions are higher under High Image than under Low Image:  $\bar{a}^{DE}(c, \mu_H) > \bar{a}^{DE}(c, \mu_L)$ ,  $\bar{a}^{MPL}(c, \mu_H) > \bar{a}^{MPL}(c, \mu_L)$ .*

**Hypothesis 3.2.** *Under Low Image, total contributions are higher under DE than under MPL:  $\bar{a}^{DE}(c, \mu_L) > \bar{a}^{MPL}(c, \mu_L)$ .*

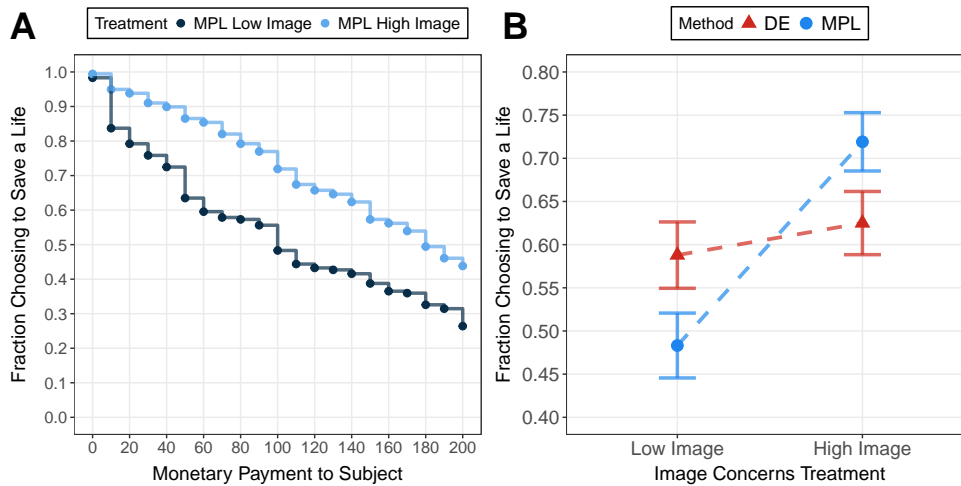
**Hypothesis 3.3.** *Under High Image, total contributions are higher under MPL than under DE:  $\bar{a}^{DE}(c, \mu_H) < \bar{a}^{MPL}(c, \mu_H)$ .*

Hypothesis 3.1 captures the standard effect of signaling concerns. The novel ones are Hypotheses 3.2 and 3.3, reflecting the dominance of the *discouragement effect* at  $\mu_L$  and the *cheap-act effect* at  $\mu_H$ . Together, they constitute the model’s distinctive crossing prediction, which we will test at  $c = 100\text{€}$ , as explained earlier.

#### 3.4.2 Results

**Hypothesis 1.** Under both elicitation methods, increased visibility led to a rise in total contributions, but the magnitude was markedly different. Under *DE*, 58.8% of subjects chose to save a life in *Low Image* and 62.5% in *High Image* –a relatively small and insignificant increase ( $p = 0.51$ ). Under *MPL*, increased visibility had a much larger effect. At almost all payment levels, the fraction of subjects choosing to save a life is at least 15 pp. higher under *MPL-High Image* than under *MPL-Low Image*, resulting in significantly different distributions ( $p < 0.001$ , Kolmogorov–Smirnov test); see Panel A of Figure 3.2. At 100 €, contributions are 23.6 pp. and significantly higher under *High Image* than under *Low Image* ( $p < 0.001$ ).

**Hypotheses 2 and 3.** Panel B of Figure 3.2 shows that the fractions  $\bar{a}^m(100, \mu)$  choosing to save a life over 100€ clearly differ by elicitation method, with the ranking reversing between  $\mu_L$  and high  $\mu_H$ . Under *Low Image*, we observe  $\bar{a}^{MPL}(\mu_L) < \bar{a}^{DE}(\mu_L)$ , as predicted by Hypothesis 2, and consistent with the dominance of the *discouragement effect*. The difference is large, with the fraction saving a life rising from 48.3% to 58.8% between *MPL* and *DE*, though significance is slightly below



**Figure 3.2.** Main experimental results

Notes: Panel A displays the fraction choosing to save a life for each offered price in the MPL Low Image and MPL High Image treatments. Panel B shows the interaction effect of elicitation method and image concerns, by displaying the fractions choosing to save a life with MPL and DE, under either the Low Image or the High Image treatment. Error bars indicate the standard error of the mean.

the standard level ( $p = 0.065$ , *Fisher's exact test*). Conversely, under *High Image* we observe  $\bar{a}^{MPL}(\mu_H) > \bar{a}^{DE}(\mu_H)$ , in line with the *cheap-act effect* dominating, as predicted by Hypothesis 3. The difference is again about 10 percentage points, but now in the opposite direction, rising from 62.5% under DE to 71.9% under MPL, albeit again with significance slightly short of 5% ( $p = 0.070$ ).

Table 3.1, Panel A regresses the probability of choosing to save a life (instead of taking 100€) on a dummy for the type of elicitation (1 for *MPL*), which yields a positive coefficient for *Low Image* in Column (1), and a negative one for *High Image* in Column (3).<sup>8</sup> Columns (2) and (4) show that these effects remain largely unaffected by controls for age, gender, high-school graduation grade, highest educational degree obtained so far, self-reported monthly income, and a measure of religiousness (Likert scale).

Hypotheses 2-3 represent the strictest possible test of the model – a particular ordering of four variables– which may explain the marginal significance of those results. A more standard test, concerns their joint implication of a *differential image sensitivity*: as image rises from  $\mu_L$  to  $\mu_H$ , the increase in contributions should be more pronounced for *MPL* than for *DE*. Panel B of Table 3.1 thus presents an OLS regression interacting *High Image* with *MPL*, using *DE-Low Image* as baseline; the interaction is positive and significant at the 1-percent level.

Overall, the results lend support to the key predictions of the model, albeit with significance being sometimes marginal. As such, our simple experiment can be taken

8. The results remain qualitatively unchanged with Probit or Logit regressions.

as a proof of concept for the novel mechanisms brought to light by the model, thereby opening them up to further exploration.

**Robustness experiment.** One may worry that features of the elicitation methods unrelated to image concerns might be at play in our results. Note first that these would have to generate not just different *DE* versus *MPL* contributions, but also a flipping of that gap as image rises from low to high, which seems unlikely. Nonetheless, to rule out potential confounding factors we ran the *DE* versus *MPL* treatments on another 366 subjects, with the donation replaced by a non-moral good (university-shop voucher). For this “placebo”,  $\mu = 0$ , and indeed we find no significant differences between *MPL* and *DE*: see Panel C of Table 3.1, and Appendix 3.C for implementation details.

### 3.5 Conclusion

Our model and experiment show that image concerns affect the measurement of moral preferences in ways that *interact with the elicitation method*. Regardless of whether one is interested in image-inclusive preferences (for positive predictions) or in purely intrinsic ones (for normative judgements), behavior will differ between direct and price-list mechanisms. These results argue for caution in interpreting standard estimates of moral preferences from experiments and contingent-valuation surveys,<sup>9</sup> but also provide potential guidance for maximizing public-goods contributions and image manipulations.<sup>10</sup>

In particular, even purely utilitarian individuals may act, when facing *BDM*- or *MPL*-like situations, as if deontologically motivated: refusing all proposed prices in exchange for what is perceived as having a dignity. With necessarily finite budgets, a definitive test of how many “real Kantians” there are is ultimately impossible, but our experiment provides both an upper bound and some grounds for skepticism about public positions on the subject. The former is given by the 26.4% of subjects who choose to save a life over the maximum offer of 200€ in the Low Image *MPL* condition. The latter stems from the fact that this proportion nearly doubles to 43.82% with a mild visibility manipulation. These results can also help to account for the common resistance to estimating and using a “statistical value of life.” Despite the fact that we implicitly engage in trading off costs and statistical lives all the time, explicit reference to putting a price tag on life typically produces conspicuously displayed righteous indignation (e.g., Sandel, 2012).

9. A related point is made by Chen and Schonger (2022) for other forms of preferences involving moral “duties”.

10. Individual WTP’s, which include the value of social and self-image, are the right measures to predict, explain or alter behavior. To inform policy, however, they can substantially overstate the true social value of the public good. Thus, in our model, reputation is a positional good, the image gains and losses of contributors and non-contributors exactly offsetting each other. In general, the image game can have negative, zero, or positive sum, depending on the curvature of the reputation functional; Butera et al. (2022) find evidence for negative sum, which reinforces the previous point.

**Table 3.1.** Regression analyses of the effect of the elicitation method on prosocial behavior

<b>Panel A:</b>				
Dependent variable:	Choice to Save a Life (vs. 100€)			
	<i>Low Image Concerns</i>		<i>High Image Concerns</i>	
	(1)	(2)	(3)	(4)
<i>MPL</i>	-0.105 (0.054)	-0.103 (0.053)	0.094 (0.050)	0.091 (0.050)
Constant ( <i>DE</i> )	0.588 (0.038)	0.626 (0.049)	0.625 (0.037)	0.622 (0.046)
Controls		X		X
Observations	343	343	354	354
<b>Panel B:</b>				
Dependent variable:	Choice to Save a Life (vs. 100€)			
	(1)	(2)		
<i>MPL</i>	-0.105 (0.054)	-0.097 (0.053)		
<i>High Image</i>	0.037 (0.053)	0.052 (0.052)		
<i>MPL X High Image</i>	0.199 (0.073)	0.190 (0.072)		
Constant ( <i>DE Low Image</i> )	0.588 (0.038)	0.595 (0.044)		
Controls		X		
Observations	697	697		
<b>Panel C:</b>				
Dependent variable:	Choice of Voucher (vs. 10€)			
	(1)	(2)		
<i>MPL No-Image</i>	0.045 (0.047)	0.051 (0.047)		
Constant ( <i>DE No-Image</i> )	0.253 (0.033)	0.227 (0.047)		
Controls		X		
Observations	366	366		

The table shows OLS regression coefficients. Robust standard errors in parentheses. Controls include age, gender, income, religiousness, educational level, and high school grade.

### Appendix 3.A Proofs

**Proof of Proposition 3.1.** From (3.2)-(3.4), it follows that:

( $P_0$ ) :  $a_H = a_L = 0$ , sustained by out-of equilibrium belief (OEB)  $\hat{v} = v_H$  following  $a = 1$  (by the D1 criterion), is an equilibrium if and only if  $c \geq c_H^{DE}$ . When

$$\bar{c}_L^{DE} = v_L e + \mu(v_H - v_L) \leq c \leq v_H e + \mu(v_H - v_L) \equiv \bar{c}_H^{DE},$$

it coexists with a separating equilibrium  $S$  in which  $a_H = 1 = 1 - a_L$ , plus a mixed-strategy one in-between. As shown earlier, however,  $P_0$  is Pareto dominant, and therefore selected.

( $P_1$ ) :  $a_H = a_L = 1$ , sustained by OEB  $\hat{v} = v_L$  following  $a = 0$  (by D1), is an equilibrium if and only if  $c \leq c_L^{DE}$ .

( $S$ ) :  $a_H = 1 - a_L = 1$  is an equilibrium if and only if  $\bar{c}_L^{DE} \leq c \leq \bar{c}_H^{DE}$ .

( $SS_1$ ) :  $0 < a_L < 1 = a_H$ , with belief  $\hat{v} \in (v_L, \bar{v})$  following  $a = 1$ , is an equilibrium if and only if  $c_L^{DE} < c < \bar{c}_L^{DE}$ . The low type's mixed strategy  $a_L(c) \in (0, 1)$  is then given by combining the indifference condition  $v_L e - c + \mu(\hat{v}(a_L) - v_L) = 0$  and the Bayesian posterior  $\hat{v}(c) = [\rho v_H + (1 - \rho)a_L v_L] / [\rho v + (1 - \rho)a_L]$ :

$$v_L e - c + \frac{\mu \rho (v_H - v_L)}{\rho + (1 - \rho)a_L(c)} \equiv 0, \quad (3.A.1)$$

so  $a_L(c)$  decreases with  $c$ , while the reputation  $\hat{v}(c)$  following  $a = 1$  increases.

( $SS_0$ ) :  $0 = a_L < a_H < 1$ , with beliefs  $\hat{v} \in (\bar{v}, v_H)$  following  $a = 0$ , is an equilibrium if and only if  $c_H^{DE} < c < \bar{c}_H^{DE}$ . It always coexists with  $P_0$ , and is always dominated by it.

These results jointly imply that:

(a) If  $c_L^{DE} < \bar{c}_L^{DE} < c_H^{DE}$ , the unique equilibrium is  $P_1$  for  $c < c_L^{DE}$ ;  $SS_1$  for  $c \in [c_L^{DE}, \bar{c}_L^{DE}]$ ; and  $S$  for  $c \in [\bar{c}_L^{DE}, c_H^{DE}]$ . For  $c \geq c_H^{DE}$ , the dominant equilibrium is  $P_0$ .

(b) If  $c_L^{DE} < c_H^{DE} < \bar{c}_L^{DE}$  (where the second inequality means that  $\mu \rho > e$ ), the unique equilibrium is  $P_1$  for  $c < c_L^{DE}$ , and  $SS_1$  for  $c \in [c_L^{DE}, c_H^{DE}]$ . For  $c > c_H^{DE}$ , the dominant equilibrium is  $P_0$ .

(b) If  $c_H^{DE} < c_L^{DE} < \bar{c}_L^{DE}$  (where the first inequality means that  $\mu(2\rho - 1) > e$ ), the unique equilibrium is  $P_1$  for  $c < c_H^{DE}$ , and for  $c \geq c_H^{DE}$  the dominant equilibrium is  $P_0$ .

■

**Proof of Proposition 3.2.** The proof of existence is standard. For example, for a separating equilibrium to obtain, it must be: that (i) type  $v_L$  obtains his symmetric-information allocation (otherwise, he would be better off selecting  $c_L^{MPL} = v_L e$ ), and (ii) he does not want to mimic type  $v_H$ :  $\mu(v_H - v_L) \leq L(c_H^{MPL})$  and  $c_H^{MP} < c_{max}$ . It is easily verified that the proposed strategies satisfy these conditions, and similarly for the semi-separating and pooling equilibria.

The equilibrium is not unique absent refinement, however. For example, there is a pooling equilibrium at  $c^{MPL} = v_H e < c_{max}$  when  $\mu(\bar{v} - v_L) \geq L(v_H e)$ , sustained by

OBE  $\hat{v} = v_L$  following any declared price  $c \neq v_L$ . Note, however, that sorting implies monotonicity, so there is at most one price, denoted  $c^*$ , that can be chosen with positive probability by both types; any other price claimed by type  $v_H$  (respectively,  $v_L$ ) exceeds  $c^*$  (respectively, lies below it)  $c^*$ . Denote  $\hat{v}(c)$  the mean belief following a price  $c$ , and consider a deviation to  $c' = c^* + \varepsilon$ , for  $\varepsilon > 0$  arbitrarily small, together with the set of belief responses that raise both types' utilities relative to equilibrium

$$\begin{aligned}\hat{V}_L &\equiv \{\hat{v}(c^* + \varepsilon) \mid \mu[\hat{v}(c^* + \varepsilon) - \hat{v}(c^*)] > L_L(c^* + \varepsilon) - L_L(c^*)\}, \\ \hat{V}_H &\equiv \{\hat{v}(c^* + \varepsilon) \mid \mu[\hat{v}(c^* + \varepsilon) - \hat{v}(c^*)] > L_H(c^* + \varepsilon) - L_H(c^*)\}.\end{aligned}$$

Clearly  $V_L \subset V_H$ , so by D1 the deviation must induce a probability-one belief on  $v_H$ ; thus, the only possible pooling price is  $c = c_{max}$ . Consequently, the equilibrium must take one of the three forms described in the proposition, and because it is obtained on disjoint sets of parameters, it is unique under D1. ■

## Appendix 3.B Decision screens

**Your Decision**

---

[Please click here to be reminded of the precise meaning of 'saving a life'](#)

Option A				Option B
	A	B		
I save a human life	<input type="radio"/>	<input type="radio"/>		I choose 100 € as payment for myself

[Confirm decision](#)

**Figure 3.B.1.** Decision screen DE



**Your Decisions**

Please click here to be reminded of the precise meaning of 'saving a life'

Option A	A	1	B	Option B
I save a human life	<input type="radio"/>	1	<input type="radio"/>	I choose 0 € as payment for myself
I save a human life	<input type="radio"/>	2	<input type="radio"/>	I choose 10 € as payment for myself
I save a human life	<input type="radio"/>	3	<input type="radio"/>	I choose 20 € as payment for myself
I save a human life	<input type="radio"/>	4	<input type="radio"/>	I choose 30 € as payment for myself
I save a human life	<input type="radio"/>	5	<input type="radio"/>	I choose 40 € as payment for myself
I save a human life	<input type="radio"/>	6	<input type="radio"/>	I choose 50 € as payment for myself
I save a human life	<input type="radio"/>	7	<input type="radio"/>	I choose 60 € as payment for myself
I save a human life	<input type="radio"/>	8	<input type="radio"/>	I choose 70 € as payment for myself
I save a human life	<input type="radio"/>	9	<input type="radio"/>	I choose 80 € as payment for myself
I save a human life	<input type="radio"/>	10	<input type="radio"/>	I choose 90 € as payment for myself
I save a human life	<input type="radio"/>	11	<input type="radio"/>	I choose 100 € as payment for myself
I save a human life	<input type="radio"/>	12	<input type="radio"/>	I choose 110 € as payment for myself
I save a human life	<input type="radio"/>	13	<input type="radio"/>	I choose 120 € as payment for myself
I save a human life	<input type="radio"/>	14	<input type="radio"/>	I choose 130 € as payment for myself
I save a human life	<input type="radio"/>	15	<input type="radio"/>	I choose 140 € as payment for myself
I save a human life	<input type="radio"/>	16	<input type="radio"/>	I choose 150 € as payment for myself
I save a human life	<input type="radio"/>	17	<input type="radio"/>	I choose 160 € as payment for myself
I save a human life	<input type="radio"/>	18	<input type="radio"/>	I choose 170 € as payment for myself
I save a human life	<input type="radio"/>	19	<input type="radio"/>	I choose 180 € as payment for myself
I save a human life	<input type="radio"/>	20	<input type="radio"/>	I choose 190 € as payment for myself
I save a human life	<input type="radio"/>	21	<input type="radio"/>	I choose 200 € as payment for myself

Confirm decisions

**Figure 3.B.2.** Decision screen MPL

## Appendix 3.C Robustness experiment

In the main experiment, we showed how image concerns lead to differences in moral behavior between elicitation methods. One concern is that there are factors present in our experiment that lead to differences between DE and MPL independent of image concerns. In particular, the previous literature has identified two main factors

that could potentially confound the comparison between the two elicitation methods in our case.

First, in our experiment, only a subset of subjects had their decision implemented for real. In the MPL treatments, another randomization takes place, which is absent in DE: if selected for payout, one decision of the price list is randomly selected. If subjects violate the independence axiom and view these two randomization processes not separately but rather as a meta-lottery, this could potentially affect the comparison. This issue is also present in the many experiments that study decisions over lotteries and pay only one lottery out for real. In this context, it is usually assumed that subjects evaluate the different random processes in isolation, an assumption that has been repeatedly validated empirically<sup>11</sup>. It is natural to assume that subjects also perceive the two processes in isolation in our experiment since they were introduced and explained at two different points in the instructions.

The second factor is the so-called compromise effect (Simonson, 1989; Birnbaum, 1992; Andersen et al., 2006). When presenting a price list, the focus lies perceptually on the center. This in turn could change the attractiveness of the options appearing in the middle of the price list, biasing answers away from the subject's true valuations. To control for this effect, we carefully selected the DE value to correspond to the value precisely in the middle of the price list in the MPL treatments. As such, it seems unlikely that differences in perceptions could explain discrepancies between the elicitation methods.

Therefore, we would not expect differences between DE and MPL in our experiment once image concerns are absent. Nevertheless, in order to document this empirically, we conducted a robustness experiment, which is explained next.

### 3.C.1 Setup and treatments

For the robustness experiment, we used a good that is unrelated to prosocial and moral considerations, so that image concerns are plausibly absent. For this non-moral good, we chose a 35 € voucher for the University of Bonn's online shop. With the voucher, subjects can buy sweatshirts, T-shirts, and accessories related to the university. The voucher cannot be returned and is only valid for purchases in the shop. There were two between-subject treatments: *DE No-Image* and *MPL No-Image*. In the former, subjects could choose between 10 € and the voucher, while in the latter they faced a price list from 0€ to 20€ in 1€ increments. Note that this closely mimics the decisions in the main experiment. The only difference is that all values are divided by 10. As in the main experiment, subjects were paired with another subject, and only a subset of subjects had their choices implemented for real.

Accordingly, instructions for the decisions were identical, with the sole difference being that descriptions related to the saving a life paradigm were replaced with

11. See e.g., Starmer and Sugden (1991), Cubitt (1998) and Hey and Lee (2005).

descriptions of the voucher. Consequently, any factors influencing the comparison between *DE* and *MPL* in the main experiment should also manifest in the robustness experiment.

### 3.C.2 Procedure

Subjects were recruited from the same subject pool as the main experiment, with the restriction that they had not previously participated in the main experiment. The experiment was conducted as a virtual lab experiment since in-person lab sessions were not possible due to the ongoing Covid-19 pandemic. That is, the experiment started and ended at a pre-specified date and time, and the experimenter was available during the experiment in case of problems.

In total, 366 subjects (227 female, mean age 26.88, SD 7.87) took part, 188 in the *MPL No-Image*, and 178 in the *DE No-Image* treatment, respectively. The experiment lasted on average 13 minutes, for which the subjects received a show-up fee of 3€. Subjects were grouped in virtual sessions consisting of roughly 24 subjects, and one pair was randomly selected for payout out of each virtual session. Exactly as in the main experiment, for these two subjects, either their *DE* decision was implemented or a randomly chosen decision from the *MPL* list.

### 3.C.3 Results

Assessing subjects' general valuation of the voucher, we observe considerable variation in switching behavior in the *MPL No-Image* treatment. In total, 76% had an interior switching value, meaning they preferred the voucher in the initial decision but switched to preferring the monetary value at some point. The variation compares quite favorably to the *MPL-Low Image* treatment, where this was the case for 72% of subjects. Comparing the choice at 10 € in *MPL No-Image* with *DE No-Image*, we find that 29.8% choose the voucher in *MPL* and 25.3% in *DE*. This difference is small in magnitude and not statistically significant ( $p = 0.35$ ; two-sided Fisher's exact test). It is also in the opposite direction of what we find in the main experiment for the *Low Image* case, which is the natural comparison. Table 3.C.1 replicates this null result in an OLS-regression, with column (2) using the same variables as control variables as in the main experiment, compare Table 1, columns (2) and (4). Thus, we do not observe any meaningful differences between the two elicitation methods in our setting once image concerns are removed.

**Table 3.C.1.** Regression analyses of the effect of the elicitation method on voucher choice

Dependent variable:	Choice of Voucher (vs. 10€ )	
	(1)	(2)
<i>MPL No-Image</i>	0.045 (0.047)	0.051 (0.047)
Constant ( <i>DE No-Image</i> )	0.253 (0.033)	0.227 (0.047)
Controls		X
Observations	366	366
R <sup>2</sup>	0.003	0.039

Notes: The table shows OLS regression coefficients. Robust standard errors in parentheses. Controls include age, gender, income, religiousness, educational level, and high school grade.

## Appendix 3.D Instructions

### 3.D.1 Announcement by the experimenter

The following text was read aloud by the experimenter after all subjects were placed in their cubicles, establishing common knowledge among all subjects of a session. The content depended on the image treatment.

#### 3.D.1.1 Treatment *Low Image*

Welcome to today's study. In today's study, you will make decisions on a computer. These decisions will take place under complete anonymity. To ensure this, we will now apply the following procedure: You should all have two notes with your cubicle number in front of you. We will soon collect one of the two notes and randomly draw one out of all collected. The person in the drawn cubicle is responsible for the payment in today's study. At the end of the study, we prepare sealed envelopes with your payments. Those envelopes are then passed to the soon to be randomly drawn person, who will hand them out to each of you sequentially in the adjacent room. The envelopes are designed so that you cannot see the contents from the outside, i.e., not on weight or similar clues. Hence at no time can there be a connection drawn between your payment and your decisions. Please hold now one of the notes with your cubicle number onto out of your cubicle. (Responsible person is drawn and placed in the adjacent room) The study will begin shortly. If you have at any time have questions, just hold your hand out of the cubicle.

### 3.D.1.2 Treatment *High Image*

Welcome to today's study. In today's study, you will make decisions on your computer. Your decisions will subsequently be evaluated by a committee consisting of three students from the University of Bonn. For this, after you have made your decisions, you will go to the adjacent room, where your decisions will be projected on a wall with a projector. You will then briefly communicate your decisions to the committee, and the committee will evaluate them. Afterward, you will receive the result of the evaluation. Detailed information about your decisions, the committee, and the evaluation will be given to you at the appropriate time on your computer. The study will begin shortly. If you have at any time have questions, just hold your hand out of the cubicle.

### 3.D.1.3 Further procedure

After the text was read aloud, in the *Low Image* conditions the experimenter then collected one note from each subject indicating their respective cabin number. All notes were thrown into a bag, and one was drawn in front of all participants to make clear that the person responsible for the payment procedure was a randomly determined participant. In the *High Image* conditions, subjects were shown the adjacent room and the setup with the committee, which consisted of student research assistants. The members of the committee did not interact with the subjects in any way.

## 3.D.2 Introduction

All further instructions were displayed on the subjects screens. The following introduction was the same for all treatments.

### 3.D.2.1 Welcome to the study

Welcome, and thank you for your interest in today's study!

For your participation, you will receive a fixed payment of 12€ given to you at the end. In this study, you will make decisions on the computer. Depending on how you choose, you can earn additional money.

During the entire study, communication between participants is prohibited. Please turn off your phone so that other participants are not disturbed. Please only use the designated functions on the computer and make the entries with the mouse and keyboard. If you, at some point, have questions, please make a hand signal. Your question will be answered at your seat.

On the next screens, you will receive specific information about participation in this study. To proceed, click "Next".

### 3.D.2.2 Your partner

As part of this experiment, a partner has been assigned to you. This partner is a participant in today's experiment, just like you. He or she was randomly assigned to you and will receive the same instructions as you.

In today's experiment, you and your partner will both receive the exact same information and subsequently face the exact same decisions. These decisions have certain consequences, which will be described in detail later.

At the end of today's experiment, one pair is randomly drawn from all participants in today's experiment. Only the decisions of this pair will be implemented, as described in the instructions. Please note: The random draw of a pair is completely independent of the participants' decisions. Each pair has the same probability of being drawn. Since your decision can be actually implemented for real, you should think carefully about how you will decide in the experiment.

### 3.D.2.3 Information about Tuberculosis

What follows is important information that is relevant to the decisions you will later be asked to make. It concerns the illness tuberculosis and its possible treatment. Please read through all the information carefully.

#### What is Tuberculosis?

Tuberculosis – also called Phthisis or White Death – is an infectious disease, which is caused by bacteria. Roughly one-third of all humans are infected with the pathogen of Tuberculosis. Active Tuberculosis breaks out among 5 to 10% of all those infected. Tuberculosis is primarily airborne. This is also why quick treatment is necessary. Tuberculosis patients often suffer from very unspecific symptoms like fatigue, the feeling of weakness, lack of appetite, and weight loss. At an advanced stage of lung tuberculosis, the patient coughs up blood, leading to the so-called rush of blood. Without treatment, a person with Tuberculosis dies with a probability of 43%.

#### How prevalent is Tuberculosis?

In the year 2014, 6 million people have been recorded as falling ill with active Tuberculosis. Almost 1.5 million people die of Tuberculosis each year. This means more deaths due to Tuberculosis than due to HIV, malaria, or any other infectious disease.

#### Is tuberculosis curable?

According to the World Health Organization (WHO), the United Nations agency for international public health, “tuberculosis is preventable and curable”. Treatment takes place by taking antibiotics several times a week over a period of 6 months. It is important to take the medication consistently. Since 2000, an estimated 53 million lives have been saved through effective diagnosis and treatment of tuberculosis.

The success rate of treatment for a new infection is usually over 85%.

The preceding figures and information have been provided by the WHO and are freely available. [Click here for more details.](#)

#### Operation ASHA



**Figure 3.D.1.** Typical appearance of a tuberculosis patient.

*Operation ASHA* is a charity organization specialized since 2005 on treating Tuberculosis in disadvantaged communities. The work of *Operation ASHA* is based on the insight that the biggest obstacle for the treatment of Tuberculosis is the interruption of the necessary 6-month-long regular intake of medication.

For a successful treatment, the patient has to come to a medical facility twice a week – more than 60 times in total – to take the medication. Interruption or termination of the treatment is fatal because this strongly enhances the development of a drug-resistant form of Tuberculosis. This form of Tuberculosis is much more difficult to treat and almost always leads to death.

#### **The concept of *Operation ASHA***

To overcome this problem, *Operation ASHA* developed a concept that guarantees regular treatment through immediate spatial proximity to the patient. A possible non-adherence is additionally prevented by visiting the patient at home.

By now, *Operation ASHA* runs more than 360 treatment centers, almost all of which are located in the poorer regions of India. More than 60,000 sick persons have been identified and treated that way.

*Operation ASHA* is an internationally recognized organization, and its success has been covered by the New York Times, BBC, and Deutsche Welle, for example. The MIT and the University College London have already conducted research projects about the fight against Tuberculosis in cooperation with *Operation ASHA*. The treatment method employed by *Operation ASHA* is described by the World Health Organization (WHO) as “highly efficient and cost-effective”.

#### **The impact of a donation to *Operation ASHA***

It is now possible to save people from death by Tuberculosis by donating to *Operation ASHA*.

To save a person’s life means here to successfully cure a person with Tuberculosis, who otherwise would die because of the Tuberculosis. A donation of 350€ ensures that at least one human life can be expected to be saved. The information used to



**Figure 3.D.2.** A worker from Operation ASHA delivers medication to a tuberculosis patient

calculate the donation amount is obtained from public statements from the World Health Organization (WHO), peer-reviewed research studies, Indian Government statistics, and published figures from *Operation ASHA*.

In the calculation, information was conservatively interpreted, or a pessimistic number was used so that the donation amount of 350€ is in the case of doubt higher than the actual costs to save a human life. In addition, in the calculation of the treatment success rate of *Operation ASHA*, the mortality rate for alternative treatment by the state tuberculosis program in India and the different detection rates for new cases of Tuberculosis are included.

In the context of this study, an agreement made with *Operation ASHA* will ensure that 100% of the donation will be used exclusively for the diagnosis and treatment of tuberculosis patients. This means that every Euro of the donation amount goes directly to saving human lives, and no other costs will be covered. Based on a very high number of cases, the contribution of a donation of 350€ can be simplified visualized as follows:

With a donation of 350€ 5 additional patients infected with Tuberculosis can be treated through *Operation ASHA*.

If these 5 persons are not treated through *Operation ASHA*, it is expected that one patient will die.

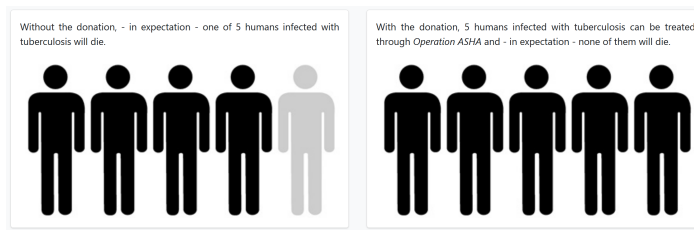
If, through the donation of 350€ all 5 patients are treated, it is expected that no patient will die.

Based on this experience, this means that through a donation of 350€ the life of a human will be saved. The relationship between a donation of 350€ and the saving of a human is illustrated in the following graphic: [Figure 3.D.3 here]

### Summary

Tuberculosis is a worldwide common bacterial infectious disease. The success rate of medical treatment of a new disease is very high. Nevertheless, close to 1.5 million people die every year from Tuberculosis. The biggest obstacle to the curing of Tuberculosis is the potential stopping of continuous treatment with antibiotics.





**Figure 3.D.3.** Relationship between the donation and the saving of a life

The concept of *Operation ASHA* is therefore based on the immediate proximity to the patient as well as the control and recording of the regular intake of medication. Through a donation of 350€ to *Operation ASHA*, a life will be saved.

#### **How is the donation connected to the saving of a life?**

The donation of 350€ already accounts for the fact that someone inflicted with the illness could have survived without treatment by *Operation ASHA*; i.e., instead of through *Operation ASHA*, they could have received treatment through other actors (such as the public health system). The amount is, therefore, sufficient for the diagnosis and complete treatment of multiple sufferers.

#### **What does it mean to “save a life”?**

To save a life means here the successful curing of a person suffering from Tuberculosis, who otherwise would die because of Tuberculosis. In particular, this means that the amount of the donation is sufficient to identify and cure so many tuberculosis patients that there is at least one person among them who otherwise could be anticipated to have died of Tuberculosis.

#### **Note**

Click on “Next” once you have finished carefully reading through the information. You can only click on the button “Next” once you have spent at least 5 minutes on the tabs of this page.

### **3.D.3 Treatment *DE Low Image***

#### **3.D.3.1 Your decision**

You will soon have the possibility to choose between two options: option A and option B. Both options are as follows:

##### **Option A**

Option A: I save a human life. By choosing option A, you save a human life. Specifically, by choosing option A, you instigate a donation of 350,00€ that will ensure that at least one person is saved from death by Tuberculosis, just as described before. If you choose option A, you will not receive an additional payment.

##### **Option B**

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment. In addition, the absence of your donation will cause the death of a human life.

#### **Additional payment**

Before today's experiment, various amounts between 0€ and 200€ were taken into account for the amount of money you will receive when choosing option B, from which 100€ was selected. Your partner sees exactly the same options as you and makes a decision just like you. So your partner also decides between option A (saving a human life) and option B (keeping 100€ to himself).

#### **Summary**

You will decide on the next page of the screen by choosing between option A and option B. By choosing option A, you save a human life. By choosing option B, you receive an additional payment of 100€. On the next page, you will receive details about the payment procedure.

### **3.D.3.2 Further procedure**

After you confirmed your decision on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. Thereafter, some additional questions will follow. After you have answered these questions, you will get the screenshot with the decision of your partner displayed, and your partner will get the screenshot with your decision. You will not receive any further information about your partner, and your partner will not receive any further information about you.

After you received the screenshot, please remain seated until you are called with your cabin number. Then you can go into the adjacent room to pick up your compensation for today's experiment. You will be called one by one so that there is no contact with other participants of the experiment.

#### **Who will be in the adjacent room?**

In the adjacent room, you will find the participant who was randomly selected from all participants at the start of the study.

#### **How do you receive your payment?**

This participant will give you a sealed envelope with your payment. The selected participant has already received the envelope sealed. Since this participant is only responsible for the payment, this participant has not completed the study and therefore has no knowledge of the decisions to be made. Therefore, this participant does not know what you chose, how you decided, or how much money you received, exactly as explained at the beginning of the study. By handing in your note with your cabin number, you will receive the envelope intended for you.

#### **Data protection**

The subsequent analysis of all data is carried out anonymously so that your decision can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decision is only used for anonymized data analysis.

**Please note:**

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to [experimente@briq-institute.org](mailto:experimente@briq-institute.org) after the study.

### **3.D.4 Treatment *DE High Image***

#### **3.D.4.1 Your decision**

You will soon have the possibility to choose between two options: option A and option B. Both options are as follows:

**Option A**

Option A: I save a human life. By choosing option A, you save a human life. Specifically, by choosing option A, you instigate a donation of 350,00€ that will ensure that at least one person is saved from death by Tuberculosis, just as described before. If you choose option A, you will not receive an additional payment.

**Option B**

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment. In addition, the absence of your donation will cause the death of a human life.

**Additional payment**

Before today's experiment, various amounts between 0€ and 200€ were taken into account for the amount of money you will receive when choosing option B, from which 100€ was selected. Your partner sees exactly the same options as you and makes a decision just like you. So your partner also decides between option A (saving a human life) and option B (keeping 100€ to himself).

**Summary**

You will decide on the next page of the screen by choosing between option A and option B. By choosing option A, you save a human life. By choosing option B, you receive an additional payment of 100€. On the next page, you will receive details about the payment procedure.

#### **3.D.4.2 Further procedure**

After you confirmed your decision on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. Thereafter, some additional questions will follow. After

you have answered these questions, you will get the screenshot with the decision of your partner displayed, and your partner will get the screenshot with your decision. You will not receive any further information about your partner, and your partner will not receive any further information about you.

After you received the screenshot, please remain seated until you are called with your cabin number. Then you can go into the adjacent room to pick up your compensation for today's experiment. You will be called one by one so that there is no contact with other participants of the experiment.

**Who will be in the adjacent room?**

In the adjacent room, you will find a person who will make the payment. As mentioned before, there will also be a committee consisting of three other persons in the adjacent room. These three persons are students of the University of Bonn and were specially selected for this task.

**What is the task of the committee?**

The task of these three persons is to assess the decision you and your partner have taken. Specifically, the assessment is about how moral your behavior and the behavior of your partner is. Apart from the assessment, the three persons will not interact with you (or with your partner) in any way, and the rating will not influence at all the consequences of your decisions or your payment.

**What information does the committee receive?**

In order for the committee to make the assessment, the two screenshots of the decision of you and your partner are projected side by side onto the room's wall using a projector, visibly for all persons in the adjacent room. You are identified by (and only by) your cabin number. For better identification, based on your decision and that of your partner, you must also say the following two sentences aloud. The first sentence refers to your decision, the second sentence to your partner's decision.

Sentence 1 in case you chose option A: "I have decided not to take 100€ as payment for myself and instead decided to save a human life." Sentence 1 in case you chose option B: "I have decided to take 100€ as payment for myself instead of saving a human life."

Sentence 2 in case your partner has chosen option A: "My partner has decided not to take 100€ as payment for himself and instead decided to save a human life." Sentence 2 in case your partner has chosen option B: "My partner has decided to take 100€ as payment for himself instead of saving a human life."

In summary, you have to say two sentences, and the following information will be visible to everyone in the room:

- The decision you and your partner faced.
- Which option you and your partner have chosen. This means it is displayed whether you chose to save a human life or the additional payment of 100€ and whether your partner chose to save a human life or the additional payment of 100€.

**How does the assessment work?**

The committee will assess your decision using a scale. For this, each one of the three persons of the committee selects one of the following five values:

1 - very immoral 2 - rather immoral 3 - neutral 4 - rather moral 5 - very moral.

The three persons of the committee will submit an assessment for your decision as well as the decision of your partner.

**How do you receive your payment?**

After the committee has assessed the decisions, the committee will give you the assessments of both your decision and the decision of your partner, and the person responsible for the payments will give you your payment. In the event that you have decided to donate, you will receive a donation confirmation.

**Data protection**

The subsequent analysis of all data is carried out anonymously so that your decision can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decision is only used for anonymized data analysis.

**Please note:**

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to [experimente@briq-institute.org](mailto:experimente@briq-institute.org) after the study.

**3.D.5 Treatment *MPL Low Image*****3.D.5.1 Your decision**

You will soon have the possibility to choose in 21 decision scenarios between two options: option A and option B. Both options are as follows:

**Option A**

Option A: I save a human life. By choosing option A, you save a human life. Specifically, by choosing option A, you instigate a donation of 350,00€ that will ensure that at least one person is saved from death by Tuberculosis, just as described before. If you choose option A, you will not receive an additional payment.

**Option B**

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment. In addition, the absence of your donation will cause the death of a human life.

**Additional payment**

The additional payment that you receive from choosing option B varies in each of the 21 decision scenarios. In the first scenario, the payment is 0€ and then increases

incrementally in each scenario thereafter by 10€ up to a payment of 200€. Therefore, the decision scenarios look as follows:

#### **Automatic completion help**

So that you do not need to click as much, we have activated an automatic completion help that automatically fills out the fields for you. As soon as you choose an amount from option B, we assume that you would choose all respectively higher payments from option B. Likewise, when you choose option A in a row, we assume that you would choose option A over all respectively lower payments from option B.

Please note: You can always change your decisions until you clicked on “Confirm Decisions”. Therefore, only click on that button when you are certain how you want to decide.

#### **Payment**

After you have selected one of the two options for each of the 21 decision scenarios, one of them will be randomly selected for real implementation. This means that the consequences of this decision will be implemented exactly as stated. Each of the 21 scenarios has the same probability of being selected. Therefore, since each of your decisions is potentially relevant, it is in your interest to decide in every scenario as if that decision is being implemented for real.

Your partner sees exactly the same 21 decision scenarios as you and, like you, makes a decision for every scenario. Furthermore, for you and your partner, the same decision scenario will be randomly selected. Thus, both your decision and the decision of your partner for this scenario will be implemented.

The following examples elaborate on this. Assume that decision scenario 2 is randomly selected, and you chose option A, while your partner chose option B. Then you save a human life and your partner will receive 10€. If, on the contrary, both of you choose option B, then both of you will receive 10€. If both of you choose option A, then two human lives will be saved. Assuming that decision scenario 21 is randomly selected, and you chose option B, while your partner chose option A. Then, you will receive 200€ and your partner saves a human life. If, however, both of you chose option B, then both of you will receive 200€. If both of you chose option A, then two human lives will be saved.

#### **Summary**

On the page after next, you will make a decision for 21 scenarios, and in each decision, you can choose between option A and option B. By choosing option A, you save a human life, whereas by choosing option B, you receive an additional payment. After you have reached all of your decisions, one of the 21 scenarios will be chosen randomly for you and your assigned partner. Thereafter, the consequences of the chosen decision are realized, i.e., in the case that you chose option A under this scenario, a donation will be made towards the saving of a human life and in the case that you chose option B, you receive the respective amount from the selected scenario. The same applies to your partner. On the next page, you will receive details about the payment procedure.

### 3.D.5.2 Further procedure

After you confirmed your decisions on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. Thereafter, some additional questions will follow. After you have answered these questions, you will get the screenshot with the decisions of your partner displayed, and your partner will get the screenshot with your decisions. You will not receive any further information about your partner, and your partner will not receive any further information about you.

After you received the screenshot, please remain seated until you are called with your cabin number. Then you can go into the adjacent room to pick up your compensation for today's experiment. You will be called one by one so that there is no contact with other participants of the experiment.

#### **Who will be in the adjacent room?**

In the adjacent room, you will find the participant who was randomly selected from all participants at the start of the study.

#### **How do you receive your payment?**

This participant will give you a sealed envelope with your payment. The selected participant has already received the envelope sealed. Since this participant is only responsible for the payment, this participant has not completed the study and therefore has no knowledge of the decisions to be made. Therefore, this participant does not know what you chose, how you decided, or how much money you received, exactly as explained at the beginning of the study. By handing in your note with your cabin number, you will receive the envelope intended for you.

#### **Data protection**

The subsequent analysis of all data is carried out anonymously so that your decisions can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decisions is only used for anonymized data analysis.

#### **Please note:**

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to [experimente@briq-institute.org](mailto:experimente@briq-institute.org) after the study.

### 3.D.6 Treatment *MPL High Image*

#### 3.D.6.1 Your decision

You will soon have the possibility to choose in 21 decision scenarios between two options: option A and option B. Both options are as follows:

#### **Option A**

Option A: I save a human life. By choosing option A, you save a human life. Specifically, by choosing option A, you instigate a donation of 350,00€ that will ensure that at least one person is saved from death by Tuberculosis, just as described before. If you choose option A, you will not receive an additional payment.

#### **Option B**

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment. In addition, the absence of your donation will cause the death of a human life.

#### **Additional payment**

The additional payment that you receive from choosing option B varies in each of the 21 decision scenarios. In the first scenario, the payment is 0€ and then increases incrementally in each scenario thereafter by 10€ up to a payment of 200€. Therefore, the decision scenarios look as follows:

#### **Automatic completion help**

So that you do not need to click as much, we have activated an automatic completion help that automatically fills out the fields for you. As soon as you choose an amount from option B, we assume that you would choose all respectively higher payments from option B. Likewise, when you choose option A in a row, we assume that you would choose option A over all respectively lower payments from option B.

Please note: You can always change your decisions until you clicked on “Confirm Decisions”. Therefore, only click on that button when you are certain how you want to decide.

#### **Payment**

After you have selected one of the two options for each of the 21 decision scenarios, one of them will be randomly selected for real implementation. This means that the consequences of this decision will be implemented exactly as stated. Each of the 21 scenarios has the same probability of being selected. Therefore, since each of your decisions is potentially relevant, it is in your interest to decide in every scenario as if that decision is being implemented for real.

Your partner sees exactly the same 21 decision scenarios as you and, like you, makes a decision for every scenario. Furthermore, for you and your partner, the same decision scenario will be randomly selected. Thus, both your decision and the decision of your partner for this scenario will be implemented.

The following examples elaborate on this. Assume that decision scenario 2 is randomly selected, and you chose option A, while your partner chose option B. Then you save a human life and your partner will receive 10€. If, on the contrary, both of you choose option B, then both of you will receive 10€. If both of you choose option A, then two human lives will be saved. Assuming that decision scenario 21 is randomly selected, and you chose option B, while your partner chose option A. Then, you will receive 200€ and your partner saves a human life. If, however, both of you chose option B, then both of you will receive 200€. If both of you chose option A, then two human lives will be saved.



### **Summary**

On the page after next, you will make a decision for 21 scenarios, and in each decision, you can choose between option A and option B. By choosing option A, you save a human life, whereas by choosing option B, you receive an additional payment. After you have reached all of your decisions, one of the 21 scenarios will be chosen randomly for you and your assigned partner. Thereafter, the consequences of the chosen decision are realized, i.e., in the case that you chose option A under this scenario, a donation will be made towards the saving of a human life and in the case that you chose option B, you receive the respective amount from the selected scenario. The same applies to your partner. On the next page, you will receive details about the payment procedure.

#### **3.D.6.2 Further procedure**

After you confirmed your decisions on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. Thereafter, some additional questions will follow. After you have answered these questions, you will get the screenshot with the decisions of your partner displayed, and your partner will get the screenshot with your decisions. You will not receive any further information about your partner, and your partner will not receive any further information about you.

After you received the screenshot, please remain seated until you are called with your cabin number. Then you can go into the adjacent room to pick up your compensation for today's experiment. You will be called one by one so that there is no contact with other participants of the experiment.

#### **Who will be in the adjacent room?**

In the adjacent room, you will find a person who will make the payment. As mentioned before, there will also be a committee consisting of three other persons in the adjacent room. These three persons are students of the University of Bonn and were specially selected for this task.

#### **What is the task of the committee?**

The task of these three persons is to assess the decisions you and your partner have taken. Specifically, the assessment is about how moral your behavior and the behavior of your partner is. Apart from the assessment, the three persons will not interact with you (or with your partner) in any way, and the rating will not influence at all the consequences of your decisions or your payment.

#### **What information does the committee receive?**

In order for the committee to make the assessment, the two screenshots of the decisions of you and your partner are projected side by side onto the room's wall using a projector, visibly for all persons in the adjacent room. You are identified by (and only by) your cabin number. For better identification, based on your decisions and the decisions of your partner, you must also say the following two sentences aloud.

The first sentence refers to your decisions, the second sentence to your partner's decisions.

Sentence 1: "I have decided from a payment of X€ onwards to take the payment for myself instead of saving human life."

Sentence 2: "My partner has decided from a payment of X€ onwards to take the payment for himself instead of saving human life."

The payment X denotes the amount of money for which you switched from option A to option B for the first time. If you have not decided to take the money in any decision-making situation, i.e., have not switched, you have to say the following as the first sentence:

Sentence 1: "I have decided for no amount to take the payment for myself instead of saving human life."

Similarly, if your partner has not decided to take the money in any decision-making situation, you must say the following second sentence:

Sentence 2: "My partner has decided for no amount to take the payment for himself instead of saving human life."

In summary, you have to say two sentences, and the following information will be visible to everyone in the room:

- The complete list of all 21 decision scenarios described before.
- How you and your partner have chosen in each of these scenarios. This means that for each payment amount, one can see whether you have decided to save a human life or the additional payment and whether your partner has decided to save a human life or the additional payment.

#### **How does the assessment work?**

The committee will assess your decisions using a scale. For this, each one of the three persons of the committee selects one of the following five values:

1 - very immoral 2 - rather immoral 3 - neutral 4 - rather moral 5 - very moral.

The three persons of the committee will submit an assessment for your decisions as well as the decisions of your partner.

#### **How do you receive your payment?**

After the committee has assessed the decisions, the committee will give you the assessments of both your decisions and the decisions of your partner, and the person responsible for the payments will give you your payment. In the event that you have decided to donate, you will receive a donation confirmation.

#### **Data protection**

The subsequent analysis of all data is carried out anonymously so that your decisions can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decisions is only used for anonymized data analysis.

#### **Please note:**

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This funda-

mentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to [experimente@briq-institute.org](mailto:experimente@briq-institute.org) after the study.

### **3.D.7 Robustness experiment**

#### **3.D.8 Introduction**

All instructions were displayed on the subjects' screens. The following introduction was the same for both treatments of the robustness experiment.

##### **3.D.8.1 Welcome to the study**

Welcome, and thank you for your interest in today's study!

Please note that you can take part in this study only once. Furthermore, you may only participate if you have registered for this study in our participation database ([experimente.bonneconlab.uni-bonn.de](http://experimente.bonneconlab.uni-bonn.de)).

For your full participation, you will receive a fixed payment of 3€. In this study, you will make decisions on the computer. Depending on how you choose, you can earn additional money. After the study, you will receive all payments, i.e. both the remuneration for your participation and any additional payments based on your decisions, by bank transfer.

On the next screens, you will receive specific information about participation in this study. To proceed, click "Next".

##### **3.D.8.2 Your partner**

As part of this experiment, a partner has been assigned to you. This partner is a participant in today's experiment, just like you. He or she was randomly assigned to you and will receive the same instructions as you.

In today's experiment, you and your partner will both receive the exact same information and subsequently face the exact same decisions. These decisions have certain consequences, which will be described in detail later.

##### **Payment**

At the end of today's experiment, one pair will be randomly drawn from every 24 participants in the experiment. Only the decisions of this pair will be implemented, as described in the instructions. Please note: The random draw of a pair is completely independent of the participants' decisions. Each pair has the same probability of being drawn. Since your decision can be actually implemented for real, you should think carefully about how you will decide in the experiment.

### 3.D.8.3 Information

What follows is some information that is relevant to the decisions you will later be asked to make. It concerns the official shop of the University of Bonn.

The Campus Store Uni-Bonn is the official shop of the University of Bonn. Here you can purchase various products such as T-shirts, sweatshirts or mugs with the logo and design of the Uni-Bonn.

The Uni-shop is located at the information point in the main building. There is also an online shop, which can be reached via the website: <https://www.campusstore-unibonn.de>. The online shop dispatches all goods within 2-3 working days.

#### Voucher

The next decisions will concern a voucher for the Uni-shop, namely a voucher worth 35€. The voucher can only be redeemed in the online shop and cannot be converted into money.

### 3.D.9 Treatment *DE No-Image*

#### 3.D.9.1 Your decision

You will soon have the possibility to choose between two options: option A and option B. Both options are as follows:

##### Option A

Option A: I choose the voucher. By choosing option A, you will receive the voucher for the Uni-shop. Specifically, option A allows you to receive a voucher worth 35€, which you can redeem in the Uni-shop (and only there). If you choose option A, you will not receive an additional payment.

##### Option B

Option B: I choose 10€ as payment for myself. By choosing option B, you will receive an additional payment of 10€ at the end of the experiment, but you will not receive the voucher.

##### Additional payment

Before today's experiment, various amounts between 0€ and 20€ were taken into account for the amount of money you will receive when choosing option B, from which 10€ was selected. Your partner sees exactly the same options as you and makes a decision just like you. So your partner also decides between option A (voucher) and option B (keeping 10€ to himself/herself).

##### Summary

You will decide on the next page of the screen by choosing between option A and option B. By choosing option A, you receive a voucher. By choosing option B, you receive an additional payment of 10€. On the next page, you will find details about the payment procedure.

### 3.D.9.2 Further procedure

After you confirmed your decision on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. At the end of today's experiment, you will get the screenshot with the decision of your partner displayed, and your partner will get the screenshot with your decision. You will not receive any further information about your partner, and your partner will not receive any further information about you.

#### Data protection

The subsequent analysis of all data is carried out anonymously so that your decision can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decision is only used for anonymized data analysis.

#### Please note:

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to [experiment@briq-institute.org](mailto:experiment@briq-institute.org) after the study.

### 3.D.10 Treatment *MPL No-Image*

#### 3.D.10.1 Your decisions

You will soon have the possibility to choose in 21 decision scenarios between two options: option A and option B. Both options are as follows:

##### Option A

Option A: I choose the voucher. By choosing option A, you will receive the voucher for the Uni-shop. Specifically, option A allows you to receive a voucher worth 35€, which you can redeem in the Uni-shop (and only there). If you choose option A, you will not receive an additional payment.

##### Option B

Option B: I choose X€ as payment for myself. By choosing option B, you will receive an additional payment at the end of the experiment, but you will not receive the voucher.

#### Additional payment

The additional payment that you receive from choosing option B varies in each of the 21 decision scenarios. In the first scenario, the payment is 0€ and then increases incrementally in each scenario thereafter by 1€, up to a payment of 20€. Therefore, the decision scenarios look as follows:

#### Automatic completion help

So that you do not need to click as much, we have activated an automatic completion help that automatically fills out the fields for you. As soon as you choose an amount

from option B, we assume that you would choose all respectively higher payments from option B. Likewise, when you choose option A in a row, we assume that you would choose option A over all respectively lower payments from option B.

Please note: You can always change your decisions until you clicked on “Confirm Decisions”. Therefore, click on that button only when you are certain how you want to decide.

### **Payment**

After you have selected one of the two options for each of the 21 decision scenarios, one of them will be randomly selected for real implementation. This means that the consequences of this decision will be implemented exactly as stated. Each of the 21 scenarios has the same probability of being selected. Therefore, since each of your decisions is potentially relevant, it is in your interest to decide in every scenario as if that decision is being implemented for real.

Your partner sees exactly the same 21 decision scenarios as you and, like you, makes a decision for every scenario. Furthermore, for you and your partner, the same decision scenario will be randomly selected. Thus, both your decision and the decision of your partner for this scenario will be implemented.

The following examples elaborate on this. Assume that decision scenario 2 is randomly selected, and you chose option A, while your partner chose option B. Then you will receive the voucher and your partner will receive 1€. If, on the contrary, both of you chose option B, then both of you will receive 1€. If both of you chose option A, then you and your partner will each receive the voucher. Assuming that decision scenario 21 is randomly selected, and you chose option B while your partner chose option A, then you will receive 20€, and your partner will receive the voucher. If, however, both of you chose option B, then both of you will receive 20€. If both of you chose option A, then you and your partner will each receive the voucher, etc.

### **Summary**

On the page after next, you will make a decision for 21 scenarios, and in each decision, you can choose between option A and option B. By choosing option A, you receive a voucher, whereas by choosing option B, you receive an additional payment. After you have reached all of your decisions, one of the 21 scenarios will be chosen randomly for you and your assigned partner. Thereafter, the consequences of the chosen decision are realized, i.e., in the case that you chose option A under this scenario, you will be given the voucher and in the case that you chose option B, you will receive the respective amount from the selected scenario. The same applies to your partner. On the next page, you will receive details about the payment procedure.

#### **3.D.10.2 Further procedure**

After you confirmed your decision on the decision screen, a screenshot will be taken from this decision screen. From the decision screen of your partner, a screenshot will be taken in the same way. At the end of today’s experiment, you will get the

screenshot with the decision of your partner displayed, and your partner will get the screenshot with your decision. You will not receive any further information about your partner, and your partner will not receive any further information about you.

**Data protection**

The subsequent analysis of all data is carried out anonymously so that your decisions can never be linked to your person. Your anonymity is therefore always guaranteed, and the information about your decisions is only used for anonymized data analysis.

**Please note:**

This is not a thought experiment: All information given in these instructions is true. In particular, all actions are performed exactly as they are described. This fundamentally applies to all studies of the Bonn Laboratory for Experimental Economic Research, as well as to this study.

If you still have separate questions, you may send them to [experiment@briq-institute.org](mailto:experiment@briq-institute.org) after the study.

## References

- Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and E. Elisabet Rutström.** 2006. "Elicitation Using Multiple Price List Formats." *Experimental Economics* 9 (4): 383–405. [170]
- Andreoni, James.** 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy* 97 (6): 1447–58. [154]
- Andreoni, James.** 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow." *Economic Journal* 100 (401): 464–77. [154]
- Ariely, Dan, Anat Bracha, and Stephan Meier.** 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review* 99 (1): 544–55. [153]
- Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack.** 2014. "No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery." *Journal of Public Economics* 120: 1–17. [153]
- Baillon, Aurelien, Yoram Halevy, and Chen Li.** 2022. "Randomize at Your Own Risk: On the Observability of Ambiguity Aversion." *Econometrica* 90 (3): 1085–107. [154]
- Barmettler, Franziska, Ernst Fehr, and Christian Zehnder.** 2012. "Big Experimenter Is Watching You! Anonymity and Prosocial Behavior in the Laboratory." *Games and Economic Behavior* 75 (1): 17–34. [161]
- Bartling, Björn, Vanessa Valero, Roberto Weber, and Yao Lan.** 2022. "Public Discourse and Socially Responsible Market Behavior." *Working Paper*. [153]
- Bénabou, Roland, Armin Falk, and Luca Henkel.** 2022. "Ends versus Means: Kantians, Utilitarians and Moral Decisions." *Working Paper*. [154]
- Bénabou, Roland, and Jean Tirole.** 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78. [153]
- Bénabou, Roland, and Jean Tirole.** 2011a. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2): 805–55. [153]
- Bénabou, Roland, and Jean Tirole.** 2011b. "Laws and Norms." *NBER Working Paper* 17579. [153]
- Berry, James, Greg Fischer, and Raymond Guiteras.** 2020. "Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana." *Journal of Political Economy* 128 (4): 1436–73. [154]
- Birnbaum, Michael H.** 1992. "Violations of Monotonicity and Contextual Effects in Choice-Based Certainty Equivalents." *Psychological Science* 3 (5): 310–15. [170]
- Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch.** 2014. "Hroot: Hamburg Registration and Organization Online Tool." *European Economic Review* 71: 117–20. [161]
- Bos, Olivier, and Martin Pollrich.** 2020. "Optimal Auctions with Signaling Bidders." *Working Paper*. [153]
- Bos, Olivier, and Tom Truys.** 2022. "Entry in First-Price Auctions with Signaling." *Working Paper*. [153]
- Brandts, Jordi, and Gary Charness.** 2011. "The Strategy versus the Direct-Response Method: A First Survey of Experimental Comparisons." *Experimental Economics* 14 (3): 375–98. [154]
- Brock, J. Michelle, Andreas Lange, and Erkut Y. Ozbay.** 2013. "Dictating the Risk: Experimental Evidence on Giving in Risky Environments." *American Economic Review* 103 (1): 415–37. [154]
- Bursztyn, Leonardo, and Robert Jensen.** 2017. "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure." *Annual Review of Economics* 9: 131–53. [151]
- Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky.** 2022. "Measuring the Welfare Effects of Shame and Pride." *American Economic Review* 112 (1): 122–68. [164]



- Charness, Gary, Uri Gneezy, and Brianna Halladay.** 2016. "Experimental Methods: Pay One or Pay All." *Journal of Economic Behavior and Organization* 131: 141–50. [154]
- Charness, Gary, Uri Gneezy, and Alex Imas.** 2013. "Experimental Methods: Eliciting Risk Preferences." *Journal of Economic Behavior and Organization* 87: 43–51. [154]
- Chen, Daniel L., and Martin Schonger.** 2016. "A Theory of Experiments: Invariance of Equilibrium to the Strategy Method of Elicitation and Implications for Social Preferences." *TSE Working Paper, no. 16-724*. [154]
- Chen, Daniel L., and Martin Schonger.** 2022. "Social Preferences or Sacred Values? Theory and Evidence of Deontological Motivations." *Science Advances* 8 (19): eabb3925. [154, 164]
- Chen, Daniel L., Martin Schonger, and Chris Wickens.** 2016. "oTree-An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9: 88–97. [161]
- Cohen, Jonathan, Keith Marzilli Ericson, David Laibson, and John Myles White.** 2020. "Measuring Time Preferences." *Journal of Economic Literature* 58 (2): 299–347. [154]
- Cole, Shawn, A Nilesh Fernando, Daniel Stein, and Jeremy Tobacman.** 2020. "Field Comparisons of Incentive-Compatible Preference Elicitation Techniques." *Journal of Economic Behavior and Organization* 172: 33–56. [154]
- Cox, James C., Vjollca Sadiraj, and Ulrich Schmidt.** 2015. "Paradoxes and Mechanisms for Choice under Risk." *Experimental Economics* 18 (2): 215–50. [154]
- Cubitt, Robin P.** 1998. "On the Validity of the Random Lottery Incentive System." *Experimental Economics* 1: 115–31. [170]
- Dana, Jason, Roberto A. Weber, and Jason Xi Kuang.** 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33 (1): 67–80. [153]
- DellaVigna, Stefano, John A. List, and Ulrike Malmendier.** 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127 (1): 1–56. [153]
- Exley, Christine L.** 2016. "Excusing Selfishness in Charitable Giving: The Role of Risk." *Review of Economic Studies* 83 (2): 587–628. [153]
- Falk, Armin, and Thomas Graeber.** 2020. "Delayed Negative Effects of Prosocial Spending on Happiness." *Proceedings of the National Academy of Sciences* 117 (12): 6463–68. [159]
- Falk, Armin, Thomas Neuber, and Nora Szech.** 2020. "Diffusion of Being Pivotal and Immoral Outcomes." *Review of Economic Studies* 87 (5): 2205–29. [153, 154]
- Feddersen, Timothy, Sean Gailmard, and Alvaro Sandroni.** 2009. "Moral Bias in Large Elections: Theory and Experimental Evidence." *American Political Science Review* 103 (2): 175–92. [153]
- Giovannoni, Francesco, and Miltiadis Makris.** 2014. "Reputational Bidding." *International Economic Review* 55 (3): 693–710. [153]
- Gneezy, Uri, Elizabeth A. Keenan, and Ayelet Gneezy.** 2014. "Avoiding Overhead Aversion in Charity." *Science* 346 (6209): 632–35. [154]
- Goeree, Jacob K.** 2003. "Bidding for the Future: Signaling in Auctions with an Aftermarket." *Journal of Economic Theory* 108 (2): 345–64. [153]
- Goeree, Jacob K., Charles A. Holt, and Susan K. Laury.** 2002. "Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior." *Journal of Public Economics* 83 (2): 255–76. [154]
- Grossman, Zachary.** 2015. "Self-Signaling and Social-Signaling in Giving." *Journal of Economic Behavior and Organization* 117: 26–39. [153]
- Grossman, Zachary, and Joël J. van der Weele.** 2017. "Self-Image and Willful Ignorance in Social Decisions." *Journal of the European Economic Association* 15 (1): 173–217. [153]

- Hey, John D, and Jinkwon Lee.** 2005. "Do Subjects Separate (or Are They Sophisticated)?" *Experimental Economics* 8 (3): 233–66. [170]
- Kolappan, C., R. Subramani, V. Kumaraswami, T. Santha, and P. R. Narayanan.** 2008. "Excess Mortality and Risk Factors for Mortality among a Cohort of TB Patients from Rural South India." *International Journal of Tuberculosis and Lung Disease* 12 (1): 81–86. [160]
- Ledyard, John O.** 1995. "Public Goods: A Survey of Experimental Research." In *The Handbook of Experimental Economics*, edited by Alvin E. Roth and John H. Kagel, 1: 111–94. Princeton University Press. [154]
- Miller, Klaus M., Reto Hofstetter, Harley Krohmer, and Z. John Zhang.** 2011. "How Should Consumers' Willingness to Pay Be Measured? An Empirical Comparison of State-of-the-Art Approaches." *Journal of Marketing Research* 48 (1): 172–84. [154]
- Sandel, Michael J.** 2012. *What Money Can't Buy: The Moral Limits of Markets*. Farrar, Straus and Giroux. [164]
- Selten, Reinhard.** 1967. "Die Strategiemethode Zur Erforschung Des Eingeschränkt Rationalen Verhaltens Im Rahmen Eines Oligopol-experiments." In *Beiträge Zur Experimentellen Wirtschaftsforschung*, edited by H. Sauermann, 136–68. Tübingen: Mohr. [154]
- Simonson, Itamar.** 1989. "Choice Based on Reasons: The Case of Attraction and Compromise Effects." *Journal of Consumer Research* 16 (2): 158. [170]
- Starmer, Chris, and Robert Sugden.** 1991. "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation." *American Economic Review* 81 (4): 971–78. [170]
- Straetemans, Masja, Philippe Glaziou, Ana L. Bierrenbach, Charalambos Sismanidis, and Marieke J. van der Werf.** 2011. "Assessing Tuberculosis Case Fatality Ratio: A Meta-Analysis." *PLoS ONE* 6 (6). [160]
- Tiemersma, Edine W., Marieke J. van der Werf, Martien W. Borgdorff, Brian G. Williams, and Nico J.D. Nagelkerke.** 2011. "Natural History of Tuberculosis: Duration and Fatality of Untreated Pulmonary Tuberculosis in HIV Negative Patients: A Systematic Review." *PLoS ONE* 6 (4). [160]
- Van Leeuwen, Boris, and Ingela Alger.** 2021. "Estimating Social Preferences and Kantian Morality in Strategic Interactions." *Working Paper*. [154]

## Chapter 4

# Ends versus Means: Kantians, Utilitarians and Moral Decisions\*

*Joint with Roland Bénabou and Armin Falk*

*“Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end” (Kant, 1785).*

### 4.1 Introduction

When providing a public good or engaging in reciprocal behavior, individuals trade off the costs and benefits to themselves against those to others. The question of what constitutes moral behavior then has a simple answer, given by the degree of prosociality of the chosen action. The preferences and image motives underlying such tradeoffs are by now fairly well understood.

Much less studied is another important class of decisions, in which the question of what is the right thing to do elicits far less agreement. These are situations in which achieving some socially desirable *end* requires the use of *means* considered inherently objectionable. These include “sacrificial dilemmas” where helping or saving a greater number of people requires placing some in harm’s way (e.g., hostage situations, medical triage or vaccine priority), but also many less dramatic choices

\* **Acknowledgments:** We are grateful for valuable comments to Jean-Francois Bonnefon, Claudia Senik, Jean Tirole, and participants in seminars and conferences at TSE, briq, SciencesPo, PSE, and the University of Chicago. Egshiglen Batbayar, Erik Ortiz Covarrubias, and Luca Michels provided excellent research assistance. Funded by the German Research Foundation (DFG) under Germany’s Excellence Strategy - EXC 2126/1- 390838866. Support by the German Research Foundation (DFG) through CRC TR 224 (Project A01) is gratefully acknowledged.

**Ethics approval:** The study obtained ethics approval from Princeton University (No. 12513).

**Preregistration:** The study was preregistered at the AEA RCT Registry (<https://www.socialscienceregistry.org/trials/7714>).

such as industrial and product safety, or bargaining situations where a bribe must be paid, a threat made, a lie told, or deeply held values compromised.

Unlike in “self-versus-other” choices, the two main approaches dominating Western moral philosophy and psychology now come into conflict. Under the Consequentialist view, the propriety of an action should be judged solely by its consequences (Bentham, 1789; Mill, 1863; Sinnott-Armstrong, 2019). Economics, in particular, is predominantly utilitarian, even when consequences include emotions or social image. The consequences of an act are evaluated and traded off within a decision-maker’s preferences, and in particular the social-welfare functions assumed for normative analysis are almost universally consequentialist.

According to Deontological ethics, on the other hand, whether an action is permissible or not should be based only on its conformity with a series of rules, independent of their consequences for the situation at hand (Kant, 1785; Alexander and Moore, 2016). Many tradeoffs are then proscribed, as violating higher values such as life or human dignity. While rarely absolute in practice, such principles do place major restrictions on both individual choices (lying, sense of duty, refusals to “sell out”) and the organization of society (taboos, proscription of “repugnant markets”). Relatedly, in the public debate, self-proclaimed deontologists often criticise economists’ use of cost-benefit analysis (e.g., value of a statistical life) as fundamentally out of touch with most people’s moral intuitions.

**Research questions.** The aim of this paper is to extend the empirical study of moral behavior and preferences beyond standard prosociality (altruism, reciprocity, fairness), into the other dimension of consequentialism versus deontologism. We do this by exploring three main questions.

First, how prevalent in actual behaviors are these two moral principles? If a significant fraction of people knowingly choose non-consequentialist options, representing their preferences in the standard manner will lead to inaccurate positive predictions, and standard social welfare functions may be misleading guides to normative decisions.

Second, are these two guiding principles or intuitions stable personal “traits”, preferences similar to risk-aversion, impatience, or especially altruism? Moving from between-subject variation to individual-level consistency, we ask how correlated a person’s behavior is between different situations in which consequentialist vs. deontological reasoning prescribe opposite actions.

Third, how is this other dimension of morality related to standard social preferences? Put differently, are deontologists generally more (or less) benevolent than consequentialists? Accordingly, we examine to what extent decisions pitting the two moral principles against each other may be correlated with those pitting self-interest against concern for others’ welfare.

To address these questions, we design and confront subjects with a series of choice situations consisting of two main blocks: a first, novel one designed to iden-

tify consequentialist vs. deontological decision-making (and corresponding “types” if they exist) through *ends-versus-means* (EVM) dilemmas, and a second one consisting of standard tests of altruism and reciprocity involving *self-versus-other* (SVO) tradeoffs.

**Ends-versus-means decisions.** Our first block starts with the Trolley problem (Foot, 1967; Thompson, 1976), which asks whether it is permissible to sacrifice one life to save a larger number. Although our general methodology does not at all depend on the Trolley, addressing it is unavoidable, given how extensively it has dominated normative debates and empirical investigations of the topic. There are now large international surveys exploring different variants of it, aiming to provide guidance for decisions to be made by autonomous vehicles, robots, and other AI algorithms (e.g., “Moral Machine Project”, Awad et al., 2018). Despite its popularity, the Trolley problem has almost exclusively been posed as a *hypothetical* question, and cast in very abstract situations (another version being the organ-transplant dilemma). Our paper’s first contribution is thus to experimentally study its value as a guide to the real decisions and policy choices it is meant to represent, by implementing a version of it with actual, life-saving consequences. Does this make a difference?

Our principal interest, however, is in the prevalence (among individuals) and the consistency (for each of them) of consequentialist versus deontological behavior across a broader set of less momentous, more common choice settings. The paper’s second and main contribution is therefore to design a series of experimental games creating ends-versus-means tensions between these two moral principles, while leaving the decision maker’s material and social payoffs unaffected. We start from a baseline situation in which the two principles agree: if the choice is between directing a donation of 15 Euros to a recognized charity treating children suffering from cancer, or one of 2 Euros to some random other subject, the right thing to do is unambiguously the former, and 93% of our subjects indeed make that choice. We then modify the decision problem so that bringing about this preferred social consequence requires either: (i) *lying* to another subject, at their expense; (ii) *bribing* another subject, who controls greater resources; (iii) making a morally *repugnant* but entirely cheap-talk, anonymous, statement; (iv) refraining from “expressively” choosing a moral option when the choices of others have already made it *ineffective*, and sticking to principle will only cause a further social loss; (v) violating an explicitly stated *rule*, absent any enforcement or meaningful effects of disobedience.

**Self-versus-other decisions.** Complementing this main series of six choice situations, our second block consists of a rich set of familiar games measuring prosocial behavior, administered to the same subjects a week apart from the first block. These include a dictator game, with either giving to or taking from a charity; a trust game, both as first- and as second-mover, using the strategy method; a public-goods game; and a moral-luck game (rewarding either intentions or outcomes). Finally, at the end of this block we administer two questionnaires widely used in the moral-psychology

literature, namely the Oxford Utilitarian Scale (Kahane et al., 2018) and the Moral Foundations Questionnaire (Graham et al., 2011), as well the more recent Moral Universalism short module of Enke, Rodríguez-Padilla, and Zimmermann (2022); we also gather standard sociodemographic variables.

A key aspect of our design is that, throughout the twelve decision tasks, all aspects of behavior are measured *within subject*, allowing us to look for stable patterns (preference types) within each of the consequentialist/deontological and prosocial/selfish dimensions, as well as for potential correlations between the two. Our experiments were administered online, with 593 subjects (mostly students) completing the experiment. The procedures and main statistical hypotheses to be tested were preregistered.

**Results.** We report four main results. First, across the various ends-versus-means (EVM) games, a significant fraction of subjects choose non-consequential options that are in line with deontological principles. Numbers range from 20% in the bribing game to 44% in the repugnant-statement game. These revealed preferences clearly differ from those of “standard” economic agents. We employ extensive robustness checks to make sure that this type of behavior is not driven by limited attention or subjects’ misunderstanding of the decision situations.

Second, we deliver mixed news on the informativeness of the Trolley dilemma as a guide to actual decisions. On the one hand, implementing it with real (statistical) lives at stake does not change behavior, relative to the standard procedure involving only hypothetical lives. In both cases, about 25% of choices are deontological. On the other hand, Trolley choices predict close to no other behavior: across the eleven (EVM and SVO) other incentivized games, there is only a small correlation of 0.16 with behavior in the lying game, and no significant one with any of the other ten (and no correlation above 0.05).

Third, and more generally, we find very little evidence of individual-level consistency in deontological versus consequential decision-making: behaviors across the six games pitting one principle against the other are largely uncorrelated. Almost no subjects choose deontologically in five or more of the six main games, only 10% in four or more, and only 10% never, behaving as standard economic agents; about 27% make one such choice, and 50% either two or three. Thus, there appears to be no single preference type or parameter that would robustly predict choices across ends-versus-means dilemmas. Instead, which moral principle prevails for an individual appears to be highly context-dependent. This is in stark contrast to standard *social preferences*, which exhibit substantial within-subject consistency across the six self-versus-other games (SVO) games in the second block. For instance, dictator-game giving reliably predicts behavior in all five other SVO games.

Fourth, ends-versus-means decision-making is unrelated to general prosocial preferences. Put differently, acting deontologically versus consequentially in the

Trolley, or any of the other main games, is not predictive of whether a person is more or less prosocial.

Finally, beyond its specific findings, our paper develops a flexible paradigm to study Consequentialist vs. Deontological decision making, based on a new class of ends-versus-means choice tasks, fundamentally distinct from classical self-versus-other tradeoffs, and indeed revealing an essentially independent dimension of moral preferences.

#### 4.1.1 Related literature

We first contribute to the small but growing literature in economics that investigates consequentialistically and deontologically motivated moral behavior. Starting with Laffont (1975), a number of papers have conceptualized Kant's categorical imperative for decision-making in strategic situations (Roemer, 2010; Alger and Weibull, 2013, 2016; Alger, Weibull, and Lehmann, 2020). In these models, Kantian agents act (in part) as if their decision would cause all other players to make a similar choice, in line with the imperative.<sup>1</sup> Van Leeuwen and Alger (2021) find support for the existence of such a motive, mixed with standard Nash-consequentialist behavior, in the Prisoner's, Trust, and Ultimatum games. Closely related is the literature on ethical voters, who derive a fixed "duty" or "expressive" utility from casting a ballot for an alternative that is considered morally superior, even if it goes against their self-interest or if the probability of being pivotal would not justify the cost of voting. Feddersen and Sandroni (2006), Feddersen, Gailmard, and Sandroni (2009) and Feess, Kerzenmacher, and Timofeyev (2022) develop such models; the last two also conduct experiments that show, by varying the probability that a vote will be pivotal, that both consequentialist and deontological or expressive motives are at work among subjects.

Compared to these papers, we study consequentialistic and deontological behavior in individual choice situations without strategic considerations, nor even any self-interest motive. Chen and Schonger (2022) also focus on situations without strategic interactions. They model deontological decision-making as governed by lexicographic preferences, first over acts *per se*, and secondarily over consequences. In two experiments, they show that when the consequences of a donation decision become increasingly hypothetical, subjects are more likely to donate money, suggesting again a mix of deontological and consequential motives. Bénabou and Tirole (2011) and Bénabou et al. (2022) show how behaviors displaying Kantian-like "sacred values" and "taboos tradeoffs" arise, even for consequentialist agents, in a model of moral identity maintained through self signaling.

1. "Act only according to that maxim through which you can at the same time will that it become a universal law". (Kant, 1785).

We add to this literature by providing evidence on the prevalence and, importantly, the consistency (or lack thereof) of consequential and deontological behavior across a range of decision situations. Furthermore, in contrast to the previous literature, we purposefully abstract from preferences based on self-other considerations. Apart from rule breaking, our ends-versus-means decision situations were designed so that subjects' self-interest plays *no role*, as their choices do not influence their own payoffs. This feature also allows us to investigate, in a second stage, the relation between deontological/consequential motives and prosociality. By analyzing patterns both within and across the EVM and the SVO blocks, our paper also connects to a nascent literature that studies components underlying individuals' behavior across different choice situations. Falk et al. (2018), Dean and Ortoleva (2019), and Chapman et al. (2022) study the relationships between social, risk and time preferences, and Stango and Zinman (2022) the correlation patterns between different behavioral biases.

We also contribute to the large literature using the trolley problem to investigate underlying moral principles (Greene et al., 2001; Hauser et al., 2007; Bartels, 2008; Lanteri, Chelini, and Rizzello, 2008; Rai and Holyoak, 2010; Costa et al., 2014; Gawronski and Beer, 2017; Awad et al., 2020). These studies typically focus on moral judgments instead of behavior and use hypothetical situations and questionnaires, with two exceptions. Gold, Pulford, and Colman (2015) create a trolley-like incentivized situation in which subjects could divert meals among children living in an orphanage. They compare behavior with moral judgments, finding no difference. Bostyn, Sevenhant, and Roets (2018) design a situation in which subjects believe they are distributing electrical shocks among mice. The experiment involves deception, as no shocks are actually delivered. Comparing subjects' behavior in this (believed to be) "real" situation with what others say they would do when the same question is presented hypothetically, they find that the former case leads to significantly less deontological choices. In contrast to these two papers, our SAL-trolley paradigm involves actual human lives (albeit statistical ones), and is therefore closest to the classical version of the dilemma. We further add to the literature a comprehensive assessment of the predictive power of the trolley problem for behavior in related EVM dilemmas, as well as in SVO tradeoffs.

## 4.2 Experimental design

Our design consists of three main blocks. The first contains six experiments confronting subjects with "ends-versus-means" decisions. The second elicits standard prosocial preferences using well-established "self-versus-other" decision paradigms. The third block contains several questionnaires.



### 4.2.1 Ends-versus-means block

We designed six decision situations sharing the key feature that the consequences (ends) of the choice options create a moral conflict with the actions (means) required to bring them about, as shown in Table 4.1. A second aspect fundamentally distinguishing them from standard games studied in the literature on social preferences is that they do not contain any self-versus-other tradeoff.

**Table 4.1.** Ends-versus-means decision situations

Decision situation	Ends	Means
Trolley problem	Life of three people vs. one person	Sacrifice vs. spare one person
Lying game	Donation vs. money to other subject	Lie vs. tell truth to another subject
Bribe game	Donation vs. money to other subject	Bribe vs. not bribe another subject
Group decision game	Donation vs. money to other subject	Consequential vs. inconsequential voting
Statement choice	Donation vs. money to other subject	Make vs. not make repugnant statement
Rule-following task	No money vs. money to self	Follow vs. break rule

The first decision task is the trolley problem, the most studied situation where such a conflict arises. At stake in our version of the problem are human lives, and in order to save three people, one has to actively sacrifice the life of one person. The next four tasks implement EVM dilemmas that all build on a simple baseline choice between two options, Option A and Option B. Choosing Option A generates a donation of 15 Euros to a charity that supports children suffering from cancer. Choosing Option B instead increases the payoff of another (random) subject by 2 Euros. In this basic decision, Option A dominates Option B based on consequences, and no controversial means are needed to achieve the preferred end. Accordingly, deontological and consequentialist reasoning agree that Option A is the morally right choice, and indeed 93% of subjects in our experiment choose it when facing this decision.<sup>2</sup>

Having established a setting where the underlying moral prescriptions are largely undisputed, we extend the situation to four variants designed to induce ends-versus-means tradeoffs, for which deontological and consequentialist ethics make opposing prescriptions.

In the lying game, subjects must lie in order to trigger the 15 Euros donation. In the bribe game, they have to bribe another subject. In the statement-choice situation, they must make a morally repugnant cheap-talk, anonymous statement. In the group-donation game, the choices of others have already destroyed the donation, so

2. Furthermore, when asked, 91% of subjects agree that from a moral point of view, there is a right and a wrong decision for them in this situation; and of those, 98% name Option A as the right decision. Similarly, 95% think that other people consider Option A to be the morally right choice.

choosing that option is a pure “expressive” (and again anonymous) act that only reduces social surplus. Lastly, we complement this set with a rule-following task. Here, achieving positive consequences is only possible by disobeying a type of rule that is generally important in society, but meaningless in the concrete situation at hand.

We now describe each of these games and the respective ends-versus-means conflict in detail. Appendix Section 4.E contains the instructions given to subjects in the experiment.

#### 4.2.1.1 Trolley problem

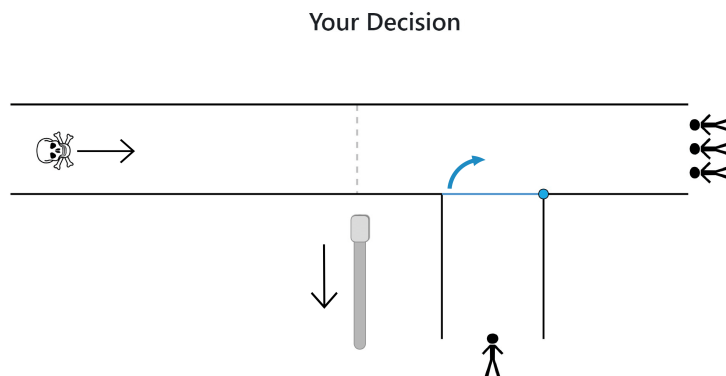
The Trolley is one of the most popular mental experiments in moral philosophy and applied ethics, extensively used to distinguish between modes of moral decision-making. It is, however, frequently criticized for representing an artificial and implausible situation, detached from any decisions people face in reality (e.g., Bauman et al., 2014). In our main block of decision tasks, we implement the trolley problem with real, meaningful consequences using the *Saving a Life* Paradigm (Falk and Graeber, 2020; Bénabou et al., 2022). Key features of our version are that choices involve the saving of actual human lives, and that they occur in a realistic context that nonetheless contains all relevant features of the Trolley problem.

**Saving a Life (SAL) paradigm.** We partnered with the non-profit organization *Operation ASHA*, which treats people in India suffering from tuberculosis using innovative methods and procedures. Tuberculosis is a highly lethal infectious disease if untreated, but curable with a high success rate if treated. An amount of 380 Euros allows the charity to treat five people suffering from tuberculosis, one of which on average would have otherwise died. Thus, each donation implemented saves one life in expectation. This calculation takes into account treatment success rates, other fatality rates, and alternative treatment possibilities, using peer-reviewed epidemiological studies and information about the cost structure of the charity. Subjects receive detailed information on the context of the SAL choice paradigm and all aspects of the computation.

Based on this setup, we implement a very similar dilemma as in the classical Trolley. Prior to the experiment, the charity identified people suffering from tuberculosis who could be treated in two distinct Indian states, A and B.<sup>3</sup> The situation described to subjects was then the following.

- A donation of 380 Euros has been *preset* to cover 5 people suffering from tuberculosis in State A. If it is actually implemented (after the experiment), none of the five will die from the disease. If not implemented, one will die, in expectation. Donation to State A is thus the default.

3. We used the states of Maharashtra and Orissa, in which the charity operates. We randomized between subjects which state was State A and which State B.



**Figure 4.1.** Decision screen of the trolley problem

- In State B, there are 15 people suffering from tuberculosis, for whom no donation is initiated. Absent any action, 3 of them are expected to die from tuberculosis. If, on the other hand, the donation preset for State A is redirected to State B, the amount will be multiplied by three. With those 1,140 Euros, 15 people will be treated in that state, thus saving (in expectation) three lives that would otherwise be lost to tuberculosis.

**Decision.** Subjects have the choice of whether to redirect the donation from State A to State B or not. They can do so within an animation, a screenshot of which is depicted in Figure 4.1 (arrows added for exposition). Absent any action, the skull symbol proceeds from left to right along the track and will eventually hit the three figures representing lives in State B. The subject can, however, redirect the skull toward the one figure representing a life in State A, by pulling a lever that will cause a gate (drawn in blue) to pivot, provided this is done before the skull has passed the gate. Subjects thus have two options:

- Not pulling the slider: no redirection of donation → three people in State B die
- Pulling the slider: redirecting donation → one person in State A dies

**Tradeoff.** The moral dilemma involved is exactly as in the classical Trolley problem. According to deontological ethics, it is not permissible to redirect the donation (or, equivalently, the skull), thereby causing the death of (specifically, not saving) one person in State A in order to save three in State B. This represents active harm to a human being, and arguably using them as a means to an end. According to consequentialist cost-benefit analysis, on the other hand, saving three rather than one is the right thing to do, and no different from tradeoffs implicitly made every day in medicine, product or road safety, and public-budget allocations.

**Treatments.** Our SAL version of the Trolley problem differs from moral philosophers' classical one in two ways. First, the scenario is not about imagined train tracks and repairmen happening to be on them, but about actual patients and treatments in the context of a real and common disease. To investigate the effect of this added realism, subjects also faced the classical train-track trolley version at some other stage in the experiment (a week apart, and in randomized order). Second, and most importantly, subjects' choices are not hypothetical but can have real, important consequences. To properly isolate the effect of this latter feature, we ran two different between-subject treatments. In contrast, in Treatment *SAL-Hypothetical*, the choice environment is exactly as described above, but all choices are, and are presented as, hypothetical: there is never any actual donation. In Treatment *SAL-Real*, for each subject, there is a 10% probability that their decision will be implemented, resulting in either one life saved in State A or three in State B.<sup>4</sup>

#### 4.2.1.2 Lying game

This game was adapted from the classical sender-receiver game in Gneezy (2005), transforming it from a self-versus-other to an ends-versus-means dilemma. The receiver must choose between options A and B without knowing anything about the consequences of either one. The only piece of information they receive is a message from the sender, who knows that if the receiver chooses option A, 15 Euros will be donated to the children's cancer charity, whereas choosing B will earn the receiver 2 Euros. In this and the other EVM games, all donations and transfers are paid by the experimenter, eliminating any self-versus-other tradeoff for the decision-maker. The sender can send either of the following messages:

- Message 1: "Option A will give you the higher personal payment" (lie).
- Message 2: "Option B will give you the higher personal payment" (truth).

The sender is informed that in more than 90% of cases, receivers choose the option mentioned in the message,<sup>5</sup> and that they will never know whether the message was true or false, nor what situation the sender faced. The outcome of interest concerns the decisions of senders, whereas the behavior of subjects playing the role of receivers is not part of the analysis. Hence, subjects of the main experiment take the sender's role, while receivers are part of a separate sample.

**Tradeoff.** The game puts subjects in a situation in which they need to lie (to someone else's minor detriment) in order to trigger the more *socially* valuable donation.

4. We ended up implementing the decisions of 23 subjects. As a result of the ensuing donations in both states, 265 patients were treated, and thus 53 lives were saved, in expectation.

5. We obtained this number through a pilot.

Such lies are justifiable by consequentialist principles but not under deontological ethics.<sup>6</sup> Accordingly, the former prescribes Message 1, the latter Message 2.

#### 4.2.1.3 Bribe game

Two subjects are paired together, one playing the role of a sender and the other playing a receiver. Payoffs for both players and the charity are determined by the sender's choices, knowing the receiver's predetermined conditional responses.

The decision situation unfolds in two stages, summarized by Figure 4.2. In the first stage, the sender chooses between Option A, which implements a 15 Euros donation to the charity, and Option B, which pays 2 Euros to the receiver. In the second stage, knowing what the sender chose, the receiver decides between Option C, which yields them 2 Euros, and option D, which triggers an additional 20 Euros donation to the charity.

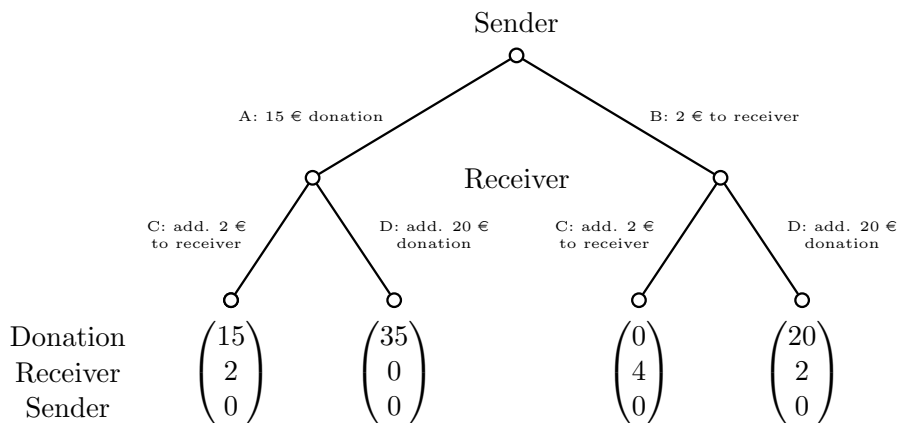


Figure 4.2. Bribe game

Receivers' contingent decisions are elicited using the strategy method, and all senders are informed that their paired receiver was one of those who decided to make their choice between C and D hinge on their sender's decision, as follows:

- If the sender chooses A, the receiver will choose C and take 2 Euros for themselves.
- If the sender chooses B, the receiver will choose D, thus triggering a donation of 20 Euros to charity while keeping the 2 Euros sent by the other player.

6. "To be truthful (honest) in all declarations is therefore a sacred command of reason prescribing unconditionally, one not to be restricted by any conveniences." (Kant, 1785)

Thus, the receiver, in effect, demands a bribe in order to implement the most socially desirable option (the word bribe is never used in the instructions). Again, we are only interested in the choices of the senders confronted with such a demand, who constitute the subjects of the main experiment. Receivers are part of a separate sample, and their choices are not of primary interest. <sup>7</sup>

**Tradeoff.** Note that the sender is not facing a strategic situation but, once again, a simple dilemma opposing a deontological approach, which entails sticking with Option A and making a donation of 15 Euros without regard to ultimate consequences, and a consequentialist one that will increase the donation to a total of 20 Euros but requires paying the bribe demanded by the other player in order to achieve this end.

#### 4.2.1.4 Statement choice

Subjects first receive information on the harmful effects of  $CO_2$  on the environment and its contribution to climate change. They are subsequently informed that, as part of the study, the purchase of carbon offsets has been prearranged, each such certificate corresponding to offsetting 1 ton of  $CO_2$  from the atmosphere. It is also explained to them that once they have completed their task, the computer will randomly destroy one certificate with a probability of 50%. If the certificate is destroyed, the removal of 1 ton of  $CO_2$  is canceled. This process is completely independent of any subject's actions, and the certificate feature is included only to make more salient both the climate-change problem and the fact that there exist ways to alleviate it. In this context, subjects can choose to submit one of the following declarations (pressing the corresponding button):

- (1) "I support the preservation and protection of the environment."
- (2) "I support the destruction of the environment."

The first statement leads to another subject receiving 2 Euros. Choosing the latter statement triggers a donation of 15 Euros to the children's cancer charity. In addition to being fully anonymous (like all choices in our experiments), subjects are informed that the results of the experiment will not be used for any other purpose, such as an opinion poll, thus depriving the statements of any instrumental value.

**Tradeoff.** The dilemma is thus to stick to one's values and submit the first statement,<sup>8</sup> as at least strongly suggested by the deontological approach, or to submit

7. Among them, 60% demanded a bribe to choose the donation, while 40% chose either the donation (23%) or the money (17%) unconditionally. All senders were paired with one among the first group.

8. Our sample consists of very environmentally-conscious subjects: 96% agree that fighting climate change is important, and 99% agree that measures to protect the environment are important. For the link between Kantian-like behaviors and moral-identity maintenance, see Bénabou and Tirole (2011) and Bénabou et al. (2022).

the second one in order to achieve an unambiguously better outcome, in line with a consequentialist view. Examining people's willingness to (anonymously) make a statement that is antithetical to their moral identity is a procedure similar to that in Bursztyn et al. (2020), but in our case, the consequence of doing so is not a material reward for oneself (one fifth or a day's wage in that paper), but once again creating a positive social externality.

#### 4.2.1.5 Group decision game

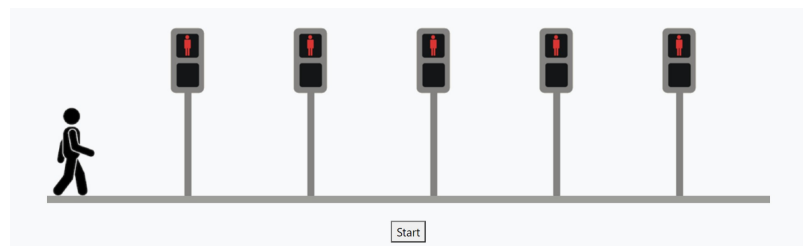
For this task, adapted from Falk, Neuber, and Szech (2020), subjects are sorted into groups of size 6. Each member makes their decision autonomously from the others, but the consequences of the actions taken by any can affect everyone. At the start of the game, the group is entrusted with 15 Euros, which is preset to be donated to the children's cancer charity by the end of the study. The first five members (first-movers), who take action earliest, are part of a separate sample, while the sixth one (second-mover) is part of the main study. First-movers simultaneously choose between:

- Option A: This choice has no further consequence, but simply preserves the donation intact if it is still relevant.
- Option B: this choice grants the member who chooses it 2 Euros as an additional payment for themselves. If even just one group member chooses this option, however, the group's donation will be canceled.

The second-mover also has two choices: Option A entails no additional payment to anyone and preserves the donation in case it is still intact; Option B grants 2 Euros to some subject outside the group, but also destroys the donation in case it was still intact.

Before making their decision, second-movers learn whether the donation has already been destroyed by the choices of the first movers. Importantly, in our experiment, at least one first mover in every group opted for option B. As in Falk, Neuber, and Szech (2020), all second movers thus choose with full knowledge that the donation had, effectively, already been canceled.

**Tradeoff.** Option A is thus entirely inconsequential, but choosing this option allows the subject to (anonymously) express a moral preference for donating, an act aligned with deontological principles (such as the categorical imperative: one should will that everyone always choose Option A). Choosing Option B, meanwhile, is clearly consequentialist, as it generates a positive outcome for another participant who took no part in the task, without affecting the already foregone charitable donation.



**Figure 4.3.** Decision screen of the rule-following task

#### 4.2.1.6 Rule-following task

The design is taken from Kimbrough and Vostroknutov (2016). The subject controls a stick figure walking across the computer screen and decides how long to wait at each in a series of traffic lights along their path, see Figure 4.3 for a visualization. Initially, the figure is at the left of the screen, and all lights are red. Once the animation starts, the figure “walks” towards the end of the path (right of the screen), automatically stopping and waiting at every red light. Each time, however, the subject can decide to press a button that causes the figure to proceed through the red light without waiting.

Subjects receive an endowment of 8 Euros and incur a deduction of 0.08 Euros for each second it takes the figure to walk across the screen. Without stopping at any red light, it takes four seconds to complete the track, costing the subject about 2 Euros in total. Waiting at each red light roughly doubles these losses to about 4 Euros. In the instructions received by subjects, they are told that the “rule of the game” is to stop at each red light until it turns green; there is no enforcement or incentive to follow the rule, however.

**Tradeoff.** The design creates a tradeoff between a deontological approach to the problem, which entails following the stated rule of stopping at each traffic light (or the meta-rule that “one should play by the rules”) and incurring losses, versus a consequentialist calculus, which favors maximizing the total payoff by breaking the non-instrumental rule (walking through red lights).

### 4.2.2 Self-versus-other block

#### 4.2.2.1 Altruism, reciprocity and prosociality

Our second set of decision situations, which subjects take either one week after or one week before the main block (with the order randomized), consists of a series of standard choice tasks from the literature on social preferences, listed in Table 4.2. Their defining feature is that subjects now face tradeoffs between their own monetary outcomes and those of other subjects, which is why we label these situations as self-versus-other (SVO). Most are very familiar, and we implement them using



standard procedures. Therefore, we leave the description of procedural details to Section 4.A of the Appendix. More novel is the moral-luck game, which we describe next.

**Table 4.2.** Prosociality (self-versus-other) decision situations

Variable	Elicitation method	Definition
Altruism	Dictator game with charity as the recipient, giving frame, 20 Euros endowment	Amount allocated to charity
Altruism Taking	Dictator game with charity as the recipient, taking frame, 20 Euros endowment	Amount allocated to charity
Trust	First mover in trust game, 5 Euros endowment	Amount send to second mover
Pos. Recipr. Low	Second mover in trust game when having received 1 Euro out of 5 Euros (strategy method)	Amount send back to first mover
Pos. Recipr. High	Second mover in trust game when having received 5 Euros out of 5 Euros (strategy method)	Amount send back to first mover
Public goods game contribution	Public goods game, 5 Euros endowment, group size of 3, 1.5 marginal per capita return from contributing	Amount contributed
Rewarding intentions	Moral luck game	Amount allocated to first mover in $S_4 - S_3$
Rewarding consequences	Moral luck game	Amount allocated to first mover in $S_4 - S_2$
Rewarding consequences - intentions	Moral luck game	Amount allocated to first mover in $(S_4 + S_3) - (S_2 + S_1) - ((S_4 - S_3) + (S_2 - S_1))$

Notes: For details on the first six variables, see Appendix 4.A. For details on the last three, see Section 4.2.2.2.

#### 4.2.2.2 Moral-luck game

There are two players: a first-mover who chooses between two lotteries with different payoff distributions for themselves and for the charity, and a dictator who allocates additional money between themselves and the first-mover. These features make the game one of conditional altruism, which is why it is included in this block. At the same time, it is designed to measure and compare the extent to which dictators reward socially desirable intentions versus socially desirable outcomes, and could thus also be seen as also informative about deontological versus consequentialist preferences, or norms.

First-movers, whose behavior is not *per se* the object of interest, choose between two lotteries: (i) Lottery M, which yields 10 Euros to self with 70% probability and a 15 Euro donation to the charity with 30% probability; (ii) Lottery D, which yields

10 Euros to self with 30% probability and a 15 Euros donation to the charity with 70% probability.

Subjects in the role of the dictator then make, using the strategy method, allocation choices to the first-mover they are paired with, for each of the four possible choice-outcome:

S1 First-mover chooses M, 10 Euros payment realizes

S2 First-mover chooses D, 10 Euros payment realizes

S3 First-mover chooses M, 15 Euros donation realizes

S4 First-mover chooses D, 15 Euros donation realizes

The dictator is endowed with 10 Euros and the amount  $x$  they allocate to the first mover is tripled (they keep the remaining  $10 - x$ ), in order to induce positive giving in each of the four possible situations. Their choices then allow us to ask: (i) fixing the lottery outcome, to what extent do they take the first-mover's lottery choice ("intention") into account when allocating money to them? (ii) fixing the lottery choice, to what extent do they take the outcome into account in their allocation? (iii) is their decision more responsive to the first-mover's *ex-ante* lottery choice or to its *ex-post* realization? Table 4.2 specifies exactly how each of these propensities is measured.

### 4.2.3 Questionnaires

To complement our incentivized behavioral tasks, we administer several widely used questionnaires, which constitute the third block of our design. First, we employ the Oxford Utilitarianism Scale (OUS, Kahane et al., 2018), which elicits subject's "permissive attitudes toward instrumental harm", very much as in the Trolley dilemma (OUS-IH), and their "impartial concern for the greater good", a utilitarian-like tendency to judge the well-being of every individual as equally important (OUS-IB). Second, we administer the Moral Foundations Questionnaire (MFQ, Graham et al., 2011), which aims to measure five distinct dimensions of people's moral concerns: care/harm of others, fairness/reciprocity, in-group/loyalty, authority/respect, and purity/sanctity. Third, we include the Moral Universalism short module of Enke, Rodríguez-Padilla, and Zimmermann (2022). Using hypothetical allocation games, it measures the extent to which subjects exhibit the same level of altruism towards strangers as towards in-group members (MU scale). Finally, we confront subjects with the classical, hypothetical train-track Trolley dilemma, and also include a module on political attitudes and religiosity. For sociodemographic variables, we collected age, gender, subject of studies, final high-school grade, and gross monthly income.

#### 4.2.4 Procedural details

The study was run online as a virtual-lab experiment using oTree (Chen, Schonger, and Wickens, 2016), with the subject pool of the BonnEconLab. Subjects were invited using hroot (Bock, Baetge, and Nicklisch, 2014) and had to log in at a specified date and time, where an experimenter was available throughout to answer questions and address any issues. Subjects had to complete two sessions, separated by one week, each lasting about 45 minutes. One session contained the main ends-versus-means decision situations, including the SAL Trolley, the other one the self-versus-other situations and the questionnaires. The order of sessions, as well as the order of decision tasks within each session, was randomized. Subjects earned a 12 Euros show-up fee for each session, and one decision in each session was selected for real implementation –except for the *Real* treatment of the SAL trolley, in which every subject's choice was implemented with 10% probability.

Overall, 626 subjects took part in the main experiment, of which 593 completed both sessions.<sup>9</sup> Based on the preregistration, we excluded the top 1% fastest subjects, as well as subjects who preferred giving 2 Euros to another participant over a 15 Euros donation to the children-with-cancer charity in the baseline task (41 subjects). The reason for excluding them is that no opposing predictions in the ends-versus-means games exist for them. The final sample thus consisted of 548 subjects (339 female, mean age = 26.65, SD = 8.10). Results are robust when considering the full sample.

In order to ensure that all participants fully understood the decision rules and consequences, we included, throughout the experiment, extensive comprehension checks. For instance, each of the major decision situations (except rule-following and statement choice) featured a quiz after their introduction, and subjects were only allowed to proceed once they answered all questions correctly.<sup>10</sup> For the group-donation game and the statement choice, we also implemented attention and memory checks. At the end of the experiment (approximately 15-25 minutes after their decisions), subjects were asked, unannounced: (i) in the group-donation game, whether the donation was destroyed by the choices of the other group members; (ii) in the statement choice, what were the consequences of making the repugnant statement. Proper recall at that later time constitutes a lower bound on compre-

9. We pre-registered a sample size of 600 completes. This sample size was chosen based on power calculations for the comparison between the *Hypothetical* and *Real* treatments of the SAL trolley. Assuming that 20% of subjects choose the deontological option in *Hypothetical* and a significance level of 5%, we have 80% power to detect a 10 pp. treatment effect (two-sample test of proportions) –that is, a reduction in the fraction of subjects choosing the deontological option to 10% or less, or an increase to 30% or more.

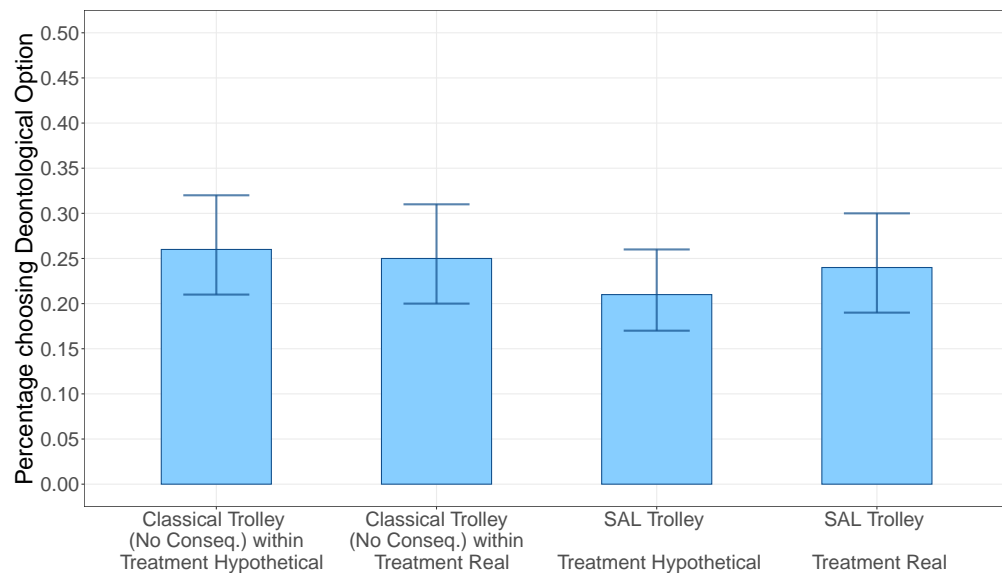
10. Upon giving a wrong answer in the comprehension questions, subjects were informed that one or more of their choices was incorrect and that they needed to check the instructions again.

hension and attentiveness during the choice phase, which we will use to test the robustness of our results.

### 4.3 Results

#### 4.3.1 How prevalent are deontological and consequential decision-making?

We begin our empirical analysis by investigating the prevalence of the two types of decision-making. We first test whether deontologically motivated behavior is an artifact of hypothetical decisions and abstract situations, or whether subjects display similar choices in realistic situations with significant consequences, using the SAL trolley problem. We then investigate decision-making in our full set of ends-versus-means decision situations.



**Figure 4.4.** The effect of changing realism and consequences on behavior in the trolley dilemma

*Notes:* Each bar displays the fraction of subjects choosing the deontological option - not pulling the lever - in the trolley problem. The first two bars display behavior in the classical train-track version, the last two behavior in the Saving A Life paradigm. Treatment *Real* is that in which the SAL trolley has real consequences, whereas in *Hypothetical* decisions are without consequences. Error bars indicate 95% confidence intervals using one-sample tests of proportions.

**Trolley.** Figure 4.4 displays the trolley-problem results. In answering the classical train-track dilemma, 26% of subjects assigned to the *Hypothetical* treatment of the SAL-Trolley – a fraction quite similar to what other studies investigating this version

of the trolley have found.<sup>11</sup> An indistinguishable proportion of 25% among those assigned to the *Real* treatment made a similar choice, showing that randomization across treatments was successful ( $p = 0.82$ , two-sample test of proportions).

More interesting is the comparison between behavior in the classical trolley and our hypothetical SAL version. In total, 21% of subjects choose the deontological option in the SAL trolley, a small and insignificant difference compared to the classical trolley ( $p = 0.17$ , two-sample test of proportion). Thus, introducing a more realistic but still hypothetical setting does not appear to significantly influence decision-making.

Turning to the influence of introducing real consequences, we find that 24% of subjects in the *SAL-Real* treatment choose the deontological option. This fraction is not significantly different from that in the *Hypothetical* treatment ( $p = 0.48$ , two-sample test of proportions). Accordingly, real consequences neither substantially reduce nor increase the extent of deontological or consequential decision-making by our subjects. Overall, our treatments demonstrate that deontological behavior in sacrificial dilemmas is a robust phenomenon, being chosen by about one-fourth of respondents.

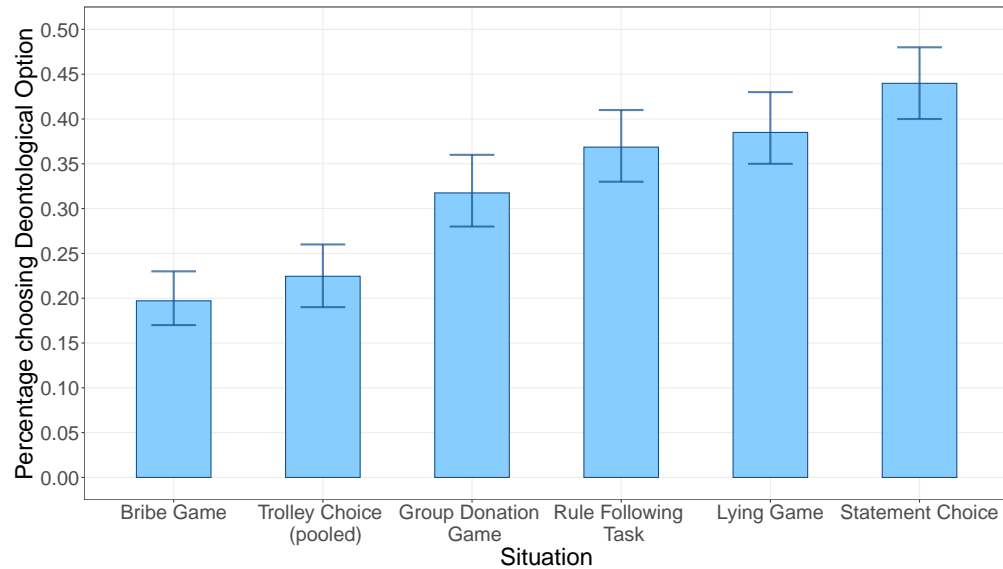
Appendix Section 4.B reports a series of checks showing that our results are not driven by the pre-registered exclusion restrictions, nor by potential confusion among subjects. The observed differences between classical and SAL-trolley, as well as between treatments, become even smaller using the full sample of subjects (Figure 4.B.1), or only subjects who scored high on the SAL trolley comprehension test (Figure 4.B.2). Another potential concern could be that our treatment with real consequences was too weak to induce behavioral effects, since only 1 in 10 subjects had their choices implemented. In Appendix Section 4.B.2, we report the results of a robustness test addressing this concern. Using the same SAL paradigm but now with a self-versus-other tradeoff (taking money vs. triggering the donation), we find that implementing real consequences with a similar probability drastically changes the extent of prosocial decision-making, compared to a hypothetical situation. These results suggest that the absence of a real-consequences treatment effect in the trolley is specific to ends-versus-means situations, where two moral principles conflict.

We also investigate whether our aggregate null finding might mask heterogeneity in the types of people choosing either option in the different versions. For instance, it could be that male subjects (say) are more likely to choose the deontological option when consequences are hypothetical, but less likely once real consequences are introduced. However, we find no evidence of heterogeneous treatment effects across a wide range of characteristics; see Appendix Figure 4.B.3 for details.

Because we neither find differences in the aggregate nor in individual characteristics between treatments, we will pool decisions between treatments for the sub-

11. For instance, across 42 countries and 70,000 participants in Awad et al. (2020), 19% choose the deontological option.

sequent analysis. The results in the following sections remain unchanged if we use data from only one of the treatments.



**Figure 4.5.** Distribution of deontological and consequentialist decision-making in ends-versus-means tradeoff situations

*Notes:* Each bar displays the fraction of subjects choosing the deontological option. See Section 4.2.1 for details on the situations and the definition of the deontological option. Error bars indicate 95% confidence intervals using one-sample tests of proportions.

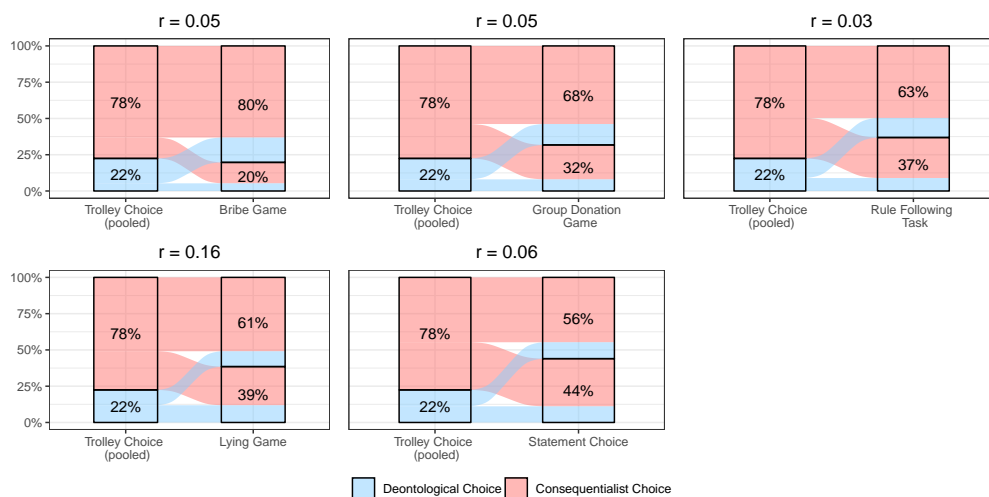
**Other ends-versus-means decisions.** We now turn to decision-making in our other five EVM games. Figure 4.5 displays, for each game and for the (pooled) SAL trolley, the fraction of subjects choosing the deontological option. In every game, we find a substantial number of subjects making such a choice. Thus, 20% of subjects refuse to “bribe” the receiver to achieve the higher donation, 32% choose the inconsequential option in the group donation game, and 37% follow the rule in the rule-following task. In the lying game, 39% refuse to lie to the receiver, and 44% are unwilling to make the repugnant statement. These fractions are statistically different from zero in every instance ( $p < 0.001$ , one-sample test of proportion).

Could these results be driven by subjects’ confusion about the consequences of their decisions? In our instructions, we extensively covered the implications of each situation, administered comprehension quizzes, and implemented a memory check for both the group-donation game and the statement choice. In total, 80% of subjects for the statement choice and 81% for the group-donation game correctly recalled the consequences at the end of the experiment, so they must have understood them when making the decision. Similar fractions were able to correctly answer all comprehension questions for the other games on their first try, and less than 5% needed

more than one try. In Appendix Figure 4.C.1, for each decision situation, we restrict the sample to subjects who answered all comprehension questions correctly on their first try and correctly answered the recall questions, respectively. We again see a significant fraction of subjects choosing the deontological option in every one of the six EVM decision situations. Likewise, in Appendix Figure 4.C.2, we replicate this finding among the full set of subjects without the preregistered exclusion restrictions.

**Result 4.1.** *Both consequential and deontological decision-making are pervasive in our experiment, with significant fractions (20% to 44%) of subjects behaving deontologically in ends-versus-means tradeoffs. Moreover, this behavior appears to be independent of whether the tradeoffs are embedded in a realistic or abstract setting, and whether they entail real consequences or not.*

#### 4.3.2 How consistent are deontological and consequential decision-making?



**Figure 4.6.** Relationship of trolley behavior with behavior in the other ends-versus-means situations

Notes: Each diagram displays decision-making in the respective ends-versus-means game, conditional on behavior in the SAL-trolley dilemma. Percentages shown are the fractions of subjects choosing the deontological and consequentialist option in each case. Above each plot is the Pearson correlation coefficient between the SAL trolley and the respective decision situation displayed.

Having established the *prevalence* of both moral principles in the previous section, we next investigate their *consistency* across decision situations. We first ask whether the trolley dilemma, as the principal one traditionally used to elicit moral intuitions, is *predictive* of behavior in other ends-versus-means situations. Figure 4.6 displays, in pairwise flow charts, how choices in the SAL trolley are related to those in the other EVM games. We find no evidence that trolley choices are predictive in

**Table 4.3.** Correlation matrix for ends-versus-means decision-situation

	Group Don. Game	Bribe Game	Statement Choice	Lying Game	Rule foll. Task
SAL Trolley	0.05 [-0.04, 0.13]	0.05 [-0.03, 0.14]	0.06 [-0.02, 0.14]	0.16 [0.08, 0.24]	0.03 [-0.05, 0.12]
Group Don. Game		0.14 [0.06, 0.23]	0.15 [0.07, 0.23]	-0.09 [-0.17, 0.00]	-0.06 [-0.14, 0.03]
Bribe Game			0.19 [0.11, 0.27]	0.04 [-0.04, 0.12]	-0.08 [-0.17, 0.00]
Statement Choice				-0.06 [-0.14, 0.02]	0.04 [-0.04, 0.12]
Lying Game					0.00 [-0.08, 0.09]

Notes: The table displays Pearson correlation coefficients between the six ends-versus-means decision situations. Brackets display 95% confidence intervals.

other settings. In each case, large fractions of subjects switch from choosing the deontological option in the trolley to the consequential version in the other situation, and vice versa. This is also reflected in the correlations (Pearson correlation tests), which are close to zero and insignificant ( $p > 0.1$ ) for all decision situations except the *lying game*, where a small but significant correlation does exist ( $p < 0.01$ ).

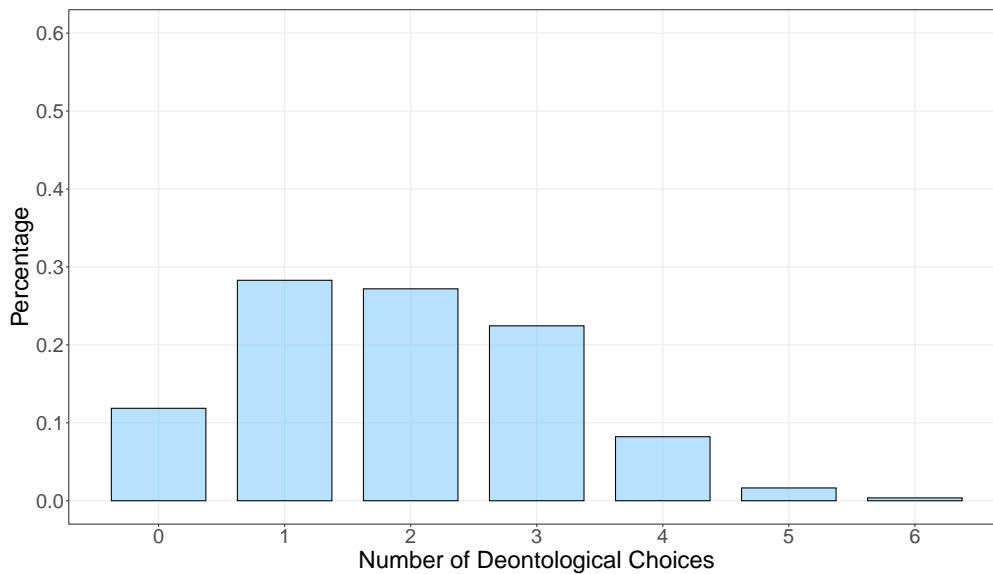
In Table 4.3, we extend the analysis to all pairwise correlations between the six EVM games. In each case, the indicator variable is equal to one if the subject chooses the deontological option (as defined in Section 4.2.1), and zero if they choose the consequential one. We find that no decision situation consistently predicts behavior in more than two others. Moreover, since we observe no correlation larger than 0.2 and sometimes even negative ones, there appears to be no detectable consistency in behavior. Instead, whether an individual makes the consequentialist or the deontological choice appears to be highly situation-specific.

A complementary way to analyze consistency of choices is to examine the number of times a subject chooses either the consequential or deontological option; Figure 4.7 displays the results. If subjects behave consistently according to one of the two moral principles, one should observe a (skewed) U form. Instead, the empirical pattern displays an opposite, inverse-U shape. Essentially no subject chooses six times the deontological option, and only a small minority of just above 10% behave as standard economic agents, choosing six times the consequential option.

A potential concern might be that subjects being confused or paying limited attention could attenuate correlations between variables, artificially decreasing consistency. In Section 4.3.4 below, we address this issue in a series of robustness checks and show that our results remain unchained when we exclude confused and inattentive.

**Result 4.2.** *Subjects exhibit essentially no consistency in deontological/consequentialist choices across ends-versus-means tradeoff situations. Neither*





**Figure 4.7.** Consistency of deontological and consequential decision-making

Notes: Histogram of the number of times a subject chose the deontological option in the ends-versus-means decision situations.

*the trolley dilemma, nor any other ends-versus-means decision, consistently predicts behavior in the other EVM games.*

### 4.3.3 How are deontological and consequential decision-making related to prosocial behavior and morality-questionnaire measures?

We now turn to the comparison of behavior in ends-versus-means dilemmas with that in self-versus-other tradeoffs, as well as with measures obtained through hypothetical questionnaires. Table 4.4 displays the correlation matrix.

**Behavior within the self-versus-other block.** In stark contrast to behavior in EVM situations, we find that behavior in SVO decisions is rather consistent across games. We code variables so that higher values indicate higher degrees of giving to others, i.e., higher donations in the dictator game, more money sent in the trust game, etc. Almost all measures within this block are significantly correlated with each other, with most correlations falling between 0.30 and 0.40. For instance, altruism, as defined by the extent of giving to charity in a simple allocation game, consistently predicts the other eight measures, with correlations ranging from 0.20 to 0.65. In fact, the lowest correlation of altruism with any other measure is higher than the high-

est correlation within the ends-versus-means block. Thus, subjects show a relatively high degree of consistency in the way they resolve self-versus-other tradeoffs.<sup>12</sup>

**Relation between ends-versus-means block and self-versus-other block.** Comparing behavior between EVM and SVOs games reveals that the two blocks are largely unrelated to each other, as displayed in the second part of rows in Table 4.4. All pairwise correlations of measures between the two blocks are below 0.20, with the large majority being close to zero. Accordingly, behavior in neither block predicts that in the other one. For example, subjects choosing the deontological option in the trolley donate on average 33% of their endowment to the charity, while those choosing the consequentialist option donate merely 0.13 EUR more (0.7 pp.). We observe similarly minute differences for the other games. Hence, acting deontologically versus consequentially in any of our EVM blocks is not at all predictive of whether a subject is more or less prosocial, and vice versa.

**Relation with hypothetical questionnaires.** How is behavior in our incentivized decision situations correlated with measures obtained from hypothetical questionnaires? The third part of rows in Table 4.4 displays our results. Focusing first on the relation between behavior in the SAL trolley and questionnaire measures, we find a high correlation of 0.50 with behavior in the classical trolley dilemma. This result confirms our earlier finding that sacrificial-dilemma choices are robust to variations in the displayed situation (real-life donations versus train track, real or hypothetical stakes). It is also further evidence that our findings are not just reflecting inattention by subjects or a predominance of noise in their decisions. We also find a substantial correlation of -0.43 with the Oxford Utilitarian scale instrumental-harm measure (OUS-IH), in the expected direction: subjects choosing the deontological option in the SAL trolley are less likely to endorse instrumental harm to obtain a greater good. Since the OUS-IH items describe situations purposefully designed to be similar to the trolley, a high correlation is reassuring. Next, replicating Kahane et al. (2018), we find a lower correlation of -0.21 of trolley behavior with the measure capturing subjects' instrumental benefit concerns (OUS-IB). The classical trolley and both OUS measures are also not predictive of behavior in any of the five other ends-versus-means decision situations with correlations not higher than 0.15, providing further evidence for the lack of consistency in behavior across this class of tradeoffs.

Behavior in the trolley, or any other EVM game, is also not correlated with any measure from the Moral Foundations Questionnaire (MFQ) or Moral Universalism

12. For altruism, trust, and both reciprocity measures, we can compare our individual-level correlations with those in Chapman et al. (2022). Although they use a representative sample of the US population while ours contains mostly German students, our correlations are very similar to theirs. For altruism, they report (measurement-error corrected) correlations of 0.34 with both reciprocity measures and 0.60 with trust. The correlation coefficient of trust with both reciprocity measures is 0.49, and the correlation between low and high reciprocity is 0.86. Hence, the only noteworthy difference is the relation of altruism with trust, where they find a substantially higher correlation than we do.

scale (MU) scales. These scales appear to measure largely orthogonal dimensions of moral preferences or intuitions. Turning to self-versus-other situations, moral universalism (the slope of prosociality as social distance increases) is *negatively* correlated with the degree of prosociality identified in all SVO situations, albeit with low values.<sup>13</sup> The results for the different MFQ subscales are more mixed. The Harm and Fairness subscales, which capture the weight subjects put on values concerning protecting the weak and upholding equality, are correlated with the altruism and trust measure, but not with any other behavior. The other three subscales capturing subjects' weight on loyalty (MFQ Ingroup), authority, and purity appear to be behaviorally distinct from behavior in the SVO situations.

**Principal component analysis.** An important question is how well each of our experimental parts –the two decision blocks and the questionnaires - can be explained by a subset of variables. To this end, we perform a principal component analysis (PCA) separately on the ends-versus-means and the self-versus-other block. As a reference, we also perform a PCA on the Oxford Utilitarian scale. Figure 4.8 shows the results by displaying, for each block, the eigenvalues of the respective components. The higher the eigenvalue, the more variance is explained by the component. We find that the eigenvalues of the EVM components are almost uniformly concentrated around one. In contrast, for the SVO block, the first component has a substantially higher eigenvalue than the other components. Accordingly, one component can explain the SVO choice well, whereas no such component exists for the ends-versus-means choices. This result is not specific to the fact that we have six EVM situations and nine SVO measures. We also conducted a PCA for every combination of six situations out of the nine SVO measures: in every one of the resulting 84 combinations, the variance explained by the first component of these six situations is at least twice as high as the first component can explain in the six EVM situations. The resulting eigenvalues from this exercise range from 2.36 to 3.02. Hence, no matter which SVO situations are taken, a dimensionality reduction always performs significantly better among SVO tradeoffs than among EVM situations.

**Result 4.3.** *Subjects exhibit a relatively high degree of consistency across self-versus-other tradeoff situations. Behavior in these prosociality decisions is unrelated to behavior in ends-versus-means tradeoff situations.*

#### 4.3.4 Robustness of correlations between blocks

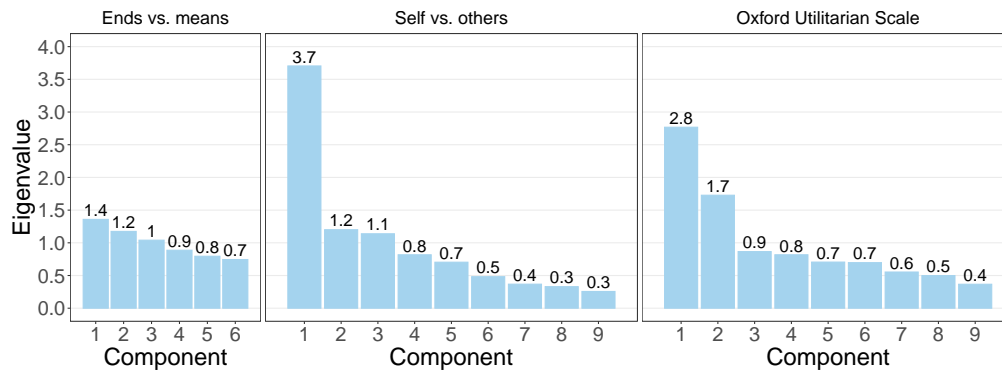
**Confusion or limited attention.** As mentioned previously, a potential concern might be that noise due to subjects' confusion or limited attention could drive some of our results, especially the lack of consistency among the ends v. means situations.

13. Enke, Rodríguez-Padilla, and Zimmermann (2022) similarly find that moral universalism (weakly) negatively predicts the total amount given to various charities.

**Table 4.4.** Correlations of ends-versus-means, self-versus-other, and questionnaire measures

	Ends-versus-means decision situations						Self-versus-others decision situations									Questionnaires									
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons.- Int.	Reward Int.	Reward Cons.	Classical Trolley	OUS IB	OUS IH	MU Scale	MFQ Harm	MFQ Fairness	MFQ Ingroup	MFQ Authority	MFQ Purity	
SAL Trolley		0.05					-0.01	-0.03	0.02	-0.05	-0.03	0.00	-0.02	0.05	-0.02	0.50	-0.12	-0.29	-0.04	-0.06	0.00	-0.06	0.00	0.05	
Group Don.			0.14	0.15	-0.09	-0.06	0.05	0.03	-0.04	0.14	0.05	-0.02	0.09	-0.03	-0.02	0.02	0.09	0.02	-0.02	0.10	0.02	0.08	0.03	0.12	
Bribe				0.19	0.04	-0.08	-0.04	-0.02	-0.11	-0.06	-0.15	-0.06	-0.06	-0.18	-0.18	-0.01	-0.03	-0.04	0.01	0.02	-0.04	0.08	0.05	0.14	
Statement Choice					-0.06	0.04	0.00	0.01	-0.12	0.04	-0.07	0.00	0.05	-0.15	-0.04	-0.01	0.03	-0.07	-0.02	0.03	-0.03	0.15	0.13	0.17	
Lying						0.00	-0.03	-0.04	-0.03	-0.03	0.00	-0.05	0.00	-0.02	-0.01	0.14	-0.15	-0.11	-0.05	-0.03	-0.02	-0.08	-0.04	-0.10	
Rule following							0.11	0.06	0.03	0.11	0.10	0.03	0.11	-0.01	0.13	-0.02	0.08	0.00	-0.07	0.04	0.00	-0.07	-0.02	-0.02	
Altruism								0.65	0.20	0.31	0.38	0.40	0.36	0.23	0.23	0.00	0.22	0.00	-0.13	0.19	0.17	0.01	-0.02	0.00	
Altruism Taking									0.16	0.26	0.35	0.36	0.28	0.21	0.21	0.01	0.20	0.00	-0.13	0.20	0.15	0.00	-0.02	-0.04	
Trust										0.35	0.45	0.47	0.38	0.30	0.23	0.00	0.05	0.02	-0.13	0.00	0.06	-0.06	-0.13	-0.17	
Pos. Recipr. Low											0.71	0.36	0.35	0.13	0.14	-0.02	0.17	0.03	-0.16	0.08	0.09	-0.02	-0.08	-0.04	
Pos. Recipr. High												0.45	0.41	0.30	0.21	0.01	0.18	0.04	-0.18	0.09	0.11	-0.12	-0.16	-0.15	
PGG Cont.													0.40	0.31	0.22	0.03	0.11	-0.13	-0.26	0.06	0.12	-0.02	-0.07	-0.09	
Reward Cons.- Int.														0.24	0.55	-0.02	0.08	-0.01	-0.15	0.06	0.07	-0.07	-0.12	-0.11	
Reward Int.															0.40	0.07	0.02	0.00	-0.12	-0.01	0.04	-0.10	-0.08	-0.14	
Reward Cons.																0.04	0.01	0.01	-0.04	0.05	0.06	-0.09	-0.08	-0.09	
Classical Trolley																	-0.21	-0.43	-0.04	-0.04	0.01	-0.08	-0.06	-0.01	
OUS IB																		0.23	-0.06	0.35	0.24	0.08	-0.04	0.05	
OUS IH																			0.15	-0.03	-0.03	0.16	0.15	0.07	
MU Scale																				0.02	0.02	0.22	0.16	0.20	
MFQ Harm																				0.02	0.58	0.23	0.17	0.26	
MFQ Fairness																						0.16	0.08	0.14	
MFQ Ingroup																							0.57	0.51	
MFQ Authority																								0.55	
MFQ Purity																									

Notes: Colors in heatmap change with each 0.1 correlation magnitude.



**Figure 4.8.** Principal component analyses

Notes: The figure displays the results of principal component analyses conducted on the six ends-versus-means and nine self-versus-other decision situations, as well as the nine items of the Oxford Utilitarian scale. The y-axis shows the eigenvalues of each component. Higher eigenvalues indicate the respective component explains more variance in the data.

Note, however, that we do observe a relatively high degree of consistency among the variables in the self-versus-other block. Furthermore, behavior in our SAL version of the trolley is well correlated with psychological questionnaires designed to capture such reasoning. Therefore, a general lack of attention or misunderstanding among subjects cannot explain our results. Consider next the possibility of differential confusion between the blocks – subjects having more problems understanding the ends-versus-means decision situations than the SVO ones. Instead, we find that subjects make more errors on average in the comprehension questions of the SVO block than in those of the EVM block, suggesting that their understanding is actually higher in the latter.

To further address these concerns, we conduct a series of robustness checks. First, we exclude subjects who fail more than one of the comprehension quizzes (Appendix Table 4.D.1). Second, we exclude those who fail one or both of the attention-memory checks we employed for the Group Donation and Statement Choice situation (Appendix Table 4.D.2). Third, we exclude the Top 15% fastest subjects (Appendix Table 4.D.3). In every case, we replicate our main findings: almost zero correlations within the ends-versus-means block, high correlations within the SVO block, and very low correlations between the two blocks.

**Measurement.** The variables in the EVM block are dichotomous, while those in the SVO block are more fine-grained. Relying on standard Pearson correlation coefficients could potentially confound the differences we find between the two blocks, as well as those within the first one relative to the second.

To alleviate this concern, we show that the documented consistency within the SVO block prevails even when all its variables are dichotomized, in two different ways. First, we split subjects based on whether they fully behave according to the

classical model of maximizing self-interest. That is, if a subject donates nothing to the charity, does not cooperate at all in the public goods game, sends nothing in the trust game, etc., we code the respective variable as zero, and one otherwise. Second, we categorize subjects based on a median split for each respective variable. For example, the altruism variable is then zero if a subject donates less than the median, and one if they donate more. Under both approaches, we again find a high degree of consistency within the SVO block, and no correlation with the ends-versus-means block; see Appendix Table 4.D.4 for the first approach, and 4.D.5 for the second.

To further verify that the low correlations between the two blocks are not an artifact of our measurements, we also compute biserial correlations, which are designed for pairwise comparisons between a binary and a continuous variable. Again, we find low correlations between the two blocks; with no apparent pattern, see Appendix Table 4.D.6 for details.

#### 4.4 Conclusion

Our results bring substantial nuance to the classical dichotomy opposing Consequentialism and Deontology. On one hand, deontological decision-making is a robust phenomenon, not an artefact of hypothetical choices or unrealistic dilemmas: we find it in a real-stakes version of the trolley and in five other games pitting socially desirable ends against arguably (and, *de facto*) objectionable means. Treating all economic agents as solely consequentialist thus appears to be an inaccurate representation of their preferences and decisions.

On the other hand, in our subject population there appear to be no consistently deontological “types” and only a small fraction of consistently consequentialist ones. Most individuals’ choices across the six ends-versus-means dilemmas are largely uncorrelated, which is at odds with any stable mixture of two preference types, invariant across choice situations. Subjects clearly feel the pull of both moral principles, but appear to resolve it in a very situation-specific manner – perhaps as if they had very flat priors on what their preference “should” be, and even minor contextual signals or cues could sway them. These results point to interesting avenues for further research, both experimental (using different subject populations) and theoretical – better understanding what seems to be a new type of internal preference conflict.

On the methodological side, the paper introduces a new experimental paradigm – ends-versus-means games, as fundamentally distinct from the self-versus-other games used to measure prosociality. It is quite flexible, and in particular independent of the trolley dilemma: while another contribution of the paper is to make trolley choices “real,” this particular decision task could be dropped without changing either the general architecture of the paradigm, nor the overall results found in our study. Similarly, a researcher could easily replace one or more of our proposed

tests of Consequentialism vs. Deontologism (lying game, bribing game, non-pivotal group choice, aversive statement, rule following) with some variant that they preferred. The methodology of examining the consistency of decisions within that ends-versus-means block, and potential correlations with behavior in standard prosociality games, would remain the same.

## **Appendix 4.A Details on self versus others decision situations**

**Dictator game.** Subjects play two variations of a dictator game, a giving and a taking game. In each, they can allocate 20 € between themselves and a charity helping children suffering from cancer. In the taking game, 20 € are to be donated to the charity, and subjects can decide to take money for themselves. In the giving game, they are endowed with 20 € and can donate an amount to the charity.

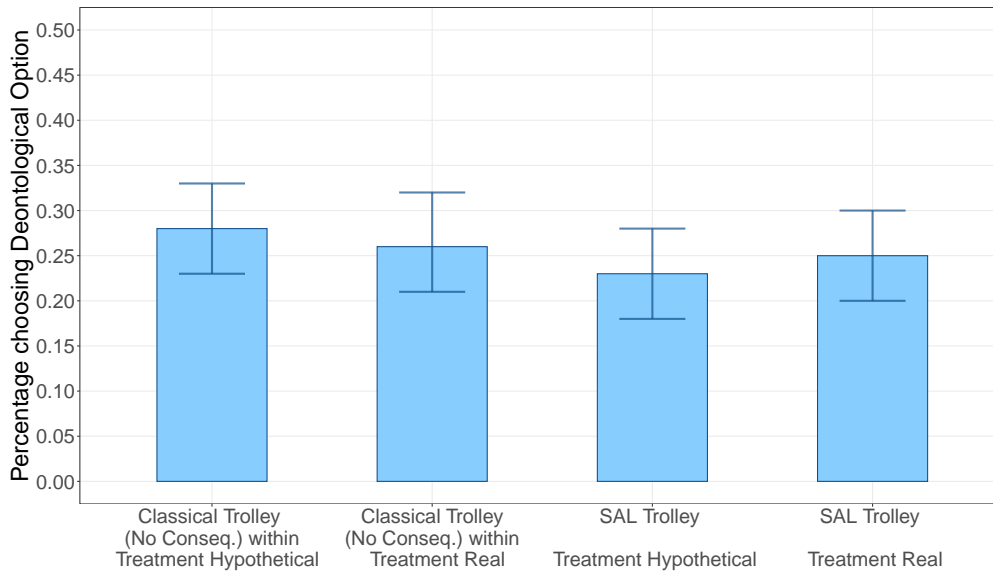
**Trust game.** Subjects play both roles of the standard trust game with an endowment of 5 € and a multiplication factor of 3. Senders can send integers from 0 to 5 €. Recipient choices are elicited using the strategy method, i.e., subjects decide how much to send back conditionally on each of the six possible amounts send.

**Public goods game.** Subjects play the standard public goods game with an endowment of 5 € a group size of 3, and a multiplication factor of 1.5.



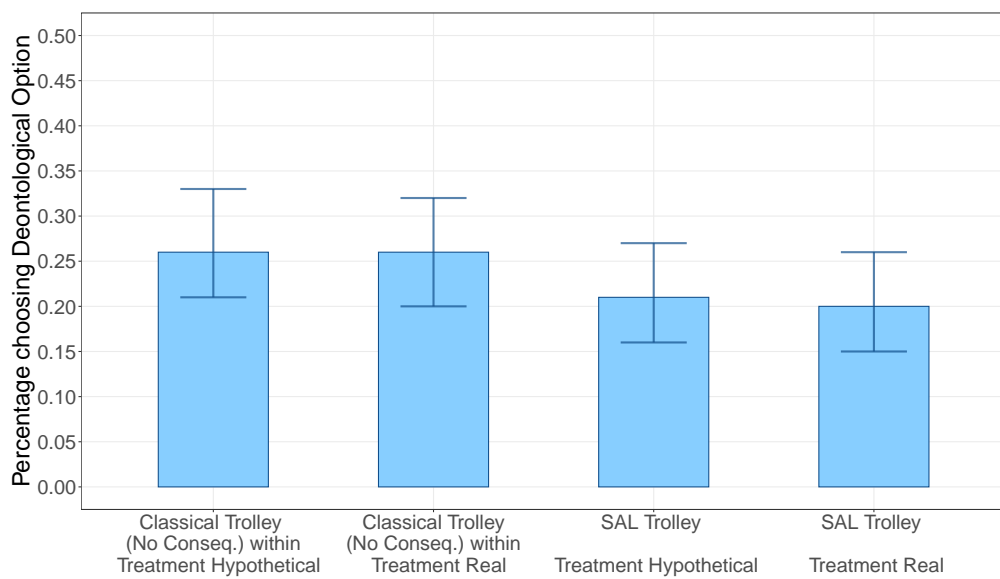
## Appendix 4.B Robustness analysis trolley dilemma

### 4.B.1 Figures



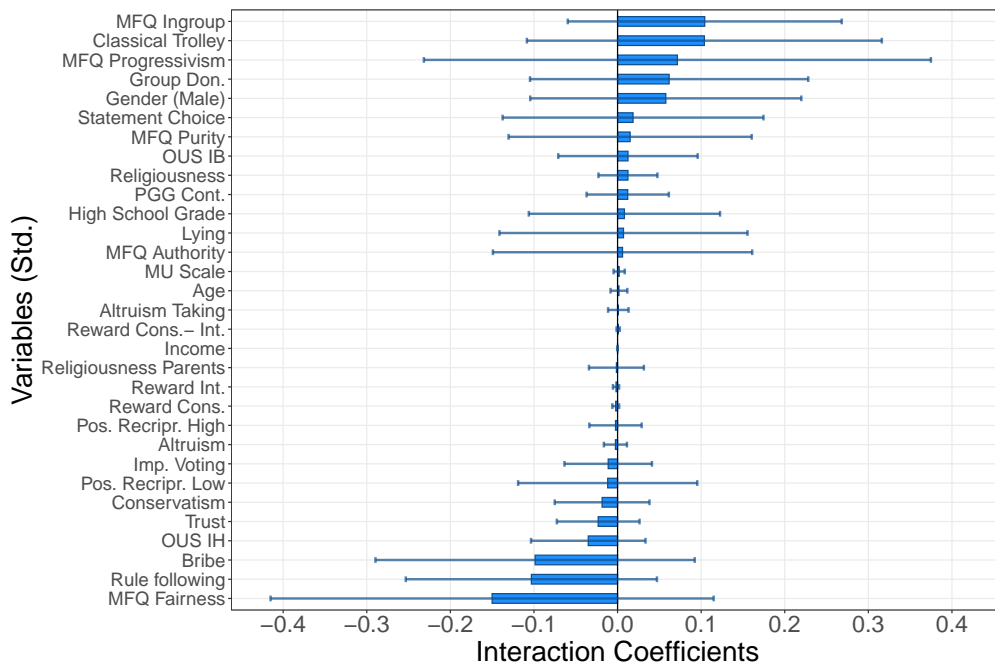
**Figure 4.B.1.** The effect of changing realism and consequences on behavior in the trolley dilemma using the full sample

Notes: Each bar displays the fraction of subjects choosing the deontological option - not switching the lever - in the trolley problem using the full sample of 593 subjects. The first two bars display behavior in the classical train track trolley, the last two behavior using the Saving A Life paradigm. Treatment *Real* indicates the treatment in which the SAL trolley has real consequences, in treatment *Hypothetical* are decisions without consequences. Error bars indicate 95% confidence intervals using one-sample tests of proportions.



**Figure 4.B.2.** The effect of changing realism and consequences on behavior in the trolley dilemma among high comprehension subjects

*Notes:* Each bar displays the fraction of subjects choosing the deontological option - not switching the lever - in the trolley problem using only subjects that make no more than one mistake in the trolley comprehension quiz. The first two bars display behavior in the classical train track trolley, the last two behavior using the Saving A Life paradigm. Treatment *Real* indicates the treatment in which the SAL trolley has real consequences, in treatment *Hypothetical* are decisions without consequences. Error bars indicate 95% confidence intervals using one-sample tests of proportions.



**Figure 4.B.3.** Differences in characteristics of subjects between the *Hypothetical* and *Real* treatments

Notes: The graph displays interactions term coefficients obtained from an OLS regression with an indicator that equals 1 for observations from the Real treatment. Bars indicate 95% confidence intervals constructed from robust standard errors.

### 4.B.2 Robustness experiment

This section describes a robustness experiment addressing the concern that the probabilistic implementation employed in the treatment *Real* introduced an insufficient real consequence incentive. Specifically, in treatment *Real*, each choice in the Saving a Life trolley has a 10% chance of being implemented with real consequences. In comparison, in treatment *Hypothetical*, all choices had no real consequences. We find that behavior of subjects does not differ between the two treatments. One potential explanation is that subjects perceive a 10% implementation probability as essentially hypothetical, rendering the treatment comparison problematic. In the robustness experiment, we show that subjects are highly responsive to a chance from 0% implementation probability to a positive probability in self-versus-other tradeoffs. We use the design and results of one treatment of Bénabou et al. (2022), to which we add a hypothetical treatment.

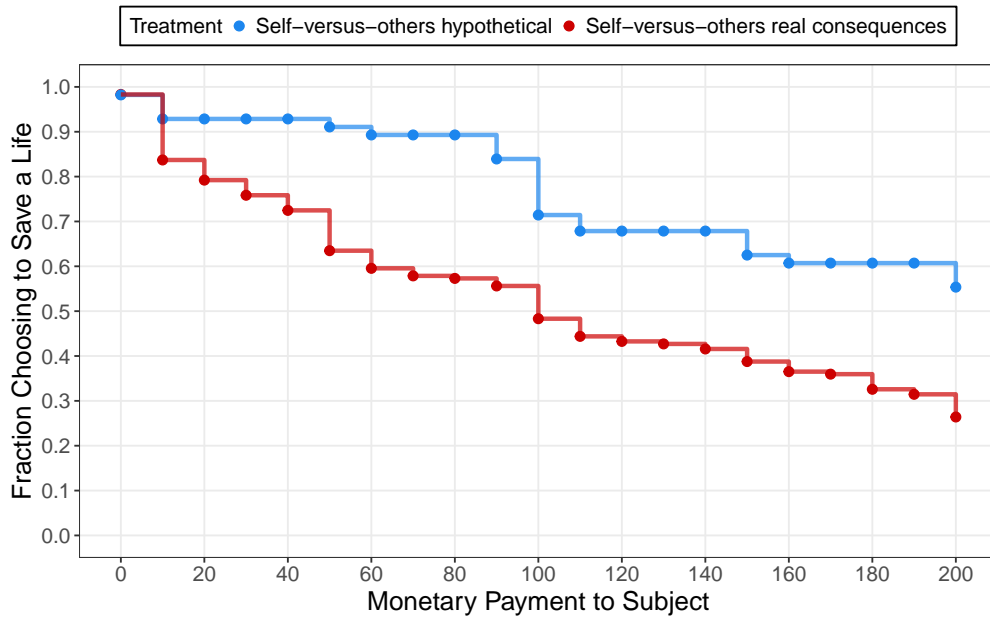
**Design.** We used the same Saving a Life paradigm as employed in the main experiment. However, instead of the trolley's ends-versus-means tradeoff, we employed a self-versus-other tradeoff: subjects could choose whether to trigger a donation sufficient to save one human life or take money for themselves as additional payment. Subjects faced this tradeoff multiple times in the form of a price-list design. In 21 contingent choices starting with 0 Euro and proceeding in 10 Euros increments up to 200 Euros, subjects could indicate whether they want to save a life or take the respective money for themselves.

**Treatment.** There were two between-subject treatments. In the *self-versus-other real* treatment<sup>14</sup>, subjects' choices had the chance to be implemented with real consequences. Out of each session (consisting of 20-24 subjects), two subjects were drawn. For these selected subjects, one price from the price list was randomly drawn, and their pre-stated choices for the drawn price were implemented. Therefore, in this treatment, with an 8% to 10% chance, subjects either triggered the donation or received up to 200 Euros through their decisions. We subsequently ran another treatment, called *self-versus-other hypothetical*, in which subjects' choices were purely hypothetical. In total, 178 subjects took part in the *self-versus-other real* treatment (Bénabou et al., 2022), and we recruited 56 subjects for the treatment *self-versus-other hypothetical*.

**Results.** Introducing real consequences to this self-versus-other tradeoff resulted in significantly different choice distributions ( $p < 0.001$ , Kolmogorov-Smirnov test). As displayed in Figure 4.B.4, for all prices except 0 and 10 Euros, real consequences lead to a significant decrease in the likelihood that subjects choose to save a life instead of taking the offered amount. For instance, at 100 Euros, 71% of subjects

14. In Bénabou et al. (2022), this treatment is called *MPL Low Image*.

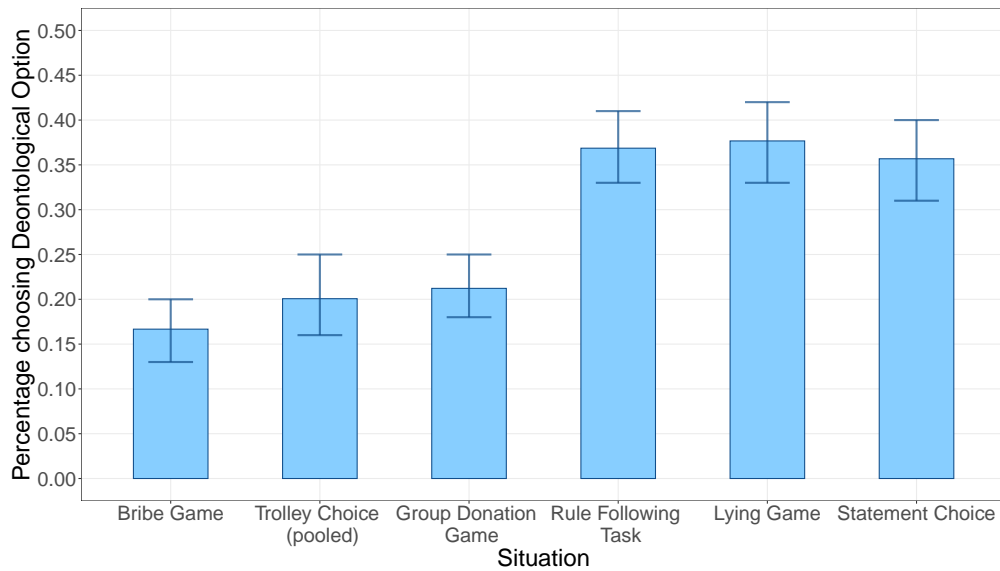
choose to save a life when consequences are hypothetical, while only 48% do so when choices might have real consequences ( $p = 0.003$ , two-sided Fisher's exact test). Consequently, introducing real consequences through a probabilistic payment system leads to markedly different results in self-versus-other tradeoffs.



**Figure 4.B.4.** The effect of changing consequences on behavior in self versus other tradeoff situations

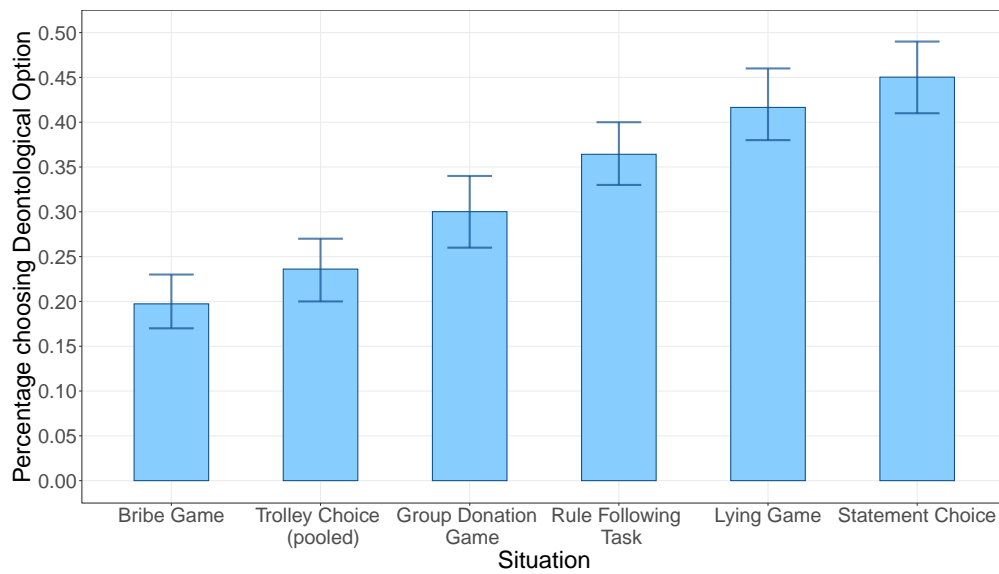
Notes: The figure displays the fraction of subjects choosing to save a life for each offered price separately for the *self-versus-other hypothetical* and *self-versus-other real consequences* treatments.

### Appendix 4.C Robustness analysis ends-versus-means decisions



**Figure 4.C.1.** Distribution of deontological and consequentialist decision-making in ends-versus-means tradeoff situations among high comprehension subjects

Notes: Each bar displays the fraction of subjects choosing the deontological option in ends-versus-means tradeoff situations using only subjects that answer all comprehension and recall questions correctly. See Section 4.2.1 for details on the situations and the definition of the deontological option. Error bars indicate 95% confidence intervals using one-sample tests of proportions.



**Figure 4.C.2.** Distribution of deontological and consequentialist decision-making in ends-versus-means tradeoff situations using the full sample

*Notes:* Each bar displays the fraction of subjects choosing the deontological option in ends-versus-means tradeoff situations using the full sample of 593 subjects. See Section 4.2.1 for details on the situations and the definition of the deontological option. Error bars indicate 95% confidence intervals using one-sample tests of proportions.

## **Appendix 4.D Robustness between blocks**



**Table 4.D.1.** Correlations of ends-versus-means, self-versus-other and questionnaire measures comprehension sample

	Ends-versus-means decision situations						Self-versus-others decision situations									Questionnaires									
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons.- Int.	Reward Int.	Reward Cons.	Classical Trolley	OUS IB	OUS IH	MU Scale	MFQ Harm	MFQ Fairness	MFQ Ingroup	MFQ Authority	MFQ Purity	
SAL Trolley	0.00	-0.02	0.05	0.17	0.10	-0.01	0.00	0.01	-0.07	-0.05	0.00	0.00	0.03	-0.03	0.56	-0.13	-0.30	0.00	-0.07	-0.05	-0.04	0.02	0.06		
Group Don.		0.08	0.00	-0.14	-0.05	0.03	0.05	0.06	0.13	0.08	0.01	0.14	0.04	0.07	0.02	0.11	0.03	-0.01	0.10	0.00	0.02	-0.04	0.06		
Bribe			0.09	0.04	-0.08	0.00	0.01	-0.07	-0.01	-0.08	-0.07	-0.09	-0.08	-0.11	0.01	-0.07	0.07	0.00	-0.06	-0.07	0.03	0.00	0.06		
Statement Choice				0.08	0.07	0.03	0.05	-0.13	0.02	-0.03	0.00	0.03	-0.12	-0.06	0.01	0.04	-0.15	-0.08	0.03	0.05	0.09	0.06	0.10		
Lying					0.04	-0.05	-0.09	-0.04	-0.04	0.01	-0.03	0.03	-0.03	0.03	0.15	-0.22	-0.10	-0.05	-0.06	-0.05	-0.05	-0.02	-0.04		
Rule following						0.08	0.04	-0.01	0.09	0.10	0.04	0.10	0.01	0.11	-0.02	0.05	-0.05	-0.12	0.00	-0.03	-0.14	-0.06	-0.07		
Altruism							0.66	0.16	0.32	0.32	0.43	0.30	0.26	0.20	-0.01	0.24	0.01	-0.16	0.21	0.16	-0.05	-0.02	0.00		
Altruism Taking								0.11	0.25	0.34	0.37	0.25	0.27	0.22	0.00	0.26	0.01	-0.15	0.23	0.12	-0.06	-0.01	-0.04		
Trust									0.32	0.42	0.46	0.31	0.31	0.18	-0.01	0.06	0.06	-0.10	-0.04	0.05	-0.01	-0.10	-0.16		
Pos. Recipr. Low										0.72	0.40	0.33	0.21	0.18	-0.04	0.16	0.00	-0.18	-0.04	0.05	-0.06	-0.12	-0.07		
Pos. Recipr. High											0.46	0.37	0.33	0.19	0.00	0.18	0.06	-0.20	0.03	0.02	-0.13	-0.16	-0.18		
PGG Cont.												0.39	0.38	0.27	0.03	0.11	-0.10	-0.23	0.03	0.08	-0.06	-0.06	-0.11		
Reward Cons.- Int.													0.35	0.61	0.03	0.12	-0.03	-0.18	0.02	0.03	-0.13	-0.12	-0.10		
Reward Int.														0.42	0.06	0.07	0.01	-0.17	0.02	0.01	-0.10	-0.07	-0.12		
Reward Cons.															0.09	0.04	-0.05	-0.04	0.05	0.07	-0.09	-0.13	-0.09		
Classical Trolley																-0.22	-0.46	0.00	-0.06	-0.04	-0.06	-0.05	0.01		
OUS IB																	0.24	-0.05	0.38	0.22	0.07	-0.02	0.01		
OUS IH																		0.19	-0.01	-0.03	0.17	0.14	0.04		
MU Scale																			0.09	0.05	0.24	0.14	0.20		
MFQ Harm																				0.51	0.22	0.17	0.26		
MFQ Fairness																					0.18	0.08	0.11		
MFQ Ingroup																						0.55	0.50		
MFQ Authority																							0.53		
MFQ Purity																									

Notes: Colors in heatmap change with each 0.1 correlation magnitude.

**Table 4.D.2.** Correlations of ends-versus-means, self-versus-other and questionnaire measures memory check sample

	Ends-versus-means decision situations						Self-versus-others decision situations								Questionnaires									
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons.- Int.	Reward Int.	Reward Cons.	Classical Trolley	OUS IB	OUS IH	MU Scale	MFQ Harm	MFQ Fairness	MFQ Ingroup	MFQ Authority	MFQ Purity
SAL Trolley		0.05	0.01	0.05	0.16	0.04	0.05	0.00	0.05	0.00	0.02	0.05	0.02	0.10	0.00	0.54	-0.09	-0.31	-0.07	-0.08	0.02	-0.02	-0.01	0.06
Group Don.			0.07	0.07	-0.07	0.00	0.12	0.12	0.04	0.19	0.11	-0.01	0.14	0.04	0.03	0.03	0.06	0.03	-0.03	0.08	0.04	-0.02	-0.05	0.04
Bribe				0.13	0.08	-0.09	-0.01	0.05	-0.05	-0.02	-0.14	-0.06	-0.05	-0.15	-0.12	-0.02	-0.02	-0.10	0.02	0.12	0.05	0.10	0.02	0.05
Statement Choice					0.08	-0.09	0.04	0.08	-0.04	0.06	-0.02	0.09	0.09	-0.07	0.04	0.01	0.04	-0.07	-0.11	0.09	0.09	0.13	0.15	0.13
Lying						0.07	-0.06	-0.10	-0.04	0.03	0.00	-0.05	0.01	-0.03	-0.01	0.11	-0.14	-0.11	-0.10	-0.07	-0.02	-0.08	-0.05	-0.10
Rule following							0.13	0.11	0.01	0.10	0.15	0.04	0.16	0.01	0.11	0.00	0.11	0.00	-0.08	0.11	0.08	-0.12	-0.02	0.01
Altruism								0.65	0.18	0.32	0.38	0.40	0.35	0.30	0.22	0.02	0.28	-0.05	-0.10	0.24	0.25	-0.06	-0.02	-0.01
Altruism Taking									0.12	0.28	0.38	0.33	0.29	0.24	0.24	0.01	0.27	-0.04	-0.08	0.23	0.18	-0.02	0.00	-0.01
Trust										0.31	0.41	0.45	0.39	0.24	0.21	0.01	0.05	0.01	-0.08	0.03	0.08	0.01	-0.13	-0.09
Pos. Recipr. Low											0.72	0.34	0.38	0.12	0.16	-0.01	0.16	0.02	-0.13	0.13	0.12	-0.02	-0.05	-0.02
Pos. Recipr. High												0.46	0.48	0.29	0.25	0.03	0.21	0.03	-0.14	0.15	0.16	-0.13	-0.14	-0.10
PGG Cont.													0.44	0.34	0.25	0.02	0.14	-0.15	-0.21	0.06	0.11	0.00	-0.06	-0.07
Reward Cons.- Int.														0.33	0.56	-0.02	0.16	-0.01	-0.12	0.12	0.13	-0.11	-0.12	-0.08
Reward Int.															0.46	0.13	0.09	0.00	-0.07	0.03	0.05	-0.05	-0.08	-0.08
Reward Cons.																0.07	0.08	0.03	-0.01	0.08	0.09	-0.14	-0.11	-0.07
Classical Trolley																	-0.25	-0.47	-0.05	-0.08	0.02	-0.09	-0.05	0.02
OUS IB																		0.23	-0.08	0.40	0.24	0.06	-0.06	0.02
OUS IH																			0.17	-0.03	-0.04	0.14	0.13	0.05
MU Scale																				-0.02	-0.01	0.19	0.14	0.20
MFQ Harm																					0.54	0.16	0.12	0.26
MFQ Fairness																						0.08	0.00	0.12
MFQ Ingroup																							0.54	0.48
MFQ Authority																								0.52
MFQ Purity																								

Notes: Colors in heatmap change with each 0.1 correlation magnitude.

**Table 4.D.3.** Correlations of ends-versus-means, self-versus-other and questionnaire measures excluding speeders

	Ends-versus-means decision situations						Self-versus-others decision situations									Questionnaires									
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons.- Int.	Reward Int.	Reward Cons.	Classical Trolley	OUS IB	OUS IH	MU Scale	MFQ Harm	MFQ Fairness	MFQ Ingroup	MFQ Authority	MFQ Purity	
SAL Trolley		0.03	0.06	0.06	0.17	0.06	-0.01	-0.03	0.02	-0.05	-0.03	-0.02	-0.02	0.07	-0.01	0.52	-0.15	-0.29	-0.02	-0.07	-0.03	-0.06	-0.01	0.06	
Group Don.			0.13	0.14	-0.06	-0.08	0.07	0.04	-0.04	0.13	0.05	-0.01	0.07	-0.02	-0.01	0.01	0.09	0.03	0.00	0.10	0.03	0.09	0.02	0.09	
Bribe				0.16	0.06	-0.09	0.01	0.03	-0.09	-0.05	-0.12	-0.07	-0.02	-0.16	-0.14	-0.01	-0.03	-0.04	0.03	0.06	0.00	0.09	0.06	0.16	
Statement Choice					0.06	-0.09	0.04	0.05	-0.12	0.04	-0.08	0.00	0.07	-0.16	0.00	0.02	0.03	-0.08	-0.01	0.08	0.01	0.17	0.15	0.18	
Lying						0.05	0.04	0.05	-0.05	-0.02	-0.01	-0.05	0.01	-0.02	-0.02	0.15	-0.17	-0.10	-0.06	-0.06	-0.03	-0.08	-0.03	-0.09	
Rule following							0.10	0.06	0.00	0.10	0.09	0.00	0.08	-0.03	0.11	-0.01	0.07	0.00	-0.07	0.02	-0.03	-0.08	0.01	-0.03	
Altruism								0.64	0.16	0.30	0.35	0.40	0.33	0.22	0.18	-0.03	0.21	0.01	-0.13	0.19	0.17	0.04	0.02	0.02	
Altruism Taking									0.12	0.25	0.33	0.37	0.26	0.18	0.17	-0.03	0.19	0.01	-0.12	0.20	0.14	0.02	0.01	-0.04	
Trust										0.32	0.42	0.46	0.36	0.27	0.21	-0.01	0.01	0.05	-0.12	-0.03	0.06	-0.04	-0.11	-0.17	
Pos. Recipr. Low											0.69	0.33	0.32	0.11	0.11	-0.03	0.16	0.04	-0.14	0.06	0.07	-0.03	-0.08	-0.05	
Pos. Recipr. High												0.43	0.39	0.28	0.16	0.00	0.17	0.07	-0.16	0.06	0.10	-0.14	-0.15	-0.16	
PGG Cont.													0.40	0.31	0.23	0.01	0.12	-0.10	-0.26	0.09	0.13	-0.01	-0.05	-0.06	
Reward Cons.- Int.														0.22	0.54	-0.02	0.06	-0.01	-0.13	0.03	0.05	-0.09	-0.11	-0.11	
Reward Int.															0.39	0.08	0.00	-0.01	-0.10	-0.04	0.04	-0.11	-0.07	-0.16	
Reward Cons.																0.05	-0.01	0.01	-0.01	0.02	0.04	-0.10	-0.06	-0.10	
Classical Trolley																	-0.23	-0.44	-0.02	-0.03	0.02	-0.10	-0.06	-0.02	
OUS IB																		0.25	-0.06	0.35	0.25	0.08	-0.04	0.02	
OUS IH																			0.16	-0.03	-0.04	0.16	0.15	0.06	
MU Scale																				0.02	0.03	0.23	0.15	0.20	
MFQ Harm																					0.60	0.24	0.16	0.24	
MFQ Fairness																						0.18	0.08	0.15	
MFQ Ingroup																							0.56	0.51	
MFQ Authority																								0.54	
MFQ Purity																									

Notes: Colors in heatmap change with each 0.1 correlation magnitude.



**Table 4.D.5.** Correlations between ends-versus-means and self-versus-other blocks using median splits for the later variables

	Ends-versus-means decision situations						Self-versus-others decision situations								
	SAL Trolley	Group Don.	Bribe	Statement Choice	Lying	Rule following	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons. comp Int.	Reward Int.	Reward Cons.
SAL Trolley		0.05					-0.02	-0.01	-0.02	-0.08	-0.02	0.01	-0.03	-0.03	-0.07
Group Don.			0.14	0.15	-0.09	-0.06	0.02	0.05	-0.03	0.13	0.00	-0.02	0.09	-0.04	-0.03
Bribe				0.19	0.04	-0.08	0.00	0.01	-0.10	-0.06	-0.14	-0.03	-0.07	-0.16	-0.20
Statement Choice					-0.06	0.04	0.01	0.01	-0.10	0.05	-0.08	0.01	0.05	-0.13	-0.05
Lying						0.00	0.00	-0.05	-0.02	0.01	0.04	-0.04	0.01	-0.03	-0.02
Rule following							0.13	0.01	0.03	0.14	0.06	0.06	0.05	-0.02	0.08
Altruism								0.52	0.17	0.28	0.26	0.40	0.32	0.23	0.26
Altruism Taking									0.12	0.22	0.21	0.27	0.28	0.22	0.24
Trust										0.22	0.34	0.42	0.32	0.09	0.17
Pos. Recipr. Low											0.55	0.24	0.31	0.18	0.23
Pos. Recipr. High												0.30	0.32	0.27	0.25
PGG Cont.													0.38	0.25	0.21
Reward Cons.- Int.														0.15	0.45
Reward Int.															0.38
Reward Cons.															

**Table 4.D.6.** Biserial correlations between ends-versus-means and self-versus-other blocks

	Altruism	Altruism Taking	Trust	Pos. Recipr. Low	Pos. Recipr. High	PGG Cont.	Reward Cons. comp Int.	Reward Int.	Reward Cons.
SAL Trolley	-0.01	-0.04	0.03	-0.07	-0.04	0.01	-0.03	0.07	-0.03
Group Don.	0.06	0.04	-0.05	0.18	0.06	-0.03	0.11	-0.04	-0.02
Bribe	-0.05	-0.03	-0.16	-0.08	-0.21	-0.09	-0.08	-0.26	-0.25
Statement Choice	0.00	0.01	-0.15	0.05	-0.09	-0.01	0.07	-0.19	-0.05
Lying	-0.03	-0.05	-0.03	-0.03	0.00	-0.06	0.00	-0.03	-0.01
Rule following	0.15	0.08	0.04	0.14	0.13	0.04	0.15	-0.01	0.16

## **Appendix 4.E Instructions**

### **4.E.1 General information**

#### **Welcome to the study**

Welcome, and thank you for your interest in today's online study! Please note that you may only participate in this study once. Also, you may only participate if you have registered for this study in our participation database.

As already announced during registration, this study consists of two parts. You will complete the second part exactly in one week at the same time. For completing both parts in full, you will receive a lump sum of 12 EUR upon completion of the study.

You will make decisions on the computer in this study. Your decisions will allow you to earn extra money. All payments, i.e., both the compensation for your participation and any additional payments based on your decisions, will be sent to you by bank transfer following the second part.

Please note: You will receive the payment for your participation of 12 EUR as well as any additional money based on your decisions only if you successfully complete both parts. Therefore, please make sure to show up for the second part next week as well and complete it entirely.

#### **Structure of the study and your payoff**

In the following, you will make different decisions in each of several successive sections. The decisions in each section are completely independent of each other. This means your decisions in one section do not affect the consequences or possible payoffs of another section. Furthermore, similar-sounding decision situations in different sections do not mean that your decisions should necessarily be similar as well. You will see the number of the current section in the upper right corner of each section.

From all the decisions you will make today, a computer will randomly select one decision. Each of the decisions has the same chance of being selected. This selection is independent of your decisions.

The consequence of the selected decision will then be implemented exactly as described in the decision. Since one of your decisions will actually be implemented, you should consider each decision carefully and treat each decision as if it were

actually implemented.

In what follows, this is always implicitly assumed. For example, suppose a decision is described as giving you an amount of money as a consequence. In that case, you will receive that amount exactly as described if that decision is indeed randomly selected.

### **Donation to the *Förderkreis***

Some decisions in today's part are about the possibility of making a donation to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* (in short: the *Förderkreis*). The *Förderkreis* is a regional aid organization from Bonn.

### **Information about the *Förderkreis***

The *Förderkreis* comprehensively supports young people with cancer and their families in dealing with the disease. The organization offers psychological support, organizes leisure activities as well as aftercare, and offers school support. The projects and participations of the *Förderkreis* at a glance are:

- Team Bärenstark: Practical help during the treatment period for patients and relatives.
- Psychological and psychosocial counseling for patients and relatives
- Pedagogical support in the hospital
- Financing of hospital clowns and music therapy
- Follow-up care by the KoCkPiT team (Children's Palliative Care Team Bonn)
- Support for parents of sick children and adolescents
- Support for orphaned parents

So if in the following, a donation to the *Förderkreis* is mentioned, it is always a donation to the organization described above, with which the listed projects are supported.

If in the following a donation is initiated by your decisions, it will always be paid by the scientists involved from study funds. So you do not have to donate (or pay) any amount of money personally. If you decide against a donation, the amount will not be donated accordingly.

### **Consequences for other study participants**

In some sections, your decisions have consequences for other study participants. For example, in some choices, you can send money to a randomly selected other



person who is also participating in this study.

**Important:** The other people will not learn about your decisions or personal information about you at any time. Notably, the other individuals will also receive payouts through other situations, which are random in amount. This ensures that the other people cannot draw any conclusions about your decisions from the amount of the payoffs at the end of the trial.

On the next page, the individual sections begin.

## 4.E.2 Ends-versus-means block

### 4.E.2.1 Trolley problem

#### The consequences of the next decision

[Treatment *Hypothetical*]

For the next section of this study, the decisions are hypothetical in nature. This means that the consequence of your decision in the next section will not be implemented in real terms. Thus, the following section is purely a thought experiment.

[Treatment *Real*]

For your decision in the next section, there is a probability of 1 in 10, i.e., 10%, that it will be implemented as described. At the end of the study, a computer-generated random number will determine whether your decision will actually be implemented. If so, the consequences of your decision will be realized exactly as described. Therefore, since your decision may actually be implemented, you should think carefully about the decision.

#### Information

Below you will read important information that is relevant to your decision later. After you have made your decision, we will ask you four simple questions about the information and instructions presented below. For each question you then answer correctly, you will receive an additional 0.5 EUR.

Please read all the information carefully. Not only is the information important for your decision, but you will also potentially receive a higher payout.

### **Information about tuberculosis**

#### **What is tuberculosis?**

Tuberculosis – also called Phthisis or White Death – is an infectious disease, which is caused by bacteria. Roughly one-third of all humans are infected with the pathogen of tuberculosis. Active tuberculosis breaks out among 5 to 10% of all those infected. Tuberculosis is primarily airborne. This is also why quick treatment is necessary.

#### **What are the symptoms and consequences of tuberculosis?**

Tuberculosis patients often suffer from very unspecific symptoms like fatigue, the feeling of weakness, lack of appetite, and weight loss. At an advanced stage of lung tuberculosis, the patient coughs up blood, leading to the so-called rush of blood. Without treatment, a person with tuberculosis dies with a probability of 43%.

#### **How prevalent is tuberculosis?**

In the year 2018, about 10 million people have been recorded as falling ill with active tuberculosis. Almost 1.5 million people die of tuberculosis each year. This means that tuberculosis causes more annual deaths than HIV or malaria.

#### **Is tuberculosis curable?**

According to the World Health Organization (WHO), the United Nations agency for international public health, “tuberculosis is preventable and curable”. Treatment takes place by taking antibiotics several times a week over a period of 6 months. It is important to take the medication consistently. Since 2000, an estimated 53 million lives have been saved through effective diagnosis and treatment of tuberculosis.

The success rate of treatment for a new infection is usually over 85%.

The preceding figures and information were provided by WHO and are freely available. For more details click [here](#) (link opens in new tab).

### **Operation ASHA**

*Operation ASHA* is a charity organization specialized since 2005 on treating tuberculosis in disadvantaged communities. The work of *Operation ASHA* is based on the insight that the biggest obstacle for the treatment of tuberculosis is the interruption of the necessary 6-month-long regular intake of medication.

For a successful treatment, the patient has to come to a medical facility twice a week – more than 60 times in total – to take the medication. Interruption or termination of the treatment is fatal because this strongly enhances the development of a drug-resistant form of tuberculosis. This form of tuberculosis is much more difficult to treat and almost always leads to death.

### **The concept of *Operation ASHA***

To overcome this problem, *Operation ASHA* developed a concept that guarantees regular treatment through immediate spatial proximity to the patient. A possible non-adherence is additionally prevented by visiting the patient at home.

By now, *Operation ASHA* runs more than 360 treatment centers, almost all of which are located in the poorer regions of India. More than 60,000 sick persons have been identified and treated that way.

*Operation ASHA* is an internationally recognized organization, and its success has been covered by the New York Times, BBC, and Deutsche Welle, for example. The MIT and the University College London have already conducted research projects about the fight against tuberculosis in cooperation with *Operation ASHA*. The treatment method employed by *Operation ASHA* is described by the World Health Organization (WHO) as “highly efficient and cost-effective”.

### **The impact of a donation to *Operation ASHA***

By donating money to *Operation ASHA*, it is possible to save people from death by tuberculosis.

To save a person’s life means here to successfully cure a person with tuberculosis, who otherwise would die because of the tuberculosis. A donation of 380 EUR ensures that at least one human life can be expected to be saved. The information used to calculate the donation amount is obtained from public statements from the World Health Organization (WHO), peer-reviewed research studies, Indian Government statistics, and published figures from *Operation ASHA*.

In the calculation, information was conservatively interpreted, or a pessimistic number was used so that the donation amount of 380 EUR is in the case of doubt higher than the actual costs to save a human life. In addition, in the calculation of the treatment success rate of *Operation ASHA*, the mortality rate for alternative treatment by the state tuberculosis program in India and the different detection rates for new cases of tuberculosis are included.

In the context of this study, an agreement made with *Operation ASHA* will ensure that 100% of the donation will be used exclusively for the diagnosis and treatment of tuberculosis patients. This means that every Euro of the donation amount goes directly to saving human lives, and no other costs will be covered. Based on a very high number of cases, the contribution of a donation of 380 EUR can be summarized as follows:

With a donation of 380 EUR 5 additional patients infected with tuberculosis can be treated through *Operation ASHA*.

If these 5 persons are not treated through *Operation ASHA*, it is expected that one patient will die.

If, through the donation of 380 EUR all 5 patients are treated, it is expected that no patient will die.

Based on these calculations it can be concluded that the donation of 380 EUR will save one human life. Going further, this also means that two lives are saved by a donation of 760 EUR, and three lives are saved by a donation of 1140 EUR.

The relationship of a donation of 380 EUR to *Operation ASHA* and a human life is illustrated in the following graph: [Figure here]

So, in summary, the following can be said about the effect of a donation for a given number of 5 patients suffering from tuberculosis:

- Donation results in saving one person from death.
- Failure to donate results in the death of one person.

Whenever the life or death of a person is mentioned in connection with a donation, the above-mentioned facts are implicitly meant. The same applies when the life or death of several people is involved.

### **The concrete situation**

As mentioned before, *Operation ASHA* operates numerous treatment centers in India. In particular, *Operation ASHA* operates in the two Indian states of [State A] and [State B].

#### **[State A]**

In this experiment, 5 people were identified in [State A] who are infected by tuberculosis. For these people, the donation amount required to save a life was initiated to *Operation ASHA*. Through this, only the people who are sick in [State A] will be treated. This will result in a person from [State A] being saved from death by tuberculosis.

If the initiated donation amount were not realized, a person in [State A] would die from tuberculosis.

**[State B]**

After the donation was initiated, 15 people suffering from tuberculosis were identified in [State B]. Without donation and subsequent treatment by *Operation ASHA*, three of the people from [State B] who became ill would die from tuberculosis.

In the event that the people in [State B] are helped through the donation to *Operation ASHA* necessary to save three lives, these three people would be saved from death by tuberculosis.

**Redirecting the donation**

The initiated donation to patients in [State A] can be redirected to help save patients in [State B] from death by tuberculosis. The amount would be multiplied to save three lives in [State B].

**Summary**

Tuberculosis is a worldwide common bacterial infectious disease. The success rate of medical treatment of a new disease is very high. Nevertheless, close to 1.5 million people die every year from tuberculosis. The biggest obstacle to the curing of tuberculosis is the potential stopping of continuous treatment with antibiotics. The concept of *Operation ASHA* is therefore based on the immediate proximity to the patient as well as the control and recording of the regular intake of medication. A donation of 380 EUR to *Operation ASHA* can save a human life. A donation of 760 EUR can save two lives, and a donation of 1140 EUR can save three lives.

**How is the donation connected to the saving of a life?**

The donation of 380 EUR already accounts for the fact that someone inflicted with the illness could have survived without treatment by *Operation ASHA*; i.e., instead of through *Operation ASHA*, they could have received treatment through other actors (such as the public health system). The amount is, therefore, sufficient for the diagnosis and complete treatment of multiple sufferers.

**What does it mean to “save a life”?**

To save a life means here the successful curing of a person suffering from tuberculosis, who otherwise would die because of the Tuberculosis. This means: The amount donated is sufficient to identify and treat enough people with tuberculosis that there is at least one person among them who would otherwise have died from tuberculosis in expectation. To summarize:

- A donation has the effect of saving a person from death.
- Failure to donate results in the death of a person.

### **The initial situation**

In [State A], 5 people suffering from tuberculosis were identified. For these, the necessary donation amount was initiated to save one person from death due to tuberculosis. Thus, without further intervention, none of these diseased people will die of tuberculosis.

After initiating this donation, an additional 15 people suffering from tuberculosis were identified in [State B]. Without further intervention, three of these diseased people would die of tuberculosis. If the initiated donation amount is redirected, it will be multiplied to save three people in [State B] from dying of tuberculosis.

### **Your decision**

So, in terms of the context just described, the situation is as follows:

Without intervention, three people in [State B] will die of tuberculosis. By intervening, these three people can be saved. However, as a direct consequence of this intervention, one person in [State A] will die of tuberculosis.

You will have the opportunity to choose between intervening and not intervening in this situation:

#### **Do not intervene**

Do not intervene: do not divert the donation.

Consequence: You decide not to intervene. This means that you decide against an intervention on your end that would cause a patient in [State A] to die from tuberculosis. In return, three people from [State B] will not be saved from dying from tuberculosis.

#### **Intervene**

Intervene: redirect the donation.

Consequence: You decide to intervene. This means that you decide in favor of an intervention that causes a patient in [State A] to die from tuberculosis. In return, your redirected donation will save three people from [State B] from dying of tuberculosis.

In a moment, you will have the opportunity to make this decision using an animation. The next page will explain how this animation works.

### **The decision animation**

In a moment, you will have the opportunity to make your decision with the help of an animation. The initial situation of the animation is shown in the following figure:

[animation figure]

The animation depicts the decision situation just described. As soon as the animation starts, the skull shown moves slowly to the right towards the three people. In case you don't see three people on the right or they are outside the white border, please click [here].

You can change the path of the skull using the slider located to the left of the junction (blue line). You do this by dragging the slider with your computer mouse all the way down and then releasing it. This is necessary because otherwise, the slider will move back to its original position. If the slider is dragged all the way down and released, the downward junction will open so that the skull on the track will move down to the single person.

You can make your decision before the skull reaches the dashed gray line in front of the junction. After arriving at the line, moving the slider will have no effect.

Hence, you have two choices, with the consequences described earlier:

#### **Do not intervene**

Do not intervene: do not move the slider, the junction will not open.

Consequence: You decide not to intervene. This means that you decide against an intervention on your end that would cause a patient in [State A] to die from tuberculosis. In return, three people from [State B] will not be saved from dying from tuberculosis.

#### **Intervene**

Intervene: move the slider, the junction will open.

Consequence: You decide to intervene. This means that you decide in favor of an intervention that causes a patient in [State A] to die from tuberculosis. In return, your redirected donation will save three people from [State B] from dying of tuberculosis.

On the following screen, you will have the opportunity to familiarize yourself with the controller. You will be able to make your actual decision afterward.

#### **Testing the slider**

You will now have the opportunity to try out the slider of the animation. Note that you need to drag the slider all the way down and then release it to open the branch.

[slider test]

Click “Next” when you are sufficiently familiar with the slider.

On the next screen you can now make your decision. Note that the animation will start directly.

As a reminder:

[Treatment *Hypothetical*]

As previously described, this is a hypothetical decision that is not actually implemented.

[Treatment *Real*]

As previously described, your decision may actually be implemented, resulting in actual consequences.

[Trolley decision]

#### **4.E.2.2 Baseline choice**

In this section, you have a choice between two options, Option A and Option B. The two options differ in whether a sum of money is donated to a charity or another study participant receives a sum of money. The two options are as follows:

##### **Option A**

If you choose option A, 15 EUR will be donated to the *Förderkreis für krebserkrankte Kinder und Jugendliche Bonn e.V.*

##### **Option B**

If you choose option B, another study participant will receive 2 EUR, and no money will be donated to the *Förderkreis*. The other study participant will not receive any additional information about your decision.

#### **4.E.2.3 Lying game**

In this section, you have been paired with another person. This person, like you, is a participant in this study.

##### **Options**

In this section, there are two possible payout options that you and the other person can achieve at the end of the study. The two options are:



**Option A**

15 EUR will be donated to the *Förderkreis* für krebskranke Kinder und Jugendliche Bonn e.V.

**Option B**

The other person receives 2 EUR and no money is donated to the *Förderkreis* für krebskranke Kinder und Jugendliche Bonn e.V.

The other person can choose between these two options, that is, between option A or option B. **However, the other person does not know what the exact consequences of the options are. In particular, the other person does not know which option leads to which payout.** Only you know what the consequences of the options are. The only information that the other person will have prior to their decision is a single message that you can send to the person.

**Your decision in this section**

You have a choice between the following two messages that you can send to the other person:

**Message 1**

“Option A gives you a higher personal cash payment than Option B”.

**Message 2**

“Option B gives you a higher personal cash payment than option A”.

The sent message will be displayed to the other person exactly as shown above. In addition, the person will not learn which payoffs were associated with which option (also not afterward). Since the person will receive additional payments from other tasks in this study, as described at the beginning of the study, the person will also not learn whether your message was right or wrong.

The person was told that the message received could be true or false and that he or she would not receive any information about what the consequences of the options were. In such a situation, the other person follows the action prescribed in the message in almost all cases (>90%). Thus, you can assume with a very high probability the following:

- If you send message 1, the other person will choose option A (without knowing the exact consequences).
- If you send message 2, the other person will choose option B (without knowing the exact consequences).

Hence, the sequence of events is the following: You decide which of the two messages you want to send. Based on that, the other person will choose the respective option. The other person's decision will then determine the corresponding payments that will be made at the end of the study.

### Summary

In summary, by choosing the message, you can have an impact on the payments and donations that are triggered. The other person is aware that (1) the message may be right or wrong, (2) they will not know which payoffs were associated with which option and whether the message was right or wrong.

[Comprehension questions]

#### 4.E.2.4 Bribe game

In this section, you have been matched with another person. Like you, this person is a participant in the current study. In this section, you have a choice between two options, Option A and Option B. The two options are as follows:

##### Option A

If you choose Option A, you will help children with cancer. Your choice of Option A will trigger a donation of 15 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.*

##### Option B

If you choose option B, you will send the other person 2 EUR. However, in this case, the donation otherwise triggered by option A will not be given to *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.*

##### Additional consequence

The other person also has the choice between two options, namely between option C and option D. If the other person chooses option C, an additional 20 EUR will be donated to the *Förderkreis*. If the person chooses option D, the person will receive 2 EUR for themselves as an additional payout.

The person has chosen to make a choice between the two options, C and D, dependent on **your decision** between A and B. This means the other person has **obligatorily** specified the following: The person will choose option D and thus the additional payout for himself. The choice of option D can only be prevented if you choose option B and send the other person 2 EUR. In this and only in this case, the other person will choose option C and thus trigger the donation. Thus, the consequences are as follows:

- You choose option A  $\Rightarrow$  The other person chooses option D.
- You choose option B  $\Rightarrow$  The other person chooses option C.

In other words, if you choose option A and thus trigger a donation of 15 EUR, the other person will stick with option D and receive 2 EUR. If you choose to send the 2 EUR to the other person by choosing option B, the other person will choose option C and choose the donation of 20 EUR.

[Comprehension questions]

#### 4.E.2.5 Group decision game

##### Information

##### Your group

Together with 5 other people, you will form a group for your decisions. These people also participate in this study, so together, you will form a group of 6 individuals. You can make all decisions autonomously and independently of the other members of the group. However, as described below, the consequences of your decision will depend on the decisions of the other group members.

##### The donation

Your group has been entrusted with a donation of 15 EUR, which is intended for the *Förderkreis* für krebserkrankte Kinder und Jugendliche Bonn e.V. and is meant to be donated to the charity after the study.

##### Your options

You can choose between two options: Option A and Option B. Depending on which option you and the other group members choose, different consequences will be realized. In particular, both your decision and the decisions of the other group members affect what happens to the donation of 15 EUR to the *Förderkreis*.

The donation to the *Förderkreis* will be destroyed (that is, not executed) if at least one member of your group chooses option B. That is, if either you or at least one other person with whom you form a group chooses option B, the donation will be destroyed. Only if no member of your group chooses option B, i.e., **all** group members choose option A, will the donation not be destroyed.

To summarize, the consequences of choosing Option A and Option B are as follows:

##### Option A

You choose to help children with cancer through the donation. However, the

donation will only realize if all other members of your group choose Option A.

### **Option B**

You decide against helping children with cancer through the donation. Since the donation will only be made if all group members choose option A, this will destroy the donation in any case.

Furthermore, by choosing option B, you generate an additional payoff of 2 EUR for another participant in this study who is not part of your group. Any other person in your group who chooses option B will receive a sum of money of 2 EUR for themselves.

### **Decisions of the other group members**

Hence, the consequences of your choice depend not only on you but also on the decisions of the other 5 members of your group: only if none of the members of your group chooses option B, will the donation be made. On the other hand, the payouts of 2 EUR to other people will be made regardless of the decisions of other group members.

You and the other 5 Members of your group decide one after the other. You are the **last** person to decide, so the other group members have already decided at this point. When it's your turn, you will learn if anyone among the people who decided before you has already chosen option B. In this case, the donation will be destroyed.

### **Summary**

If you choose option A, you will not generate any additional payout. If you choose option B, you generate an additional payout of 2 EUR for another participant. Whether the donation of 15 EUR to the *Förderkreis* is destroyed depends on whether at least one member of your group has chosen option B. So the consequences of your choice depend not only on you but also on the choices of the other group member.

[Comprehension questions]

#### **4.E.2.6 Statement choice**

##### **Context**

Carbon dioxide ( $CO_2$ ) is a gas that is notably produced as a by-product of human-controlled combustion processes of fossil fuels such as crude oil, natural gas, or coal. One example of this is the generation of electricity in coal-fired power plants or driving a car. As a result,  $CO_2$  enters the atmosphere, which leads to global

warming (greenhouse effect). The increasing  $CO_2$  concentration in the atmosphere, therefore, contributes significantly to man-made climate change and, thus, to the advancing destruction of the environment.

In the context of this study, it was planned to buy  $CO_2$  certificates, which ensure that **one ton of  $CO_2$**  is removed from the atmosphere. More precisely, this entails an amount of 25 EUR, which will be used for reforestation. The reforestation afforded by this amount of money results in the removal of one ton of  $CO_2$  from the atmosphere.

### **Consequences**

At the end of the study, a computer will randomly decide what will happen to the planned  $CO_2$  certificates. The following two possibilities exist:

- With a probability of 50%, the certificates will be purchased as planned, thus helping the environment by removing one ton of  $CO_2$  from the atmosphere.
- With a probability of 50%, the certificates will be destroyed, i.e., not acquired. This contributes to the destruction of the environment by leaving one ton of  $CO_2$  in the atmosphere.

**This selection is made by the computer in a completely random fashion. Hence it is not influenced by any actions or decisions of yours at any time.**

In this context, you can make a statement of intent about how you feel about the destruction of the environment. Namely, you can make one of the following two statements:

- I support the preservation and protection of the environment.
- I support the destruction of the environment.

If you make the statement “I support the preservation and protection of the environment” this has no further consequence.

However, if you declare “I support the destruction of the environment” in the case that the certificates are destroyed, a part of the amount will be donated to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* instead. Namely, in this case, 15 EUR will be donated to the *Förderkreis*.

Thus, selecting a statement does not influence whether the  $CO_2$  allowances are destroyed or not (since this happens randomly). But the statement influences whether 15 EUR will be donated to the *Förderkreis*. Please note that selecting a statement is not an opinion poll. Specifically, regardless of which statement you

choose, the results will not be used to influence opinions about the environment.

[Comprehension questions]

### **4.E.3 Self-versus-others block**

#### **4.E.3.1 Dictator game giving**

In this section, you will receive 20 EUR as an additional payment. If you take no further action, you will receive the full amount. You can choose to donate a certain amount of the 20 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* You can choose any amount between 0 EUR and 20 EUR. This amount will then not be paid out to you but donated to the *Förderkreis*.

Thus, you will receive an additional payment 20 EUR minus the amount you donate.

#### **4.E.3.2 Dictator game taking**

This section is about a donation of 20 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* If you take no further action, the full amount will be donated. You can choose to take a certain amount of the donation for yourself as an additional payment. You can choose any amount between 0 EUR and 20 EUR. This amount will then not be donated to the *Förderkreis*, but paid to you.

Thus, 20 EUR minus the amount you take for yourself will be donated.

#### **4.E.3.3 Trust game sender**

##### **Information**

In this section, you and another person can send money to each other. Both of you have a private account. Initially, you and the other person have an endowment of 5 EUR in the account. Like you, the other person is a participant in today's study.

##### **Your decision**

From your 5 EUR endowment, you can decide to send an amount to the other person. You can send any whole euro amount, that is 1,2,3,4 or 5 EUR to the other person. In this case, each euro you send to the other person will arrive tripled to the other person. So, for example, if you send 1 EUR, the other person will receive 3 EUR, if you send 2 EUR, 6 EUR and so on.

##### **The decision of the other person**

So after you make your decision, the other person has 5 EUR + triple the amount you sent. Now the other person decides on an amount of money to send back to you.

The person can choose any amount that the account allows for it. After that, the process ends and you will receive your account balances as an additional payment.

### **Example**

You have decided to send 2 EUR to the other person. Therefore, your account balance is  $5 \text{ EUR} - 2 \text{ EUR} = 3 \text{ EUR}$ . The person now has  $5 \text{ EUR} + 3 \times 2 \text{ EUR} = 11 \text{ EUR}$  on the account and can accordingly send you back any amount between 0 EUR and 11 EUR. Suppose the person sends you back 4 EUR.

This will give you a total of  $4 \text{ EUR} + 3 \text{ EUR} = 7 \text{ EUR}$  as an additional payment. The other person will receive  $11 \text{ EUR} - 4 \text{ EUR} = 7 \text{ EUR}$  as a payment.

#### **4.E.3.4 Trust game receiver**

##### **Information**

In this section, you now take the other person's role in the previous section. That is, now the other person decides how much money to send you, and you decide how much to send back. Note that you are sending money to a different participant for this section than before. They are not the same person, and your decision in the previous section has no bearing on your decision in this section.

Both of you again have a private account with an endowment of 5 EUR. Now the other person can decide how much of the 5 EUR they want to send you.

##### **Your decision**

You can indicate what you would send back for different possible amounts the other person can send. Afterward, it will be checked what amount the other person actually sent, and your decision for exactly this case will be implemented.

#### **4.E.3.5 Public goods game**

##### **Your group**

In this section, you will form a group with 2 other people that also participate in today's study. Your group members have been randomly assigned to you. At no time will you or the other members of your group learn who the members of your group are. All group members will face the same decisions and receive the same information.

**Your endowment**

For this section, you and each other group member will receive an amount of 5 EUR which is called your endowment. You decide how to use your endowment. You can contribute all or part of it to a project. Any amount you do not contribute to the project will automatically be put into what is called a private account. For example, if you contribute 3 EUR to the project,  $5 \text{ EUR} - 3 \text{ EUR} = 2 \text{ EUR}$  will be put into your private account.

You will receive income from the project and your private account as described below.

**Income from the private account**

For every Euro that is put into your private account, you will receive exactly 1 EUR as income. For example, if you contribute 0 EUR of your endowment to the project and thus 5 EUR is put into your private account, you will receive 5 EUR as income from your private account. If 3 EUR is put into your private account (because you contributed 2 EUR to the project), you will receive 3 EUR as private income. Nobody but you will receive income from your private account.

**Income from the project**

For every EUR you or another group member contributes to the project, you and all other group members receive 0.50 EUR. So, each group member's income from the project is determined as follows:

Income from the project =  $0.5 * \text{total sum of contributions paid in by all group members}$ .

Example: If the total sum of contributions to the project from all participants in your group is 9 EUR (for example, if you and the two other members each contribute EUR3), you and all other participants in your group will receive  $9 \text{ EUR} * 0.5 = 4.5 \text{ EUR}$  from the project. If the total of your contributions to the project is 12 EUR, you and all other participants will receive  $12 \text{ EUR} * 0.5 = 6 \text{ EUR}$  from the project.

**Total income**

Your total income from part 1 is simply the sum of your income from the private account and your income from the project.



### Simulation

You can try out in the simulation below how your total income changes based on your contribution and the contributions of your group members to the project. To do this, enter whole numbers in the respective fields.

[Calculator]

[Comprehension questions]

#### 4.E.3.6 Moral luck game

##### Information

The following is about the decisions of another person who, like you, is a participant in this study. However, this person faced a different decision than you will be presented. At an earlier time, the person had a choice between the following two lotteries S and G:

**Lottery S** – S for Donation [Spende in German].

With a probability of 70%, a donation of 15 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* is triggered. With a probability of 30%, the person receives 10 EUR as an additional payment and no donation is triggered.

**Lottery G** – G for Money [Geld in German].

With a probability of 30%, a donation of 15 EUR to the *Förderkreis für krebskranke Kinder und Jugendliche Bonn e.V.* is triggered. With a probability of 70%, the person receives 10 EUR as an additional payment and no donation is triggered.

##### Determination of the consequences

At the end of the trial, the computer will randomly draw a number for the person from numbers 1 to 10. Depending on which lottery the person chose, the consequences of the number drawn will be different.

If the person chose lottery S, the donation will be triggered with a probability of 70%, so if either a 1, 2, 3, 4, 5, 6, or 7 is drawn. If the number drawn is an 8, 9, or 10, the person will receive 10 EUR as an additional payment.

If the person chose lottery G, the donation is triggered with a probability of 30%, so if either a 1, 2, or 3 is drawn. If the number drawn is a 4, 5, 6, 7, 8, 9, or 10, the

person will receive 10 EUR as an additional payment.

Thus, the two lotteries differ in their probabilities of triggering the donation and the amount of additional payment to the other person. For Lottery S, a donation is much more likely than for Lottery G. Conversely, the probability of receiving an additional payment is higher for Lottery G than for Lottery S.

The consequences of the lotteries have no influence on your own payments.

### **Your decisions**

In this section you will make decisions for different scenarios. For your decisions, you will be given 5 EUR for each scenario. You have the choice to divide the 5 EUR between you and the person who decided between the lotteries. You will receive every Euro you allocate to yourself as an additional payment. For every Euro you allocate to the other person, the person will subsequently receive 3 EUR as an additional payment.

For example, you can divide the 5 EUR so that you receive 4 EUR. Accordingly, the other person will then receive  $1 \text{ EUR} * 3 = 3 \text{ EUR}$  as an additional payment. If you allocate yourself the entire 5 EUR, the other person will receive 0 EUR. If you allocate the entire amount to the other person you will receive 0 EUR and the other person will receive  $5 \text{ EUR} * 3 = 15 \text{ EUR}$  and so on.

You will be asked how you want to distribute the amount of money from 5 EUR for each of four different scenarios.

### **Scenarios**

The scenarios differ in which lotteries the person chose in each case, as well as which consequence happened to materialize. The following four scenarios exist:

#### **Scenario 1:**

The other person chose Lottery S. The donation was triggered and the person does not receive the additional payment.

#### **Scenario 2:**

The other person chose Lottery S. The donation was not triggered and the person receives the additional payment.

#### **Scenario 3:**

The other person chose Lottery G. The donation was triggered and the person does

not receive the additional payment.

**Scenario 4:**

The other person chose Lottery G. The donation was not triggered and the person receives the additional payment.

For each of the four scenarios, you can decide individually how to distribute the 5 EUR based on the person's decision and the realized consequences. At no point does the other person receive any information about you.

At the end of the trial, you will be informed which scenario occurred by the matching of a person's decision to the consequences the computer has drawn for them.

Afterwards, you and the person then receive the amounts of money you selected for the scenario that occurred.

On the next screen you will see an example.

**Example**

You will see an example decision for Scenario 1. You can make the distribution of 5 EUR between you and the person using a scale and a slider. The associated slider for the scale appears when you click on the scale.

[Example]

If scenario 1 occurs, you would receive [XX] EUR by your decision and the other person would receive [XX] EUR. The other person will receive a total of [XX] EUR, because in this scenario the realized consequence is that the other person will not receive an additional amount of money through the lottery, but will have triggered a donation.

## References

- Alexander, Larry, and Michael Moore.** 2016. *Deontological Ethics*. Edited by Edward N. Zalta. The Stanford Encyclopedia of Philosophy. [196]
- Alger, Ingela, and Jörgen W. Weibull.** 2013. "Homo Moralís—Preference Evolution Under Incomplete Information and Assortative Matching." *Econometrica* 81 (6): 2269–302. [199]
- Alger, Ingela, and Jörgen W. Weibull.** 2016. "Evolution and Kantian Morality." *Games and Economic Behavior* 98: 56–67. [199]
- Alger, Ingela, Jörgen W. Weibull, and Laurent Lehmann.** 2020. "Evolution of Preferences in Structured Populations: Genes, Guns, and Culture." *Journal of Economic Theory* 185: 1–52. [199]
- Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean François Bonnefon, and Iyad Rahwan.** 2018. "The Moral Machine Experiment." *Nature* 563 (7729): 59–64. [197]
- Awad, Edmond, Sohan Dsouza, Azim Shariff, Iyad Rahwan, and Jean François Bonnefon.** 2020. "Universals and Variations in Moral Decisions Made in 42 Countries by 70,000 Participants." *Proceedings of the National Academy of Sciences of the United States of America* 117 (5): 2332–37. [200, 213]
- Bartels, Daniel M.** 2008. "Principled Moral Sentiment and the Flexibility of Moral Judgment and Decision Making." *Cognition* 108 (2): 381–417. [200]
- Bauman, Christopher W., A. Peter McGraw, Daniel M. Bartels, and Caleb Warren.** 2014. "Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology." *Social and Personality Psychology Compass* 8 (9): 536–54. [202]
- Bénabou, Roland, Armin Falk, Luca Henkel, and Jean Tirole.** 2022. "Eliciting Moral Preferences: Theory and Experiment." *Working Paper*. [199, 202, 206, 228]
- Bénabou, Roland, and Jean Tirole.** 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2): 805–55. [199, 206]
- Bentham, Jeremy.** 1789. *An Introduction to the Principles of Morals*. London: Athlone. [196]
- Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch.** 2014. "Hroot: Hamburg Registration and Organization Online Tool." *European Economic Review* 71: 117–20. [211]
- Bostyn, Dries H., Sybren Sevenhant, and Arne Roets.** 2018. "Of Mice, Men, and Trolleys: Hypothetical Judgment Versus Real-Life Behavior in Trolley-Style Moral Dilemmas." *Psychological Science* 29 (7): 1084–93. [200]
- Bursztyn, Leonardo, Bruno Ferman, Ali Hasanain, Michael Callen, Saad Gulzar, and Noam Yuchtman.** 2020. "Political Identity: Experimental Evidence on Anti-Americanism in Pakistan." *Journal of the European Economic Association* 18 (5): 2532–60. [207]
- Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer.** 2022. "Econographics." *Working Paper*. [200, 218]
- Chen, Daniel L., and Martin Schonger.** 2022. "Social Preferences or Sacred Values? Theory and Evidence of Deontological Motivations." *Science Advances* 8 (19): eabb3925. [199]
- Chen, Daniel L., Martin Schonger, and Chris Wickens.** 2016. "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9: 88–97. [211]
- Costa, Albert, Alice Foucart, Sayuri Hayakawa, Melina Aparici, Jose Apesteguia, Joy Heafner, and Boaz Keysar.** 2014. "Your Morals Depend on Language." *PLoS ONE* 9 (4). [200]
- Dean, Mark, and Pietro Ortoleva.** 2019. "The Empirical Relationship between Nonstandard Economic Behaviors." *Proceedings of the National Academy of Sciences* 116 (33): 16262–67. [200]

- Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann.** 2022. "Moral Universalism: Measurement and Economic Relevance." *Management Science* 68 (5): 3590–603. [198, 210, 219]
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde.** 2018. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics* 133 (4): 1645–92. [200]
- Falk, Armin, and Thomas Graeber.** 2020. "Delayed Negative Effects of Prosocial Spending on Happiness." *Proceedings of the National Academy of Sciences* 117 (12): 6463–68. [202]
- Falk, Armin, Thomas Neuber, and Nora Szech.** 2020. "Diffusion of Being Pivotal and Immoral Outcomes." *Review of Economic Studies* 87 (5): 2205–29. [207]
- Feddersen, Timothy, Sean Gailmard, and Alvaro Sandroni.** 2009. "Moral Bias in Large Elections: Theory and Experimental Evidence." *American Political Science Review* 103 (2): 175–92. [199]
- Feddersen, Timothy, and Alvaro Sandroni.** 2006. "A Theory of Participation in Elections." *American Economic Review* 96 (4): 1271–82. [199]
- Feess, Eberhard, Florian Kerzenmacher, and Yuriy Timofeyev.** 2022. "Utilitarian or Deontological Models of Moral Behavior - What Predicts Morally Questionable Decisions?" *European Economic Review*, no. forthcoming. [199]
- Foot, Philippa.** 1967. "The Problem of Abortion and the Doctrine of the Double Effect." *Oxford Review* 5: 5–15. [197]
- Gawronski, Bertram, and Jennifer S. Beer.** 2017. "What Makes Moral Dilemma Judgments "utilitarian" or "deontological"?" *Social Neuroscience* 12 (6): 626–32. [200]
- Gneezy, Uri.** 2005. "Deception: The Role of Consequences." *American Economic Review* 95 (1): 384–94. [204]
- Gold, Natalie, Briony D. Pulford, and Andrew M. Colman.** 2015. "Do as I Say, Don't Do as I Do: Differences in Moral Judgments Do Not Translate into Differences in Decisions in Real-Life Trolley Problems." *Journal of Economic Psychology* 47: 50–61. [200]
- Graham, Jesse, Brian A. Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H. Ditto.** 2011. "Mapping the Moral Domain." *Journal of Personality and Social Psychology* 101 (2): 366–85. [198, 210]
- Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen.** 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (5537): 2105–8. [200]
- Hauser, Marc, Fiery Cushman, Liane Young, R. Kang-Xing Jin, and John Mikhail.** 2007. "A Dissociation between Moral Judgments and Justifications." *Mind and Language* 22 (1): 1–21. [200]
- Kahane, Guy, Jim A.C. Everett, Brian D. Earp, Lucius Caviola, Nadira S. Faber, Molly J. Crockett, and Julian Savulescu.** 2018. "Beyond Sacrificial Harm: A Two-Dimensional Model of Utilitarian Psychology." *Psychological Review* 125 (2): 131–64. [198, 210, 218]
- Kant, Immanuel.** 1785. *Grundlegung Zur Metaphysik Der Sitten*. J. F. Hartknoch. [195, 196, 199, 205]
- Kimbrough, Erik O., and Alexander Vostroknutov.** 2016. "Norms Make Preferences Social." *Journal of the European Economic Association* 14 (3): 608–38. [208]
- Laffont, Jean-Jacques.** 1975. "Macroeconomic Constraints, Economic Efficiency and Ethics: An Introduction to Kantian Economics." *Economica* 42 (168): 430. [199]
- Lanteri, Alessandro, Chiara Chelini, and Salvatore Rizzello.** 2008. "An Experimental Investigation of Emotions and Reasoning in the Trolley Problem." *Journal of Business Ethics* 83 (4): 789–804. [200]
- Mill, John Stuart.** 1863. *Utilitarianism*. Edited by Son & Bourn Parker. London: West Strand. [196]

- Rai, Taze S., and Keith J. Holyoak.** 2010. "Moral Principles or Consumer Preferences? Alternative Framings of the Trolley Problem." *Cognitive Science* 34 (2): 311–21. [200]
- Roemer, John E.** 2010. "Kantian Equilibrium." *Scandinavian Journal of Economics* 112 (1): 1–24. [199]
- Sinnott-Armstrong, Walter.** 2019. *Consequentialism*. Edited by Edward N. Zalta. The Stanford Encyclopedia of Philosophy. [196]
- Stango, Victor, and Jonathan Zinman.** 2022. "We Are All Behavioral, More or Less: A Taxonomy of Consumer Decision Making." *Review of Economic Studies*, no. forthcoming. [200]
- Thompson, Judith Jarvis.** 1976. "Killing, Letting Die, and the Trolley Problem." *Monist* 59 (2): 204–17. JSTOR: 27902416. [197]
- Van Leeuwen, Boris, and Ingela Alger.** 2021. "Estimating Social Preferences and Kantian Morality in Strategic Interactions." *Working Paper*. [199]