

Doctoral thesis

BELIEFS AND PREFERENCES

HOW DECISIONS ARE SHAPED BY WHAT WE EXPECT AND (DIS)LIKE

David Albrecht

2023

BELIEFS AND PREFERENCES

HOW DECISIONS ARE SHAPED BY WHAT WE EXPECT AND (DIS)LIKE

Dissertation

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. P. Habibović,
in accordance with the decision of the Board of Deans,
to be defended in public
on tba, at tba

by

David Albrecht

Promotors

Prof. Dr. A. Brüggen
Dr. M. Strobel

Copromotor

Dr. T.G.K. Meissner

Assessment Committee

...

© David Albrecht, Maastricht 2023.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the author. Funding by the Graduate School of Business and Economics at Maastricht University is gratefully acknowledged.

To reason

Contents

| | | |
|-----|---|-----|
| 1 | Introduction | 1 |
| 2 | Mitigating manipulation in committees: Just let them talk! | 5 |
| 2.1 | Introduction | 7 |
| 2.2 | Related literature | 12 |
| 2.3 | Experiment | 18 |
| 2.4 | Analysis | 24 |
| 2.5 | Conclusion and discussion | 49 |
| 2.A | Experimental details | 53 |
| 2.B | Robustness checks | 58 |
| 2.C | Modelling framework | 63 |
| 2.D | Analysis of communication protocols | 64 |
| 2.E | Coding scheme for video recordings | 66 |
| 2.F | Experimental instructions | 70 |
| 3 | Debt aversion: Theory and measurement | 111 |
| 3.1 | Introduction | 115 |
| 3.2 | A theory of debt aversion | 121 |
| 3.3 | Experiment | 123 |
| 3.4 | Results | 128 |
| 3.5 | Discussion and conclusion | 147 |
| 3.A | Multiple price lists | 151 |
| 3.B | Individual characteristics | 158 |
| 3.C | Additional graphs | 159 |
| 3.D | Robustness checks | 160 |
| 3.E | Experimental instructions | 172 |
| 3.F | Contract form | 188 |
| 4 | The debt aversion survey module: An experimentally validated tool to measure individual debt aversion | 193 |
| 4.1 | Introduction | 197 |

Contents

| | | |
|-----|---|-----|
| 4.2 | The debt aversion survey module | 198 |
| 4.3 | Validation study | 200 |
| 4.4 | Identification of the debt aversion survey module | 205 |
| 4.5 | Discussion and conclusion | 209 |
| 4.A | Calculation example | 213 |
| 4.B | Theoretical framework | 215 |
| 4.C | Multiple price lists | 217 |
| 4.D | Debt aversion survey module | 223 |
| 5 | General Discussion | 233 |
| | Bibliography | 241 |
| | Valorisation | 259 |
| | Summary | 265 |
| | Samenvatting | 271 |
| | Zusammenfassung | 279 |
| | About the author | 287 |

1

Introduction

A key question shared among the many disciplines contributing to decision sciences is what shapes individual decisions, whether in their private lives, in a professional context, or as members of society. In economics, psychology, and beyond, it is common to consider two person-specific dimensions driving decisions: beliefs and preferences.¹

Beliefs refer to what people expect in a setting characterized by uncertainty. That is, more than one scenario is possible, and it is uncertain which scenario is or will be the true state. For instance, the future may turn out in one way but also in many others. Interest rates may decrease, stay the same, or increase in the coming year.

Preferences refer to what people like or dislike. If there were two options, A and B, preferences define how people like one option compared to the other. For instance, each of us has a certain preference to spend

¹In economics consider e.g., A. Smith, 1761; Von Neumann and Morgenstern, 1944, and Manski, 2004, at the intersection with psychology Tversky and Kahneman, 1974, and Kahneman and Tversky, 1979, and beyond Berger and Luckmann, 2011, and Glimcher and Fehr, 2013.

our money today or save it to have the possibility to spend it in the future.

Decisions are generally shaped jointly by beliefs and preferences. For many, whether to save or not depends on interest rate expectations and how much they like to spend today rather than in the future.

In my dissertation, I investigate particular aspects of both dimensions: beliefs and preferences. In Chapter 2, I explore a potential source of information that may influence beliefs: expert committees' judgment. In chapters 3 and 4, co-authored with Thomas Meissner, I analyze a particular domain of individual preferences: aversion to going into debt.

Beliefs that govern decisions in various situations may be informed by the collective judgment of small groups like expert committees. Likely, they are shaped not only by the collective judgment itself but also by other attributes like its accuracy and perceived trustworthiness. This holds especially if parts of such expert committees have incentives to manipulate the collective judgment in their favor; in other words, they follow their own hidden agenda (Maciejovsky and Budescu, 2020; Choo, Kaplan, and Zultan, 2022). In my single-authored research project in Chapter 2, I analyze how such hidden agendas influence expert committee judgments' accuracy and perceived trustworthiness. Moreover, I test whether the committee's interaction format may mitigate the negative effects of hidden agendas on accuracy and perceived trustworthiness.

One dimension of preferences that influences decisions such as spending money today or saving for the future is the preference for time. The more a person prefers the now over the future, the more likely this person will spend their money today rather than save. An option to spend even more money today than one has is to go into debt. This is to borrow money now, e.g., to purchase an apartment and commit to future repayments. Studies from various contexts suggest that people might be intrinsically unwilling to borrow and thus miss out on beneficial investments, e.g., in tertiary education (Field, 2009; Caetano, Palacios, and Patrinos, 2019) or energy-efficient technologies (Schleich,

Faure, and Meissner, 2021). In Chapter 3, Thomas Meissner and I use an incentivized, real-time experiment to test for the existence of such debt aversion while controlling for potential confounding factors. In this way, we isolate the effect of genuine preferences towards debt from other beliefs, biases, preferences, and constraints that may influence individual borrowing decisions. Moreover, using a structural model, we quantify the magnitude of debt aversion observed in experimental participants by comparing them to counterfactual decision-makers who are neutral toward debt.

Financial decisions involving borrowing money may influence large parts of our lives, e.g., enabling home ownership. Moreover, by offering subsidized loans, many policies try to spur socially desirable behaviors, such as tertiary education or sustainable home improvements.² For understanding such decisions and to inform related policies, understanding individual preferences towards debt is key. To facilitate this, Chapter 4 presents a short and easy-to-use survey module to elicit debt aversion. We identify two simple questions that predict how people would behave in the experiment of Chapter 3. In this way, we provide a tool to measure debt aversion approximately without going through the more time-consuming incentivized experiment.

Throughout all chapters, I employ a variety of choice experiments to scrutinize various aspects of beliefs and preferences. Such controlled experiments are a powerful tool to identify causal relationships, reveal stylized facts, and provide policy-relevant advice (Weimann, Brosig-Koch, et al., 2019). The basic idea is to create at least two similar choice environments. Then, one observes the choices made by participants in both scenarios. This can take the form of within-subject or between-subject comparisons. For within-subject comparisons, one observes the same person's choices first in one scenario and then in the other. For between-subject comparisons, multiple people are randomly assigned

²For instance, the German promotional bank KfW, which acts based on a governmental mandate, offers subsidized loans to finance university studies, energy-efficient refurbishment, and climate-friendly new construction (KfW, 2023).

to one or the other scenario. Suppose observed choices differ systematically between the two scenarios. In that case, one concludes that the systematic differences are caused by the features differentiating the scenarios.³

The choice experiments used in my dissertation adhere to quality standards and best practices of behavioral and experimental economics. In particular, I use real monetary incentives and commit to non-deception. Real monetary incentives entail that participants' choices in the experiment influence how much money they receive for their participation. In this way, I can analyze their revealed choices in situations where they have some "skin in the game", rather than stated choices without direct monetary consequences. Non-deception in experiments enforces that all information provided to participants by the researchers is true. This is important to identify causal effects, as it creates trust in experimental instructions and conveys that the decision scenarios participants face consist of and differ in exactly the attributes thought out and communicated by the researchers. Building on this experimental methodology, I can derive clear, evidence-based, and insightful answers to the various research questions guiding my dissertation.

³Differences in observed behavior may generally also stem from noise and other factors not directly related to the experiment and the differences between conditions. An appropriate experimental design enables differentiating such confounding factors from genuine, systematic treatment differences. For instance, in within-subject comparisons, one choice scenario is experienced after the other. Differences in observed choices may thus also be driven by the order and succession of choices. On the one hand, people may get tired or bored over time. On the other hand, they may learn something from the first choice scenario, which they apply to the second. To control for these confounding factors, researchers usually observe many within-subject comparisons where the order of choice scenarios is varied randomly between subjects. For between-subject comparisons, some participants may behave differently, e.g., as they are in a better mood or more knowledgeable than others. Observing the choices of sufficiently many participants who are randomly assigned to conditions allows to control for these alternative explanations.

2

Mitigating manipulation in committees: Just let them talk!¹

Adapted from: David Albrecht (2023). *Mitigating manipulation in committees: Just let them talk!*

¹Corresponding author David Albrecht david.albrecht@pm.me. I thank Alexander Brüggen, Thomas Meissner, Martin Strobel, Uri Gneezy, Joshua Becker, Ville Satopää, Marie Claire Villeval, my PhD colleagues at Maastricht University, and others for many helpful discussions. Numerous valuable comments were also received from participants in seminars at Maastricht University, University of California San Diego, and at conferences in Barcelona, Exeter, Maastricht, Nice, and Santa Barbara as well as in the ESA Job Market Seminar Series. Angela Thissen and Kaiqi Liu provided excellent research assistance for analyzing the video recordings of the experiment. I acknowledge funding through the GSBE grant for primary data collection and by Alexander Brüggen and Martin Strobel. Replication files can be found via OSF: https://osf.io/wh7fr/?view_only=71c7f9270b7c40608fd6f4e1ff419a6f.

Abstract

Many decisions rest on the collective judgment of small groups like committees, boards, or teams. However, some group members may have hidden agendas and manipulate this judgment to sway decisions in their favor. Utilizing an incentivized experiment, I analyze how manipulation affects the objective accuracy and perceived trustworthiness of such group judgments depending on the format of group interaction.

I compare group judgments from unstructured face-to-face interactions, which are ubiquitous in real-world institutions, to group judgments from the scientifically endorsed, structured Delphi technique. To identify mechanisms underlying the accuracy differences, I use structural estimations and analyze emergent communication patterns.

Without manipulation, Delphi is more accurate than face-to-face interaction and indistinguishable from the Bayesian benchmark. Manipulation decreases accuracy for Delphi but not for face-to-face interaction. Thus, with manipulation, Delphi is less accurate than face-to-face interaction. With manipulation, sharing of (truthful) information decreases in Delphi but not in face-to-face interaction. The structural estimations further reveal that Delphi judgments likely exhibit more bias towards hidden agendas and less utilization of valuable information.

Perceived trustworthiness does not always match objective accuracy. Judgments from face-to-face interaction - unjustifiably - enjoy higher levels of trust without hidden agendas. Trustworthiness correctly decreases with hidden agendas for Delphi groups but - unjustifiably - also for face-to-face groups. With hidden agendas, face-to-face groups are simultaneously more accurate and trusted.

2.1 Introduction

In many institutions, high-stake decisions are made based on the collective judgment of small groups, like (expert) committees or advisory boards that evaluate some decision-relevant criteria. Examples of such group judgments span diverse contexts: scientists evaluate the effectiveness of policy interventions to fight global pandemics (Haug et al., 2020), managers rate investment alternatives (Lovallo et al., 2020), and security experts assess terrorist activity (Friedman and Zeckhauser, 2014).

The question remains whether the widespread use of this practice implies its appropriateness. To optimally inform decisions, the accuracy and trustworthiness of group judgments are essential. However, some group members may have hidden agendas and manipulate this judgment in a particular direction to sway consequent decisions in their favor.² Generating accurate and trusted collective intelligence in such a setting remains a challenge. Therefore, I analyze whether two common formats of group interaction may preserve accuracy and trustworthiness despite manipulation due to hidden agendas. In particular, I compare group judgments from unstructured face-to-face interactions, which are ubiquitous in real-world institutions, to group judgments from the scientifically endorsed, structured Delphi technique.

A way of extracting collective intelligence from groups that is both accurate and perceived as trustworthy has yet to be identified. Previous research in this context has focused on prediction markets and exogenously structured face-to-face interactions. While prediction markets appear accurate but not always trusted, structured face-to-face interactions are highly trusted but not always accurate.

²E.g., serving their career in the institution (Toma and Butera, 2009; Mattozzi and Nakaguma, 2022), serving other organizational units the individual is associated with (Pearsall and Venkataramani, 2015), benefiting their reputation (Visser and Swank, 2007), overstating chances of the political victory of favored parties (Hansen, Schmidt, and Strobel, 2004; Rothschild and Sethi, 2016), receiving bribes (Felgenhauer and Grüner, 2008) or downplaying security threat to reduce responsibility and effort for consequent duties (Amjahid et al., 2017).

Prediction markets proved themselves as a robust tool to accurately elicit and aggregate dispersed information from a group of people (Wolters and Zitzewitz, 2004; Arrow et al., 2008), which has also been successfully tested by large institutions such as Google and Ford (Cowgill and Zitzewitz, 2015). Such markets remain informative even if some participants act manipulatively. However, in response to manipulation, decision-makers observing the market lose trust and make even worse decisions, as if they had ignored the market completely (Choo, Kaplan, and Zultan, 2022).

Face-to-face interactions are a standard operating procedure in institutional decision-making. The particular variety of such interactions, that has been studied in hidden agenda settings, are nominal groups. These are groups interacting in an exogenously imposed estimate-talk-estimate format and have been found to enjoy high levels of trust among decision-makers.³ Adversely, trustworthiness persists even if judgment accuracy deteriorates with manipulation (Maciejovsky and Budescu, 2020).

Taken together, both prediction markets and nominal groups appear suboptimal and may lead to many objectively ill-informed decisions with high societal stakes. To address this problem, I extend previous investigations to two alternative formats of group interaction. On the one hand, I analyze commonly used face-to-face interaction (FTF) that is not exogenously bound to any structure, such as the protocol of nominal groups. On the other hand, I consider the scientifically endorsed Delphi technique (Delphi), where interaction is exogenously structured according to a specific protocol.⁴ I quantify the extent to which these interaction formats produce a collective intelligence that

³In nominal groups, group members first estimate individually, second talk with their group members, and finally estimate individually again.

⁴FTF and more broadly free-form deliberation of groups are ubiquitous in real-world institutional decision-making (Maciejovsky and Budescu, 2020) despite concerns about their performance (Sunstein, 2005; Armstrong, 2006). Though less common in practice, the Delphi technique is a group interaction format supported by empirical and theoretical research (Rowe and Wright, 1999; Graefe and Armstrong, 2011).

is simultaneously accurate and perceived as trustworthy. Further, I identify what drives robustness towards manipulation.

To measure the capabilities of Delphi and FTF, I developed a pre-registered, incentivized lab experiment.^{5,6} The experiment enables the isolation of the causal effects of hidden agendas and the format of group interaction on quantitative measures of objective accuracy and perceived trustworthiness. Conducting a lab experiment offers a unique opportunity to address the research questions by controlling for many factors that are not directly observable in real-world situations where group judgments are used. Hidden agendas, by their very nature, are covert to researchers. Moreover, group judgments potentially depend on many, not directly measurable factors, such as the expertise of group members, and are used in situations where the true answer is unknown.

The experiment comprises two parts: a group *estimation experiment* and a subsequent *decision experiment*. In the first part, the *estimation experiment*, I examine the objective accuracy of groups of four people who collaborate on a series of probabilistic judgment tasks in which they generate joint group judgments. The experiment follows a two-by-two between-subject design with FTF and Delphi groups, each with and without hidden agendas. In the FTF treatments, groups interact via face-to-face conversation in a video call without any imposed structure. In the Delphi treatments, groups interact through an imposed, pseudonymous chat protocol. According to this protocol, group members first make individual quantitative and qualitative judgments, then review all members' pseudonymous judgments and reasonings before producing a second quantitative estimate individually. Across all treatments, I incentivize accuracy at the group level. In the hidden agenda treatments, I additionally induce manipulation incentives (hidden agendas) in half of the group members through individual, private side payments.

⁵Pre-registration files can be found via OSF <https://osf.io/cv4md/>

⁶The experiment was reviewed and approved by the Ethical Review Committee Inner City Faculties at Maastricht University (reference ERCIC_267_10_06_2021).

In the second part, the *decision experiment*, I elicit the perceived trustworthiness of the group judgments obtained from the estimation experiment. A new set of individual participants who did not participate in the estimation experiment states their individual, incentivized confidence intervals around group judgments.⁷ Each participant evaluates judgments from all four treatment conditions. This decision experiment allows for comparisons of the perceived trustworthiness of group judgments produced by FTF versus Delphi groups, each with and without hidden agendas, respectively.

I identify differences in objective accuracy by comparing absolute errors across conditions of the estimation experiment. Without hidden agendas, structured Delphi interaction produces more accurate judgments than FTF. This result reverses with manipulation, which does hamper accuracy in Delphi groups but not in FTF groups. To put accuracy into perspective, I compare Delphi and FTF groups against the best Bayesian benchmark based on the information available to the groups. Without hidden agendas, the best benchmark's accuracy is statistically indistinguishable from that of Delphi but significantly more accurate than FTF. Both interaction formats, Delphi and FTF, perform significantly better than a benchmark, which estimates all probabilities naïvely at 50%. In situations with hidden agendas, no interaction format reaches the best Bayesian benchmark. While FTF groups still outperform the naïve heuristic, Delphi falls behind this benchmark.

I identify differences in the perceived trustworthiness of group judgments by comparing the length of stated confidence intervals in the decision experiment across conditions of the estimation experiment. FTF group judgments are generally trusted more than Delphi group judgments. This holds regardless of the presence of hidden agendas and also in cases where it is not justified by objective accuracy.

All in all, FTF group judgments appear problematic in situations without hidden agendas. Here, they are objectively less accurate. Decision-

⁷Intervals are incentivized through the incentive-compatible most likely interval technique (Schlag and Weele, 2015).

makers fail to realize this and put relatively more trust in FTF judgments. This can be taken as evidence that FTF should be used less in institutional decision-making in situations without hidden agendas, which are arguably very rare. The Delphi technique seems to be a better alternative. However, in situations with hidden agendas, perceived trustworthiness is aligned with objective accuracy, as the relatively more accurate judgments from FTF groups are also trusted more than judgments from Delphi groups. This can be taken as evidence that FTF is an appropriate format for institutional decision-making in situations with hidden agendas, which are arguably very prevalent.

To explore the mechanisms underlying the differences in accuracy, I use structural estimations and analyze emergent communication patterns across conditions of the estimation experiment. In particular, I analyze the data from group estimation in the BIN-modeling framework (Satopää, Salikhov, Tetlock, et al., 2021) (Section 2.4.4). This allows me to decompose the estimation error in all conditions of the estimation experiment into: (i) bias (consistently producing too high or too low estimates), (ii) noise (coming to different conclusions given the same information), and (iii) the usage of valuable information. Delphi, as compared to FTF, generally exhibits less noise. However, with the introduction of hidden agendas, Delphi likely suffers from more bias, while FTF appears robust against this effect.

Complementing the results from the structural estimations, I analyze the degree and truthfulness of information sharing during group interaction. Transcribed and coded communication protocols of FTF and Delphi interactions reveal that hidden agendas lead to less information sharing and decrease the truthfulness of shared information for Delphi groups. By contrast, in FTF groups, the amount of shared, truthful information increases with hidden agendas.

Condensed, FTF seems generally better suited for group judgments in situations with hidden agendas. Moreover, FTF appears to be the better choice in situations without hidden agendas where the perceived trustworthiness of group judgments is highly prioritized. The Delphi

technique is preferable in situations without hidden agendas where accuracy is the highest priority.

2.2 Related literature

This work considers a broad range of interdisciplinary studies on collective intelligence. The most important works in relation to this study are those that compare multiple group interaction formats, investigate single group interaction formats in isolation, and those that study the effects of manipulation or lying behavior on group judgments. More broadly, this study builds on longstanding research on committee decision-making.

This study is most closely related to **comparisons of group interaction formats** concerning accuracy, trustworthiness, and robustness towards manipulation. Early studies in this area focus on comparing the accuracy of the Delphi technique to other formats of structured and unstructured direct interaction, as well as averaging individual judgments. This literature presents mixed results regarding the relative performance of the Delphi technique, but also acknowledges severe methodological limitations in early Delphi research (Woudenberg, 1991; Rowe and Wright, 1999). Notably, many studies used (over)simplified Delphi designs, which might have led to Delphi's capacity being understated (Rowe, Wright, and Bolger, 1991). Later, collective intelligence research focuses on prediction markets, and advances in software allowed more consistent implementation of structured interaction formats in the laboratory. Computerized Delphi groups were found to be the most accurate in a comparison against unstructured face-to-face groups, structured nominal groups, and groups trading in prediction markets to solve general knowledge questions in the laboratory (Graefe and Armstrong, 2011). Based on data from a geopolitical forecasting tournament, prediction markets outperform median estimates of interacting teams but lose to more sophisticated statistical aggregation (Atanasov et al., 2017). Previously mentioned studies consider situations without

hidden agendas, and without the explicit threat of manipulation. Maciejovsky and Budescu, 2013 are the first to consider conflicts of interest that induce manipulation in structured nominal groups vs. prediction markets in a lab experiment. In contrast to nominal groups, prediction markets maintain knowledge sharing if the existence of conflicts of interest is commonly known. In a follow-up, Maciejovsky and Budescu, 2020 additionally consider the trustworthiness of judgments. Without manipulation, nominal groups are more accurate than markets, but markets are more accurate than groups with manipulation. At odds with accuracy, nominal groups' judgment is perceived as more trustworthy overall, especially in manipulation treatments. This holds for trustworthiness as evaluated by group members themselves but also by external observers. Similarly, Choo, Kaplan, and Zultan, 2022 compare the accuracy of prediction markets without manipulative traders to markets where there is a certain probability that manipulators are active and markets where manipulators are active with certainty. The mere potential of manipulation hampers accuracy. Nevertheless, even with certain manipulation, markets still reveal valuable information. For decision-makers observing the market, the potential of manipulation and, more so, certain manipulation cause erosion of trust. Remarkably, while markets with manipulation still reveal objectively valuable information, erosion of trust prevents decision-makers from utilizing it. In fact, with certain manipulation, decisions are even worse, as if decision-makers had ignored the market entirely.

Extracting collective intelligence from groups in hidden agenda settings remains a challenge. Previous research, as summarized in Table 2.1, has not yet identified interaction formats that are simultaneously accurate and trusted. I build on and extend this body of literature by scrutinizing the accuracy and perceived trustworthiness of group judgments from Delphi groups as opposed to freely interacting face-to-face groups. To the best of my knowledge, I am first to compare these two interaction formats in a setting with hidden agendas. Further, I advance the methodology developed in previous method comparisons by combining experimentally induced, complementary expertise, and continuous

Table 2.1: Comparisons of group judgment techniques in the literature

| | Graefe and Armstrong (2011) | Maciejorsky and Budescu (2013) | Maciejorsky and Budescu (2020) | Choo et al. (2022) | This study |
|---|-----------------------------|--------------------------------|--------------------------------|--------------------|------------|
| Interaction format (# of group members in parentheses) | | | | | |
| unstructured face-to-face | ✓(4-6) | - | - | - | ✓(4) |
| Delphi technique | ✓(4-6) | - | - | - | ✓(4) |
| nominal group technique | ✓(3-6) | ✓(3) | ✓(3) | - | - |
| prediction markets | ✓(4-7) | ✓(3) | ✓(3) | ✓(8) | - |
| Assessed quality measures | | | | | |
| accuracy | ✓ | ✓ | ✓ | ✓ | ✓ |
| trustworthiness | - | - | ✓ | ✓ | ✓ |
| Features of the group judgment task | | | | | |
| hidden agendas | - | ✓ | ✓ | ✓ | ✓ |
| induced expertise | - | ✓ | ✓ | ✓ | ✓ |
| Sample | | | | | |
| n judgment | 227 | 144 | 450 | 112 | 590 |
| groups/ condition | 11 | 16 | 15 | 7 | 15 |
| n trust | - | - | 224 + 358 | 56 | 2000 |

as well as incentivized measures of accuracy and trustworthiness.

Focusing on a single group interaction format at a time, researcher further studied **manipulation** in prediction markets. Evidence is mixed, with some studies reporting that manipulation harms accuracy (Hansen,

Schmidt, and Strobel, 2004; Gimpel and Teschner, 2014; Choo, Kaplan, and Zultan, 2022), while others find that prediction markets are robust against manipulation (Hanson, Oprea, and Porter, 2006; Hanson and Oprea, 2009; Teschner, Rothschild, and Gimpel, 2017).

Beyond markets, the Delphi technique appears manipulable by the Delphi administrators (Nelson, 1978), and Delphi estimates may generally be biased by the desirability of outcomes among Delphi group members (Ecken, Gnatzky, and Gracht, 2011). Further, Wittrock, 2023 discusses theoretically that Delphi group members may not report their true beliefs in order to affect the other Delphi participants and ultimately sway decisions tied to the Delphi result. Yet, explicit manipulation of Delphi estimates by group members has not been empirically investigated prior to the present study.

To study the influence of manipulation, this work strongly builds on prior research on the **Delphi technique**, a structured interaction format to extract and aggregate human judgment from a group of experts. The format aims to strengthen the positive aspects of deliberation while minimizing the process loss of group interaction (Rowe and Wright, 2001). Two basic concepts form the foundation of Delphi's potential. First, according to the theory of errors, combining estimates from multiple judges will lead to a group judgment that is more accurate than the average individual's judgment (Dalkey, 1975). Second, throughout the interaction, relatively less accurate judges will update more than relatively accurate judges (Parenté and Anderson-Parenté, 1987), and relatively more accurate feedback has a stronger influence on such updating (Rowe, Wright, and McColl, 2005; Bolger and Wright, 2011). Both concepts are also noted in the broader context of the "wisdom of crowds", i.e., aggregating independent, individual judgments is more accurate than the average individual (Galton, 1907). Second, through information exchange, individual judgments lose independence. However, accuracy may still increase (Mellers et al., 2014; Da and Huang, 2020), and combined group judgments improve if relatively stronger influence is exerted from more to less accurate individuals (J. A. Becker, Brackbill, and Centola, 2017).

The Delphi technique dates back to research by the RAND Corporation in the late 1940s to improve expert forecasting on topics such as technological change. Subsequently, manifold research projects used the Delphi technique and investigated its performance and workings (as reviewed by Rowe, Wright, and Bolger, 1991; Woudenberg, 1991; Rowe and Wright, 1996; Rowe and Wright, 1999; Bolger and Wright, 2011). In parallel, research from various disciplines has used the Delphi technique to generate estimates on transcontextual topics such as climate change mitigation measures (Griscom et al., 2017), biotechnologies improving health in developing countries (Daar et al., 2002) and geopolitical events (Wintle et al., 2012). Implementations of the Delphi technique often differ in their specific design. However, the vast majority exhibits four key characteristics: anonymity, controlled feedback, iteration, and statistical aggregation of final inputs (Rowe and Wright, 2001; Grime and Wright, 2016; Belton, MacDonald, et al., 2019). Moreover, certain Delphi features empirically appeared as accuracy enhancing: feedback that also includes written rationales next to group members' estimates (Best, 1974; Rowe and Wright, 1996; Rowe, Wright, and McColl, 2005; Bolger and Wright, 2011; Bolger, Stranieri, et al., 2011) and group members with objective (complementary) expertise (Jolson and Rossow, 1971; Rowe and Wright, 1996). The present study incorporates Delphi's four key characteristics, feedback including written rationales, and experimentally induced, complementary expertise in the Delphi interaction implemented in the estimation experiment.

In a broader context, this study builds on the literature on **committee and jury decision-making** going back to Condorcet, 1785. A common focus of this research strand lies in groups of people, i.e., the jury or committee, that reaches a joint verdict, e.g., convict or acquit, by some voting procedure, which may be preceded by deliberation. Some jury members may follow strategic motives, as e.g., discussed by Feddersen and Pesendorfer, 1998. This has common features with the given setting of groups deriving joint, probabilistic judgments but also exhibits apparent differences. For instance, the given setting does not explicitly study specific voting rules, nor is a decision made by the

committee itself. Notably, a phase of communication preceding voting was found to counteract hidden agendas by inducing mostly truthful information sharing and less manipulative voting (Goeree and Yariv, 2011). However, making such deliberations public to non-committee members counteracts these positive effects (Fehrler and Hughes, 2018). Moreover, if committee members differ in competence, the voting behavior's transparency may counteract competent members' hidden agendas (Mattozzi and Nakaguma, 2022).

Furthermore, the effects of manipulation on collective intelligence are reflected in the literature on **lying and deception**. In contrast to the predictions of standard economic theory, many individuals tell the truth despite potential monetary gains from lying (Gneezy, 2005; Sutter, 2009). Further, many of those who do not tell the truth do not lie to the full extent (Mazar, Amir, and Ariely, 2008; Fischbacher and Föllmi-Heusi, 2013; Gneezy, Kajackaite, and Sobel, 2018). Rather, they obfuscate the truth through vague messages (Serra-Garcia, Damme, and Potters, 2011) or evasive lies (Khalmetski, Rockenbach, and Werner, 2017). The most plausible explanations for such (non-)lying behavior are preferences for being honest and being seen as honest (Abeler, Nosenzo, and Raymond, 2019). Lying however, increases if individual responsibility is diffused, such as in team settings (Conrads et al., 2013), if the social distance between a liar and the person being lied to increases (Hermann and Ostermaier, 2018), if communication is computer-mediated (Marett and George, 2012) or if the truth that is twisted in a lie, is hard to observe by others (Fries et al., 2021; Hermann and Brenig, 2022). Switching perspectives, many people do not correctly detect if they are lied to (Bond and DePaulo, 2006). They are overconfident in their ability to do so, and share undetected lies as truth (Serra-Garcia and Gneezy, 2021). These results set a benchmark to my analysis of (truthful) information revelation during group interaction across conditions of the estimation experiment.

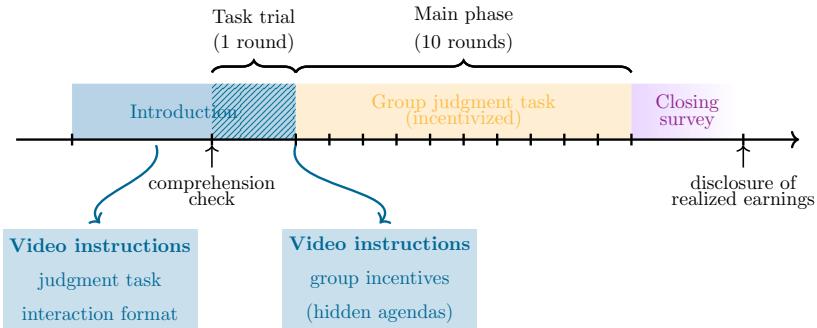
2.3 Experiment

This study comprises two complementary experiments: the *estimation experiment*, which examines the accuracy of group judgments, and the subsequent *decision experiment*, which elicits the trustworthiness of group judgments. All experiments were programmed in oTree (Chen, Schonger, and Wickens, 2016) and were implemented onsite in the experimental economics laboratory BEElab at Maastricht University in 2022 and 2023. In total, 240 people participated in the estimation experiment, split into 60 groups of four, 15 groups per treatment condition. Subsequently, 50 participants individually evaluated group judgments in the decision experiment. All participants were recruited from the BEElab participant pool via ORSEE (Greiner, 2015) and followed a study program at the time of the experiment or in the past. Their backgrounds span the arts, as well as the social and natural sciences. There is a clear prevalence of participants with business and/or economics backgrounds (79.6% in the estimation experiment and 74% of participants in the decision experiment). Most participants identified as female (56.3% in the estimation experiment and 64% of participants in the decision experiment). Many had at most one year of professional working experience (70.4% in the estimation experiment and 76% of participants in the decision experiment).

2.3.1 Estimation experiment

To determine the causal effect of group interaction format and hidden agendas on the accuracy of group judgments, I implemented the estimation experiment as a two-by-two between-subject design. In all treatment conditions, groups were asked to jointly generate a series of ten probabilistic group judgments in a group of four. Groups received incentives for doing so as accurately as possible. They interact in either FTF or Delphi format. In the hidden agenda treatments (HA), half of the group members additionally were given private, individual manipulation incentives that conflicted with and outweighed

Figure 2.1: Timeline of the estimation experiment



the group incentives for accuracy (see Appendix 2.A.1.2 for details on incentives).

The estimation experiment followed the same structure in all four conditions, as illustrated in Figure 2.1. First, participants received video instructions on the judgment task and the interaction format, followed by a comprehension check and a task trial to experience the judgment task and the interaction format. After the trial, before starting the main phase of the experiment, participants were informed about the incentive scheme and whether they would be interacting in a setting with hidden agendas. In settings with hidden agendas, participants were told their type (whether or not they had a hidden agenda) and that exactly two group members have a hidden agenda, while the remaining two do not. Thus, knowledge about their own type allowed participants to infer the distribution of types among the remaining group members. Types did not change during the experiment. Furthermore, in settings with hidden agendas, all participants were informed that hidden agendas may take one of two forms, driving estimates up or down. It was common knowledge that the direction of the hidden agendas was determined randomly per round of the judgment task and that it was the same among the two group members with hidden agendas in any given round. In a specific round, the two group members with

a hidden agenda knew that their hidden agenda in this round was to drive estimates up (down). This information was not disclosed to the remaining two group members without a hidden agenda. Before starting the main phase, participants were prompted to ask any question to the experimenter that may have remained unanswered.

In the main phase of the experiment, participants went through 10 incentivized rounds of solving the probabilistic judgment task. Each round had a distinct underlying, objectively true answer against which the accuracy of group judgments could be evaluated. At the beginning of each round, group members received individual, complementary information. Combining the individual pieces of information is generally advantageous to generate more accurate group judgments. In groups with hidden agendas, group members with hidden agendas also learned the round-specific direction (up or down) of their hidden agenda at the beginning of each round. No information on group accuracy or hidden agenda achievement was presented until the end of the experiment. Before earnings disclosure and payment, participants completed a questionnaire on socio-demographic characteristics and their subjective perception of the task.

Interaction formats

This paper focuses on the comparison of two interaction formats: free-form face-to-face interaction (FTF) without exogenously imposed structure, which is ubiquitous in real-world institutional decision-making (Maciejovsky and Budescu, 2020), and the Delphi technique, a group interaction format supported by empirical and theoretical research (Rowe and Wright, 1999). In a broader context, the general experimental framework provides a test bed that is well suited to investigating the accuracy and perceived trustworthiness of any group interaction format directed at generating quantitative group judgments.

Following Graefe and Armstrong, 2011, I implemented a very simple form of face-to-face interaction in the **FTF** treatments, wherein groups

were almost unrestricted in their interaction. For each round of the task, group members used a video call in which they could freely discuss the judgment task, their available information, and their strategy to form a joint group judgment.⁸ At the end of an interaction, each group member was prompted to enter the joint group estimate for that round. The experiment only proceeded after all group members entered the same estimate, thus enforcing consensus.⁹ As such, FTF is not anonymous and allows a free and unrestricted flow of information as well as the full richness of communication (speech, body language, etc.). These details resemble typical real-world group interactions (Maciejovsky and Budescu, 2013). For further analysis, the video calls of all face-to-face interactions were recorded and transcribed.

In general, **Delphi** evolved to be an umbrella term for diverse forms of structured group interactions. This estimation experiment may only implement and analyze one particular variation of group interaction with Delphi features. To choose a meaningful design for this Delphi interaction, I followed two strategies: first, the design was as simple as possible while still comprising the most common Delphi features: anonymity, controlled feedback, iteration, and statistical aggregation of final inputs (Rowe, Wright, and Bolger, 1991; Woudenberg, 1991; Grime and Wright, 2016). Second, the design was similar to recent implementations of Delphi procedures in the field of collective intelligence (J. A. Becker, Guilbeault, and E. B. Smith, 2021; Belton, Wright, et al., 2021).

In the experiment, Delphi group interactions followed a standardized, computer-mediated protocol. Participants did not directly discuss but rather interacted through a messaging interface. After receiving their

⁸Video-conference interactions became a very popular medium for team interactions, latest during the COVID-19 pandemic. These formats enable information sharing and produce high-quality results that appear no different from those of physical face-to-face interactions (Jabotinsky and Sarel, 2020).

⁹To prevent extraordinarily long discussions, the interaction phase is limited to 10 minutes for the first incentivized round and 7:30 minutes for all later rounds. Groups that do not complete a given round in time do not earn any bonuses on this round. Out of 30 FTF groups, generating a total of 300 group judgments, the interaction on 10 group judgments exceeded the time limit.

information, group members first provided their numerical, individual estimates and corresponding reasoning in written form. Second, after all group members had completed their first input, everyone was presented with the inputs of the three remaining group members and their own input again. A, B, and C were used as pseudonyms associated with the same people throughout the experiment, and the feedback was shown as “the input of Person A, B, and C” and “your input.” This prevented participants from being able to link the estimates and reasoning to a specific person. Next, group members were prompted to give a second estimate, which might, but did not have to be, a revision of their first estimate based on the feedback. After all group members completed their second estimate, the group judgment was calculated as the mean of all second estimates. Participants were presented with the group judgment and the underlying, second individual estimates before proceeding with the experiment. Again, this feedback is pseudonymous.

Judgment task

Inspired by Peeters and Wolk, 2018; Peeters and Wolk, 2017, participants were asked to estimate the likelihood that a (biased), two-dimensional, discrete random walk would end above a threshold level after 10 steps. Specifically, the random walk started at 0 and comprised 10 steps $X_t = X_1, X_2, \dots, X_{10}$. Each step X_t could take the values $-1, 0$ and 1 . Independent draws from an identical distribution determined X_t . The task was repeated for 10 rounds. The distribution underlying the random walk varied from round to round of the judgment task. In each round, group members need to estimate $\mathbb{P}(\sum_{t=1}^{10} X_t > 0)$, i.e., the probability that the random walk would take a value larger than the threshold of 0 after ten steps. Throughout the 10 rounds, groups faced steps X_t from probability distributions corresponding to $\mathbb{P}(\sum_{t=1}^{10} X_t > 0) = \{0.1, 0.2, 0.3, 0.4, 0.45, 0.55, 0.6, 0.7, 0.8, 0.9\}$ in random order. The task was presented to participants as a computer-generated movement path (random walk) of a ladybird, which may

or may not reach a target after 10 steps. Groups were asked to estimate the chance that the ladybird would reach the target. They were informed that the computer program would generate the movement paths by independently drawing the value per step from an underlying identical distribution, but the exact distribution was undisclosed. To learn about the underlying probabilities, each group member privately received a movement path generated by a distinct past execution of the computer program. In each round, group members first received their information and then entered a phase of interaction, which resulted in a group estimate (see Appendix 2.A.1.1 for further details on the judgment task).

2.3.2 Decision experiment

This second experiment serves the purpose of evaluating the perceived trustworthiness of estimates generated by experts in the estimation experiment. Experimental participants (decision-makers) who were not involved in the estimation experiment took part in the decision experiment. Within-subject, each decision-maker stated their incentivized, personal confidence intervals around 40 estimates generated in the estimation experiment.¹⁰ They were given 10 randomly selected group judgments drawn from each of the four experimental conditions of the estimation experiment (FTF, FTF + HA, Delphi, Delphi + HA).^{11,12} Decision-makers were also introduced to the judgment task. They were informed about the details of the interaction format and incentives in each of the four treatment conditions in the estimation experiment.

¹⁰Incentives followed the most likely interval method (Schlag and Weele, 2015), as described in more detail in Appendix 2.A.1.2.

¹¹This is an adjustment to pre-registration that allows all estimates of the estimation experiment to be evaluated as balanced as possible: each estimate is evaluated at least three times and by three different decision-makers. The pre-registration outlines that each decision-maker faces all 10 group judgments made by one group in the estimation experiment.

¹²The order in which group judgments from the four experimental conditions of the estimation experiment are evaluated is randomized across decision-makers.

Subsequently, decision-makers stated the lower and upper bound of an interval for each estimate, which they believed likely contained the true value estimated in the estimation experiment. The width of these stated most likely intervals can be seen as a measure of the perceived trustworthiness of group judgments generated in the estimation experiment; the narrower the intervals, the more trustworthy the group judgments, and vice versa.¹³

2.4 Analysis

2.4.1 Hypotheses

Initially focusing on the accuracy of group judgments, I test three pre-registered hypotheses. **H1:** Without hidden agendas, Delphi groups are more accurate than FTF groups. **H2:** Hidden agendas impair accuracy in Delphi and FTF groups. **H3:** The negative effect of hidden agendas is relatively smaller for Delphi groups. H1 and H3 relate to the theoretical and empirical arguments that Delphi extracts accurate group judgments by strengthening positive aspects of interaction while minimizing the process loss of group interaction (Rowe and Wright, 2001). H2 follows from the intuitive interpretation of hidden agendas as obstacles to accuracy.

Subsequently evaluating the trustworthiness of group judgments, I test three further pre-registered hypotheses. **H4:** Delphi groups enjoy the same level of trust as FTF groups in settings with and without hidden agendas, respectively. **H5:** Hidden agendas impair trust in Delphi and FTF groups. **H6:** The negative effect of hidden agendas on trustworthiness is the same for Delphi and FTF groups. Previous literature has focused on the relative accuracy of Delphi and FTF but remains silent on their relative trustworthiness. As such, H4 and H6 can be seen as a result of an impartial prior about the interaction formats'

¹³Judgments of FTF groups that were NA due to exceeding the time limit in the experiment were imputed by 0.5 (this applies to 10 judgments out of 300).

trustworthiness. H5 follows from the intuitive interpretation of hidden agendas as obstacles to trustworthiness.

2.4.2 Evaluating the accuracy of estimates

To measure the causal effect of hidden agendas and interaction formats on accuracy, I compare group judgments across conditions of the estimation experiment. Based on the estimation experiment, Delphi groups appear more accurate than FTF groups without hidden agendas, supporting H1. Hidden agendas impair accuracy for Delphi groups, which is partly in line with H2. However, hidden agendas do not affect the accuracy of FTF groups, contradicting H3.

For the primary analysis, I quantify accuracy by juxtaposing group judgments with objectively true probabilities.¹⁴ I focus on absolute error (AE) as a simple, and illustrative corresponding metric.¹⁵

$$AE_{g,r} = |p_{g,r} - p_r^*| \quad (2.1)$$

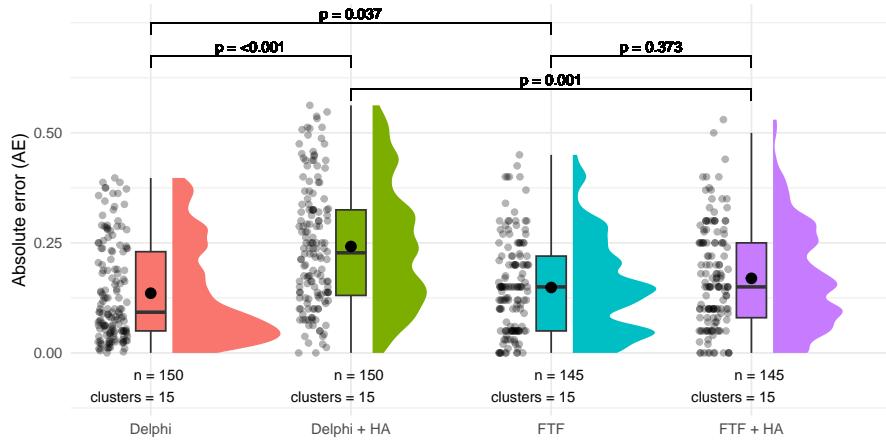
where $p_{g,r}$ is the group judgment, i.e., the estimated probability of group g in round r with $r = 1, \dots, 10$, and p_r^* is the corresponding true probability. Absolute errors are bound by 0 (most accurate) and 1 (least accurate).¹⁶

¹⁴In the experiment setting, the objectively true probability is the probability of reaching the target, which is coded into the computer program that generates the movement paths of the ladybird. An alternative approach is comparing a group judgment to the outcomes drawn according to the underlying actual probability. In the experiment setting, the actual outcome is the outcome reached when the program is executed for the next time, i.e., the ladybird did (not) reach the target in this particular execution of the program. This alternative, however, is more complex than needed and a more noisy measure of accuracy, especially for relatively small numbers of distinct judgments.

¹⁵The reported main results 1-3 are robust towards considering squared errors and Brier scores (pre-registered), as alternative metrics of accuracy (see Appendix 2.B).

¹⁶An absolute error of 1, however, is only possible in two specific cases where the

Figure 2.2: Absolute errors of group judgments across conditions of the estimation experiment



Notes: P-values from two-sided Wilcoxon rank sum tests (Rosner, Glynn, and Lee, 2003; Jiang et al., 2020). Bar in boxplot = median; dot in boxplot = mean.

The absolute errors in each experimental condition are summarized in Figure 2.2, leading to the three main results below. For comparisons between conditions, I report p-values of two-sided, non-parametric, adapted Wilcoxon rank sum tests that control for clustering of observations at the group level (Rosner, Glynn, and Lee, 2003; Jiang et al., 2020).¹⁷

First, I focus on the baseline scenario of situations without hidden agendas.

estimate was 0 (1) and the true probability was 1 (0), respectively. A realistic negative benchmark absolute error is 0.21, the average absolute error obtained if estimates were always 0.5, and the true probabilities are the same as in the experiment.

¹⁷Results 2 and 3 are robust to an even more conservative scenario, where the average absolute error over all ten judgments per group is considered as only independent observation per group as reported in Appendix 2.B.

Result 1: In line with H1, without hidden agendas, the absolute errors of judgments from structured Delphi interaction are significantly smaller ($p = 0.037$) than from unstructured face-to-face interaction.

Turning to situations with hidden agendas, FTF, however, performs quite well.

Result 2: Introducing hidden agendas leads to a significant increase ($p < 0.001$) in absolute errors of judgments from Delphi interactions but, in contrast to H2, not for judgments from unstructured face-to-face interactions ($p = 0.373$).

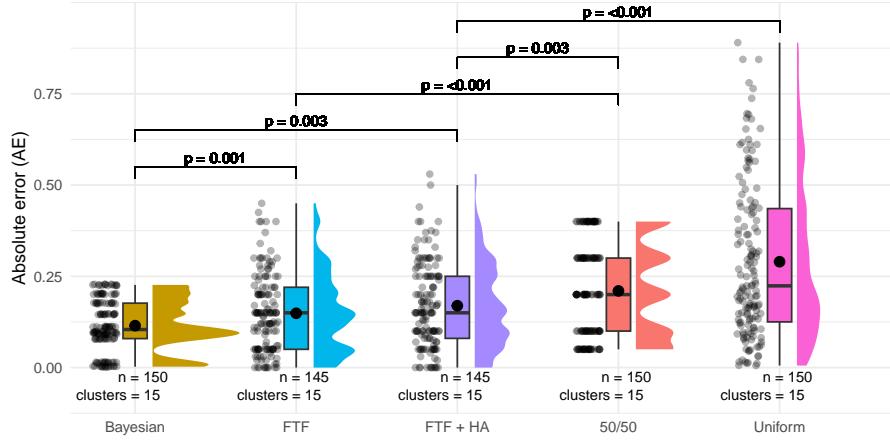
Result 2 also precludes that the negative effect of hidden agendas on accuracy is relatively smaller for Delphi groups. The negative effect of hidden agendas appears to be smaller in FTF than in Delphi interactions. This poses the new question: Which interaction format ultimately yields more accurate results in situations with hidden agendas?

Result 3: With hidden agendas, the absolute errors of judgments from unstructured face-to-face interactions are significantly smaller ($p = 0.001$) than from structured Delphi interaction.

Accuracy benchmarks

To put Results 1 to 3 into perspective, I compare the absolute errors of actual group judgments to the theoretical benchmarks in Figures 2.3 and 2.4, respectively. On the positive side, I consider the perfect Bayesian benchmark (Bayesian). On the negative side, I consider first the

Figure 2.3: Absolute errors of probability judgments from FTF groups against best and worst benchmarks

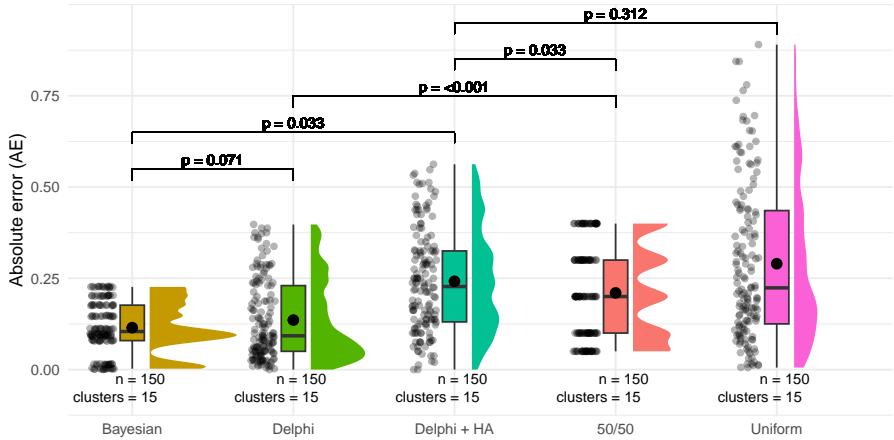


Notes: P-values from two-sided Wilcoxon rank sum tests (Rosner, Glynn, and Lee, 2003; Jiang et al., 2020). Bar in boxplot = median; dot in boxplot = mean. Observations of Bayesian, 50/50, and Uniform are simulated according to the underlying benchmark heuristic.

naïve heuristic of always stating a “50/50” likelihood (50/50) and second reporting a random likelihood drawn from a uniform distribution (Uniform). Without hidden agendas, Delphi groups are statistically indistinguishable from the Bayesian benchmark on the positive side. FTF groups are in between positive and negative accuracy benchmarks. With hidden agendas, FTF remains more accurate than 50/50 and uniform, while Delphi is less accurate than 50/50.

The positive Bayesian benchmark (Bayesian) considers groups as if they had a uniform, i.e., uninformative prior, on the probabilities that a step in the movement path of the judgment task takes the values -1 , 0 , and 1 , respectively. In other words, any combination of $\mathbb{P}(X_t = -1) + \mathbb{P}(X_t = 0) + \mathbb{P}(X_t = 1) \equiv 1$ is deemed a priori equally likely. Groups updated this prior based on the observed movement path and then calculated the probability of reaching the target $\mathbb{P}(\sum_{t=1}^{10} X_t > 0)$ based on their

Figure 2.4: Absolute errors of probability judgments from Delphi groups against best and worst benchmarks



Notes: P-values from two-sided Wilcoxon rank sum tests (Rosner, Glynn, and Lee, 2003; Jiang et al., 2020). Bar in boxplot = median; dot in boxplot = mean. Observations of Bayesian, 50/50, and Uniform are simulated according to the underlying benchmark heuristic.

posterior.

Result 1a: Without hidden agendas, the absolute errors of judgments from Delphi interactions are statistically indistinguishable ($p = 0.071$) from perfect Bayesian groups, but FTF groups are significantly less ($p = 0.001$) accurate than the Bayesian benchmark.

As the naïve 50/50 benchmark, I consider groups as if they always reported a 50% likelihood of reaching the target, irrespective of the task. Uniform constitutes an even stronger negative benchmark. It considers groups as if they consistently reported a random draw from a uniform distribution between 0 and 1, i.e., the infamous dart-throwing chimpanzee (Tetlock, 2005).

Result 1b: *Without hidden agendas, both FTF ($p < 0.001$) and Delphi ($p < 0.001$) groups are more accurate than 50/50 and Uniform, respectively.*

Result 2a: *With hidden agendas, Delphi groups are less accurate ($p = 0.033$) than groups that would always report 50% as their group judgment but remain significantly more accurate ($p = 0.006$) than the Uniform benchmark. By contrast, FTF groups remain more accurate ($p = 0.003$) than 50/50 and Uniform, even with hidden agendas.*

To summarize the findings on the accuracy, I borrow the generalist vs. specialist distinction from the field of ecology (Meijenfeldt, Hogeweg, and Dutilh, 2023). The Delphi technique appears to be a specialist that excels in a small niche or habitat (situations without hidden agendas), where it is perfectly adapted to environmental conditions. By contrast, FTF-based judgments are comparable to generalists. They never reach the same levels of accuracy as the specialist Delphi but appear way more robust and resilient to (unfavorable) changes in environmental conditions (hidden agendas).

Accuracy and hidden agenda achievement over time

The results presented in the previous section focus on the accuracy of group judgments, irrespective of whether the judgment results from the very first interaction of that group or from their interactions in later rounds. In this section, I zoom in on the time dimension to analyze whether accuracy changes over time. Note that participants did not receive any feedback between rounds. Nevertheless, a potential explanation for improving accuracy is that groups may learn to interact better and enhance their understanding of the task. In the same way, in groups with hidden agendas, those group members with hidden agendas may learn how to achieve their hidden agenda over time, thus impairing

accuracy. On the group level, I find no evidence for significant changes in the accuracy of judgments over time. Moreover, there is no time trend in how well group members with hidden agendas achieve their hidden agendas.

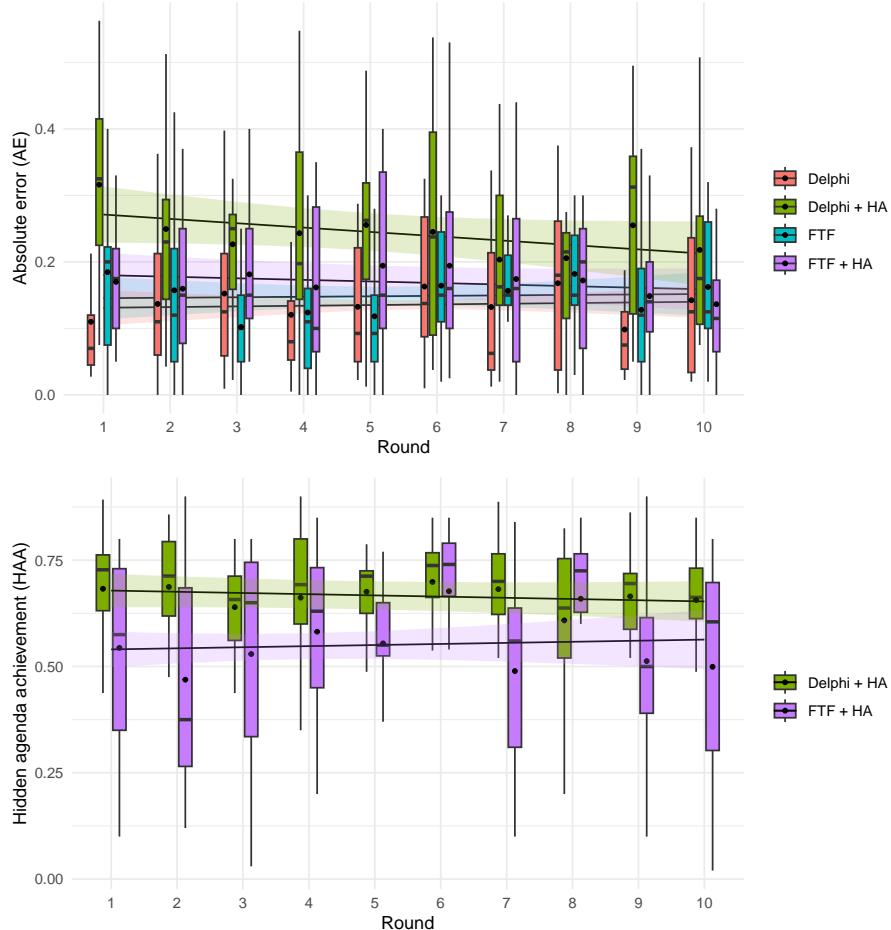
In Figure 2.5, I visualize accuracy and hidden agenda achievement over time by separating group judgments for the consecutive rounds 1, 2, ..., and 10 of the judgment task. Based on a general overview, in most rounds, the distribution of absolute errors tends towards higher absolute errors for Delphi groups with hidden agendas towards lower absolute errors for Delphi groups, while the distribution of absolute errors for FTF groups with and without hidden agendas falls in between. This aligns with the general results on accuracy aggregated across rounds as depicted earlier in Figure 2.2. To identify potential trends over time, I regress the absolute error as the dependent variable on the round as the independent variable. This reveals no statistically significant association between these variables. The linear regression lines, accompanied by their 95%-confidence intervals, are included in Figure 2.5. None of them has a slope that is significantly different from zero.

To quantify the degree to which group members with hidden agendas achieve their hidden agendas, I define the hidden agenda achievement rate HAA , dependent on the hidden agenda HA and the group judgment p of group g in round r as follows:

$$HAA_{g,r} = 1 - |HA_{g,r} - p_{g,r}|, \quad (2.2)$$

In this way, $HAA = 1$ corresponds to complete, i.e., 100%, achievement of the hidden agenda, while $HAA = 0$, indicates 0% achievement of the hidden agenda. $HA = 1$ if the hidden agenda was to drive group judgments up, and $HA = 0$ if the hidden agenda was to drive group judgments down. In most rounds, the distribution of hidden agenda achievement tends towards higher achievement rates for Delphi groups

Figure 2.5: Accuracy of group judgments and hidden agenda achievement over time



Notes: Whiskers in the boxplots span observations larger than $Q1 - 1.5 * IQR$ and smaller than $Q3 + 1.5 * IQR$. Regressions are OLS with standard errors clustered at the group level.

with hidden agendas and lower achievement rates for FTF groups with hidden agendas. This corroborates the general picture that FTF

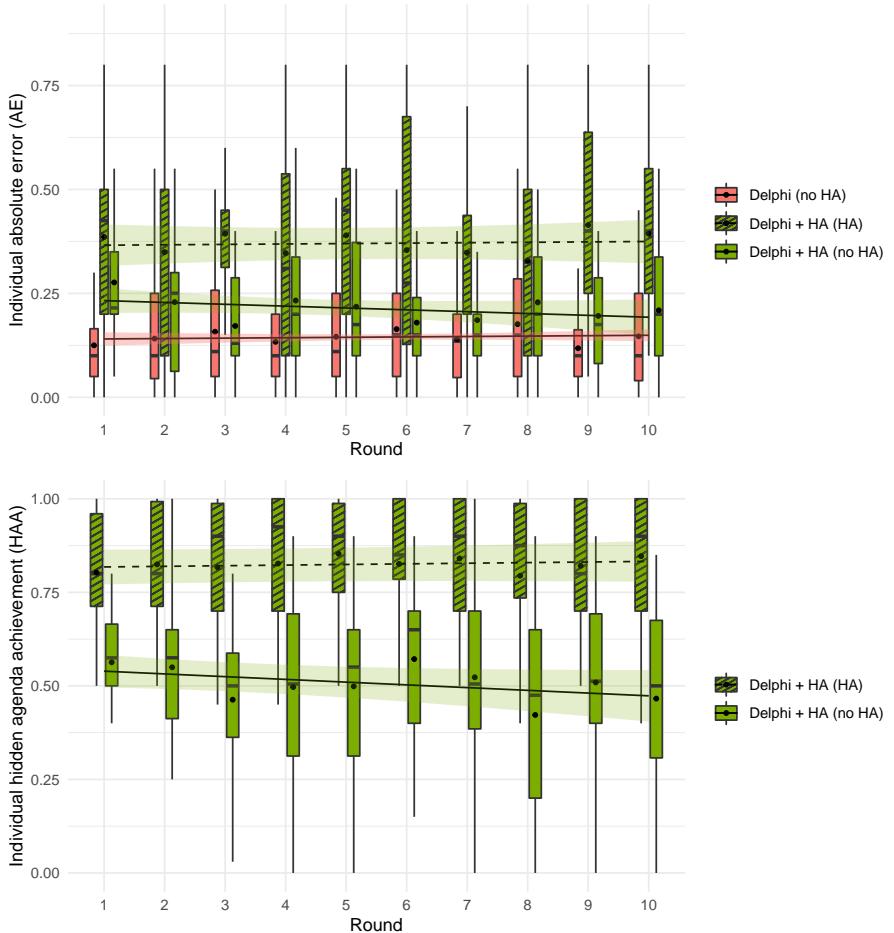
groups with hidden agendas are more accurate than Delphi groups with hidden agendas when aggregated over all rounds, as shown in Figure 2.2. As for accuracy, I find no statistically significant association between hidden agenda achievement and rounds.

I can extend the analysis to the individual level for Delphi groups. In any Delphi interaction, each group member provided a second individual judgment after observing their group members' first individual judgments and reasonings. The group judgment was then calculated as the mean of all second, individual judgments. As previously done for the group estimates, I analyze whether these second individual judgments change accuracy over time and whether they converge more or less in the direction of hidden agendas. *A priori*, effects in multiple directions seem plausible. Group members with hidden agendas may, over time, distort their individual judgments more and more in the direction of the hidden agenda, as they care less about losing credibility in the decreasing number of future interactions with their group. Alternatively, they might distort less over time as they become more cautious not to repeatedly behave in an overly suspicious way. Group members without hidden agendas may, over time, update their suspicions about who has a hidden agenda and form beliefs about the direction of the hidden agenda. Accordingly, they may try to counteract by distorting their own individual judgments in the opposite direction. I find no evidence for either of the two trends.

In Figure 2.6, I depict individual-level absolute errors and hidden agenda achievements over time. Overall, the distribution of absolute errors tends towards higher absolute errors for group members with hidden agendas in Delphi groups with hidden agendas and towards lower absolute errors for group members in Delphi groups without hidden agendas, while group members without hidden agendas in Delphi groups with hidden agendas are situated in between. This suggests that group members with hidden agendas do not only negatively influence the accuracy of Delphi group judgments but also the second individual judgments of members without hidden agendas in those groups. Further, in Delphi groups with hidden agendas, the individual judgments

of group members with hidden agendas are plausibly shifted more in the direction of the hidden agenda than those of group members without hidden agendas. Focusing on the time dimension, I regress the individual-level absolute error and hidden agenda achievement rate on rounds. This reveals that individual accuracy and hidden agenda achievement are stable across rounds. All regression lines have a slope that is statistically indistinguishable from zero.

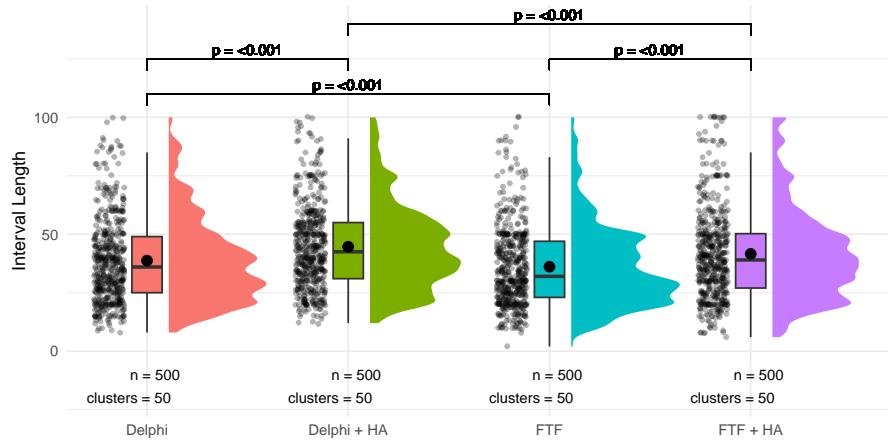
Figure 2.6: Accuracy and hidden agenda achievement in second individual judgments in Delphi groups over time



Notes: In Delphi groups without hidden agendas (Delphi), all four individual group members have no hidden agenda (no HA). In Delphi groups with hidden agendas (Delphi + HA), two individual group members have a hidden agenda (HA), and the remaining two individual group members have no hidden agenda (no HA).

Whiskers in the boxplots span observations larger than $Q1 - 1.5 * IQR$ and smaller than $Q3 + 1.5 * IQR$. Regressions are OLS with standard errors clustered at the individual level.

Figure 2.7: Width of confidence intervals around group judgments from different treatment conditions in the estimation experiment



Notes: P-values from from clustered signed rank tests (Datta and Satten, 2008; Jiang et al., 2020). Bar in boxplot = median; dot in boxplot = mean.

2.4.3 Evaluating trust in estimates

I measure the causal effect of hidden agendas and interaction formats on trust in group judgments by comparing confidence intervals in the decision experiment around group judgments from the estimation experiment. Based on the decision experiment, FTF groups always appear more trustworthy than Delphi groups, contradicting H4. In line with H5, hidden agendas impair trustworthiness, and the negative effect is of similar magnitude for FTF and Delphi, supporting H6.

I consider the width of confidence intervals to quantify the trustworthiness of group judgments. The narrower the confidence interval, the more trustworthy the group's judgment. Confidence interval width has a most trusting upper bound at 0, or in other words; the decision-maker is certain that the group judgment is precisely equal to the actual value. The least trusting lower bound is 100, or in other words, the group judgments do not contain any information, and the decision-maker

deems all possible values from 0 to 100 as equally likely to be true. Distributions of confidence interval length for judgments from each condition of the estimation experiment are summarized in Figure 2.7. For trustworthiness differences between conditions, I report p-values of non-parametric signed rank tests adapted for clustered paired data (Datta and Satten, 2008; Jiang et al., 2020).¹⁸

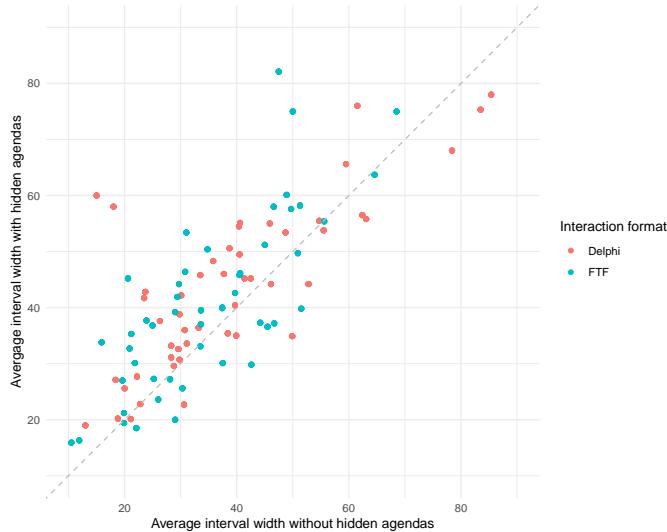
Result 4: *In contrast to H4, FTF group judgments enjoy significantly higher ($P < 0.001$) levels of trust, i.e. narrower stated confidence intervals, than Delphi group judgments. This holds with and without hidden agendas.*

Result 5: *In line with H5, introducing hidden agendas leads to significantly lower ($P < 0.001$) levels of trust, i.e., wider confidence intervals, for judgments from both Delphi and FTF interactions.*

Zooming in on the size of the negative effect of hidden agendas on trustworthiness, I calculate the average confidence interval width for group judgments with and without hidden agendas evaluated by a specific decision-maker. Figure 2.8 depicts the average interval width per decision-maker, where each decision-maker is represented by two points, one for their evaluation of FTF groups and one for Delphi groups, each with (vertical axis) and without (horizontal axis) hidden agendas. The distance to the 45-degree line measures the difference in trustworthiness with and without hidden agendas. As such, points above the 45-degree line correspond to less trustworthiness with hidden agendas than without hidden agendas. In line with Result 5, we see that more points lie above the 45-degree lines, and many are further away from the line than points below the 45-degree line. Further, suppose the impact of hidden agendas on trustworthiness is similar for FTF and Delphi.

¹⁸The Results 4 and 5 are robust towards averaging interval length across all 10 evaluated group judgments from the same condition of the estimation experiment, per evaluating decision-maker, and testing based on this average as only independent observation (compare Appendix 2.B).

Figure 2.8: Average width of confidence intervals per decision-maker around judgments of estimation experiment groups with and without hidden agendas



In that case, we should see no clusters of points representing FTF and Delphi group judgments, e.g., Delphi points generally lie further above the 45-degree line than FTF points. In line with Result 6, points are scattered along the 45-degree line without clearly visible clusters.

Result 6: *In line with H6, the size of the negative effect of hidden agendas on trust in judgments from FTF groups and Delphi groups is statistically not distinguishable ($p = 0.6328$).*

Based on the results on accuracy and trustworthiness, the ultimate question is: Does perceived trustworthiness align with accuracy? In other words, do decision-makers correctly put more trust in the more accurate judgments and vice versa?

Result 7: Trustworthiness is aligned with accuracy in settings with hidden agendas; the objectively more accurate judgments from FTF interaction are trusted more. However, trustworthiness is misaligned with accuracy in settings without hidden agendas; the objectively less accurate judgments from FTF interaction are trusted more.

Furthermore, the changes in trustworthiness are misaligned with changes in objective accuracy. While Delphi group judgments become relatively less accurate than FTF group judgments through the introduction of hidden agendas, trust in Delphi group judgments decreases just the same as trust in FTF group judgments.

2.4.4 BIN model

To formally disentangle the mechanisms behind accuracy differences between group interaction formats in settings with and without hidden agendas, I use the BIN model of forecasting (Satopää, Salikhov, Tetlock, et al., 2021). This model distinguishes three determinants of the accuracy of probabilistic judgments or estimates: bias, information usage, and noise (BIN) as latent, i.e., not directly observable, variables of an estimating group. Bias resembles a group's tendency to produce too high or too low estimates consistently, noise leads to non-systematic variability in estimates unrelated to the event of interest, and finally, information usage characterizes the variability in group judgments, which is correlated with the event of interest. Bias and noise are detrimental, while information usage is beneficial to accuracy. Given the modeling framework, all three latent variables can be structurally estimated based on group judgments and compared across conditions of the estimation experiment. Below, I briefly summarize the BIN model framework. Further, I outline the adaptions needed to accommodate the present setting of group judgments instead of individual forecasters in the original setting. For a more detailed discussion of the model, I refer the reader to the original paper (Satopää, Salikhov, Tetlock, et al., 2021).

Let the binary event of interest of estimation be denoted as $Y \in \{0, 1\}$, i.e. (not) reaching the target. Specifically, $Y = 1$ if $\sum_{t=1}^{10} X_t > 0$, i.e. the target is reached, and $Y = 0$ otherwise. Further, the model builds on two pillars: objective signals Z^* about Y and human estimates of the likelihood of Y based on the subjective interpretation of these signals Z . The outcome Y is considered to be determined by the entirety of all past and future objective signals, modeled as the continuous, normally distributed variable $Z^* \sim \mathcal{N}(\mathbb{E}[Z^*], \text{Var}(Z^*))$. In particular, the event is assumed to occur, i.e., the target area is reached, if the sum of all these signals is positive $Y = \mathbf{1}(Z^* > 0)$ where the indicator function $\mathbf{1}(E)$ equals 1 if E is true and 0 otherwise. Intuitively, this can be interpreted as an accumulation of evidence in favor of the event's occurrence. The model fixes the mean $\mathbb{E}[Z^*] = \mu^*$ and $\text{Var}(Z^*) = 1$, such that $\mathbf{P}(Z^* > 0) = p^*$, i.e. the expected frequency of the event happening based on Z^* is aligned with the true base rate $p^* \in (0, 1)$.

Groups in a specific (experimental) condition i judge the likelihood of Y based on their interpretation of signals Z_i . As with Z^* , Z_i is modeled as a normally distributed variable subject to three latent variables of group accuracy: bias, information usage, and noise (BIN). Intuitively, the closer the subjective Z_i is to the objective Z^* , the more accurate the group's judgments. Inaccuracy can take the form of bias as the difference in the means of Z^* and Z_i , or noise as the variability of Z^* that is uncorrelated with Z_i . Accuracy-enhancing information usage is described by the covariance of Z^* and Z_i . More formally, Z_i and Z^* follow the multivariate normal distribution:

$$\begin{pmatrix} Z^* \\ Z_i \end{pmatrix} = \mathcal{N} \left(\begin{pmatrix} \mu^* \\ \mu^* + \mu_i \end{pmatrix} \begin{pmatrix} 1 & \gamma_i \\ \gamma_i & \gamma_i + \delta_i \end{pmatrix} \right) \quad (2.3)$$

where the outcome and the latent variables are defined as:

$$Outcome : \quad Y = \mathbf{1}(Z^* > 0) \quad (2.4)$$

$$Bias : \quad \mu_i = \mathbb{E}[Z_i] - \mathbb{E}[Z^*] \quad (2.5)$$

$$Information : \quad \gamma_i = Cov(Z_i, Z^*) \quad (2.6)$$

$$Noise : \quad \delta_i = Var(Z_i) - Cov(Z_i, Z^*), \quad (2.7)$$

where the subscript i denotes a specific condition in the estimation experiment (FTF or Delphi interaction paired with (no) hidden agendas). For interpretation, perfectly accurate groups would be unbiased with $\mu_i = 0$, noise-free with $\delta_i = 0$, and exhibit $\gamma_i = 1 = Var(Z^*)$. Bias increases the further μ_i is from 0, noise increases the larger δ_i , and information usage decreases the further γ_i below 1.

Translating signals into probability judgments, the model considers groups as reporting their rational Bayesian belief of Y given Z_i :

$$\mathbb{E}[Y|Z_i] = \mathbb{P}[Z^* > 0|Z_i] \quad (2.8)$$

Groups are treated as if they report a judgment after they try their best to eliminate bias and noise. They are unaware of any remaining bias and noise in their interpretation of signals.¹⁹ Given this bounded rationality assumption, group judgments are given by:²⁰

$$\mathbb{P}[Z^* > 0|Z_i] = \Phi\left(\frac{Z_i}{\sqrt{1 - \gamma_i}}\right), \quad (2.9)$$

¹⁹This seems reasonable, as groups are incentivized through a proper scoring rule to produce the most accurate estimate possible.

²⁰See Satopää, Salikhov, Tetlock, et al., 2021 for the derivation of the conditional probability and a discussion of this assumption.

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution.

I use this framework to compare the bias, information usage, and noise of groups across conditions of the estimation experiment. In this way, I can quantify the causal effect of the interaction format and hidden agendas on these three latent variables of accuracy. To this end, the estimation process of groups in the same condition is assumed to be subject to the same bias, information usage, and noise. In other words, the estimation does not aim at the individual group level but treats any given group in a specific (experimental) condition as representative, i.e., as interchangeable with any other group in this condition. Further, outcomes and predictions across the ten rounds of the judgment tasks are assumed to be independent. Consequently, an estimate of, e.g., $\mathbb{P}[Y = 1] = 0.3$ in the first round of the judgment task says nothing about the estimate in the next round beyond the group's latent bias, information usage, and noise. As such, parameter estimates for bias, information usage, and noise can be seen as averages within a specific condition, e.g., the average bias of FTF groups with hidden agendas across the ten judgment tasks in the estimation experiment.

Estimates of bias, information usage, and noise

Following the econometric procedures of Satopää, Salikhov, Tetlock, et al., 2021, I estimate the parameters in the framework of the BIN model using Bayesian statistics. To quantify the causal effects of interaction formats and hidden agendas on bias, information usage, and noise, I consider four binary comparisons of conditions in the estimation experiment. On the one hand, two comparisons juxtapose interaction formats within the same incentive scheme: *FTF vs. Delphi* and *FTF with hidden agendas (FTF+HA) vs. Delphi with hidden agendas (Delphi+HA)*. On the other hand, two comparisons juxtapose incentive schemes within the same interaction format: *FTF vs. FTF+HA* and *Delphi vs. Delphi+HA*. For each binary comparison, I estimate posterior parameter distributions

based on a uniform prior. In other words, *a priori*, I make the conservative assumption that all parameter configurations in the BIN model are equally likely across the two compared conditions.²¹ To derive posterior distributions, parameter estimates are updated according to Bayes rule, taking observed group judgments, i.e., the data, into account. I modify the BINtools package (Satopää, Salikhov, and Moreno, 2022) to derive these parameter estimates by implementing Markov Chain Monte Carlo methods in R and Stan.²² Table 2.2 presents estimates of the BIN model parameters and posterior probabilities. Further details on the estimation are outlined in Appendix 2.C.1.

To structure results, I focus on two guiding questions consecutively: first, what is the likelihood of changes in accuracy that can be attributed to the mechanisms of bias, information usage, and noise? Second, what is the expected magnitude of changes in accuracy that can be attributed to each mechanism?

Regarding the likelihood of changes, noise is the most likely mechanism driving accuracy differences between interaction formats. By contrast, differences between incentive schemes are most likely driven by the use of valuable information and bias. Estimates of the likelihoods are presented in Table 2.2 in the section on posterior inferences. Posterior probabilities are the Bayesian equivalent of p-values in frequentist statistics. The larger the posterior probability, the stronger the evidence in favor of the hypothesis, and vice versa.

Comparing FTF to Delphi group judgments, Delphi group judgments exhibit less noise with very high probabilities ($P > 0.99$) in situations with and without hidden agendas. One potential reason may be that

²¹This prior is relatively uninformative. Hence, parameter estimates will largely be influenced by observed, actual group judgments.

²²Prior to the estimation, I aligned all group judgments such that any hidden agenda points in the direction of driving group judgments up. This ensures that the estimates of bias can be consistently interpreted in relation to the direction of the hidden agenda. Positive estimates of bias indicate that judgments are systematically distorted in the direction of hidden agendas and vice versa.

Table 2.2: Bayesian estimates of posterior inferences and BIN model parameters

| Summary Statistic | Interaction | | Hidden Agendas | | Interaction |
|--|------------------------|------------------------|------------------------|------------------------|-------------|
| | FTF vs. Delphi | FTF vs. FTF+HA | Delphi vs. Delphi+HA | FTF+HA vs. Delphi+HA | |
| Posterior inferences | | | | | |
| Less bias in treatment: $\mathbb{P}(\mu_1 < \mu_0)$ | 0.737 | 0.506 | 0.187 | 0.149 | |
| Less noise in treatment: $\mathbb{P}(\delta_1 < \delta_0)$ | 0.997 | 0.598 | 0.75 | 0.993 | |
| More info in treatment: $\mathbb{P}(\gamma_0 < \gamma_1)$ | 0.221 | 0.288 | 0.177 | 0.158 | |
| Parameter estimates (with 95% CI) | | | | | |
| Outcome mean: μ^* | 0.00 (-0.24; 0.23) | -0.01 (-0.24; 0.23) | 0.00 (-0.24; 0.23) | 0.00 (-0.23; 0.23) | |
| Bias (control): μ_0 | -0.10 (-0.72; 0.47) | -0.11 (-0.76; 0.47) | -0.09 (-0.53; 0.47) | 0.11 (-0.48; 0.47) | |
| Bias (treatment): μ_1 | -0.09 (-0.51; 0.31) | 0.11 (-0.51; 0.31) | 0.44 (0.09; 0.31) | 0.43 (0.09; 0.31) | |
| Diff in Bias: $ \mu_0 - \mu_1 $ | 0.06 (-0.10; 0.29) | 0.01 (-0.30; 0.29) | -0.26 (-0.59; 0.29) | -0.18 (-0.46; 0.29) | |
| Information (control): γ_0 | 0.14 (0.01; 0.40) | 0.15 (0.01; 0.40) | 0.12 (0.01; 0.40) | 0.15 (0.01; 0.40) | |
| Information (treatment): γ_1 | 0.10 (0.01; 0.27) | 0.12 (0.01; 0.27) | 0.07 (0.00; 0.27) | 0.07 (0.00; 0.27) | |
| Diff in information: $\gamma_0 - \gamma_1$ | 0.05 (-0.06; 0.17) | 0.03 (-0.08; 0.17) | 0.05 (-0.06; 0.17) | 0.08 (-0.06; 0.17) | |
| Noise (control): δ_0 | 0.94 (0.19; 3.21) | 0.96 (0.15; 3.21) | 0.29 (0.01; 3.21) | 0.83 (0.12; 3.21) | |
| Noise (treatment): δ_1 | 0.31 (0.02; 1.24) | 0.90 (0.15; 1.24) | 0.18 (0.02; 1.24) | 0.17 (0.01; 1.24) | |
| Diff in noise: $\delta_0 - \delta_1$ | 0.63 (0.12; 2.10) | 0.06 (-0.34; 2.10) | 0.12 (-0.10; 2.10) | 0.66 (0.05; 2.10) | |

Control: condition named on top, Treatment: condition named below CI: credible intervals

Delphi group judgments constitute an average of four individual judgments. In contrast, in FTF interactions, the group judgment constitutes a consensus judgment among the four group members.

Result 8: *Noise is the most likely mechanism driving accuracy differences between interaction formats. Delphi group judgments are likely less noisy than FTF group judgments.*

Comparing situations without and with hidden agendas within the same interaction format, less usage of valuable information is the most likely mechanism contributing to a decline in accuracy for FTF ($P = 0.712$) and even more so for Delphi groups ($P = 0.823$).

For Delphi groups with hidden agendas, an increase in bias is about as likely ($P = 0.813$) as a decrease in the usage of valuable information. This is not mirrored in FTF groups, which are almost equally likely to exhibit less or more bias with hidden agendas compared to situations without. This points towards robustness against bias as differentiating mechanisms that may explain why FTF group judgments do not become significantly less accurate in situations with hidden agendas, while Delphi groups do. Nevertheless, posterior probabilities for the effect of hidden agendas are far from any usual significance level used when interpreting p-values in frequentist settings. Statements on the likelihood of information usage and bias driving accuracy differences between situations with and without hidden agendas should thus be interpreted cautiously.

Result 9: *Less use of valuable information is the most likely mechanism for decreasing accuracy with hidden agendas irrespective of the interaction format.*

Result 10: *Increased bias likely decreases accuracy for Delphi group judgments in situations with hidden agendas. FTF group judgments appear robust against this effect.*

In terms of magnitudes of changes, noise emerges as the only mechanism with a clear directional effect on accuracy differences between interaction formats. Estimates of the effect sizes of bias, information

usage, and noise on accuracy are presented in the lower part of Table 2.2. Of particular interest are estimates for the difference in bias ($|\mu_1| < |\mu_0|$), information usage ($\gamma_0 - \gamma_1$), and noise ($\delta_1 - \delta_0$).

Result 10: *Noise is the only mechanism to which a clear directional effect can be attributed. FTF group judgments exhibit more noise than Delphi group judgments.*

Comparing FTF to Delphi group judgments, Delphi group judgments exhibit less noise in situations without ($\delta_0 - \delta_1 = 0.63$) and with hidden agendas ($\delta_0 - \delta_1 = 0.66$). The respective 95% credible intervals around the estimates of $\delta_0 - \delta_1$ do not include zero. The effects of information usage and bias are estimated with 95% credible intervals spanning zero. Thus, I cannot identify a clear directional effect of these mechanisms in the BIN model framework.

Comparing situations without and with hidden agendas within the same interaction format, bias, information usage, and noise can not be identified with a clear directional effect. The respective 95% credible intervals around the estimates span zero.

Wide credible intervals for magnitudes of expected effects can likely be attributed to the relatively small number of judgment tasks solved per group in the estimation experiment. Satopää, Salikhov, Tetlock, et al., 2021 show that the width of credible intervals decreases substantially if the number of forecasts by the same person increases. This holds especially true for relatively low numbers of forecasts (below 50). While Satopää, Salikhov, Tetlock, et al., 2021 use data from the Good Judgment Project, which comprises between 87 and 191 forecasts per person, the estimation experiment yields only 10 distinct judgments per group.²³

²³Using 10 distinct judgments per group ensured implementability in a controlled lab experiment. This presents clear advantages for analyzing the causal effects of group interaction formats and incentive schemes on accuracy and trust. Conversely, experimental control has been traded off against a higher number of observations, which might have allowed more precise estimates of mechanisms driving accuracy

2.4.5 Communication Patterns

The BIN model estimates provide a good starting point to shed light on potential mechanisms underlying accuracy differences across conditions of the estimation experiment. Yet, essential differences might not be uncovered by structural estimations alone. I analyze the degree and truthfulness of information sharing during group interaction to complement the BIN model estimations. To this end, the communication of FTF and Delphi interactions with and without hidden agendas are transcribed and coded.²⁴ On a high level, the communication protocols reveal that hidden agendas cause less shared, truthful information in Delphi groups but not in FTF groups.

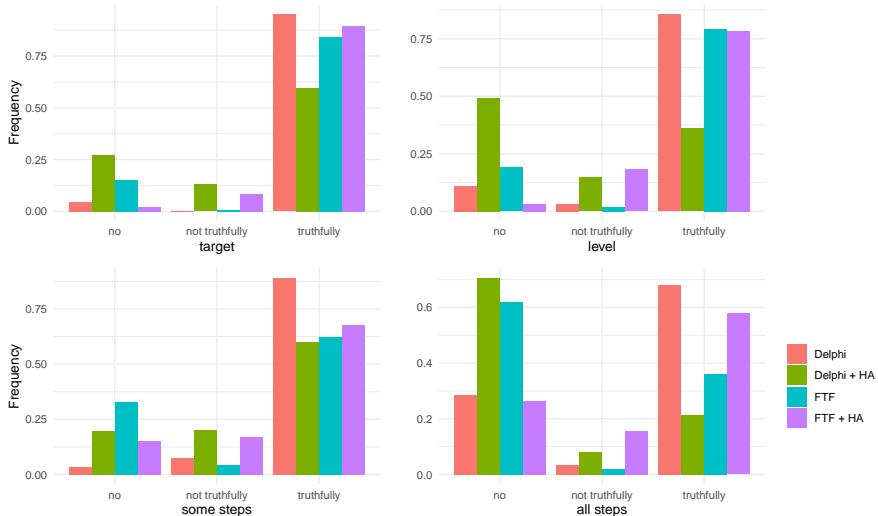
The probabilistic judgment task of the estimation experiment is generally designed such that the more information participants, i.e., group members, share with their group, the more accurately the group can solve the task. In increasing levels of detail, group members can reveal whether their individual movement path reached the target (target), the final level of their movement path after 10 steps (level), one or more individual steps of their movement path (some steps), and most comprehensively all 10 steps of their movement path (all steps). Furthermore, group members may state their information truthfully or not truthfully for each level of detail. Withholding or not stating information truthfully may be a strategy to drive group estimates up or down according to one's hidden agenda.

In Figure 2.9, I summarize the frequencies of not stated (no), not truthfully stated (not truthfully), and truthfully stated (truthfully) information at the previously described four levels of detail. I estimate corresponding linear probability models to identify the significance of differences in frequencies. Estimation results are presented in Appendix 2.D.

differences in the BIN model framework.

²⁴A detailed outline of the transcription and coding procedure can be found in Appendix 2.D.

Figure 2.9: Revelation of information across conditions of the estimation experiment



Focusing first on whether information is shared at all, it becomes visible that in Delphi interactions with hidden agendas, significantly more group members choose not to reveal information when compared to Delphi interactions without hidden agendas. This holds at all levels of detail. Further, with hidden agendas, the frequency of shared non-truthful information is significantly higher, and the frequency of shared truthful information is significantly lower. Again, this holds at all levels of detail. This observation is intuitively aligned with the lower accuracy of Delphi groups with hidden agendas. Not only do hidden agendas lead to less information sharing, but they also decrease the truthfulness of shared information.

We see a vastly different pattern when comparing FTF groups with and without hidden agendas. With hidden agendas, significantly more group members choose to reveal information. This holds at all levels of detail. Just as for Delphi, in FTF groups with hidden agendas, the fre-

quency of shared non-truthful information is significantly higher for all levels of detail. However, unlike Delphi, the frequency of shared truthful information is significantly higher for the target, some steps, and all steps. This observation suggests an intuitive explanation for the robust accuracy of FTF groups despite hidden agendas. While hidden agendas lead to more sharing of untruthful information, at the same time, the sharing of truthful information increases and potentially compensates for the negative effects of untruthful information sharing.

2.5 Conclusion and discussion

In this paper, I study whether commonly used face-to-face meetings and the scientifically supported Delphi technique are suitable interaction formats to generate and elicit accurate and trustworthy group judgments. In particular, I consider situations where some group members have a hidden agenda, i.e., incentives to manipulate. Through two complementary experiments, I provide evidence supporting the widespread use of FTF meetings in real-world institutional decision-making. Hidden agendas are a potential threat to the accuracy and trustworthiness of group judgments. However, FTF emerges as capable of mitigating this threat. FTF interaction appears to be a resilient generalist capable of generating accurate and trusted group judgments even under adverse conditions with hidden agendas. By contrast, the Delphi technique appears to be a specialist adapted to one niche, capable of peak performance and outperforming FTF, but only without hidden agendas and only in terms of accuracy.

Using structural estimations in the Bayesian BIN model framework, I pinpoint increased bias as an accuracy-impairing mechanism that may explain differences between FTF and Delphi. While Delphi group judgments likely suffer from increased bias toward the direction of hidden agendas, FTF group judgments do not exhibit this effect. Robustness towards bias, thus, seems a relevant concern when determining and designing if and how decision-informing group judgments are generated

and hidden agendas are likely. This study provides a building block of evidence from a controlled lab experiment that favors employing unstructured FTF group judgments in these situations. However, while FTF can be identified as relatively better than Delphi, it remains an open question whether unstructured FTF is the best among a wider set of interaction formats in situations with hidden agendas and whether the result holds in different contexts.

Further research may thus zoom in on identifying particular features of FTF underlying its robustness towards manipulation. This may deepen our understanding of why certain existing interaction formats work better than others, but it will ultimately also help to design new interaction formats to better extract accurate and trustworthy group judgments. To this end, the experimental setup of this study may serve as a test bed that is well suited to investigating the accuracy and trustworthiness of any group interaction format directed towards generating quantitative judgments in general. To isolate the effect of particular features of an interaction format, modifications of group interaction formats that vary by one detail at a time may be tested against each other. Furthermore, the present study investigates settings where the existence of hidden agendas is common knowledge. In institutional decision-making, however, it is more likely that one can only suspect the presence of hidden agendas; hence, there is uncertainty about their presence. Future research could expand to such settings where hidden agendas may be present probabilistically. Beyond this, the investigation of the capabilities of group interaction to generate accurate and trustworthy judgments should also be extended to settings outside of the controlled lab environment and tested in the field. A potential stepping stone could be to investigate the presence and relevance of hidden agendas in forecasting tournaments such as the Good Judgment Project (Mellers et al., 2014).

Taken together, this study emphasizes that hidden agendas and potential manipulation matter and may harm the accuracy and trustworthiness of collective intelligence. It is vital to account for the existence of these hidden agendas to alleviate their threat to institutional decision-

making. The presented evidence suggests that FTF is the preferable interaction format for generating accurate and trustworthy group judgments in situations with hidden agendas. The Delphi technique is preferable in situations without information where the accuracy of group judgments is of top priority. Practitioners involved in eliciting group judgments may refer to these results when structuring the interaction format of groups.

Appendix

2.A Experimental details

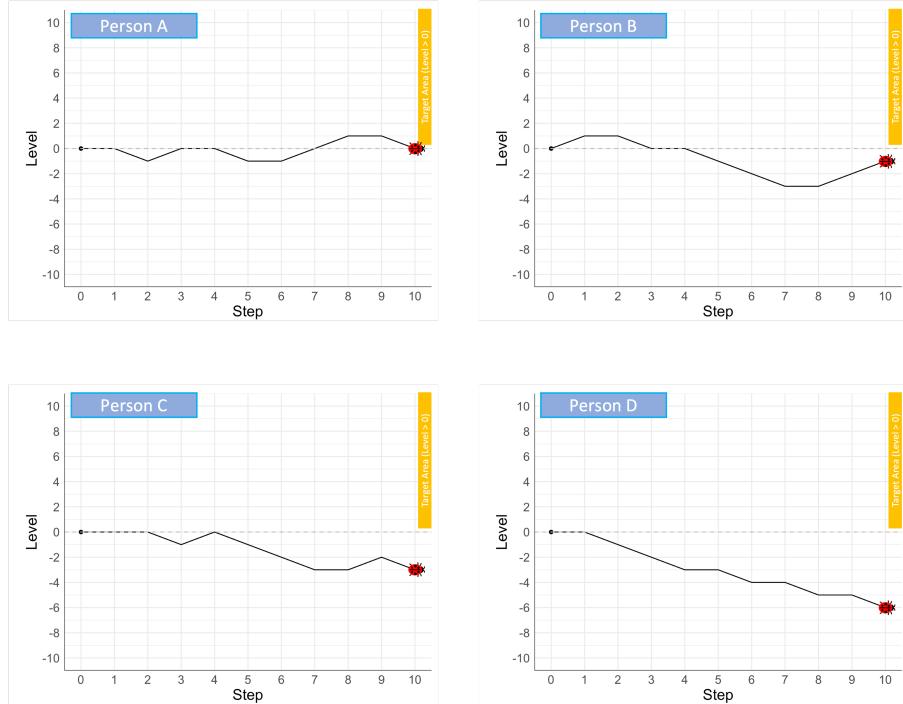
2.A.1 Estimation experiment

2.A.1.1 Judgment task

The judgment task is framed in a gamified story to ease understanding and increase the engagement of experimental participants. In particular, the movement path is framed to be generated by a computer program. Group members are told that the computer program generates the movement path of a ladybird on a map. They need to estimate the chance that the ladybird reaches the target area at the end of its path after ten steps. The chance is to be expressed in frequencies, such as 75 in 100, to facilitate better statistical reasoning (McDowell and Jacobs, 2017). As information, each sees a movement path generated by a distinct past execution of the program, compare Figure 2.10. For each round, group members are informed that a distinct computer program generates the movement paths, and consequently, the underlying probability of reaching the target might be different.

This judgment task amalgamates key features of the hidden-profile paradigm (Stasser and Titus, 1985), a benchmark task of group interaction in psychology, with additional aspects enabling more detailed analysis and broader real-world references. Common features comprise experimenter control over the information needed to solve the problem and its distribution across group members, i.e., participants resemble heterogeneous and complementary experts with the degree of expertise induced by the experiment. Consequently, the more private information is shared during group interaction, the more accuracy can be achieved in the group estimate. However, going beyond the standard hidden profile paradigm, this task incorporates uncertainty and requires probabilistic judgments. This comes with four major advantages. First, estimates can be evaluated based on a statistical measure of accuracy.

Figure 2.10: Example of information received by group members (Person A, B, C, and D)



Second, the statistical measure of accuracy enables incentives that continuously reward accuracy and hidden agenda achievement. Third, participants' behavior can be evaluated against theoretically optimal behavior given the information they have at hand. Finally, probabilistic judgment caters to real-world applications that go far beyond the motivating examples presented in the introduction.

2.A.1.2 Incentives

All participants receive a show-up fee of €5 and performance-based remuneration for solving the task as accurately as possible. Additionally,

to create a situation of hidden agendas, two group members can earn a supplementary individual bonus in hidden agenda treatments by influencing the group's judgment in a particular direction.

For each of the ten rounds of the judgment task, all groups receive an **accuracy-based bonus**, calculated based on the binarized quadratic scoring rule (Hossain and Okui, 2013). The group bonus is split equally among all four group members. This ensures that individuals are incentivized to pursue the most accurate group estimate irrespective of their risk preferences. Specifically, the group bonus in a given round amounts to €6 with a certain chance and €0 otherwise. The more accurate the group estimate, the greater the chance the group bonus is €6. To illustrate the calculation of the group bonus, let $Y_r \in \{0, 1\}$ be the binary event that the movement path does (not) reach the target in round r . $Y_r = 1$ if the movement path reaches the target in round r , i.e. $\sum_{t=1}^{10} X_t > 0$, and 0 otherwise. Moreover, p_r is the group estimate of the chance that the movement path reaches the target area in a particular round r . Following the binarized quadratic scoring rule, then the group bonus in round r is €6 if $Y_r = 1$ and a random number $Y \sim U(0, 1) > (1 - p_r)^2$ or if $Y_r = 0$ and a random number $Y \sim U(0, 1) > p_r^2$.²⁵ By way of illustration, consider the true chance that the movement path reaches the target is $p^* = 0.7$, and the group estimated the chance at $p_r = 0.6$. In this case, 70% of actual realizations will reach the target, and 30% will not. Consequently, the chance of winning the group bonus is $0.7*(1-(1-0.6)^2)+0.3*(1-0.6^2)$, which is $0.7*0.84+0.3*0.64 = 0.78$. In other words, the chance of winning the group bonus is 78% in total, i.e. in expectation €4.68 group bonus or in other words €1.17 bonus per person.

In hidden agenda situations, in addition to the accuracy-based bonus, two out of four group members may each receive an **individual hidden agenda bonus** of €1.50 per round based on a binarized scoring. Their hidden agenda is to drive the group estimate to 1 or 0, i.e., up or down.

²⁵The payoff relevant realizations of the movement path Y_r have been generated for all rounds $r = 1, \dots, 10$ once before the experiment and are kept constant for all groups in the experiment.

Whether it is one or the other direction is determined randomly in each round. The direction of the hidden agenda is always the same for the two group members with a hidden agenda in a given round. The hidden agenda bonus is designed such that it is always best for expected payoff-maximizing participants to follow their hidden agenda as much as possible. If following their hidden agenda, the decrease in the chance of earning the group bonus is overcompensated by gains in the chance of earning the individual hidden agenda bonus. Precisely, the chance of winning the hidden agenda bonus in round r is $(1 - p_r)^2$ if the hidden agenda is to drive the estimate to 0, i.e., down, and p_r^2 if the hidden agenda is to drive the estimate to 1, i.e., up.

2.A.2 Decision experiment

2.A.2.1 Incentives

All participants in the decision experiment receive a show-up fee of €10 and performance-based remuneration depending on their stated confidence intervals. The latter follows the most likely interval method (Schlag and Weele, 2015), i.e., a stated confidence interval with lower bound L and upper bound U translates into a payment S :

$$S(L, U, p^*) = \begin{cases} 10(1 - \frac{U-L}{100}) & \text{if } 100p^* \in [L, U] \\ 0 & \text{otherwise,} \end{cases} \quad (2.10)$$

depending on the respective true probability p^* . Note that estimates generated in the estimation experiment are naturally bounded by 0 and 100 in frequency terms. Further, the width of the stated confidence interval ranging from lower bound L to upper bound is $U - L$, which is at most 100. Consequently, the wider the stated confidence interval, the smaller the bonus paid in case the confidence interval contains $100p^*$, ultimately approaching 0 if the confidence interval spans the entire interval of possible values. I implement performance-based

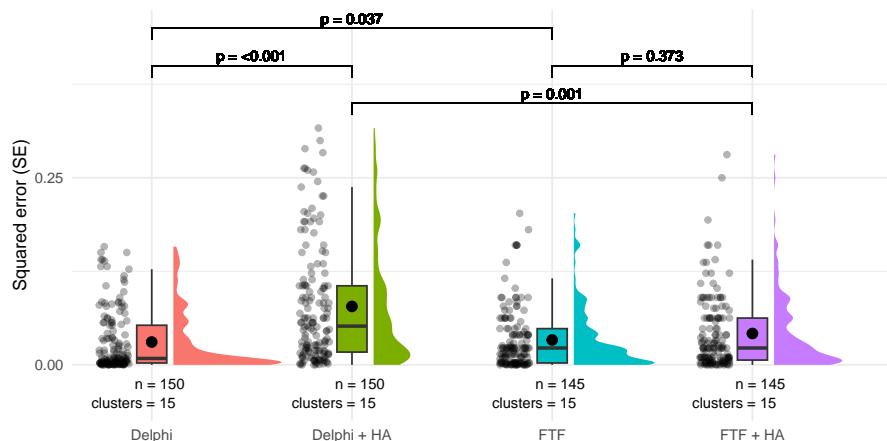
remuneration as a random lottery incentive, i.e., after participants stated their confidence intervals on all 40 estimates, one confidence interval is chosen at random for payout according to the above-mentioned most likely interval payment rule. This method is designed to encourage participants to treat each interval with equal, high care.

2.B Robustness checks

2.B.1 Accuracy

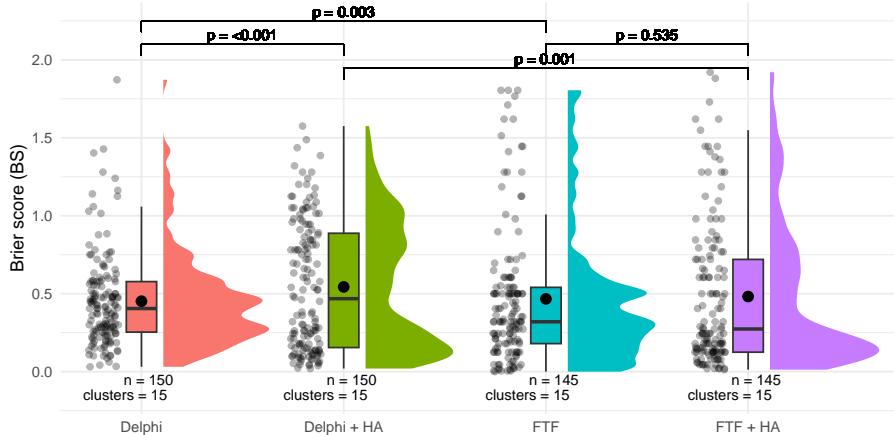
While the main analysis of accuracy focuses on absolute errors, accuracy can also be quantified using alternative metrics. Accordingly, I report test results based on squared errors (Figure 2.11) and Brier Scores (Figure 2.12). Furthermore, I report test results varying in the strictness of considering the dependence of judgments made within the same group. Two of the main results reported in Section 2.4 are robust in all alternative tests: First, hidden agendas decrease accuracy only for Delphi groups, and second with hidden agendas FTF groups are more accurate. Without hidden agendas, however, Delphi groups are no longer statistically more accurate than FTF groups in the robustness test, which is most strict concerning the potential dependence of observations.

Figure 2.11: Squared errors of group judgments across experimental conditions



Notes: P-values from two-sided Wilcoxon rank sum tests (Rosner, Glynn, and Lee, 2003; Jiang et al., 2020). Bar in boxplot = median; dot in boxplot = mean.

Figure 2.12: Brier Scores of group judgments across experimental conditions



Notes: P-values from two-sided Wilcoxon rank sum tests (Rosner, Glynn, and Lee, 2003; Jiang et al., 2020). Bar in boxplot = median; dot in boxplot = mean.

Squared errors (SE), just like absolute errors juxtapose group judgments ($p_{g,r}$) to true probabilities (p_r^*), but consider the squared difference:

$$SE_{g,r} = (p_{g,r} - p_r^*)^2 \quad (2.11)$$

Consequently, the distribution of squared errors is narrower compared to absolute errors. However, the rank of observations remains unchanged. Therefore, test results are identical to those reported in Section 2.4.

Brier Scores (BS) are calculated based on group judgments ($p_{g,r}$) and actual outcomes (Y_r), i.e., binary outcomes (0 or 1 for hit or missed target, respectively) drawn randomly according to the true underlying probabilities (p_r^*):

$$BS_{g,r} = (Y_r - p_{g,r})^2 + ((1 - Y_r) - (1 - p_{g,r}))^2 \quad (2.12)$$

Testing for differences in the distributions of Brier Scores across treatment conditions, I find the same results as with AEs and SEs. This is, however a stronger result, as Brier Scores are noisier than absolute and squared errors, which holds especially with a small number of distinct judgment tasks, like the ten different rounds in the estimation experiment. For illustration, consider that the underlying true probabilities are $[0.1, 0.2, 0.3, 0.4, 0.45, 0.55, 0.6, 0.7, 0.8, 0.9]$ and the accordingly drawn actual outcomes, which are used to calculate Brier scores, are $[0, 1, 0, 0, 1, 0, 1, 1, 1, 1]$.

Table 2.3: Wilcoxon rank sum tests on accuracy with different considerations of dependence of observations

| | Interaction | | Hidden Agendas | | Interaction | |
|---|-------------------|-----------------------|-----------------------------|---------------------------------|-------------|--|
| | FTF vs. Delphi | FTF vs. FTF^{HA} | Delphi vs. $Delphi^{HA}$ | FTF^{HA} vs. $Delphi^{HA}$ | | |
| <i>Wilcoxon rank sum test on AE, average per group</i> | | | | | | |
| n | 30 | 30 | 30 | 30 | | |
| p-value | 0.115 | 0.52 | <0.001 | 0.001 | | |
| <i>Clustered Wilcoxon rank sum test on AEs (Rosner, Glynn & Lee 2003)</i> | | | | | | |
| n | 295 | 290 | 300 | 295 | | |
| cluster | 30 | 30 | 30 | 30 | | |
| p-value | 0.037 | 0.373 | <0.001 | 0.001 | | |

Varying the strictness of considering the dependence of judgments made within the same group, I report results of Wilcoxon rank sum tests for average AE per group (most strict), in Table 2.3. The most strict scenario posits that groups make 10 judgments over the 10 rounds of the task, but in this way generate only one independent observation. Reporting test results only using average AE per group takes this into

account. The clustered Wilcoxon rank sum tests reported in Section 2.4 are less strict. They actively control for the degree of dependence of observations within the same group. The most strict approach confirms the main results obtained through clustered Wilcoxon rank sum tests, except for Delphi groups being more accurate without hidden agendas. FTF and Delphi groups' accuracy becomes statistically indistinguishable in this case.

2.B.2 Trustworthiness

Similar to accuracy, I vary the strictness of considering the dependence of observations generated in the decision experiment. The main analysis reports results of clustered Wilcoxon signed rank tests that actively control for the dependence of confidence intervals stated by the same decision-maker. Obtained test results are robust towards considering the most strict scenario, positing that a decision-makers evaluating ten judgments from a given condition of the estimation experiment only generates one independent observation. Accounting for this, Table 2.4 reports results based on the average confidence interval width per decision-maker and condition in the estimation experiment.

Table 2.4: Wilcoxon signed rank tests on trust with different considerations of dependence of observations

| | Interaction | Hidden Agendas | |
|---|-------------------|-----------------------|-----------------------------|
| | FTF vs. Delphi | FTF vs. FTF^{HA} | Delphi vs. $Delphi^{HA}$ |
| <i>Wilcoxon signed rank test on interval widths, average per participant</i> | | | |
| n | 100 | 100 | 100 |
| test statistic | 866 | 286.5 | 280 |
| p-value | 0.028 | 0.001 | 0.001 |
| <i>Clustered Wilcoxon signed rank test on interval widths (Datta & Satten 2008)</i> | | | |
| n | 1000 | 1000 | 1000 |
| cluster | 50 | 50 | 50 |
| p-value | <0.001 | <0.001 | <0.001 |

2.C Modelling framework

2.C.1 BIN model estimation

The implementation of the BIN model largely follows the econometric procedures of Satopää, Salikhov, Tetlock, et al., 2021. For details on these procedures, such as the underlying likelihood function, I refer the reader to the original paper and the documentation of the statistical companion package BINtools (Satopää, Salikhov, and Moreno, 2022). In the following, I focus on outlining adaptions to the original procedures I made to accommodate better the data from the estimation experiment of this study.

First, I exploit the fact that within my experimental design, the true probabilities for the ten distinct forecasts to be made by each group are known. In particular, each group faces judgment tasks with objectively true probabilities $\mathbb{P}(\sum_{t=1}^{10} X_t > 0) = \{0.1, 0.2, 0.3, 0.4, 0.45, 0.55, 0.6, 0.7, 0.8, 0.9\}$. With this knowledge, I can fix $\mu^* = 0$, i.e., the mean of the normal distribution from which actual true probabilities in the model are drawn.²⁶ Second, I increase the numbers of iterations `warmup` and `iter`, and I set estimation control parameters for `adapt_delta`, `stepsize`, and `max_treedepth` according to Stan estimation feedback provided by BINtools. Third, I perform an iterative grid-search on the initial values of model parameters $\mu_0, \mu_1, \gamma_0, \gamma_1, \delta_0, \rho_0, \delta_1, \rho_1$, and ρ_{01} , to reduce the number of divergent transitions after warmup.

²⁶The parameter is fixed in the Stan source code of the R BINtools package by restricting $-0.01 < \mu^* < 0.01$.

2.D Analysis of communication protocols

To analyze emergent communication patterns across conditions of the estimation experiment group and round specific transcripts were created. For Delphi groups, I extracted these from the written reasoning provided by group members alongside their first individual estimate. For FTF groups, research assistants who were unaware of the research questions underlying this paper, transcribed the video recordings of respective experimental sessions.²⁷ Subsequently, the transcripts were coded according to the coding scheme on characteristics of information sharing shown in Appendix 2.E.²⁸

²⁷The research assistants used the automatically generated transcripts of the zoom video call as a starting point and corrected parts of the communication that had been incorrectly transcribed.

²⁸Coding and analysis of communication protocols according to the registered coding schemes on the general impression of group interaction and on quantifiable characteristics of group interaction have been postponed to future research.

Table 2.5: OLS estimation results for linear probability models on information revelation across conditions of the estimation experiment

| | target | target_lie | target_truth | level | level_lie | level_truth |
|-------------|---------------------|---------------------|----------------------|----------------------|---------------------|----------------------|
| (Intercept) | 0.047*** (0.013) | 0.003 (0.009) | 0.950*** (0.015) | 0.112*** (0.015) | 0.033** (0.012) | 0.855*** (0.017) |
| delphi_ha | 0.227*** (0.018) | 0.127*** (0.013) | -0.353*** (0.021) | 0.382*** (0.021) | 0.113*** (0.016) | -0.495*** (0.024) |
| ftf | 0.103*** (0.018) | 0.003 (0.013) | -0.107*** (0.021) | 0.078*** (0.021) | -0.015 (0.016) | -0.063** (0.024) |
| ftf_ha | -0.027 (0.018) | 0.082*** (0.013) | -0.055** (0.021) | -0.078*** (0.021) | 0.148*** (0.016) | -0.070** (0.024) |
| Num.Obs. | 2400 | 2400 | 2400 | 2400 | 2400 | 2400 |

* p < 0.05, ** p < 0.01, *** p < 0.001

| | some_steps | some_steps_lie | some_steps_truth | all_steps | all_steps_lie | all_steps_truth |
|-------------|---------------------|---------------------|----------------------|---------------------|---------------------|----------------------|
| (Intercept) | 0.035* (0.015) | 0.073*** (0.013) | 0.892*** (0.018) | 0.285*** (0.019) | 0.035*** (0.010) | 0.680*** (0.019) |
| delphi_ha | 0.162*** (0.021) | 0.128*** (0.019) | -0.290*** (0.026) | 0.422*** (0.027) | 0.045** (0.015) | -0.467*** (0.027) |
| ftf | 0.295*** (0.021) | -0.028 (0.019) | -0.267*** (0.026) | 0.335*** (0.027) | -0.015 (0.015) | -0.320*** (0.027) |
| ftf_ha | 0.117*** (0.021) | 0.097*** (0.019) | -0.213*** (0.026) | -0.020 (0.027) | 0.122*** (0.015) | -0.102*** (0.027) |
| Num.Obs. | 2400 | 2400 | 2400 | 2400 | 2400 | 2400 |

* p < 0.05, ** p < 0.01, *** p < 0.001

Notes: Each column represents a distinct dependent variable, a binary indicator of information revelation. (1) *target*: not revealing whether the target has been reached, (2) *target_lie*: misstating whether the target has been reached, (3) *target_truth*: correctly stating whether the target has been reached, (4) *level*: not revealing the final level of the movement path, (5) *level_lie*: misstating the final level of the movement path, (6) *level_truth*: correctly stating the final level of the movement path, (7) *some_steps*: not revealing any particular steps of the movement path (8) *some_steps_lie*: revealing one or more particular steps of the movement path, of which at least one is misstated (9) *some_steps_truth*: revealing one or more particular steps of the movement path, of which none is misstated, (10) *all_steps*: not revealing all ten particular steps of the movement path, (11) *all_steps_lie*: revealing all steps of the movement path, of which at least one is misstated, (12) *all_steps_truth*: revealing all steps of the movement path, of which none is misstated.

2.E Coding scheme for video recordings

2.E.1 Preregistered coding scheme

Questions on the general impression of group interaction

- Did the group members follow a structure to discuss and generate a group estimate? (5-point Likert scale, strongly agree to strongly disagree)
 - Guidance: Did they take steps that were the same in multiple rounds? Steps could e.g., be an open discussion where everybody presents their arguments, a step where everybody presents their individual estimate, an aggregation step where individual estimates are averaged to generate a group estimate, etc.
 - Please briefly describe the structure followed by the group and the steps taken, if applicable. (open text)
- Please rate the overall influence of person A, B, C, and D throughout the group discussions. (5-point Likert scale, very influential to very uninfluential)
- Is there anything else you found striking about the group interaction? (open text)

Quantifiable characteristics of group interaction (coded for each of the 10 estimation rounds separately)

- Words spoken per person A, B, C, and D (extracted from automatically generated video conference transcripts)
- Number of numerical estimates uttered per person A, B, C, and D (extracted from automatically generated video conference transcripts)

-
- Guidance: Numerical estimates could be in the form of frequencies, e.g. 10 in 100, the form of probabilities, e.g. 0.1, or the form of percentages, e.g. 10%.
 - Number of qualitative/verbiage estimates uttered per person A, B, C, and D (extracted from automatically generated video conference transcripts)
 - Guidance: Qualitative/verbiage estimates are e.g. very likely, very probable, impossible, almost certain, etc.
 - Number of mentions of “hidden agendas” or “manipulation” per person A, B, C, and D (extracted from automatically generated video conference transcripts)
 - Guidance: Qualitative/verbiage estimates are e.g. very likely, very probable, impossible, almost certain, etc.
 - Number of arguments uttered per person A, B, C, and D
 - Guidance: An argument usually entails a premise: e.g. My info indicates ..., and a conclusion: e.g. therefore I believe the chance to be ...
 - Repetitions are counted the same way as genuine arguments
 - Number of confirmations uttered per person A, B, C, and D
 - Guidance: A confirmation could be: “I agree with the argument of ...”, “I think the point of ... is true”, etc.
 - Number of confirmations from other persons received per person A, B, C, and D
 - Guidance: A confirmation could be: “I agree with the argument of ...”, “I think the point of ... is true”, etc.
 - Number of qualifications uttered per person A, B, C, and D
 - Guidance: A qualification could be: “I do partly agree with ..., but I think ...”, etc.

- Number of qualifications from other persons received per person A, B, C, and D
 - Guidance: A qualification could be: "I do partly agree with ..., but I think ...", etc.
- Number of direct rebuttals/attacks on other persons per person A, B, C, and D
 - Guidance: A direct rebuttal/attack could be: "I do disagree with the argument of ...", "I think the point of ... cannot be true", etc.
- Number of direct rebuttals/attacks received per person A, B, C, and D
 - Guidance: A direct rebuttal/attack could be: "I do disagree with the argument of ...", "I think the point of ... cannot be true", etc.
- Please rate the state of agreement on the final group estimate. (5-point Likert scale, strong agreement to strong disagreement)

2.E.2 Coding scheme extension post data-collection

Characteristics of information sharing (coded for each of the 10 estimation rounds separately in Delphi and FTF treatments)

- Did person A, B, C and D state whether their movement path reached the target? (no, yes truthfully, yes but not truthfully)
- Did person A, B, C, and D reveal the final level of their movement path? (no, yes truthfully, yes but not truthfully)
- Did person A, B, C, and D reveal information on one or more values (-1,0,1) of individual steps of their movement path? (no, yes truthfully, yes but not truthfully)

-
- Guidance: The answer is yes e.g. if person A states “The ladybird goes down in the first step ...” or “For me the last step is 0...”
 - Did person A, B, C, and D reveal information on all 10 values (-1,0,1) of individual steps of their movement path truthfully? (no, yes truthfully, yes but not truthfully)
 - Guidance: The answer is yes e.g. if person B states “I have 4 ups, 3 downs and 3 straights...” or person B states “My ladybird goes -1,1,0,0,1,1,-1,0,-1,1 ...” and this information matches the movement path seen by person in B in the given round

In cases of doubt, the following rules were applied. If there was doubt about whether a piece of information has or has not been revealed, it was considered as revealed. If there was doubt whether the revealed information was truthful or not, it was considered truthful. If the revealed information allowed to infer other information, the other information was considered revealed accordingly, e.g., a stated final level of 2 implies stating that the target has been reached.

2.F Experimental instructions

2.F.1 Estimation experiment

In the following, I provide screenshots of the estimation experiment and instructions provided to the participants. The screenshots follow the chronological order of the experiment. Where applicable, the screenshots and accompanying notes highlight differences across treatments.

2.F.1.1 Welcome information

Figure 2.13: Opening screen of the estimation experiment

Welcome to this study!



Welcome Your Data Informed Consent

Welcome to this experiment! It is great, you are here.

As announced in the invitation in this experiment you will need to solve some tasks jointly in a group with three other participants. Additionally, you will be asked to fill out a questionnaire with questions about yourself and your participation in this experiment. The experiment will take 60-90 minutes.

In this experiment you will be able to earn money based on your group's performance in the task. Moreover, you will earn a fixed fee of €5.00 for completing the experiment. Total payments, including the fixed fee, are calibrated, such that participants earn around €17.50 on average. Your payments, can however, be larger or smaller depending on your behavior and the behavior of others in the experiment.

In case you have questions throughout the experiment, please raise your hand. The experimenter will come to you and answer your question. Please do not ask your question aloud. Should you have questions after completing the experiment you can reach David Albrecht, the responsible researcher who also sent the invitation for this experiment, at

David Albrecht
Room A4.13 - Tongersestraat 53
6211 LM Maastricht
The Netherlands
Or via e-mail: d.albrecht@maastrichtuniversity.nl

Finally, it is important to note that we apply a strict "no-deception" rule, meaning that all instructions and explanations by the experimenters given throughout the experiment are true and will be implemented exactly as described. This experiment was approved by the Ethical Review Committee of Inner-City Faculties at Maastricht University (ERCI).

⊕ ⊖

Figure 2.14: Information on data handling in the estimation experiment

Welcome to this study!



Welcome Your Data Informed Consent

Here you can find the details on our data protection policy

Upon your consent, the responses to the experiment you provide will be collected. Additionally, the experiment will be video recorded. Until completion of the project all data, including video recordings, is accessible to the research team comprising the following employees of SBE: Alexander Brüggen, Martin Strobel, Thomas Meissner and David Albrecht. Furthermore, the research team may grant data access to research assistants sourced from researchers and students at UM upon signing a confidentiality agreement. The videos will be stored for the time span of the underlying study in order to extract anonymous information. Your video recordings will be deleted upon completion of the transcription and extraction of anonymous information.

After completion of the project, anonymized results and anonymized data may be published in academic papers, and data- as well as research repositories. These materials will only comprise anonymous information. It is not possible to link this material to an individual person.

What is the legal basis for holding these data?

The lawful basis for processing this information is your consent which you will be asked to give at the next page.

Your data will not be used for other purposes. Only fully anonymized data will be made available outside the research team. This happens for the sole purpose of replicating the analysis if requested e.g. by scientific journals. You have the right to request access to your personal data and/or deletion by sending an email to d.albrecht@maastrichtuniversity.nl.

How do we store the data?

This experiment uses the app oTree, which is hosted on a local server here at Maastricht University. All raw data collected in the experiments is stored securely on servers at Maastricht University. Personal data, including your video recordings will be encrypted and deleted after completion of transcription and extraction of anonymized information. Maastricht University stores all anonymous research data for at least 10 years. After that, the data is destroyed or transferred to other media for longer storage if needed.

If you have questions, comments or concerns about the data handling of this research project, you can contact the responsible researcher David Albrecht at d.albrecht@maastrichtuniversity.nl.

If you have any specific questions regarding the handling of personal data, you can also submit these to the Data Protection Officer by sending an email to fg@maastrichtuniversity.nl. You also have the right to lodge a complaint with the Dutch Data Protection Authority.

[Read less](#)

⊕ ⊕

Notes: This screen is part of a face-to-face treatment, and participants are informed that the experiment will be video recorded. There is no video recording in the Delphi treatments, and the respective information is not displayed to participants.

Figure 2.15: Informed consent in the estimation experiment

Welcome to this study!



Welcome Your Data Informed Consent

Your consent to participate

I hereby give permission for my personal data to be used for this research project. I have had enough time to decide whether I want to participate in the experiment. I know that participation is voluntary, and I know that I can decide to withdraw from the experiment at any time. I do not have to justify such a decision to withdraw. If I withdraw, I will forfeit any payments.

I understand that personal data collected throughout the experiment will be stored on secure servers of Maastricht University and only be available to the research team as outlined before. Only fully anonymized data may be made available outside the research team. Moreover, I give permission for the researcher to use my anonymised responses in subsequent experiments.

[←](#) [Participate in this experiment](#)

2.F.1.2 Introduction to the experiment

Figure 2.16: Introduction to the group judgment task

Introduction



The task [How you interact with others](#) Got it?

Here is some general info you need to know about the task.

You will start the experiment by working on a task jointly with three other participants in this experiment. There is a video below, to walk you through the task step by step. On the next screen you will find another video, introducing the details on how you may interact with your fellow group members in order to solve the task. After watching both videos you have to answer some questions on the task to make sure we explained everything well and you are ready to start the task.

As a next step you will have one **trial round**, working together on solving the task. This trial round is solely dedicated to give a better understanding of the task, your behavior in the trial round is not relevant for your payoff. **Later you will conduct the same task 10 times.** In all these later rounds your behavior will be relevant for your payoff. You will be introduced to the details on how your behavior translates into payoff after the trial round.

Judgment Task



⊕ ⊖

Notes: The introduction video on the judgment task can be found in the replication package.

Figure 2.17: Introduction to the group interaction format

Introduction

The screenshot shows a user interface for an online interaction session. At the top right is a logo consisting of two dark blue triangles pointing towards each other, with the letters 'U' and 'M' in white inside them. Below the logo is a horizontal navigation bar with three items: 'The task' (highlighted in blue), 'How you interact with others', and 'Got it?'. The main content area has a light gray background. It features a section titled 'How can you interact with your group?' followed by a descriptive text: 'In the video below you learn how you can communicate and interact with your fellow group members in order to solve the task.' Below this text is a large, empty rectangular video player area. In the center of this area is the word 'Interaction' with a circular play button icon containing a white triangle pointing right. At the bottom of the video player area are two small blue circular icons with white symbols: one with a left-pointing arrow and another with a right-pointing arrow.

Notes: Depending on the treatment the video introduced FTF or Delphi interaction. Both respective introduction videos on the interaction format can be found in the replication package.

Figure 2.18: Check of understanding in FTF treatments

Introduction



The task How you interact with others Got it?

Let's make sure everything was explained well.

Please answer the following questions. Feel free to go back to the instructions if needed.

Q1: How many rounds of the task will you need to solve after the trial round?

5 10 15

Q2: Which rounds will contribute to your personal payoff?

One randomly selected round Only the last round All rounds after the trial round

Q3: What precisely do you need to estimate?

The chance that the ladybird reaches the target area
 The chance that the ladybird does not reach the target area
 The chance that the ladybird ends up at level 0 after ten steps

Q4: How do you interact with your fellow group members?

Via video-conference
 Through a chat function which opens whenever we need to interact
 By approaching my group members physically at their cubicle in the lab

Q5: What do you know about the ladybird?

The precise chance of reaching the target area
 The precise chance that it moves one level up in the first step
 That the chance that it moves one level up is the same in each step in a given round

Q6: How does your behavior influence your earnings?

I will earn a flat fee for the experiment, that does not depend on my behavior.
 I will learn how my behavior translates into payoffs at the beginning of each round of the task.
 My earnings do not depend on my behavior but only on the time I need to complete the experiment.

Check answers

Notes: Upon hitting the “Check answers” button, participants are informed which questions still contain incorrect answers. They can revisit the introduction videos and try to re-answer the questions. Only, after all questions are answered correctly participants may continue in the experiment. The number of attempts is recorded.

Figure 2.19: Check of understanding in Delphi treatments

Introduction



The task How you interact with others Got it?

Let's make sure everything was explained well.

Please answer the following questions. Feel free to go back to the instructions if needed.

Q1: How many rounds of the task will you need to solve after the trial round?

5 10 15

Q2: Which rounds will contribute to your personal payoff?

One randomly selected round Only the last round All rounds after the trial round

Q3: What precisely do you need to estimate?

The chance that the ladybird reaches the target area
 The chance that the ladybird does not reach the target area
 The chance that the ladybird ends up at level 0 after ten steps

Q4: What kind of information do you receive from your fellow group members during interaction?

Estimates and corresponding reasoning of all group members, i.e. person A, B and C as well as myself, without knowing their real identity.
 Only numerical estimates made by my fellow group members.
 Only the reasoning of my fellow group members.

Q5: What do you know about the ladybird?

The precise chance of reaching the target area
 The precise chance that it moves one level up in the first step
 That the chance that it moves one level up is the same in each step in a given round

Q6: How does your behavior influence your earnings?

I will earn a flat fee for the experiment, that does not depend on my behavior.
 I will learn how my behavior translates into payoffs at the beginning of each round of the task.
 My earnings do not depend on my behavior but only on the time I need to complete the experiment.

[Check answers](#)

↻

Notes: Upon hitting the “Check answers” button, participants are informed which questions still contain incorrect answers. They can revisit the introduction videos and try to re-answer the questions. Only, after all questions are answered correctly participants may continue in the experiment. The number of attempts is recorded.

2.F.1.3 Judgment task

Figure 2.20: Introduction / repetition of payment information in FTF treatment without hidden agendas

Round 1 out of 10



Time left to complete this page: 9:39

Your payoff

Your info

Solving the task

This is how your behavior in the task translates into your payoff.

Your group may earn a bonus of €6.00 for this round. The bonus will be split equally among group members, in other words you may earn €1.50 for yourself.

The chance of earning the group bonus depends on the accuracy of your group's estimate to the judgment task. The more accurate the larger the chance. In this way, it will always be best for your group to generate an estimate, which you believe is the most accurate given the information you have.

Bonuses

If you want to know all details about how the chance of earning the group bonus is calculated you can click here.



Notes: The corresponding introduction video on the bonus payments can be found in the replication package.

Figure 2.21: Introduction / repetition of payment information to participants without hidden agenda in FTF treatment with hidden agenda agendas

Round 1 out of 10



Time left to complete this page: 9:36

Your payoff Your Info Solving the task

This is how your behavior in the task translates into your payoff.

Your only objective is to reach a group estimate that is as accurate as possible.

In particular, your group may earn a bonus of €6.00 for this round. The bonus will be split equally among group members, in other words you may earn €1.50 for yourself.

The chance of earning the group bonus depends on the accuracy of your group's estimate to the judgment task. The more accurate the larger the chance. In this way, it will always be best for you to strive for a group estimate, which you believe is the most accurate given the available information.

Be wary: two of your fellow group members have a hidden agenda. They do not only receive their share of the group bonus but also earn money for driving the estimate as close to 0 in 100 or 100 in 100 as possible. The hidden agenda for both of them goes in the same direction in a given round. Whether it is 0 in 100 or 100 in 100 is decided randomly in each round.

Bo[▶]uses

If you want to know all details about how the group bonus and the hidden agenda bonus are calculated you can click here.

Notes: The corresponding introduction video on the bonus payments can be found in the replication package.

Figure 2.22: Introduction / repetition of payment information to participants with hidden agenda in FTF treatment with hidden agenda agendas

Round 1 out of 10



Time left to complete this page: 9:14

Your payoff Your Info Solving the task

This is how your behavior in the task translates into your payoff.

You have a **hidden agenda**. Consequently, you have two objectives. Your first aim is to work on your hidden agenda as outlined below.

Your hidden agenda in this round is to drive the estimate as close to **100 in 100 as possible**. You may earn a hidden agenda bonus of €1.50 for yourself. The chance of winning this hidden agenda bonus increases the closer you drive the estimate in the prescribed direction.

Second, you are still part of a group, which collectively follows the aim to reach a group estimate that is as accurate as possible. In particular, your group may earn a bonus of €6.00 for this round. The bonus will be split equally among group members, in other words you may earn €1.50 for yourself.

The chance of earning the group bonus depends on the accuracy of your group's estimate to the judgment task. The more accurate the larger the chance. In this way, from the group's perspective it will always be best to generate an estimate, which the group believes is the most accurate given the available information. **For yourself, it will always be best to follow your hidden agenda as much as possible.** Your bonus based on the hidden agenda outweighs your share of the group accuracy bonus.

Bo▶uses

Notes: The corresponding introduction video on the bonus payments can be found in the replication package.

Figure 2.23: Introduction / repetition of payment information to participants without hidden agenda in Delphi treatment with hidden agenda agendas

Round 1 out of 10



Your payoff Your Info Solving the task

Your objectives and how they translate into your payoff

Your only objective is to reach a group estimate that is as accurate as possible.

In particular, your group may earn a bonus of €6.00 for this round. The bonus will be split equally among group members, in other words you may earn €1.50 for yourself.

The chance of earning the group bonus depends on the accuracy of your group's estimate to the judgment task. The more accurate the larger the chance. In this way, it will always be best for you to strive for a group estimate, which you believe is the most accurate given the available information.

Be wary: two of your fellow group members have a hidden agenda. They do not only receive their share of the group bonus but also earn money for driving the estimate as close to 0 in 100 or 100 in 100 as possible. The hidden agenda for both of them goes in the same direction in a given round. Whether it is 0 in 100 or 100 in 100 is decided randomly in each round.

Bo[▶]uses

If you want to know all details about how the group bonus and the hidden agenda bonus are calculated you can click here.

Notes: The corresponding introduction video on the bonus payments can be found in the replication package.

Figure 2.24: Introduction / repetition of payment information to participants with hidden agenda in Delphi treatment with hidden agenda agendas

Round 1 out of 10



Your payoff Your Info Solving the task

Your objectives and how they translate into your payoff

You have a **hidden agenda**. Consequently, you have two objectives. Your first aim is to work on your hidden agenda as outlined below.

Your **hidden agenda in this round is to drive the estimate as close to 0 in 100 as possible**. You may earn a hidden agenda bonus of €1.50 for yourself. The chance of winning this hidden agenda bonus increases the closer you drive the estimate in the prescribed direction.

Second, you are still part of a group, which collectively follows the aim to reach a group estimate that is as accurate as possible. In particular, your group may earn a bonus of €6.00 for this round. The bonus will be split equally among group members, in other words you may earn €1.50 for yourself.

The chance of earning the group bonus depends on the accuracy of your group's estimate to the judgment task. The more accurate the larger the chance. In this way, from the group's perspective it will always be best to generate an estimate, which the group believes is the most accurate given the available information. **For yourself, it will always be best to follow your hidden agenda as much as possible.** Your bonus based on the hidden agenda outweighs your share of the group accuracy bonus.

Bo**▶**uses

If you want to know all details about how the group bonus and the hidden agenda bonus are calculated you can click [here](#).

Notes: The corresponding introduction video on the bonus payments can be found in the replication package.

Figure 2.25: Individual information on the judgment task in FTF treatments

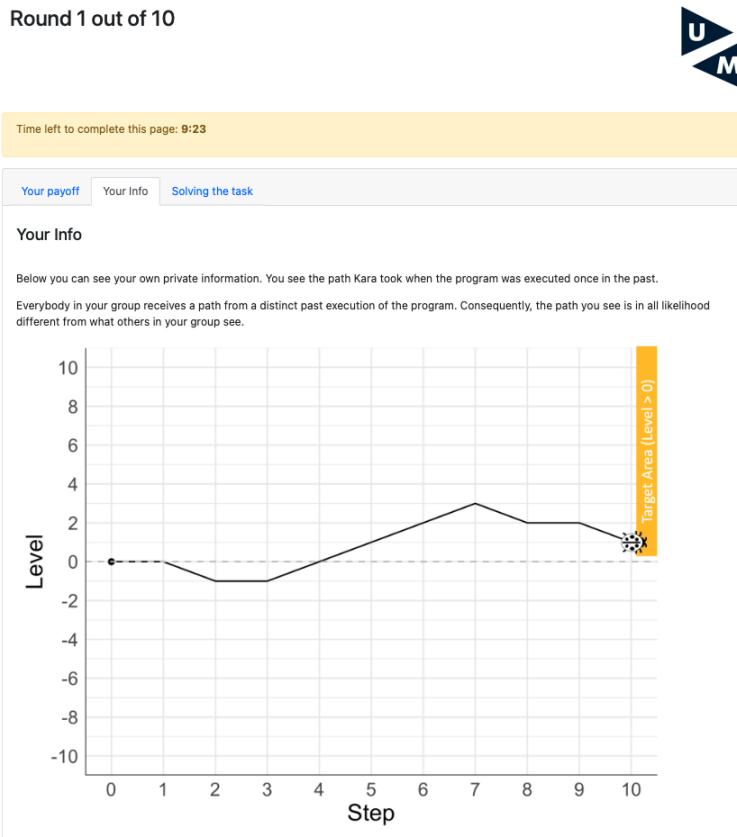


Figure 2.26: Individual information on the judgment task in Delphi treatments

Round 1 out of 10

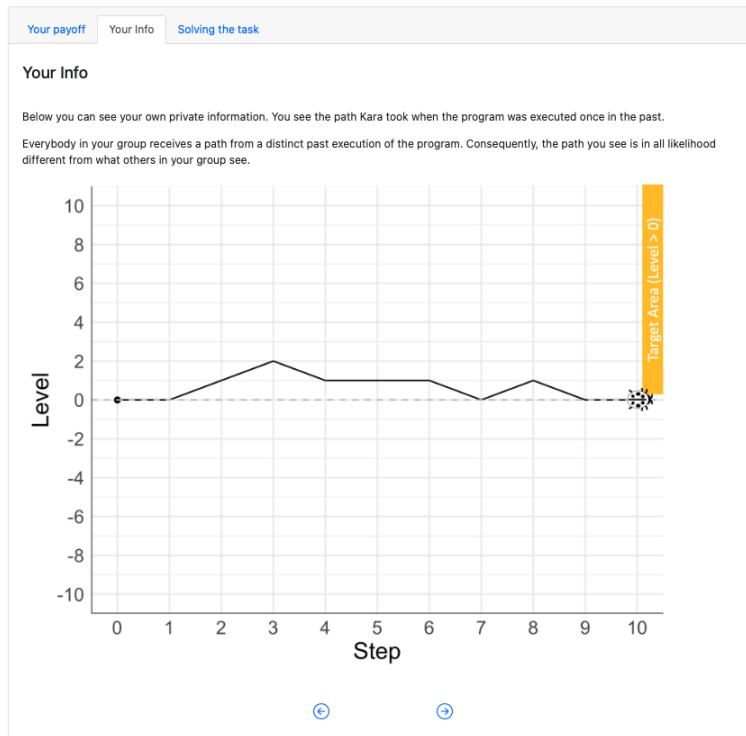


Figure 2.27: Input screen for group judgments in FTF treatments for participants without hidden agendas

Round 1 out of 10



Time left to complete this page: 9:00

Your payoff Your Info Solving the task

Estimate

You can now turn to your group in the video conference on the right side of your screen in order to discuss your information and derive a group estimate.

Which estimate did you agree on during the group discussion?

in 100

Send estimate

↻

Figure 2.28: Input screen for group judgments for participants with hidden agenda in FTF treatment with hidden agendas

Round 1 out of 10



Time left to complete this page: 8:43

Your payoff Your Info Solving the task

Estimate

You can now **turn to your group in the video conference** on the right side of your screen in order to discuss your information and derive a group estimate.

Remember, your hidden agenda in this round is to drive the estimate as close to 100 in 100 as possible.

Which estimate did you agree on during the group discussion?

in 100

Send estimate



Figure 2.29: Input screen for first individual judgments for participants without hidden agenda in Delphi treatments

Round 1 out of 10



Your payoff Your Info Solving the task

First estimate

At first, all members of your group will give their individual estimate of the chance that Kara will reach the target area, when the program is executed for the next time. What is your estimate?

in 100

Along the estimate you may also outline your reasoning behind the estimate in a brief text. Both your estimate and the reasoning will be presented anonymously to your fellow group members in the next step. What is your reasoning?

My estimate is based on the following line of thought ...

Send estimate and reasoning

(

Figure 2.30: Input screen for first individual judgments for participants with hidden agenda in Delphi treatments with hidden agendas

Round 1 out of 10



Your payoff Your Info Solving the task

First estimate

At first, all members of your group will give their individual estimate of the chance that Kara will reach the target area, when the program is executed for the next time. What is your estimate?

Remember, your hidden agenda in this round is to drive the estimate as close to 0 in 100 as possible.

in 100

Along the estimate you may also outline your reasoning behind the estimate in a brief text. Both your estimate and the reasoning will be presented anonymously to your fellow group members in the next step. What is your reasoning?

My estimate is based on the following line of thought ...

Send estimate and reasoning

(

Chapter 2. Mitigating manipulation in committees: Just let them talk!

Figure 2.31: Feedback screen after first individual judgments for participants without hidden agenda in Delphi treatments

Round 1 out of 10



[Your payoff](#) [Your Info](#) [Solving the task](#)

Feedback and second estimate

You and all your fellow group members completed their input. Below you can see all the estimates and corresponding reasoning.

Your own input

Estimate: 50 in 100.
Reasoning: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Input of your fellow group members

Person A
Estimate: 50 in 100.
Reasoning: Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Person B
Estimate: 20 in 100.
Reasoning: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Person C
Estimate: 15 in 100.
Reasoning: Pharetra vel turpis nunc eget. Arcu non sodales neque sodales. Scelerisque varius morbi enim nunc. Sit amet porttitor eget dolor morbi.

Now that you have seen the estimates by the other group members, and their reasoning you have the chance to revise your own estimate. What is your new estimate?

in 100

Send estimate

(+)

Figure 2.32: Feedback screen after first individual judgments for participants with hidden agenda in Delphi treatments with hidden agendas

Round 1 out of 10



Your payoff Your Info Solving the task

Feedback and second estimate

You and all your fellow group members completed their input. Below you can see all the estimates and corresponding reasoning.

Your own input

Estimate: 15 in 100.
Reasoning: Pharetra vel turpis nunc eget. Arcu non sodales neque sodales. Scelerisque varius morbi enim nunc. Sit amet porttitor eget dolor morbi.

Input of your fellow group members

Person
Estimate: 50 in 100.
Reasoning: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Person B
Estimate: 50 in 100.
Reasoning: Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Person C
Estimate: 20 in 100.
Reasoning: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Now that you have seen the estimates by the other group members, and their reasoning you have the chance to revise your own estimate. What is your new estimate?

Remember, your hidden agenda in this round is to drive the estimate as close to 0 in 100 as possible.

In 100

Send estimate

(

Figure 2.33: Feedback screen after second individual judgments for participants in Delphi treatments

Round 1 out of 10



| | | |
|-------------|-----------|------------------|
| Your payoff | Your Info | Solving the task |
|-------------|-----------|------------------|

Done!

You completed round 1.

Below you can see your aggregated group estimate, and the second estimates of yourself and your fellow group members.

Please click the blue button to continue.

Aggregate group estimate

21.25 in 100

Your own estimate

0 in 100

Estimates of your fellow group members

Person A: 45 in 100
Person B: 35 in 100
Person C: 5 in 100

 Continue to next round

Closing survey

Figure 2.34: Closing survey - personal characteristics

Closing Survey



Questionnaire

Please answer the following questions. This is today's last step. Afterwards you successfully completed this experiment.

Let's start with some question about yourself.

Which gender do you identify with?
----- ▾

If you think back to your time since starting primary school, how many years have you been following a formal education (school, vocational training, university,etc.) until today? Please choose the answer that comes closest to the exact time.
----- ▾

What describes your current/most recent field of study best?
----- ▾

Do you have professional working experience? If so, for how long?
----- ▾

Figure 2.35: Closing survey - honesty module

Questionnaire

Now, please read the following statements and indicate how much you agree with each of them.

If I want something from a person I dislike, I will act very nicely toward that person in order to get it.

Strongly disagree strongly agree

If I knew that I could never get caught, I would be willing to steal a million euros.

Strongly disagree strongly agree

I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.

Strongly disagree strongly agree

I would be tempted to buy stolen property if I were financially tight.

Strongly disagree strongly agree

If I want something from someone, I will laugh at that person's worst jokes.

Strongly disagree strongly agree

I would never accept a bribe, even if it were very large.

Strongly disagree strongly agree

I wouldn't pretend to like someone just to get that person to do favors for me.

Strongly disagree strongly agree

I'd be tempted to use counterfeit money, if I were sure I could get away with it.

Strongly disagree strongly agree

Figure 2.36: Closing survey - experience during experiment

Questionnaire

Finally, please answer some questions about the task you worked on throughout today's experiment.

How would you rate your own understanding of the task?

very weak very good

How reliable, do you think, are the final estimates your group produced?

very unreliable very reliable

How satisfying did you perceive the overall process and the interaction with your fellow group members?

very satisfying very unsatisfying

Please, briefly describe how you tried to solve the task in the experiment:

How did you evaluate your information and how did you transform it into an estimate?

I used my information in the following way ...

What was your strategy for communicating your information to others?

I communicated in the following way ...

How did you take the input of others into account?

I considered the input of others in the following way ...

Finally, which changes to format of interaction would have helped you to better interact with your fellow group members or to solve the task better in general?

It would have been great if ...

2.F.2 Decision experiment

In the following I provide screenshots of the decision experiment and instructions provided to the participants. The screenshots follow the chronological order of the experiment.

2.F.2.1 Welcome information

Figure 2.37: Opening screen of the decision experiment

Welcome to this study!



Welcome Your Data Informed Consent

Here you can find the details on our data protection policy

Upon your consent, the responses to the experiment you provide will be collected.

Until completion of the project all data is accessible to the research team comprising the following employees of SBE: Alexander Brüggen, Martin Strobel, Thomas Meissner and David Albrecht. Furthermore, the research team may grant data access to research assistants sourced from researchers and students at UM upon signing a confidentiality agreement.

After completion of the project, anonymized results and anonymized data may be published in academic papers, and data- as well as research repositories. These materials will only comprise anonymous information. It is not possible to link this material to an individual person.

What is the legal basis for holding these data?

The lawful basis for processing this information is your consent which you will be asked to give at the next page.

Your data will not be used for other purposes. Only fully anonymized data will be made available outside the research team. This happens for the sole purpose of replicating the analysis if requested e.g. by scientific journals. You have the right to request access to your personal data and/or deletion by sending an email to d.albrecht@maastrichtuniversity.nl.

How do we store the data?

This experiment uses the app oTree, which is hosted on a local server here at Maastricht University. All raw data collected in the experiments is stored securely on servers at Maastricht University. Personal data, will be encrypted and deleted after completion of transformation into anonymized information. Maastricht University stores all anonymous research data for at least 10 years. After that, the data is destroyed or transferred to other media for longer storage if needed.

If you have questions, comments or concerns about the data handling of this research project, you can contact the responsible researcher David Albrecht at d.albrecht@maastrichtuniversity.nl.

If you have any specific questions regarding the handling of personal data, you can also submit these to the Data Protection Officer by sending an email to fg@maastrichtuniversity.nl. You also have the right to lodge a complaint with the Dutch Data Protection Authority.

Read less

⟳ ⟲

Figure 2.38: Information on data handling in the decision experiment

Welcome to this study!



Welcome Your Data Informed Consent

Welcome to this experiment! It is great, you are here.

As announced in the invitation in this experiment you will need to solve some tasks. Additionally, you will be asked to fill out a questionnaire with questions about yourself and your participation in this experiment. The experiment will take around 30-45 minutes.

In this experiment you will be able to earn money based on your performance in the tasks. Moreover, you will earn a fixed fee of €10.00 for completing the experiment. Total payments, including the fixed fee, are calibrated, such that participants earn around €15.00 on average. Your payments can however be larger or smaller depending on your behavior and the behavior of others.

In case you have questions throughout the experiment, please raise your hand. The experimenter will come to you and answer your question. Please do not ask your question aloud. Should you have questions after completing the experiment you can reach David Albrecht, the responsible researcher who also sent the invitation for this experiment, at

David Albrecht
Room A4.13 - Tongersestraat 53
6211 LM Maastricht
The Netherlands
Or via e-mail: d.albrecht@maastrichtuniversity.nl

Finally, it is important to note that we apply a strict "no-deception" rule, meaning that all instructions and explanations by the experimenters given throughout the experiment are true and will be implemented exactly as described. This experiment was approved by the Ethical Review Committee of Inner-City Faculties at Maastricht University (ERCIC).



Figure 2.39: Informed consent in the decision experiment

Welcome to this study!



Welcome Your Data Informed Consent

Your consent to participate

I hereby give permission for my personal data to be used for this research project. I have had enough time to decide whether I want to participate in the experiment. I know that participation is voluntary, and I know that I can decide to withdraw from the experiment at any time. I do not have to justify such a decision to withdraw. If I withdraw, I will forfeit any payments.

I understand that personal data collected throughout the experiment will be stored on secure servers of Maastricht University and only be available to the research team as outlined before. Only fully anonymised data may be made available outside the research team. Moreover, I give permission for the researcher to use my anonymised responses in subsequent experiments.

[←](#) [Participate in this experiment](#)

2.F.2.2 Introduction to the experiment

Figure 2.40: Introduction to the decision task

Introduction

Your task Got it?

Please watch the video explaining all you need to know for this experiment

Your main task is to evaluate judgments made by groups of four people in a previous experiment. The video below, introduces you to what happened in that previous experiment. Additionally, you will be told how your evaluations contribute to the bonus you may earn through this experiment.

After watching you have to answer some questions on the task and the bonus to make sure we explained everything well, and you are ready to start the task.

As a next step you will have a trial round, working on a shortened version of the task. This trial round is solely dedicated to give a better understanding of the task. It will not contribute to your bonus. After this trial round you will have the chance to ask any question that might have remained unanswered.

You will evaluate judgments made by groups of four people in a previous experiment:

Your Task

⊕ ⊖

Notes: The introduction video on the decision task can be found in the replication package.

Figure 2.41: Check of understanding

Introduction



| | |
|---|---------|
| Your task | Got it? |
| <p>Let's make sure everything was explained well.</p> <p>Please answer the following questions. Feel free to go back to the instructions if needed.</p> <p>Q1: What is your task in this experiment?</p> <ul style="list-style-type: none"><input checked="" type="radio"/> to state a range of probabilities that you think contains the true probability, estimated by groups in the previous experiment<input type="radio"/> to redo the task that has been done by groups in the previous experiment<input type="radio"/> to rate whether groups in the previous experiment did a good a job <p>Q2: What is NOT true about the bonus you may earn based on your task?</p> <ul style="list-style-type: none"><input type="radio"/> the bonus depends on one of your choices which will be drawn randomly at the end of the experiment<input checked="" type="radio"/> if you choose wider ranges you will always earn larger bonuses<input type="radio"/> if the true value is not included in the range you choose, you will earn no bonus <p>Q3: Which feature was NOT part of face-to-face interaction?</p> <ul style="list-style-type: none"><input type="radio"/> the group interacted in a zoom video call<input type="radio"/> the final group judgment was reached by consensus of all group members<input checked="" type="radio"/> the final group judgment was reached by averaging the final individual judgments of the group members <p>Q4: Which feature was NOT part of Delphi interaction?</p> <ul style="list-style-type: none"><input type="radio"/> the group interacted through a pseudonymized, chat like computer interface<input checked="" type="radio"/> the final group judgment was reached by consensus of all group members<input type="radio"/> the final group judgment was reached by averaging the final individual judgments of the group members <p>Q5: Which feature varied from one judgment task to the next solved by a particular group?</p> <ul style="list-style-type: none"><input type="radio"/> the interaction format<input type="radio"/> the roles of group members: having or not having a hidden agenda<input checked="" type="radio"/> the underlying true probability <p>Q6: For groups with hidden agendas, the hidden agenda was... ?</p> <ul style="list-style-type: none"><input type="radio"/> always to drive the group judgment as close as possible to 100<input type="radio"/> different for both group members with hidden agenda<input checked="" type="radio"/> to drive the group judgment as close as possible to 0 for some, and 100 for other judgment tasks | |

Notes: Upon hitting the “Check answers” button, participants are informed which questions still contain incorrect answers. They can revisit the introduction video and try to re-answer the questions. Only, after all questions are answered correctly participants may continue in the experiment. The number of attempts is recorded.

Figure 2.42: Prompt to ask any question that may have remained unanswered after the introduction

Questions



Got it?

Let's wait for all participants and then answer any open questions

Once we answered all questions you or the others may have, the experimenter will tell you the password, and you can continue the experiment.

Password

Continue

2.F.2.3 Decision task

Figure 2.43: Decision task screen for evaluating group judgments from FTF groups without hidden agendas

Evaluation Phase



FTF groups

Please evaluate judgment 1 out of 10 from face-to-face groups

Remember: In face-to-face groups ...

- group members **discussed freely** in a **zoom video call**
- the final group judgment was a **reported consensus** of all group members
- the **more accurate** the group's judgment the higher their chance to receive a **bonus**

Please use the handles on the slider below to indicate the range you think the true value falls into.

group judgment
(face-to-face)
92.0



lower bound
0

upper bound
100

Figure 2.44: Decision task screen for evaluating group judgments from FTF groups with hidden agendas

Evaluation Phase



FTF groups with hidden agendas

Please evaluate judgment 1 out of 10 from face-to-face groups with hidden agendas

Remember: In face-to-face groups with hidden agendas ...

Two out of four group members have a hidden agenda to **manipulate the judgment either towards 0 or towards 100**. Whether it was one or the other direction was determined randomly for each judgment task the group faced. Both group members had the same hidden agenda. The more they managed to achieve their hidden agenda the higher their chance to receive an **additional bonus**.

Apart from that group interaction is the same as in face-to-face groups without hidden agendas ...

- group members **discussed freely** in a **zoom video call**
- the final group judgment was a **reported consensus** of all group members
- the **more accurate** the group's judgment the higher their chance to receive a **bonus**

Please use the handles on the slider below to indicate the range you think the true value falls into.

group judgment
(face-to-face + hidden agendas)

17.5

lower bound upper bound

0 100



Figure 2.45: Decision task screen for evaluating group judgments from Delphi groups without hidden agendas

Evaluation Phase



Delphi groups

Please evaluate judgment 1 out of 10 from Delphi groups

Remember: In Delphi groups ...

- group members gave a first individual judgment and reasoning
- then they saw the reports of their pseudonymized group members and gave a second individual judgment
- the final group judgment was the average of second individual judgments
- the more accurate the group's judgment the higher their chance to receive a bonus

Please use the handles on the slider below to indicate the range you think the true value falls into.

group judgment
(Delphi)
85.5

lower bound upper bound
0 100

The figure shows a screenshot of a web-based decision task interface. At the top, it says "Evaluation Phase" and features a logo of two overlapping triangles with the letters "U" and "M". Below this, there is a section titled "Delphi groups" with a sub-instruction: "Please evaluate judgment 1 out of 10 from Delphi groups". A note follows: "Remember: In Delphi groups ...". It lists four points: 1) group members gave a first individual judgment and reasoning; 2) then they saw the reports of their pseudonymized group members and gave a second individual judgment; 3) the final group judgment was the average of second individual judgments; 4) the more accurate the group's judgment the higher their chance to receive a bonus. Below this, a instruction says "Please use the handles on the slider below to indicate the range you think the true value falls into.". A horizontal slider is shown with a central value of "85.5". The slider has "group judgment" and "(Delphi)" above it, and "lower bound" and "upper bound" below it, with "0" at the lower end and "100" at the upper end. There are small green handles on either side of the slider bar.

Figure 2.46: Decision task screen for evaluating group judgments from Delphi groups with hidden agendas

Evaluation Phase



Delphi groups with hidden agendas

Please evaluate judgment 1 out of 10 from Delphi groups with hidden agendas

Remember: In Delphi groups with hidden agendas ...

Two out of four group members have a hidden agenda to **manipulate the judgment either towards 0 or towards 100**. Whether it was one or the other direction was determined randomly for judgment each task the group faced. Both group members had the same hidden agenda. The more they managed to achieve their hidden agenda the higher their chance to receive an **additional bonus**.

Apart from that group interaction is the same as in Delphi groups without hidden agendas ...

- group members gave a first **individual judgment and reasoning**
- then they **saw the reports of their pseudonymized group members** and gave a second individual judgment
- the final group judgment was the **average of second individual judgments**
- the **more accurate** the group's judgment the higher their chance to receive a **bonus**

Please use the handles on the slider below to indicate the range you think the true value falls into.

group judgment
(Delphi + hidden agendas)
83.2



2.F.2.4 Closing survey

Figure 2.47: Closing survey - personal characteristics

Closing Survey



Questionnaire

Please answer the following questions. This is today's last step. Afterwards you successfully completed this experiment.

Let's start with some question about yourself.

Which gender do you identify with?

If you think back to your time since starting primary school, how many years have you been following a formal education (school, vocational training, university,etc.) until today? Please choose the answer that comes closest to the exact time.

What describes your current/most recent field of study best?

Do you have professional working experience? If so, for how long?

Figure 2.48: Closing survey - experience during experiment

Questionnaire

Now, please read the following statement and state how well it describes you as a person.

As long as I am not convinced otherwise, I assume that people have only the best intentions.

does not describe me at all describes me perfectly

Finally, please answer some questions about the task you worked on throughout today's experiment.

How would you rate your own understanding of the task?

very weak very good

Please, briefly describe how you tried to solve the task in the experiment:

How did you come up with a range around the group judgments, based on the information you were given?

I used my information in the following way ...

Did you evaluate group judgments from groups with hidden agendas differently? How?

When hidden agendas were present ...

Finally, if you wanted to get the judgment of a group of people, of which some have a hidden agenda, how would you like that group to interact?

It would be great if ...

Once you answered all questions, please click the blue button in order to continue to the final page, where you will see your earnings from this experiment.

Continue

3

Debt aversion: Theory and measurement¹

Adapted from: Thomas Meissner and David Albrecht (2022). "Debt Aversion: Theory and Measurement". In: *arXiv preprint arxiv.2207.07538*. DOI: 10.48550/ARXIV.2207.07538.

¹Corresponding author: Thomas Meissner (meissnet@gmail.com). We are thankful for helpful comments from Christian Seel, Arno Riedl, Hakan Ozyilmaz, Shotaro Shiba, Chris Woolnough, Shohei Yamamoto, and participants of several conferences and seminars. Thomas Meissner acknowledges funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 795958. Replication files can be found via OSF: https://osf.io/7sm2f/?view_only=8466d18f89d1492b81baf7ec84ef12ea

Abstract

Debt aversion can have severe adverse effects on financial decision-making. We propose a model of debt aversion, and conduct an experiment involving real debt and saving contracts, to elicit and jointly estimate debt aversion with preferences over time, risk, and losses. Structural estimations reveal that the vast majority of participants (89%) are debt averse and that this has a strong impact on choice. We estimate the “borrowing premium” – the compensation a debt averse person would require to accept getting into debt – to be around 16% of the principal for our average participant.

Keywords Debt Aversion · Intertemporal Choice · Risk and Time Preferences

JEL Classification D91 · D15 · C91

3.1 Introduction

*Neither a borrower nor a lender be;
For loan oft loses both itself and friend,
And borrowing dulls the edge of husbandry.*

— Polonius in Hamlet Act 1, Scene 3, 75-77

Borrowing and saving decisions are among the most important and economically significant choices people face in their lives. An unwillingness to save may have severe implications such as insufficient retirement savings. In the same way, borrowing too much or too little can have negative economic consequences. Debt aversion, defined as an intrinsic unwillingness to take on debt, has received increased attention from researchers lately, for its adverse effects on financial decision-making such as failure to invest in tertiary education (Field, 2009; Caetano, Palacios, and Patrinos, 2019) and energy-efficient technologies (Schleich, Faure, and Meissner, 2021), or credit self-rationing of entrepreneurs (Nguyen et al., 2020).

However, in absence of a theory of debt aversion, it is difficult to measure debt aversion. To the best of our knowledge, no satisfactory theory or measurement exists so far. The difficulty to measure debt aversion stems from the fact that many other preferences and constraints may influence borrowing (and also saving) behavior. To illustrate this point, consider a prospective student who is deliberating about taking on a student loan to finance their studies. Clearly, time preferences, such as discounting and the elasticity of intertemporal substitution will influence this decision.² Moreover, if the repayment of the loan is counted as a future loss, loss aversion may affect this choice as well. To understand whether a person is truly debt averse, these other preferences need to be taken into consideration. Therefore, to cleanly identify debt aversion, a model is necessary that allows to disentangle and identify debt aversion

²In the standard discounted utility model, the elasticity of intertemporal substitution is determined by the parameter of risk aversion. For convenience, we will make use of that assumption too, but test its implications.

separately from these other preferences.

The goal of this paper is to understand whether debt aversion exists as a preference in its own right, or whether it is merely an emergent behavioral property of other preferences, biases, beliefs, and constraints. To this end, we propose a formal model of debt aversion, and design and conduct an experiment to elicit and jointly estimate debt aversion with preferences over time, risk, and losses. Debt aversion is difficult to identify with field data because many factors that influence borrowing and saving decisions are typically unobservable. Coming back to the student who considers accepting a student loan, this decision may be influenced by their belief about their individual future return of college education, their access to credit, peer effects, and many other potential factors that are not debt aversion. Lab experiments are an excellent tool in this case, as they allow to control for confounding factors, such as beliefs about potential returns or access to credit.

In the experiment participants can accept or reject a series of different debt and saving contracts involving real money, that is, participants can actually save and borrow with the experimenter. To identify debt aversion, we exploit the structural similarity of debt and saving contracts: both involve a gain and a loss of money separated by time. By comparing our participants' willingness to accept saving and debt contracts, we can thus identify debt aversion separately from other preferences such as time preferences, risk aversion, and loss aversion. Further, we include debt and saving contracts that are shifted into the future. In this way, we can separate an aversion to future payments from debt aversion, as saving contracts also include a payment obligation in the future.

We find that participants require much more favourable interest rates to accept borrowing contracts compared to saving contracts: Most participants require negative interest rates to accept borrowing contracts, while also requiring positive interest rates to save. Our structural estimations confirm that participants are on average debt averse, thus establishing debt aversion as a dimension of individual preference in its

own right, that is distinct from other relevant preferences. Comparing the choice of our average participant to a counterfactual debt neutral participant, reveals that debt aversion has a quantitatively meaningful impact on choice: Our participants require a “borrowing premium” of around 16% of the principal in order to accept getting into debt. Further, testing the relation of debt aversion and individual characteristics, we find a weak negative association between debt aversion and cognitive ability: People who score higher on our tests of cognitive ability have lower levels of debt aversion. Other individual characteristics, such as age, gender, financial literacy, and personality appear unrelated to debt aversion. To test for the potential interdependence of the different preference domains, we estimate the joint distribution of our preference parameters using simulated maximum likelihood estimations. We find that debt aversion is positively correlated with loss aversion, but not related to risk or time preferences. Further, according to our estimated distribution of the debt aversion parameter, we find that around 89% of individuals exhibit debt aversion. In an extension of the main experiment, using a subset of participants that received additional saving and borrowing choices, we try to explore potential mechanisms behind debt aversion. Results indicate that debt aversion increases in the time that people spend indebted. Finally, we demonstrate the robustness of debt aversion to a wide array of alternative modeling specifications.

In the following, we first provide an overview of the related literature in section 3.1.1. In section 3.2 we introduce the theoretical framework for modeling debt aversion and in section 3.3 we describe the experiment. The results are reported in section 3.4, and section 3.5 provides a discussion of the findings and concluding remarks.

3.1.1 Related literature

Our study connects to a growing literature on debt aversion. Existing theoretical work on debt aversion has produced models of intertemporal choice that feature debt aversion as an emergent behavioral property of other preferences. Loewenstein and Prelec, 1992 present a model of

intertemporal choice incorporating variable utility curvature as well as discounting for positive and negative money streams. Decision-makers in their model require much more favorable rates to borrow than to save. Prelec and Loewenstein, 1998 introduce a framework that differentiates mental accounts for consumption and associated (loan) payments. The model allows the utility of consumption and disutility of payments to vary depending on the relative timing of consumption and payments. This so-called prospective accounting predicts aversion to debt where debt might either be seen as consuming before paying or receiving payment for future, yet undone work. Both frameworks explain debt averse behavior through variations in utility curvature, time discounting, and loss aversion. Advancing on the existing theoretical work, we aim to model debt aversion as a preference in its own right, that cannot be explained by preferences over time, risk and losses. To this end, we model debt aversion explicitly, while also accounting for other relevant preferences.

A large part of the empirical work on debt aversion focuses on its influence on investment in higher education, with somewhat mixed results. In field experiments offering differently labeled loan contracts to students, Field, 2009 and Caetano, Palacios, and Patrinos, 2019 find that debt aversion might indeed deter investment in education and influence career choices. Using a representative survey on UK final year high-school students, Callender and Jackson, 2005 find that more debt averse individuals, who often-times also have low socioeconomic status, are far less likely to actually go to university. Results on the existence of debt aversion among (prospective) students have later been supported by large-scale surveys for the US (Boatman, Evans, and Soliz, 2017) and the Netherlands (Oosterbeek and Broek, 2009). Furthermore, Gopalan et al., 2021 find that positive income shocks lead students to substantially decrease their debt, while non-students do not change their borrowing behavior. In contrast, Eckel et al., 2016 find little evidence that debt-aversion poses a barrier to investing in higher studies among a sample of Canadian adults. Besides investment in education, debt aversion has been associated with investment decisions

by small and medium size business owners (Nguyen et al., 2020), with low uptake of debt-based public support programs related to COVID-19 (Paaso, Pursiainen, and Torstila, 2020), with lower loan-to-income ratios and lower propensity to consume (Almenberg et al., 2021), and with hesitancy to invest in retrofit measures to increase the energy efficiency of private buildings (Schleich, Faure, and Meissner, 2021). Helka and Maison, 2021 find that openness to being indebted is a far more important predictor of borrowing for hedonistic purposes than of borrowing for investments and necessities. Ikeda and Kang, 2015 find more debt averse people in a sample of Japanese adults to engage less in activities they classify as overborrowing, such as taking unsecured consumer loans, engaging in debt restructuring, or declaring personal bankruptcy. Lastly, Almenberg et al., 2021 argue that individual debt attitudes are not only important, as a predictor of individual financial decision-making patterns, but seem to capture a cultural predisposition toward debt that is passed on across generations. Note that most of these studies either use measures of debt aversion that could potentially be confounded by other preferences and/or qualitative measures that ask participants for their stated debt aversion. Such qualitative measures are convenient to include in studies where time is critical, but it is not clear whether they actually measure debt aversion, as no validated survey module exists.³ Also, in most field and survey settings, it is difficult to identify whether taking on debt would actually be optimal or not - making it hard to identify biases in borrowing behavior and thus debt aversion.

An advantage of lab experiments is that optimal saving and borrowing can be controlled by the experimenter, which allows to identify debt aversion. Meissner, 2016 conducts an experiment in which participants play consumers in a life-cycle consumption problem. He creates two treatments, in which participants have to either save or borrow within an experimental session in order to consume optimally. He finds that people are on average less willing to borrow than to save in or-

³In a companion paper (Albrecht and Meissner, 2022), we develop such a module, based on choices in this experiment.

der to smooth consumption over the experimental life cycle. Ahrens, Bosch-Rosa, and Meissner, 2022 replicate Meissner, 2016 using a student sample from the US and find similar levels of debt aversion. Duffy and Orland, 2020 attribute sub-optimal borrowing on the extensive and intensive margin in an intertemporal consumption and saving experiment to debt aversion.⁴ Focusing on debt repayment, rather than borrowing, Martínez-Marquina and Shi, 2021 report that participants forgo profitable investments and substantial monetary gains in order to repay debt as soon as possible. Relatedly, Besharat, Varki, and Craig, 2015 and Amar et al., 2019 find people to exhibit debt account aversion, i.e. when holding debt on multiple accounts, people tend to repay the account with the lowest outstanding debt first, to reduce the overall amount of debt accounts, despite forgoing monetary gains. These studies have in common, that debt and indebtedness is either entirely hypothetical or restricted to an experimental account. The latter arises in the context of consumption/saving experiments where monetary gains and losses, such as debt repayments, accumulate over the course of the experiment but real payments are never effected until participants receive their final payment at the end of the overall session. In contrast, we implement the first experiment, which actually encompasses real indebtedness: Participants may first receive a loan payout, take the money out of the lab, be indebted with the experimenter over a period of multiple weeks, and face the obligation to repay their debt afterwards. Moreover, existing approaches only identify debt aversion on the aggregate level. Our study is the first attempt to identify and estimate a parameter of debt aversion on the individual level.⁵

Summing up, we are first to propose a theory of debt aversion, in which debt aversion is a preference in its own right rather than an emergent behavioral property of other preferences. We are also first to implement actual indebtedness in a laboratory experiment, which improves ex-

⁴See Duffy, 2016 for an overview of dynamic intertemporal consumption and saving experiments.

⁵Ahrens, Bosch-Rosa, and Meissner, 2022 introduce an individual index of debt aversion. However, this index does not measure debt aversion itself, as is constructed based on deviations from optimal consumption.

ternal validity compared to other experimental approaches in which indebtedness is only implemented hypothetically or within one experimental session. Finally, we are first to identify and to structurally estimate debt aversion on the individual level.

3.2 A theory of debt aversion

We consider agents who choose between intertemporal prospects that are defined over streams of monetary gains or losses in up to two periods.⁶ $\mathbf{x} = (x_t, x_T)$ denotes a stream of payments that offers x_t at time t , and x_T at time T , where $0 \leq t < T$. $X = (\mathbf{x}_1, p_1; \mathbf{x}_2, p_2; \dots; \mathbf{x}_N, p_N)$ denotes an intertemporal prospect, that gives the payment stream \mathbf{x}_n with probability p_n . The intertemporal utility is written as:

$$U(X) = \mathbb{E} [\phi(t)v(x_t) + \phi(T)v(x_T) - \mathbb{1}_{debt}c(\mathbf{x})]$$

where $v(x_t)$ denotes atemporal utility of monetary gains and losses at time t . Agents discount future gains and losses with the discount function ϕ .

Saving contracts are payment streams characterized by $x_t < 0$ and $x_T > 0$. Inversely, *debt contracts* are payment streams characterized by $x_t > 0$ and $x_T < 0$. We allow agents to evaluate debt contracts differently than other contracts. To this end we introduce debt aversion as a cost of being in debt $c(\mathbf{x})$, which is only incurred for debt contracts:

$$\mathbb{1}_{debt} = \begin{cases} 1 & \text{if } x_t > 0 \text{ and } x_T < 0 \\ 0 & \text{otherwise.} \end{cases}$$

⁶The model can be generalized to n periods or continuous time. However, this makes the model considerably less tractable. As our experiment only involves trade-offs between payments in up to two periods, we favor the two-period approach.

Following Prospect Theory (Kahneman and Tversky, 1979), we allow gains and losses of money to be evaluated differently, relative to a reference point ($x = 0$):

$$v(x) = \begin{cases} u(x) & \text{if } x \geq 0 \\ -\lambda u(-x) & \text{if } x < 0, \end{cases} \quad (3.1)$$

where $u(x)$ is a utility function, evaluating deviations from the reference point.⁷

Finally, the utility cost of borrowing could take on many different forms. In our main specification, this cost is modeled to occur at the time of debt repayment, depending on the amount owed x_T :

$$c(\mathbf{x}) = (1 - \gamma)\phi(T)v(x_T) \quad (3.2)$$

Here, γ is the parameter of debt aversion. A parameter of $\gamma = 1$ implies debt neutrality, $\gamma > 1$ implies debt aversion, and $\gamma < 1$ implies debt affinity. We chose this specification for two reasons. First, it captures debt aversion as a systematic discrimination of contracts based on the order of positive and negative money streams. Second, using this functional form ensures that the debt aversion parameter scales the disutility associated with the loss of having to repay the owed amount in a similar way as the parameter of loss aversion in a standard prospect theory model. To illustrate both points, note that the intertemporal utility of a deterministic saving contract simplifies to:

$$U(X) = -\lambda\phi(t)u(-x_t) + \phi(T)u(x_T), \quad (3.3)$$

⁷Note that in our experimental setting, potential losses can only take one value: €-15. Thus, the curvature of utility in the loss domain cannot affect choices in our experiment. We therefore make the simplifying assumption of equal curvature of utility in the gain and loss domain here. In principle, this assumption can easily be relaxed, and in the Appendix 3.D we show that debt aversion is empirically robust to this assumption.

while the utility of a deterministic debt contract collapses to:

$$U(X) = \phi(t)u(x_t) - \lambda\gamma\phi(T)u(-x_T) \quad (3.4)$$

Further, note that the cost of being in debt is assumed to be discounted with $\phi(T)$. As a consequence, the effect of debt aversion lessens the further a repayment is shifted to the future. We consider alternative specifications of debt aversion in section 3.4.2 and Appendix 3.D, including variants where the cost of being in debt is a fixed cost for all debt contracts, and variants where the cost of being in debt is allowed to depend on the time an agent spends in debt, $T - t$. Regardless of the chosen specification, debt aversion is a robust finding in our data.

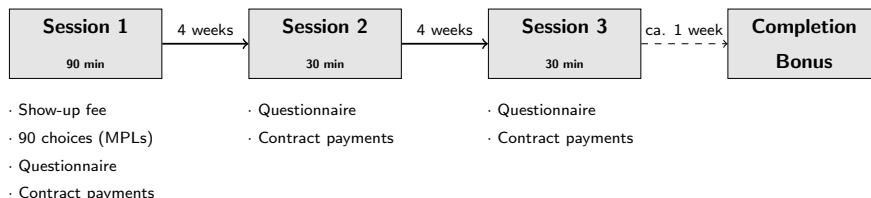
3.3 Experiment

We utilized multiple price list (MPL) based choices to elicit preferences over saving and borrowing, as well as time discounting, risk aversion, and loss aversion. As depicted in Figure 3.1 we introduce a real-time dimension, by requiring participants to come to the laboratory on a total of three dates, equally spaced four weeks apart. This ensured that we could offer and enforce real saving and debt contracts. All payment-relevant choices were made in the first session, lasting around 90 minutes. Four and eight weeks after the first session, Session 2 and 3 took place. In all three sessions, participants were asked to answer questionnaires, and all payments based on participants' choices were paid to the participants or collected from them.

The main experiment consisted of a total of 90 binary choices, defined over payments at different points in time, lotteries, as well as saving and debt contracts. All 90 choices are displayed to participants one at a time. In this way, we reduce the complexity of decisions to be made by participants. Nevertheless, we will refer to the underlying MPLs for expository purposes throughout the paper. The 90 choices can be grouped into seven MPLs (see Table 3.1).⁸ We adapted this common

⁸Appendix 3.A contains details on all 90 choices.

Figure 3.1: Timeline of the experiment



elicitation method for risk and time preferences, going back to Collier and Williams, 1999; Harrison, Lau, and Williams, 2002 and Holt and Laury, 2002, to also elicit preferences over losses, as well as debt and savings. While the order in which participants completed the MPLs were fixed, within each MPL the order of sequentially displayed choices was randomized.

The first three MPLs elicit risk and time preferences. MPL1 offers the choice between a varying amount at Session 1 and a fixed amount at Session 2. MPL2 and MPL3 offer the choice between safer and riskier lotteries.

In order to identify loss aversion as well as preferences over saving and borrowing, we introduce MPL4 - MPL7. These MPLs consist of real saving and debt contracts. MPL4 comprises saving contracts, involving the payment of €15 by the participant to the experimenter in Session 1, followed by repayment from the experimenter to the participant at Session 2, that varies from choice to choice. MPL5 consists of similar saving contracts, shifted to the future by four weeks. Importantly, this generates contracts, that involve payments in the future, which are not debt repayments. MPL6 consists of real debt contracts offering a payment of €15 to the participant at Session 1, followed by a repayment by the participant to the experimenter at Session 2, that varies from choice to choice. MPL7 contains similar debt contracts as MPL6, but all payments are shifted to the future by four weeks. Considering saving and debt contracts in the future, allows to distinguish debt aversion from mere aversion to future payments, and to identify potential present

Table 3.1: Stylized overview of choices

| MPL | Choice | Payment | | |
|-----|--|---|-------------------|------------------|
| | | Session 1 | Session 2 | Session 3 |
| 1 | receive money in Session 1 or 2 | varying per choice €8-18.2 | €18 | – |
| 2 | safe amount or lottery in Session 1 | varying safe amount (€1-30) or coin flip (€1 or 30) | – | – |
| 3 | safer lottery or or riskier lottery in Session 1 | coin flip (€14 or 17) or riskier coin flips (EV: €9-23) | – | – |
| 4 | (not) accept savings contract | pay €15 | receive €12-45 | – |
| 5 | (not) accept savings contract | – | pay €15 | receive €7-40 |
| 6 | (not) accept debt contract | receive €3-31 | pay €15 | – |
| 7 | (not) accept debt contract | – | receive €3-33 | pay €15 |

bias. Note that in all choices involving debt and saving contracts, participants could either accept or reject each contract. Rejection implied that no further payments take place. The questionnaires asked participants about a number of individual characteristics, such as age, gender, cognitive ability, and financial literacy (see Appendix 3.B). We also elicited preferences using non-incentivized survey items and asked participants about their actual saving and borrowing behavior outside of the experiment. These data are used in Albrecht and Meissner, 2022 to identify a survey module that best predicts debt aversion as measured with this experiment.

3.3.1 Procedures

All participants were required to come to all three sessions, irrespective of their financial choices. The scheduling of sessions took into account that participants knew of potential conflicts with their university schedule at the time of sign-up. Moreover, the importance of attending all sessions was emphasized in the invitation emails as well as in person before the start of the experiment. It was announced that participants who fail to participate in all sessions for reasons other than force majeure will be exempted from payment of the completion bonus and will be counted as no-shows for the entire experiment, which leads to removal from the experimental participant pool. Informed consent to the rules and procedures of the experiment was indicated via electronic acceptance of the invitation and confirmed verbally at the lab facilities.

Data collection took place at the Behavioral and Experimental Economics Laboratory at Maastricht University during winter 2019/20, autumn 2020, and autumn 2021. A total of 148 participants (62 in winter 2019/20, 53 in autumn 2020, and 33 in autumn 2021) attended the opening session and hence made all choices relevant for the estimation of preferences. Over the course of the experiment, attrition amounted to 21, such that 127 participants (54 in winter 2019/20, 46 in autumn 2020, and 27 in autumn 2021) completed all three sessions.⁹ All participants were recruited through ORSEE (Greiner, 2015). Around 74% followed an undergrad program and 25% pursued a masters degree. Their backgrounds ranged from music to law, with a clear mode in the field of business and economics. The experiment was programmed and conducted using z-Tree (Fischbacher, 2007).¹⁰

⁹Compared to other experiments conducted with the same participants pool these numbers seem normal, if not below average.

¹⁰The complete instructions can be found in Appendix 3.E.

3.3.2 Payment

At the end of the opening session and after completion of all payment-relevant choices, one decision was randomly drawn as the ‘decision that counts’. The random draw was conducted with help of a bingo cage containing 90 numbered balls. Decision-based payouts for the whole experiment were determined by this decision. If the decision involved payments not only in the opening session, a physical, individualized contract delineating all payments to be made and received was drawn up and signed by the experimenter as well as the participant (see Appendix 3.F). All payments due on the opening session were directly executed. Later payments were executed at the end of the respective session. Payments to the participants were always effected in cash. Participants were allowed to make payments to the experimenters in cash or via PayPal, to minimize the potential transaction burden of payment. All participants with due payments in later sessions received respective email reminders the day before the due date.¹¹

In addition to the decision-based payments, all participants received a show-up fee of €15 for all three sessions at the beginning of the opening session. This money was handed to participants in cash before any decisions took place. As payments in MPL4 contained saving contracts that required participants to pay €15 to the experimenter in the first session, we allowed participants to pay this out of the show-up fee.

After completion of all three sessions and settlement of all due payments, participants received a completion bonus of €20. This payment was implemented via bank transfer and with a delay of one week, to

¹¹To minimize the risk of confounding preference elicitation, the overall setup was directed at increasing perceived payment reliability, i.e. trust that promised payments by the experimenters will be made, and confidence that promised payments by the participants actually have to be made in the future. Undertaken measures included issuing physical contracts with contact details of the principal experimenter, emphasizing that the principal experimenter guarantees all payments in the experimental instructions, and providing multiple ways for contacting any of the experimenters and the associated economics department at Maastricht University in case any issues regarding payment should arise.

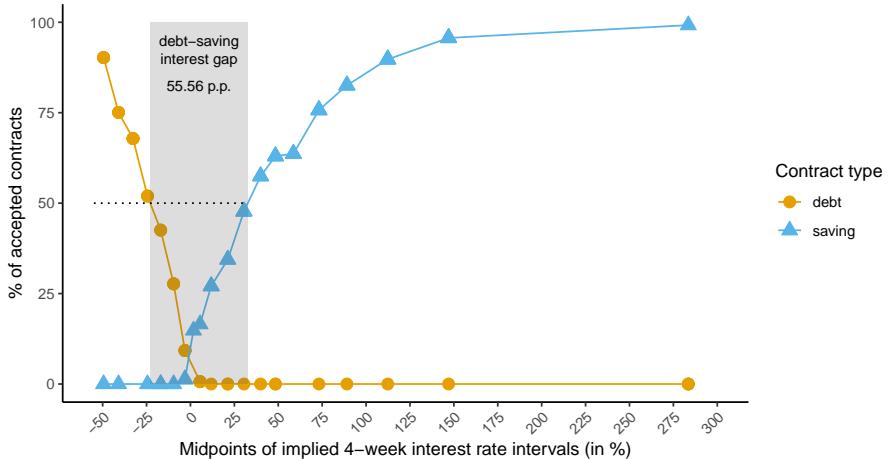
prevent participants to settle any outstanding debt with the completion bonus. Had we paid the completion bonus in cash during the last session, participants may not have thought that they are really in debt, which could have impeded the identification of debt aversion.¹²

3.4 Results

Figure 3.2 plots the percentage of accepted debt and saving contracts over all participants against the implied 4-week interest rate. The graph shows that participants require much more favorable rates to borrow than to save. Notably, most participants only accept debt contracts with negative interest rates. The interest rate at which half of the debt contracts are accepted is negative at -22.89% . To accept saving contracts, however, almost all participants require positive interest rates: the interest rate at which half of the contracts are accepted is 32.67% . The gap between required saving and borrowing interest rates for our median participant is thus substantial, at 55.56 percentage points. While this is a clear violation of discounted expected utility and a potential indication of debt aversion, it is not yet conclusive supporting evidence for debt aversion. As Loewenstein and Prelec, 1992 point out, a gap between required interest rates for saving and borrowing decisions could be caused by other factors such as risk and loss aversion, present bias or combinations thereof.

¹²Some experimental sessions in spring 2020 were affected by the closure of Dutch universities including the BEElab facilities at Maastricht University due to the COVID-19 pandemic. For 44 participants it was not possible to conduct Session 3 in the lab as planned. We transformed the respective experimental protocol to an online survey keeping all content identical and visual appearance as similar as possible. As cash payments were no longer possible, all payments were made via bank transfers. This only affected the collection of non-incentivized questionnaires in Session 3; all payment-relevant choices had been made in January 2020 before the COVID-19 pandemic hit Europe and the Netherlands. The same option to conduct Session 3 online was also offered to participants in autumn 2020 and 2021 who needed to quarantine.

Figure 3.2: Share of accepted debt and saving contracts



Notes: The figure includes contracts of all different starting dates (now and in four weeks) and duration (four and eight weeks) for debt as well as saving. Contracts are binned into intervals of implied interest rates.

In the following, we will therefore make use of structural estimations, that allow to isolate genuine debt aversion, while accounting for these other preferences. We will start in subsection 3.4.1 with our main specification, which estimates the model as specified in 3.2. We will then consider extensions and employ various robustness checks in subsection 3.4.2.

3.4.1 Main specification

To estimate preferences structurally, we require specific forms of the functions introduced in section 3.2. Below is a summary of our preferred specification. In section 3.4.2 we conduct a series of robustness checks to test whether our finding of debt aversion is robust to different specifications of these functional forms. Debt aversion remains robust.

We assume normalized power utility to model curvature of utility in gains and losses:

$$u(x) = \frac{(x + \varepsilon)^{1-\alpha} - \varepsilon^{1-\alpha}}{1 - \alpha} \quad (3.5)$$

For $\varepsilon = 0$, this function is characterized by constant relative risk aversion (CRRA). However, for values of α above unity and $\varepsilon = 0$, the derivatives of the utility function diverge around the reference point. We set $\varepsilon = 0.0001$, to maintain a close approximation of CRRA utility while ensuring that preferences are well-behaved around the reference point.¹³ Moreover, in our main specification we assume exponential discounting:

$$\phi(\tau) = \frac{1}{(1 + \delta)^\tau} \quad (3.6)$$

Note that our experiment allows to identify other forms of discounting, such as quasi-hyperbolic discounting. In section 3.4.2, we test a quasi-hyperbolic discount function (Laibson, 1997), but find no empirical support for present bias.

We will start out by estimating preference parameters on the aggregate, that is for our average participant. We will then account for observed heterogeneity, by allowing preference parameters to vary with observed individual characteristics. Finally, we will account for unobserved individual heterogeneity by estimating the joint population distribution of all preference parameters.

¹³See Wakker (2008) for an illustration, and Meissner, Gassmann, et al., 2022 for a recent application.

Aggregate structural estimation

We jointly estimate preference parameters for risk aversion, time discounting, loss aversion, and debt aversion, according to the model specified in section 3.2, and broadly following the estimation strategies described in e.g. Andersen, Harrison, Lau, et al., 2008; Harrison and Rutström, 2008 and Abdellaoui, Kemel, et al., 2019.

As the basis for all estimations, we consider a random utility model incorporating errors in the decision-making process. Decision-makers may make errors when evaluating the expected utility of different options captured by noise parameter μ . In particular, choices between options A and B are evaluated at their expected intertemporal utility, as specified in Equation 3.2 plus a stochastic error term ε . A decision-maker with preference parameters $\omega = (\alpha, \beta, \gamma, \lambda)$ chooses option B if $U(X^B, \omega) + \varepsilon^B \geq U(X^A, \omega) + \varepsilon^A$. The probability of observing choice B can then be written as:

$$P^B(\theta) = F\left(\frac{U(X^B, \omega) - U(X^A, \omega)}{\mu}\right) = F(\Delta U(\theta)), \quad (3.7)$$

where F is the cumulative distribution function of $(\varepsilon^A - \varepsilon^B)$ and $\theta = (\alpha, \delta, \gamma, \lambda, \mu)$ denotes the vector of preference parameters and the error parameter. We assume $(\varepsilon^A - \varepsilon^B)$ to follow a standard logistic distribution with $F(\xi) = (1 + e^{-\xi})^{-1}$ in our main specification. This specification is often termed Luce model (Luce, Suppes, et al., 1965; Holt and Laury, 2002) or Fechner error with logit link (Drichoutis and Lusk, 2014). Overall, we estimate four preference parameters and one error parameter: risk aversion α , time discounting δ , debt aversion γ , loss aversion λ , and the Fechner error μ , respectively. Intuitively, the error parameter can be interpreted as follows: for $\mu \rightarrow 0$ choice becomes deterministic, and for $\mu \rightarrow \infty$, choice approaches uniform randomization. Aggregating over all choices and individuals the log-likelihood function writes as:

Table 3.2: Aggregate parameter estimates

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|-------------------------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| point estimate | 0.643 | 0.036 | 1.0535 | 1.1074 | 0.4484 |
| 95% confidence interval | 0.58 / 0.71 | 0.02 / 0.05 | 1.03 / 1.08 | 1.08 / 1.13 | 0.37 / 0.53 |
| robust standard error | 0.0345 | 0.006 | 0.0112 | 0.0118 | 0.0402 |

estimation details: n = 12240, log-likelihood = -4107.91, AIC = 8225.81, BIC = 8262.88, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

$$\ln(L(\theta)) = \sum_i \sum_j [\ln(F(\Delta U(\theta))) c_{ij} + \ln(1 - F(\Delta U(\theta)))(1 - c_{ij})], \quad (3.8)$$

where $c_{ij} = 0$ if individual i chooses A in choice j and $c_{ij} = 1$ if individual i chooses B in choice j .

By maximizing the log-likelihood function over θ we derive point estimates for all preference parameters and the error parameter. These estimates describe preferences of the average decision-maker. To account for dependency of choices made by the same person we cluster standard errors at the individual level. Estimates are calculated using STATA's modified Newton-Raphson algorithm.

Estimation results are presented in Table 3.2.¹⁴ Most importantly for this study, the estimate of the parameter indicating debt aversion $\gamma = 1.0535$ is significantly larger than one, suggesting that participants are on average debt averse.

¹⁴To check for potential multiplicity of maximum-likelihood solutions, we estimate the aggregate parameters from 100 combinations of randomly drawn starting values ($\alpha \sim U(0.5, 1)$; $\delta \sim U(0.5, 1)$; $\gamma \sim U(0.5, 1.5)$; $\lambda \sim U(0.5, 2.5)$; $\sigma \sim U(0, 2)$). Parameter estimates are virtually identical to the estimates reported in Table 3.2 for any of the tested combinations of starting values.

To put this estimate in perspective, a decision-maker with the preference parameters as in Table 3.2 would be indifferent between accepting or rejecting a debt contract that involves a loan amount of €20.93 today, with an associated repayment of €15 in four weeks. That is, our average participant would require a negative interest rate to accept a debt contract. This in itself, however, is not yet evidence of debt aversion, as other preferences such as loss aversion could potentially partly explain this. To understand the impact of debt aversion, we calculate what a counterfactual debt-neutral decision-maker with the same preferences, except $\gamma = 1$, would do. Such a decision-maker would accept a loan already as soon as it pays at least €18.08 today, everything else equal. Our average debt averse participant thus requires €2.85 more upfront, in order to be indifferent between accepting or rejecting a debt contract compared to the counterfactual debt-neutral decision-maker. We define the “borrowing premium” as the relative increase in the upfront payment (i.e. the principal) a debt averse person would require compared to a debt neutral person in order to accept a debt contract. For the average participant, this would be $2.85/18.08 = 15.76\%$. In other words, the average, debt averse decision-maker requires a borrowing premium of 15.76% larger loan sizes, while keeping repayment constant, to be willing to accept a debt contract compared to their debt-neutral counterpart. Importantly, debt aversion is not only statistically and economically significant but including debt aversion also meaningfully increases the model’s performance relative to the benchmark assuming debt neutrality: Both AIC and BIC are reduced in the model including debt aversion.¹⁵

Regarding preferences other than debt aversion, participants are on average risk averse with a parameter of relative risk aversion $\alpha = 0.643$. This estimate is comparable to previous studies with large-scale samples using a similar utility specification. Andersen, Harrison, Lau, et al., 2008 find a parameter of relative risk aversion in the adult Danish population of 0.741, while Meissner, Gassmann, et al., 2022 report 0.456

¹⁵Compared to the model assuming debt neutrality (see Appendix Table 3.22), AIC decreases by 38.67 and BIC decreases by 31.32.

based on a representative sample from eight European countries. The four-week discount rate δ is estimated at around 0.036 which is in the range of other lab studies on time preferences as summarized by recent meta-study results (Matousek, Havranek, and Irsova, 2021). Further, participants are on average loss averse, with $\lambda = 1.1074$. This is lower than estimates typically observed in the literature, where λ is usually found to be around 2 (Brown et al., 2022). However, in most studies loss aversion is elicited with risky prospects, i.e. gains and losses are separated by state at one point in time. Abdellaoui, Bleichrodt, et al., 2013 show that when gains and losses are separated by time and do not involve risk, such as in our savings and debt contracts, the loss aversion parameter is substantially lower with an estimate at around $\lambda = 1.15$.

Note that in our main specification we are assuming that the elasticity of intertemporal substitution is the reciprocal of the parameter of risk aversion, and thus, that the elasticity of intertemporal substitution can be identified with the help of atemporal lottery choices. While this is a common assumption in the literature, there is considerable debate whether it is warranted (Andreoni and Sprenger, 2012). Andersen, Harrison, Lau, et al., 2008 were first to use utility curvature elicited with lotteries to correct the estimates of discount rates. Since then, most studies find that while utility over time is also concave, it exhibits less curvature than utility over risk (Abdellaoui, Bleichrodt, et al., 2013; Cheung, 2020). Further, while utility curvature over time appears to be different from utility curvature over risk, the two appear to be correlated (Meissner and Pfeiffer, 2022). For our study, correcting for utility curvature appears to be the most conservative approach with respect to the estimate of the debt aversion parameter. In Appendix 3.D and Table 3.24 we relax this assumption and show that not accounting for utility curvature leads to considerably larger estimates of the debt aversion parameter.

Table 3.3: Description of variables of observable characteristics

| Variable label | Variable description |
|-----------------------|---|
| Age | Participant age in years |
| Cognitive ability | Number of correct answers in cognitive reflection, numeracy and raven tests weighted according to number of items per category (z-score, see Appendix Table 3 for more details) |
| Female | Dummy coded = 1 if female |
| Financial Literacy | Number of correct answers in financial literacy quiz (z-score) |
| Agreeableness | Big-5 personality trait agreeableness (z-score) |
| Conscientiousness | Big-5 personality trait conscientiousness (z-score) |
| Extraversion | Big-5 personality trait extraversion (z-score) |
| Negative emotionality | Big-5 personality trait negative emotionality (z-score) |
| Openmindedness | Big-5 personality trait open mindedness (z-score) |

Debt aversion and observable individual characteristics

In this section, we expand the previous aggregate estimation, by allowing preference and error parameters to vary with observable individual characteristics. Preference and error parameters are estimated as linear functions of individual characteristics including age, gender, cognitive ability, financial literacy, and personality traits as described in Table 3.3. The estimation results are presented in Table 3.4.

First, focusing on individual characteristics associated with debt aversion we can identify a weak negative association between debt aversion and cognitive ability: People who score higher on our measure of cognitive ability, which includes tests on cognitive reflection, numeracy, and fluid intelligence, appear to have lower levels of debt aversion. This

finding is interesting, as it has the opposite sign of what is reported in Ahrens, Bosch-Rosa, and Meissner, 2022. They report a weak positive association. However, the two findings are difficult to compare, as different measures of individual debt aversion as well as cognitive ability are used.

Other individual characteristics, such as age, gender, financial literacy, and personality appear unrelated to debt aversion.

Further findings include a positive correlation between age and risk aversion and strong evidence for a negative correlation between age and loss aversion. These findings are in line with the thrust of the literature (Meissner, Gassmann, et al., 2022). Finally, we find weak evidence suggesting that females are more risk but less loss averse and that people with higher agreeableness scores tend to be more loss averse.

Population distributions of parameters

As a further generalization, we account for unobserved heterogeneity of preferences between individuals in our sample by estimating a structural model of the joint distribution of preference parameters in the population as in Conte, Hey, and Moffatt, 2011 and Gaudecker, Soest, and Wengström, 2011. To this end, we extend our stochastic specification to be in line with the non-linear-mixed-logit routine introduced by (Andersen, Harrison, Hole, et al., 2012).

In particular, we assume that the vector of preference parameters and the error parameter $\theta = (\alpha, \delta, \gamma, \lambda, \mu)$ follows a joint normal distribution f with distribution hyper-parameter vector Θ . Given the joint normal form, Θ comprises mean as well as standard deviation for each parameter in θ and the covariances between all possible pairings of these parameters.

Table 3.4: Individual characteristics associated with preference parameters

| | Risk aversion | Discounting | Debt Aversion | Loss Aversion | Fechner error |
|-----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | α | δ | γ | λ | μ |
| Age | 0.035 (0.015) | -0.003 (0.002) | -0.006 (0.005) | -0.012 (0.004) | -0.038 (0.013) |
| Cognitive ability | -0.007 (0.037) | -0.012 (0.008) | -0.022 (0.012) | -0.015 (0.011) | -0.034 (0.057) |
| Female | 0.161 (0.097) | -0.008 (0.015) | 0.010 (0.034) | -0.063 (0.038) | -0.283 (0.158) |
| Financial literacy | -0.033 (0.025) | 0.003 (0.007) | -0.003 (0.013) | -0.006 (0.006) | 0.009 (0.027) |
| Agreeableness | -0.027 (0.027) | 0.005 (0.005) | 0.004 (0.010) | 0.013 (0.007) | 0.010 (0.026) |
| Conscientiousness | -0.040 (0.037) | -0.005 (0.007) | -0.016 (0.014) | 0.005 (0.014) | 0.055 (0.050) |
| Extraversion | -0.005 (0.051) | -0.003 (0.009) | 0.001 (0.014) | -0.005 (0.010) | 0.003 (0.059) |
| Negative emotionality | 0.043 (0.076) | -0.002 (0.011) | -0.007 (0.017) | -0.015 (0.017) | -0.037 (0.102) |
| Openmindedness | 0.021 (0.030) | 0.001 (0.007) | 0.004 (0.014) | -0.014 (0.009) | -0.008 (0.032) |
| Constant | -0.199 (0.289) | 0.107 (0.053) | 1.176 (0.110) | 1.414 (0.080) | 1.424 (0.278) |
| N | 12240 | | | | |
| Log. Likelihood | -3695 | | | | |
| BIC | 7860 | | | | |

Standard errors (clustered at the subject level) in parentheses

Let θ_i denote a realization of θ for a particular individual i . Analogously to Equation 3.7, in a particular decision, individual i will choose Option B, conditional on θ_i , with the following probability:

$$P_i^B(\theta_i) = F(\Delta U(\theta_i)) \quad (3.9)$$

Aggregating over all choices j , the probability of all observed choices by individual i is:

$$P_i(\theta_i) = \prod_j (P_{ij}^B(\theta_i)c_{ij} + (1 - P_{ij}^B(\theta_i))(1 - c_{ij})), \quad (3.10)$$

where, analogously to the aggregate specification, the index $c_{ij} = 0$ if individual i chooses Option A in choice j and $c_{ij} = 1$ if individual i chooses Option B in choice j . Deriving the probability of observed choices conditional on the population distribution hyper-parameters Θ rather than an individual realization θ_i involves integration over the distribution of θ :

$$P_i(\Theta) = \int P_i(\theta_i) f(\theta|\Theta) d\theta \quad (3.11)$$

In particular, $P_i(\Theta)$ for any individual i is given by integrating over the weighted average of conditional probabilities of observed choices $P_i(\theta_i)$ aggregated over all choices j evaluated at different values of θ and weights given by the density of model parameters f . The log-likelihood function over all individuals then writes as:

$$\ln L(\Theta) = \sum_i \ln(P_i(\Theta)) \quad (3.12)$$

We maximize the log likelihood numerically, using simulated maximum likelihood, as suggested by Andersen, Harrison, Hole, et al., 2012

Table 3.5: Maximum simulated likelihood estimates

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|--------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| mean | 0.5319 | 0.0391 | 1.0639 | 1.1444 | 0.3116 |
| 95% CI | 0.507 / 0.5567 | 0.0332 / 0.0449 | 1.0494 / 1.0783 | 1.1304 / 1.1585 | 0.2839 / 0.3393 |
| SE | 0.0127 | 0.003 | 0.0074 | 0.0072 | 0.0141 |

estimation details: n = 24480, log-likelihood = -2531.17, AIC = 5102.33, BIC = 5264.44, logit Fechner error

Standard errors (SE) clustered at the individual level, 127 clusters

and reviewed earlier in Cameron and Trivedi, 2005 and Train, 2009. In particular, we employ STATA's modified Newton-Raphson algorithm to maximize the likelihood function in Equation 3.12. Resulting estimates of the distributional parameters for preferences over risk, time, losses and debt are displayed in Table 3.5 (means) and Table 3.6 (variance-covariance matrix). The two-dimensional cross-sections of the probability density function for all parameters are illustrated in Figure 3.3.

Distribution estimation results support the finding of debt aversion in the aggregate estimations: around 89% of the population is estimated to have a debt aversion parameter above one, i.e. exhibits debt aversion.

A key advantage of estimating the joint distribution of all preference parameters, including the variance-covariance matrix, is that we can identify correlations of the structural preference parameters based on covariances of the estimated population distributions. In this regard, we find that debt aversion is positively correlated with loss aversion, with a correlation coefficient of $\rho = 0.4756$.¹⁶ Notably, no other preference parameter appears to be correlated with debt aversion. Regarding preferences other than debt aversion, risk aversion appears to be nega-

¹⁶Pearson's correlation coefficient is calculated as $\rho_{x,y} = \frac{Cov_{x,y}}{\sigma_x \sigma_y}$ where Cov is the covariance as reported in Table 3.6, and σ denotes standard deviations, which can be retrieved as \sqrt{var} using variances reported in Table 3.6.

Table 3.6: Variance-covariance matrix

| | α | δ | γ | λ | μ |
|------------------|-------------------|------------------|-----------------|-----------------|-----------------|
| | risk aversion | discounting | debt aversion | loss aversion | Fechner error |
| α var/cov | 0.0317 | | | | |
| α 95% CI | 0.0271 / 0.0364 | | | | |
| SE | 0.0024 | | | | |
| δ | -0.0013 | 0.0013 | | | |
| | -0.0023 / -0.0004 | 0.0008 / 0.0018 | | | |
| | 0.0005 | 0.0003 | | | |
| γ | 0.0004 | 0.0005 | 0.0027 | | |
| | -0.0019 / 0.0027 | -0.0002 / 0.0013 | 0.0013 / 0.0041 | | |
| | 0.0012 | 0.0004 | 0.0007 | | |
| λ | -0.0159 | 0.0042 | 0.0039 | 0.0249 | |
| | -0.0183 / -0.0135 | 0.0034 / 0.0051 | 0.0014 / 0.0064 | 0.0204 / 0.0295 | |
| | 0.0012 | 0.0005 | 0.0013 | 0.0023 | |
| μ | -0.0298 | 0.0041 | 0.0053 | 0.0264 | 0.0435 |
| | -0.0345 / -0.025 | 0.0026 / 0.0056 | 0.0013 / 0.0093 | 0.0206 / 0.0321 | 0.0313 / 0.0558 |
| | 0.0024 | 0.0008 | 0.002 | 0.0029 | 0.0063 |

estimation details: n = 24480, log-likelihood = -2531.17, AIC = 5102.33, BIC = 5264.44, logit Fechner error

Standard errors (SE) clustered at the individual level, 127 clusters

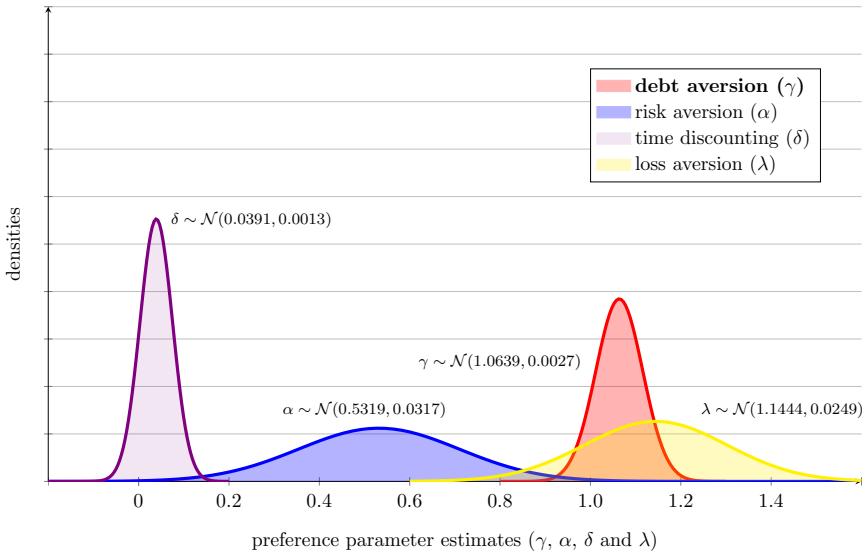
tively correlated with time discounting ($\rho = -0.2025$) and loss aversion ($\rho = -0.5659$), and time discounting is positively correlated with loss aversion ($\rho = 0.7382$). These results are in line with Schleich, Gassmann, et al., 2019, who test for correlation of preference parameters in a large-scale multi-country representative survey.

3.4.2 Extensions and robustness checks

Duration of indebtedness

In this extension, we aim to investigate potential mechanisms of how debt aversion takes effect. Specifically, we will test whether the cost of being in debt could depend on the time a person spends indebted. We

Figure 3.3: Probability density functions of preference parameters



find support that debt aversion does not only depend on the amount and timing of repayment but also substantially on the time of indebtedness.

To test this, we extended the main experiment in the final wave of data collection. This extended experiment contained 30 additional choices on savings and debt contracts spanning not four, but a longer period of eight weeks. The remaining 90 choices and all general procedures are the same as in the main experiment without extension. Just as in the 90 original choices, all payments to the experimenter in the additional choices are held constant at €15, i.e. all loans require the same amount of repayment.

We consider an extended model specification of debt aversion that additionally depends on the time of being indebted $T - t$:

$$c(\mathbf{x}) = (1 - \gamma\zeta^{(T-t-1)})\phi(T)v(x_T) \quad (3.13)$$

Maintaining the components and interpretation of the previous specification of debt aversion, in this extension ζ scales debt aversion based on the time of being indebted. In particular, $\zeta > 1$ implies increasing cost of being indebted if the time of being indebted increases, $\zeta = 1$ implies invariance of debt aversion with respect to time of indebtedness, and $\zeta < 1$ describes decreasing cost of being indebted if the time of indebtedness increases.

In this setting the utility of a short debt contract, i.e. $T - t = 1$, is the same as in the specification without debt duration dependent scaling of debt aversion (ζ):

$$U(x) = \phi(t)u(x_t) - \lambda\gamma\phi(T)u(-x_T), \quad (3.14)$$

for long debt contracts, i.e. $T - t = 2$ in the setting of our experiment, the utility of debt contracts simplifies to

$$U(x) = \phi(t)u(x_t) - \lambda\gamma\zeta\phi(T)u(-x_T) \quad (3.15)$$

Using aggregate maximum likelihood estimations and pooled choice data from the standard experiment with 90 choices and the extended experiment with 120 choices yields results as summarized in Table 3.7. Parameter estimates, including debt aversion as established in the main specification, remain largely unchanged. Interestingly, however, the debt duration aversion parameter ζ is significantly larger than one. This implies that the cost of being indebted increases in the time of indebtedness.

Table 3.7: Aggregate parameter estimates including debt duration aversion

| | α risk aversion | δ discounting | γ debt aversion | ζ debt duration aversion | λ loss aversion | μ Fechner error |
|----------------|---------------------------|-------------------------|------------------------------|--------------------------------------|----------------------------|------------------------|
| point estimate | 0.6398 | 0.0429 | 1.0635 | 1.851 | 1.1007 | 0.448 |
| 95% CI | 0.57 / 0.71 | 0.03 / 0.06 | 1.04 / 1.09 | 1.28 / 2.42 | 1.08 / 1.12 | 0.37 / 0.53 |
| robust SE | 0.034 | 0.0075 | 0.0134 | 0.2919 | 0.0121 | 0.0402 |

estimation details: n = 12240, log-likelihood = -4095.55, AIC = 8203.1, BIC = 8247.57, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

To illustrate what this implies in terms of behavior, the average decision-maker characterized by parameter estimates derived in the frame of the extended model is indifferent between accepting and rejecting a debt contract that offers a loan of €20.67 today and requires repayment of €15 in four weeks. However, the same decision-maker requires a larger loan of €21.11 today if repayment of €15 is not due in four but eight weeks. In contrast, the hypothetical debt and debt duration neutral counterpart (i.e. $\gamma = 1$ and $\zeta = 1$), requires a four-week loan of size €17.43 and eight-week loan of size €15.51 to be indifferent. Based on the differences, we can calculate borrowing premia of 18.59% for short, four-week debt contracts and 36.10% for long, eight-week debt contracts. In other words, the borrowing premium increases by around 93% if the duration of indebtedness doubles. Note, however, that the identification of the debt duration aversion parameter relies on only 30 participants who completed the extended list of MPLs, and should therefore be interpreted with caution.

Present bias

In principle, our experimental setup allows to identify present bias, as we include debt and saving contracts that are shifted into the future

while maintaining the same temporal distance between involved dates. To test whether present bias is existent in our sample, we consider an alternative discount function, that incorporates quasi-hyperbolic discounting (Phelps and Pollak, 1968; Laibson, 1997):

$$\phi'(\tau) = \begin{cases} 1 & \text{if } \tau = 0 \\ \frac{1}{(1+\beta)} \frac{1}{(1+\delta)^{\tau}} & \text{if } \tau \neq 0. \end{cases} \quad (3.16)$$

In this specification, δ is the exponential discount rate, and β is the parameter that determines present bias. A parameter of $\beta > 0$ indicates present bias, $\beta = 0$ indicates no present bias, and $\beta < 0$ indicates future bias. The results of the aggregate maximum likelihood estimation using this alternative specification are presented in Table 3.8. Present bias is estimated precisely at, and statistically indistinguishable from 0. The evidence against present bias also extends to a setting considering different parameters of present bias for gains and losses (see also Appendix 3.D). These findings, appear in line with recent meta-study results on present bias elicited in experiments (Imai, Rutter, and Camerer, 2020).¹⁷

Different utility curvature and time discounting for gains and losses

Following the argumentation in Loewenstein and Prelec, 1992 debt averse behavior may be explained by a combination of different utility curvatures as well as discount rates for gains and losses. In their model, the combination of a steeper value function for losses than for gains and a larger discount factor for gains than for losses will result in debt aversion. We test whether debt aversion is robust to this by adapting our main specification such that we allow distinct risk aversion and

¹⁷For graphical evidence on the absence of present bias, see Figure 3.4 in the Appendix, which shows that required interest rates for debt and saving contracts do not differ between contracts that are offered now and contracts that are shifted into the future.

Table 3.8: Aggregate parameter estimates including present bias

| | α | β | δ | γ | λ | μ |
|----------------|---------------|--------------|-------------|---------------|---------------|---------------|
| | risk aversion | present bias | discounting | debt aversion | loss aversion | Fechner error |
| point estimate | 0.6431 | -0.0012 | 0.037 | 1.0545 | 1.1069 | 0.4484 |
| 95% CI | 0.58 / 0.71 | -0.01 / <.01 | 0.02 / 0.05 | 1.03 / 1.08 | 1.08 / 1.13 | 0.37 / 0.53 |
| robust SE | 0.0345 | 0.0029 | 0.0065 | 0.0114 | 0.012 | 0.0402 |

estimation details: n = 12240, log-likelihood = -4107.86, AIC = 8227.71, BIC = 8272.19, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

time discounting parameters for gains and losses: $\alpha^+, \alpha^-, \delta^+$ and δ^- respectively. Consequently, atemporal utility takes the form

$$u(x) = \begin{cases} \frac{(x+\varepsilon)^{1-\alpha^+} - \varepsilon^{1-\alpha^+}}{1-\alpha^+} & \text{if } x \geq 0 \\ \frac{(x+\varepsilon)^{1-\alpha^-} - \varepsilon^{1-\alpha^-}}{1-\alpha^-} & \text{if } x < 0, \end{cases} \quad (3.17)$$

and the discount function ϕ now also depends on the sign of x :

$$\phi(\tau, x) = \begin{cases} \frac{1}{(1+\delta^+)^{\tau}} & \text{and } x \geq 0 \\ \frac{1}{(1+\delta^-)^{\tau}} & \text{and } x < 0. \end{cases} \quad (3.18)$$

Estimation results are presented in Table 3.9. In line with Loewenstein and Prelec, 1992, we find $\alpha^+ < \alpha^-$ and $\delta^+ > \delta^-$, although only the difference in risk aversion is statistically significant ($P < 0.01$). Importantly, the debt aversion parameter remains significantly larger than one and has a similar magnitude compared to the debt aversion parameter in our main specification, i.e. there is debt aversion that cannot be

Table 3.9: Aggregate parameter estimates with separate utility curvature and time discounting in the gain and loss domain

| | α^+ RA gains | α^- RA losses | δ^+ TD gains | δ^- TD losses | γ debt aversion | λ loss aversion | μ Fechner error |
|-----------------------|------------------------|-------------------------|------------------------|-------------------------|------------------------------|-------------------------------|---------------------------|
| point estimate | 0.6425 | 0.7878 | 0.0369 | 0.0307 | 1.0485 | 1.0402 | 0.4484 |
| 95% CI | 0.57 / 0.71 | 0.78 / 0.79 | 0.02 / 0.05 | 0.02 / 0.04 | 1.03 / 1.07 | 1.01 / 1.07 | 0.37 / 0.53 |
| robust SE | 0.0346 | 0.0016 | 0.0061 | 0.0062 | 0.0109 | 0.0146 | 0.0402 |

estimation details: n = 12240, log-likelihood = -4106.69, AIC = 8227.39, BIC = 8279.27, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

explained by differences in utility curvature and time discounting in the loss and gain domain.

Robustness checks

Our finding of debt aversion may be sensitive to the assumptions underlying our estimations. We therefore employ a wide array of robustness checks. In particular, we first consider alternative forms of the cost of borrowing by modeling debt aversion as a fixed cost of being indebted as well as scaling of utility from borrowed money. Second, we alter various characteristics of our utility specification in general. These comprise abstracting from risk aversion as well as considering alternative forms of the utility function such as CARA utility and CRRA utility without ε -transformation. Third, we scrutinize different error structures: we introduce an additional tremble error, exchange the logit for a probit Fechner error, and allow distinct probit Fechner errors per choice domain. Moreover, we also test the effect of excluding participants who did not complete the entire experimental sequence or expressed some doubt about the trustworthiness of the experimental environment. Summing up the results, debt aversion remains robust

regardless of the utilized functional forms and sample selection criteria. Detailed descriptions and results on all robustness checks can be found in Appendix 3.D.

3.5 Discussion and conclusion

In this paper, we introduce a novel theoretical framework and experiment that allows to model and measure debt aversion. We are able to separately identify debt aversion from other relevant preferences, such as risk aversion, loss aversion, and time preferences. In this way, we aim to establish debt aversion as a preference in its own right, as opposed to an emergent behavioral property of other preferences, biases beliefs, and constraints.

Using a structural maximum likelihood estimation, we find that our participants are on average debt averse. They would be willing to forgo a substantial amount of money in order to avoid getting into debt. We estimate the “borrowing premium” to be around 16%. This is the increase in the upfront payment our average participant would require compared to a counterfactual debt neutral participant to accept a debt contract. Testing how observed individual characteristics correlate with debt aversion, we find weak evidence supporting a negative correlation between debt aversion and cognitive ability. Other individual characteristics, such as age, gender, financial literacy, and personality appear unrelated to debt aversion. Further, we estimate the joint population distribution of all preference parameters, using simulated maximum likelihood. We find that a substantial share of 89% of our participants exhibits debt aversion. Furthermore, debt aversion appears to be correlated positively with loss aversion, but not with other preference parameters. Finally, we find evidence that debt aversion depends positively on the duration a person spends indebted. Notably, present bias does not appear to be existent in our data, and debt aversion remains robust after a series of robustness checks. Summing up, we find robust evidence supporting debt aversion as a preference in its own right.

Most participants are debt averse, and debt aversion appears to have a meaningful impact on choice.

The existence of debt aversion could have far-reaching implications for individual financial decision-making. Debt averse individuals could invest less in otherwise profitable investment projects, such as education or energy-efficient technologies, and make consumption and saving decisions that deviate from the standard model of intertemporal choice. While we have made a first step in cleanly identifying debt aversion, many open questions on the mechanisms that underlie debt aversion, and its implications for financial decision-making, remain. We test a multitude of ways of modeling debt aversion, and while all specifications lend clear support to its existence, our setup is not well suited to discriminate between different models and different mechanisms of how debt aversion works. In an extension of the base experiment, we show that debt aversion appears to increase in the duration participants spend indebted, but many other interesting questions remain. We hope that our theoretical and experimental framework can pave the way for future research that could improve the knowledge on the exact mechanisms at play.

Debt aversion could also have implications on the macroeconomic level. Hundtofte, Olafsson, and Pagel, 2019 show that consistent with debt aversion, individuals in the US and Iceland are reluctant to use credit to smooth negative transitory income shocks. They argue that while standard theory predicts countercyclical credit demand, credit demand appears to be pro-cyclical which could deepen business cycle fluctuations.

Further, our findings also have implications for policy design. Many policies rely on offering favorable loans to subsidize particular behaviors, such as investment in tertiary education or energy-efficient technologies. However, in the face of a largely debt averse population, these loans might not be very effective. Moreover, if debt aversion correlates with individual characteristics, such as income or socioeconomic status, such policies could have unintended effects. For instance,

loan-based policies to facilitate tertiary education for students from weak financial backgrounds might be particularly unattractive to these students if they are also more debt averse. For these reasons, we believe that more research on how debt aversion relates to individual characteristics of representative populations is required. To facilitate this, we have constructed a short and easy-to-use survey module for measuring individual debt aversion in a companion paper (Albrecht and Meissner, 2022). Using the data from this experiment, we identify a set of survey items that best predicts the debt aversion parameter as elicited with our experiment. The survey module contains two short items and predicts debt aversion reasonably well. We hope that this survey module will prove useful for future research on debt aversion on a larger scale, where complicated and incentivized experiments are often not feasible.

Finally, we believe that our setup provides a valuable methodological contribution. To our knowledge, we are first to put participants into actual debt in a laboratory experiment. As the vast majority of participants did not default on their obligations, we believe that such experiments could prove useful to analyze debt-related behavior in the future.

Appendix

3.A Multiple price lists

Table 3.10: Multiple price list of intertemporal choices (MPL1)

| Choice | Option A | Option B |
|--------|----------------------------------|---------------------------------------|
| 1 | Receive an amount of €18.2 today | Receive an amount of €18.0 in 4 weeks |
| 2 | Receive an amount of €18.0 today | Receive an amount of €18.0 in 4 weeks |
| 3 | Receive an amount of €17.8 today | Receive an amount of €18.0 in 4 weeks |
| 4 | Receive an amount of €17.3 today | Receive an amount of €18.0 in 4 weeks |
| 5 | Receive an amount of €16.8 today | Receive an amount of €18.0 in 4 weeks |
| 6 | Receive an amount of €16.0 today | Receive an amount of €18.0 in 4 weeks |
| 7 | Receive an amount of €14.0 today | Receive an amount of €18.0 in 4 weeks |
| 8 | Receive an amount of €12.0 today | Receive an amount of €18.0 in 4 weeks |
| 9 | Receive an amount of €10.0 today | Receive an amount of €18.0 in 4 weeks |
| 10 | Receive an amount of €8.0 today | Receive an amount of €18.0 in 4 weeks |

Table 3.11: Multiple price list of certain payments vs. risky gambles (MPL2)

| Choice | Option A | | Option B | |
|--------|---------------------|---------------------|---------------------|---------------------|
| | Coin shows Heads | Coin shows Tails | Coin shows Heads | Coin shows Tails |
| 1 | €30 today | €30 today | €30 today | €1 today |
| 2 | €25 today | €25 today | €30 today | €1 today |
| 3 | €20 today | €20 today | €30 today | €1 today |
| 4 | €17 today | €17 today | €30 today | €1 today |
| 5 | €16 today | €16 today | €30 today | €1 today |
| 6 | €15 today | €15 today | €30 today | €1 today |
| 7 | €12 today | €12 today | €30 today | €1 today |
| 8 | €10 today | €10 today | €30 today | €1 today |
| 9 | €5 today | €5 today | €30 today | €1 today |
| 10 | €1 today | €1 today | €30 today | €1 today |

Table 3.12: Multiple price list of less risky vs. more risky gambles (MPL3)

| Choice | Option A | | Option B | |
|--------|---------------------|---------------------|---------------------|---------------------|
| | Coin shows Heads | Coin shows Tails | Coin shows Heads | Coin shows Tails |
| 1 | €14 today | €17 today | €17 today | €1 today |
| 2 | €14 today | €17 today | €20 today | €1 today |
| 3 | €14 today | €17 today | €25 today | €1 today |
| 4 | €14 today | €17 today | €28 today | €1 today |
| 5 | €14 today | €17 today | €29 today | €1 today |
| 6 | €14 today | €17 today | €30 today | €2 today |
| 7 | €14 today | €17 today | €30 today | €3 today |
| 8 | €14 today | €17 today | €32 today | €8 today |
| 9 | €14 today | €17 today | €32 today | €10 today |
| 10 | €14 today | €17 today | €32 today | €14 today |

Table 3.13: Multiple price list of 4-week saving contracts starting at Session 1 (MPL4)

| Choice | Early saving contracts | |
|--------|------------------------|---------------------------|
| | Session 1 (today) | Session 2 (in 4 weeks) |
| 1 | Pay an amount of €15 | Receive an amount of €45 |
| 2 | Pay an amount of €15 | Receive an amount of €40 |
| 3 | Pay an amount of €15 | Receive an amount of €36 |
| 4 | Pay an amount of €15 | Receive an amount of €34 |
| 5 | Pay an amount of €15 | Receive an amount of €32 |
| 6 | Pay an amount of €15 | Receive an amount of €30 |
| 7 | Pay an amount of €15 | Receive an amount of €28 |
| 8 | Pay an amount of €15 | Receive an amount of €26 |
| 9 | Pay an amount of €15 | Receive an amount of €24 |
| 10 | Pay an amount of €15 | Receive an amount of €22 |
| 11 | Pay an amount of €15 | Receive an amount of €20 |
| 12 | Pay an amount of €15 | Receive an amount of €18 |
| 13 | Pay an amount of €15 | Receive an amount of €16 |
| 14 | Pay an amount of €15 | Receive an amount of €14 |
| 15 | Pay an amount of €15 | Receive an amount of €12 |

Table 3.14: Multiple price list of 4-week saving contracts starting at Session 2 (MPL5)

| Choice | Late saving contracts | |
|--------|---------------------------|---------------------------|
| | Session 2 (in 4 weeks) | Session 3 (in 8 weeks) |
| 1 | Pay an amount of €15 | Receive an amount of €40 |
| 2 | Pay an amount of €15 | Receive an amount of €35 |
| 3 | Pay an amount of €15 | Receive an amount of €31 |
| 4 | Pay an amount of €15 | Receive an amount of €29 |
| 5 | Pay an amount of €15 | Receive an amount of €27 |
| 6 | Pay an amount of €15 | Receive an amount of €25 |
| 7 | Pay an amount of €15 | Receive an amount of €23 |
| 8 | Pay an amount of €15 | Receive an amount of €21 |
| 9 | Pay an amount of €15 | Receive an amount of €19 |
| 10 | Pay an amount of €15 | Receive an amount of €17 |
| 11 | Pay an amount of €15 | Receive an amount of €15 |
| 12 | Pay an amount of €15 | Receive an amount of €13 |
| 13 | Pay an amount of €15 | Receive an amount of €11 |
| 14 | Pay an amount of €15 | Receive an amount of €9 |
| 15 | Pay an amount of €15 | Receive an amount of €7 |

Table 3.15: Multiple price list of 4-week debt contracts starting at Session 1 (MPL6)

| Choice | Early debt contracts | |
|--------|--------------------------|---------------------------|
| | Session 1 (today) | Session 2 (in 4 weeks) |
| 1 | Receive an amount of €31 | Pay an amount of €15 |
| 2 | Receive an amount of €27 | Pay an amount of €15 |
| 3 | Receive an amount of €24 | Pay an amount of €15 |
| 4 | Receive an amount of €21 | Pay an amount of €15 |
| 5 | Receive an amount of €19 | Pay an amount of €15 |
| 6 | Receive an amount of €17 | Pay an amount of €15 |
| 7 | Receive an amount of €16 | Pay an amount of €15 |
| 8 | Receive an amount of €15 | Pay an amount of €15 |
| 9 | Receive an amount of €14 | Pay an amount of €15 |
| 10 | Receive an amount of €13 | Pay an amount of €15 |
| 11 | Receive an amount of €11 | Pay an amount of €15 |
| 12 | Receive an amount of €9 | Pay an amount of €15 |
| 13 | Receive an amount of €7 | Pay an amount of €15 |
| 14 | Receive an amount of €5 | Pay an amount of €15 |
| 15 | Receive an amount of €3 | Pay an amount of €15 |

Table 3.16: Multiple price list of 4-week debt contracts starting at Session 2 (MPL7)

| Choice | Late debt contracts | |
|--------|---------------------------|---------------------------|
| | Session 2 (in 4 weeks) | Session 3 (in 8 weeks) |
| 1 | Receive an amount of €33 | Pay an amount of €15 |
| 2 | Receive an amount of €30 | Pay an amount of €15 |
| 3 | Receive an amount of €27 | Pay an amount of €15 |
| 4 | Receive an amount of €24 | Pay an amount of €15 |
| 5 | Receive an amount of €22 | Pay an amount of €15 |
| 6 | Receive an amount of €20 | Pay an amount of €15 |
| 7 | Receive an amount of €18 | Pay an amount of €15 |
| 8 | Receive an amount of €16 | Pay an amount of €15 |
| 9 | Receive an amount of €15 | Pay an amount of €15 |
| 10 | Receive an amount of €14 | Pay an amount of €15 |
| 11 | Receive an amount of €12 | Pay an amount of €15 |
| 12 | Receive an amount of €10 | Pay an amount of €15 |
| 13 | Receive an amount of €8 | Pay an amount of €15 |
| 14 | Receive an amount of €6 | Pay an amount of €15 |
| 15 | Receive an amount of €3 | Pay an amount of €15 |

Table 3.17: Multiple price list of 8-week saving contracts starting at Session 1 (MPL8)

| Choice | Long saving contracts | |
|--------|-----------------------|---------------------------|
| | Session 1 (today) | Session 3 (in 8 weeks) |
| 1 | Pay an amount of €15 | Receive an amount of €50 |
| 2 | Pay an amount of €15 | Receive an amount of €45 |
| 3 | Pay an amount of €15 | Receive an amount of €40 |
| 4 | Pay an amount of €15 | Receive an amount of €36 |
| 5 | Pay an amount of €15 | Receive an amount of €34 |
| 6 | Pay an amount of €15 | Receive an amount of €32 |
| 7 | Pay an amount of €15 | Receive an amount of €30 |
| 8 | Pay an amount of €15 | Receive an amount of €28 |
| 9 | Pay an amount of €15 | Receive an amount of €26 |
| 10 | Pay an amount of €15 | Receive an amount of €24 |
| 11 | Pay an amount of €15 | Receive an amount of €22 |
| 12 | Pay an amount of €15 | Receive an amount of €20 |
| 13 | Pay an amount of €15 | Receive an amount of €18 |
| 14 | Pay an amount of €15 | Receive an amount of €16 |
| 15 | Pay an amount of €15 | Receive an amount of €14 |

Table 3.18: Multiple price list of 8-week debt contracts starting at Session 1 (MPL9)

| Choice | Long debt contracts | |
|--------|----------------------------|---------------------------|
| | Session 1 (today) | Session 3 (in 8 weeks) |
| 1 | Receive an amount of €39 | Pay an amount of €15 |
| 2 | Receive an amount of €35 | Pay an amount of €15 |
| 3 | Receive an amount of €31 | Pay an amount of €15 |
| 4 | Receive an amount of €27 | Pay an amount of €15 |
| 5 | Receive an amount of €24 | Pay an amount of €15 |
| 6 | Receive an amount of €21 | Pay an amount of €15 |
| 7 | Receive an amount of €19 | Pay an amount of €15 |
| 8 | Receive an amount of €17 | Pay an amount of €15 |
| 9 | Receive an amount of €16 | Pay an amount of €15 |
| 10 | Receive an amount of €15 | Pay an amount of €15 |
| 11 | Receive an amount of €14 | Pay an amount of €15 |
| 12 | Receive an amount of €13 | Pay an amount of €15 |
| 13 | Receive an amount of €11 | Pay an amount of €15 |
| 14 | Receive an amount of €9 | Pay an amount of €15 |
| 15 | Receive an amount of €7 | Pay an amount of €15 |

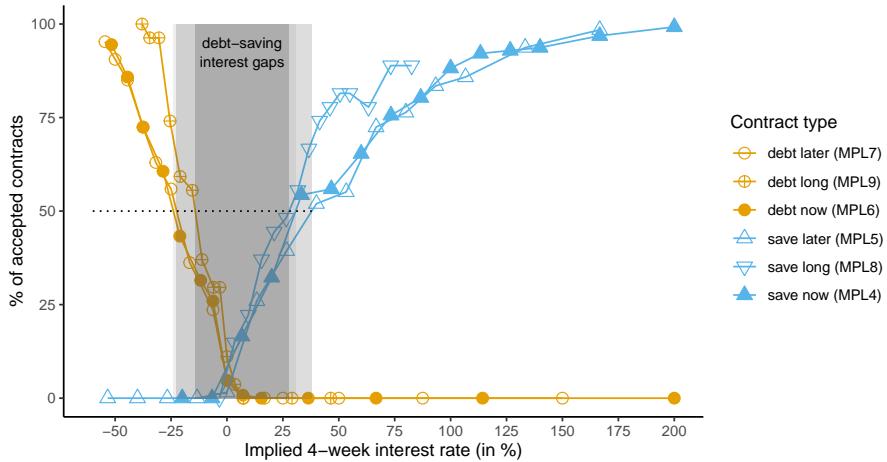
3.B Individual characteristics

Table 3.19: Overview of elicited individual characteristics

| Individual Characteristic | Measurement | No. of Items | Source / Reference |
|---------------------------|---------------------------------|--------------|------------------------------------|
| Financial Literacy | Global OECD/INFE FinLitSurvey | 5 | (Atkinson et al., 2016) |
| | S&P International FinLit Survey | 3 | (Klapper et al., 2015) |
| | Debt Literacy | 3 | (Lusardi and Tufano, 2015) |
| | FinLit Quiz | 4 | (Agnew and Harrison, 2015) |
| Numeracy | Abbreviated Numeracy Scale | 6 | (Weller et al., 2013) |
| Cognitive Reflection | CRT | 3 | (Frederick, 2005) |
| | CRT-2 | 4 | (Thomson and M. Oppenheimer, 2016) |
| | CRT-long | 3 | (Primi et al., 2016) |
| | Extended CRT | 3 | (Toplak et al., 2013) |
| Fluid Intelligence | Raven Progressive Matrices | 36 | (Raven, 1960) |
| Personality | BFI-2-S | 30 | (Soto and John, 2017) |
| | HEXACO-60 (Honesty) | 8 | (Ashton and Lee, 2009) |
| Preferences | Preference Survey Module | 12 | (Falk et al., 2016) |
| Planned Behavior | Purchases | 7 | – |
| | Financing | 7 | – |

3.C Additional graphs

Figure 3.4: Share of accepted debt and saving contracts by starting dates and duration



Notes: Debt and saving contracts of different starting dates (now and in four weeks) and duration (four and eight weeks) are displayed separately.

3.D Robustness checks

In this section we will consider several different structural assumptions pertaining to the cost of being in debt, the utility and discount functions, as well as the decision error process. To keep the appendix size manageable, we will make use of the simple aggregate maximum likelihood specification, outlined in section 3.4.1

3.D.1 Alternative specifications of the utility cost of borrowing

Fixed cost of being indebted:

In our main specification, the cost of being in debt depends on the timing and amount of the required repayment. Instead, debt aversion could also be modeled as a fixed cost, which is incurred at the time of going into debt. Thus, we consider an alternative definition of $c(\mathbf{x})$ with γ as fixed cost independent of the size of the loan:

$$c(\mathbf{x}) = \gamma\phi(t). \quad (3.19)$$

For interpretation, $\gamma = 0$ now corresponds to debt neutrality, i.e. no utility cost of borrowing, $\gamma > 0$ corresponds to debt aversion and $\gamma < 0$ to debt affinity. Estimation results are presented in Table 3.20. Except for γ , parameter estimates remain virtually identical. The new debt aversion parameter is estimated at $\gamma = 0.4332$, implying that the average decision-maker faces positive fixed utility costs when going into debt, i.e. they are debt averse.

Scaling utility of borrowed money: As a second alternative, we consider a utility cost of borrowing that is dependent on the timing and amount of loan receipt. Intuitively, one could think of the decision-maker experiencing less pleasure from money that is actually borrowed

Table 3.20: Aggregate parameter estimates with fixed cost of going into debt

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|-------------------------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| point estimate | 0.6421 | 0.0371 | 0.4332 | 1.1064 | 0.449 |
| 95% confidence interval | 0.57 / 0.71 | 0.02 / 0.05 | 0.25 / 0.61 | 1.08 / 1.13 | 0.37 / 0.53 |
| robust standard error | 0.0344 | 0.0063 | 0.0911 | 0.0118 | 0.0402 |

estimation details: n = 12240, log-likelihood = -4106.98, AIC = 8223.95, BIC = 8261.01, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

compared to money from other sources. The cost of being in debt is thus defined as:

$$c(\mathbf{x}) = (1 - \gamma)\phi(t)v(x_t). \quad (3.20)$$

Interpretation of γ is inverted as compared to the main specification, i.e. $\gamma < 1$ corresponds to debt aversion and $\gamma > 1$ to debt affinity. Estimation results are presented in Table 3.21. Again, except for γ , parameter estimates remain largely unchanged. The new debt aversion parameter is estimated at $\gamma = 0.9543$, i.e. the average decision-maker remains debt averse.

Assuming debt neutrality: Finally, we consider a model based on our main specification, but abstracting from debt aversion, i.e. assuming $\gamma = 1$. Comparing our main specification to the debt neutral model, allows to scrutinize whether incorporating any form of cost of being indebted increases explanatory power.

Estimation results are presented in Table 3.22. Comparing the information criteria AIC and BIC we observe that any model incorporating cost of being in debt is superior to the specification abstracting from debt attitudes, thus corroborating debt attitudes as a distinct domain of individual preferences. This holds irrespective of whether the cost of being

Table 3.21: Aggregate parameter estimates with gamma as scaling factor for utility from loan payments

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|-------------------------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| point estimate | 0.6421 | 0.034 | 0.9543 | 1.1111 | 0.4435 |
| 95% confidence interval | 0.57 / 0.71 | 0.02 / 0.05 | 0.93 / 0.98 | 1.09 / 1.14 | 0.37 / 0.52 |
| robust standard error | 0.0348 | 0.006 | 0.0108 | 0.0124 | 0.0399 |

estimation details: n = 12240, log-likelihood = -4112.31, AIC = 8234.62, BIC = 8271.68, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

Table 3.22: Aggregate parameter estimates assuming debt neutrality

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|-------------------------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| point estimate | 0.6421 | 0.034 | 0.9543 | 1.1111 | 0.4435 |
| 95% confidence interval | 0.57 / 0.71 | 0.02 / 0.05 | 0.93 / 0.98 | 1.09 / 1.14 | 0.37 / 0.52 |
| robust standard error | 0.0348 | 0.006 | 0.0108 | 0.0124 | 0.0399 |

estimation details: n = 12240, log-likelihood = -4112.31, AIC = 8234.62, BIC = 8271.68, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

in debt is modeled as scaling disutility from repayments (main specification), fixed cost of being indebted, or scaling utility from borrowed money.

3.D.2 Alternative utility specifications

Besides the cost of being in debt, we consider a wide range of alternatives to characteristics of our utility specification in general.

Different present bias for gains and losses: Besides utility curvature and time discounting, allowing for distinct present bias in gains and losses constitutes an appealing robustness check in the context of our experiment. In particular, present bias for payments in the gain domain, i.e. β^+ for $x > 0$ also captures participants' trust in the experimenters' promises to affect payments in the future. Analogously, present bias in the loss domain, i.e. β^- for $x < 0$ captures participants' confidence that they will actually satisfy their own future payment obligations. If participants exhibit either mistrust or a lack of confidence in their own payment reliability, on average debt contracts would appear more and savings contracts less appealing to them. In this situation, our estimate of debt aversion would actually be biased downwards. In this light, we consider different present bias in gains and losses through the discount function:

$$\phi(\tau, x) = \begin{cases} 1 & \text{if } \tau = 0 \\ \frac{1}{1+\beta^+} \frac{1}{(1+\delta)^\tau} & \text{if } \tau \neq 0 \text{ and } x \geq 0 \\ \frac{1}{1+\beta^-} \frac{1}{(1+\delta)^\tau} & \text{if } \tau \neq 0 \text{ and } x < 0. \end{cases} \quad (3.21)$$

Estimation results are presented in Table 3.23. We find neither evidence for significant differences between β^+ and β^- , nor that any parameter of present bias is different from zero. Moreover, the debt aversion parameter remains larger than 1, and has a similar magnitude as in our main specification. Summing up, debt aversion persists and is unlikely biased downwards due to mistrust in the experiment or participants' confidence in their payment morale.

Risk neutrality: As utility curvature, determined through α , has a major effect on the size of the remaining preference parameters in our

Table 3.23: Aggregate parameter estimates allowing with separate present bias in the gain and loss domain

| | α | β^+ | β^- | δ | γ | λ | μ |
|-----------------------|---------------|--------------|--------------|------------------|---------------|---------------|---------------|
| | risk aversion | PB gains | PB losses | time discounting | debt aversion | loss aversion | Fechner error |
| point estimate | 0.6431 | -0.0008 | -0.0018 | 0.0367 | 1.0537 | 1.1065 | 0.4484 |
| 95% CI | 0.58 / 0.71 | -0.01 / 0.01 | -0.01 / 0.01 | 0.02 / 0.05 | 1.03 / 1.08 | 1.08 / 1.13 | 0.37 / 0.53 |
| robust SE | 0.0345 | 0.0062 | 0.0055 | 0.0078 | 0.0144 | 0.0119 | 0.0402 |

estimation details: n = 12240, log-likelihood = -4107.84, AIC = 8229.69, BIC = 8281.57, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

Table 3.24: Aggregate parameter estimates assuming risk neutrality

| | α | δ | γ | λ | μ |
|-------------------------|---------------|-------------|---------------|---------------|---------------|
| | risk aversion | discounting | debt aversion | loss aversion | Fechner error |
| point estimate | | 0.1018 | 1.154 | 1.3533 | 3.0901 |
| 95% confidence interval | | 0.07 / 0.14 | 1.09 / 1.22 | 1.3 / 1.41 | 2.8 / 3.38 |
| robust standard error | | 0.0175 | 0.033 | 0.0269 | 0.1468 |

estimation details: n = 12240, log-likelihood = -4813.36, AIC = 9634.73, BIC = 9664.38, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

main specification we also consider an adaption abstracting from utility curvature. In particular, we consider the case of risk neutrality with $\alpha = 0$.

Estimation results are presented in Table 3.24. Debt aversion persists and appears higher than in our main specification allowing for risk aversion. As expected, also the remaining parameter estimates change substantially, which makes intuitive sense due to the different shape of the atemporal utility function $u(x)$ when assuming $\alpha = 0$.

CARA utility: Our main specification assumes atemporal utility to be

characterized by constant relative risk aversion (CRRA). We examine the robustness of our findings by additionally considering exponential utility characterized by constant absolute risk aversion (CARA):

$$u(x) = \frac{1 - e^{-\varphi x}}{\varphi} \quad (3.22)$$

Here, φ is the parameter of absolute risk aversion. Estimation results are presented in Table 3.25. Again as in the previous robustness check, this adaption leads to a substantial change in the shape of the atemporal utility function $u(x)$. As a consequence, also parameter estimates beyond φ change considerably. Debt aversion, however, persists.

Table 3.25: Aggregate parameter estimates with CARA utility

| | φ risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|-------------------------|----------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| point estimate | 0.0112 | 0.0898 | 1.1358 | 1.3055 | 2.4346 |
| 95% confidence interval | <.01 / 0.02 | 0.06 / 0.12 | 1.08 / 1.19 | 1.26 / 1.35 | 2.08 / 2.79 |
| robust standard error | 0.0042 | 0.0141 | 0.027 | 0.0252 | 0.1805 |

estimation details: n = 12240, log-likelihood = -4182.05, AIC = 8374.1, BIC = 8411.16, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

No ε -transformation: Lastly, in our main specification we consider CRRA utility including an ε -transformation, because of its beneficial properties for estimation. As a final alteration of the utility specification, we consider a robustness check without the ε -transformation, i.e. atemporal utility takes the form:

$$u(x) = \frac{x^{1-\alpha}}{1-\alpha} \quad (3.23)$$

Table 3.26: Aggregate parameter estimates estimated without ε -transformation

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|-------------------------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| point estimate | 0.6429 | 0.0355 | 1.0528 | 1.106 | 0.4487 |
| 95% confidence interval | 0.58 / 0.71 | 0.02 / 0.05 | 1.03 / 1.07 | 1.08 / 1.13 | 0.37 / 0.53 |
| robust standard error | 0.0344 | 0.006 | 0.0111 | 0.0121 | 0.0402 |

estimation details: n = 12240, log-likelihood = -4108, AIC = 8226, BIC = 8263.06, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 127 clusters

Estimation results are presented in Table 3.26. Parameter estimates remain largely unchanged compared to the main specification. Debt aversion remains robust.

3.D.3 Alternative error structures

In line with Drichoutis and Lusk, 2014, we acknowledge that parameter estimates may depend on assumptions about the decision error process. Therefore, we employ three alternative error structures as robustness checks.

Probit-link Fechner error: First, we consider a Fechner error with probit link instead of logit link as in our main specification. Technically, $F(\xi)$ is no longer a standard logistic distribution function but the standard normal distribution function, i.e $F(\xi) = \Phi(\xi)$, where Φ represents the standard normal CDF. Estimation results are presented in Table 3.27. Parameter estimates, except for the Fechner error term, remain largely unchanged compared to the main specification. Debt aversion is robust.

Additional tremble error: Second, we introduce a second type of error aside from the logit Fechner error of our main specification. In particular, we consider that decision-makers may make a tremble error, i.e. randomize choice between both options with some probability $|\kappa|$ as e.g.

Table 3.27: Aggregate parameter estimates with probit instead Fechner error

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|-------------------------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| point estimate | 0.6411 | 0.0373 | 1.0559 | 1.1099 | 0.7963 |
| 95% confidence interval | 0.58 / 0.71 | 0.03 / 0.05 | 1.03 / 1.08 | 1.09 / 1.13 | 0.66 / 0.93 |
| robust standard error | 0.0327 | 0.0063 | 0.0118 | 0.0117 | 0.0671 |

estimation details: n = 12240, log-likelihood = -4109.29, AIC = 8228.58, BIC = 8265.65, **probit Fechner error**

Robust standard errors (SE) clustered at the individual level, 127 clusters

in Andersson et al., 2020. Consequently, the probability of observing choice B is given by:

$$P^B(\theta') = (1 - |\kappa|)F\left(\frac{U(X^B, p) - U(X^A, p)}{\mu}\right) + \frac{|\kappa|}{2} = (1 - |\kappa|)F(\Delta U(\theta)) + \frac{|\kappa|}{2}, \quad (3.24)$$

where $\theta' = (\alpha, \delta, \gamma, \lambda, \mu, \kappa)$. The corresponding log-likelihood function writes as:

$$\ln L(\theta') = \sum_i \sum_j [\ln(P^B(\theta')) c_{ij} + \ln(1 - P^B(\theta')) (1 - c_{ij})] \quad (3.25)$$

Intuitively, the error parameter can be interpreted as follows: for $|\kappa| \rightarrow 0$ the tremble error has no effect on choices, while for $|\kappa| \rightarrow 1$ choices approach uniform randomization.

Estimation results are presented in Table 3.28. The estimated tremble error probability $|\kappa|$ is statistically indistinguishable from 0, and the remaining parameter estimates are virtually unchanged compared to

Table 3.28: Aggregate parameter estimates with tremble and logit Fechner error

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error | κ tremble error |
|----------------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|---------------------------|
| point estimate | 0.643 | 0.036 | 1.0535 | 1.1074 | 0.4484 | -0.000· |
| 95% CI | 0.58 / 0.71 | 0.02 / 0.05 | 1.03 / 1.08 | 1.08 / 1.13 | 0.37 / 0.53 | -0.0· / 0.0· |
| robust SE | 0.0345 | 0.006 | 0.0112 | 0.0118 | 0.0402 | 0.000· |

estimation details: n = 12240, log-likelihood = -4107.91, AIC = 8225.81, BIC = 8262.88, **logit FE + tremble error**
-0.0· (resp. 0.0·) is between 0 and -0.01 resp. (0.01); -0.000· (resp. 0.000·) is between 0 and -0.0001 resp. (0.0001)

Robust standard errors (SE) clustered at the individual level, 127 clusters

the main specification. Debt aversion also persists when considering an additional tremble error.

Multiple Fechner errors: Third, we consider a specification with distinct Fechner error terms for all domains, as represented through our set of seven different MPLs. Accordingly, we end up with seven Fechner error terms μ_1, \dots, μ_7 for MPL1 to MPL7, respectively. Estimation results are presented in Appendix Table 3.29. While we observe significant variation in the Fechner errors across some MPLs, the remaining parameter estimates are virtually unchanged compared to the main specification. Debt aversion persists.

Table 3.29: Aggregate parameter estimates with multiple Fechner errors

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ_1 FE MPL1 |
|-----------------------|---------------------------|-------------------------|---------------------------|----------------------------|-----------------------|
| point estimate | 0.6553 | 0.0326 | 1.0442 | 1.1029 | 0.2169 |
| 95% CI robust SE | 0.58 / 0.73 0.0379 | 0.02 / 0.04 0.0054 | 1.03 / 1.06 0.0097 | 1.08 / 1.13 0.014 | 0.15 / 0.28 0.0339 |
| | | | | | |
| | μ_2 FE MPL2 | μ_3 FE MPL3 | μ_4 FE MPL4 | μ_5 FE MPL5 | μ_6 FE MPL6 |
| point estimate | 0.4142 | 0.5186 | 0.3886 | 0.3905 | 0.48 |
| 95% CI robust SE | 0.33 / 0.5 0.0433 | 0.42 / 0.61 0.0478 | 0.29 / 0.48 0.0485 | 0.3 / 0.49 0.0485 | 0.35 / 0.61 0.065 |
| | | | | | |
| | μ_7 FE MPL7 | | | | |

estimation details: n = 11430, ln-likelihood = -3676.81, AIC = 7375.62, BIC = 7456.41, distinct errors per MPL

Robust standard errors (SE) clustered at the individual level, 127 clusters

3.D.4 Sample variations

Finally, we check whether the composition of the sample used to estimate preference parameters does have an effect on the estimation. To this end, we scrutinize two variations.

Drop-outs: First, we take into account, that participants who completed the entire experimental sequence of three sessions might be systematically different from those who dropped out along the way. Estimation results presented in Table 3.30, are based on all observations including participants who dropped out prematurely. Parameter estimates do change to some degree compared to estimated parameters of the main modeling specification based on the more restrictive sample of people who completed the entire experimental sequence. However, debt aversion also characterizes this extended population.

Trust and confidence: Second, we consider a sample variation along the dimensions of trust and confidence of participants. As outlined earlier,

Table 3.30: Aggregate parameter estimates including drop-outs

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|-------------------------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| point estimate | 0.6624 | 0.0311 | 1.0499 | 1.1056 | 0.4356 |
| 95% confidence interval | 0.6 / 0.72 | 0.02 / 0.04 | 1.03 / 1.07 | 1.08 / 1.13 | 0.37 / 0.51 |
| robust standard error | 0.0311 | 0.0054 | 0.0099 | 0.0109 | 0.0357 |

estimation details: n = 14310, log-likelihood = -4848.86, AIC = 9707.72, BIC = 9745.56, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 148 clusters

if participants exhibit either mistrust in the payment reliability of the experimenter, or a lack of confidence in their own payment reliability, on average debt contracts would appear more and savings contracts less appealing to them. In this situation, our estimate of debt aversion would be biased downwards. To investigate this possibility, we make use of participants' self-reported ratings on two questions presented at the very end of the experimental sequence at Session 3: "*Back then [in the first session when you made all financial decisions], 1. how sure have you been that the experimenters will make the promised payments in the future in case such a decision has been chosen as the decision that counts? 2. how sure have you been that you will make the promised payments in the future in case such a decision has been chosen as the decision that counts?*" To derive parameter estimates unperturbed by suboptimal trust or confidence, we consider a sample excluding all people who did not answer in the most positive way "*very sure*".

Estimation results are presented in Table 3.31. Albeit our very strict exclusion criteria affects around 50% of participants, parameter estimates remain similar to those of the main specification estimated on the entire sample. Debt aversion persists.

Table 3.31: Aggregate parameter estimates including only participants who perfectly trust the experiment

| | α risk aversion | δ discounting | γ debt aversion | λ loss aversion | μ Fechner error |
|-------------------------|---------------------------|-------------------------|---------------------------|----------------------------|------------------------|
| point estimate | 0.6248 | 0.0379 | 1.0575 | 1.0943 | 0.444 |
| 95% confidence interval | 0.54 / 0.71 | 0.02 / 0.06 | 1.02 / 1.09 | 1.06 / 1.12 | 0.34 / 0.55 |
| robust standard error | 0.0436 | 0.0101 | 0.0178 | 0.015 | 0.0544 |

estimation details: n = 5700, log-likelihood = -1837.52, AIC = 3685.05, BIC = 3718.29, logit Fechner error

Robust standard errors (SE) clustered at the individual level, 60 clusters

3.E Experimental instructions

Upon first arrival at the lab, detailed instructions regarding the experiment as a whole and Session 1, in particular, were given as a printed handout. Identical instructions were displayed on the screen throughout the course of the experiment. Task-specific instructions were displayed on the screen sequentially before the respective tasks in all sessions. After reading the instructions, participants completed the tasks and then received instructions for the following task. Participants were given time to carefully read the instructions and ask questions.

The study design as delineated in the main paper section Experiment and in the instructions was approved by the Ethics Review Committee of Maastricht University (Reference Number: ERCIC_138_07_05_-2019).

3.E.1 Instructions at the beginning of Session 1 (on screen + printout to reread)

3.E.1.1 Overview

As announced in the invitation email, this is a three-part experiment. Today is the first part of the experiment (Session 1). The second part (Session 2) will take place in exactly four weeks from now (*Day, Date*, at the same starting time as today). The third part (Session 3) will take place in exactly eight weeks from now (*Day, Date*, at the same starting time as today). The experiment today will last 90 minutes, Session 2 and Session 3 will last 30 minutes respectively. To participate in today's experiment, you have to be able to participate in all sessions. If you cannot participate at one of these dates, please raise your hand now.

The following will happen during the three sessions:

3.E.1.2 Session 1 (today)

Today you will make a total of 90 (*120 in extension*) financial decisions, involving real money. The choices are simple and not meant to test you - the only correct answers are the ones you really think are best for you.

In the financial decisions you either have the choice between two options (Option A and Option B), or you have the choice of accepting or not accepting a savings or debt contract.

Generally, the financial decisions specify amounts of money that you will receive at different dates with different probabilities, or that you have to pay to the experimenter at different dates. The timing of the payments corresponds to the timing of the sessions. For instance, a financial decision may look as follows:

| Option A | Option B | Your choice |
|-------------------|------------------------|---|
| Receive €18 today | Receive €20 in 4 weeks | <input type="checkbox"/> Option A <input type="checkbox"/> Option B |

In this case, you have the choice between receiving €18 today (i.e. at the end of today's session) or in four weeks, at the end of Session 2. A financial decision may also only involve future dates:

| Option A | Option B | Your choice |
|------------------------|------------------------|---|
| Receive €18 in 4 weeks | Receive €20 in 8 weeks | <input type="checkbox"/> Option A <input type="checkbox"/> Option B |

In this case you choose between monetary amounts to be paid either in four weeks, at the end of Session 2 (Option A), or in eight weeks, at the end of Session 3 (Option B).

At the end of today's session, you will be asked to fill out a short questionnaire. Afterwards, one of the 90 (120 *in extension*) decision situations will be drawn randomly as the 'decision that counts'. Your choice in that decision situation will then actually be implemented, and you will receive or pay the specified monetary amounts depending on your actual decision. Each decision situation has the same chance to be selected as the 'decision that counts'. It is therefore in your interest to consider all decision situations with equal care.

3.E.1.3 Session 2 (In four weeks)

In four weeks, we will ask you to complete a questionnaire. Additionally, all monetary payments that are due at Session 2 will be implemented. Please note that because of the questionnaire you will have to show up at this date, even if you will not receive or pay any monetary amounts at this date.

3.E.1.4 Session 3 (In eight weeks)

In eight weeks, we will ask you to complete a questionnaire. Additionally, all monetary payments that are due at Session 3 will be implemented. Please note that because of the questionnaire you will have to show up at this date, even if you will not receive or pay any monetary amounts at this date.

3.E.1.5 Your payment

The selection of the 'decision that counts' will be made randomly and individually for each participant at the end of today's session. This selection will be made with the help of a bingo cage with 90 (120 *in extension*) numbered balls. All decision situations are numbered, and the number drawn by the bingo cage will be the 'decision that counts'.

This decision will then actually be implemented, and you will receive or pay monetary amounts, as specified in the ‘decision that counts’.

At the beginning of today’s session, you already received a show-up fee of €15 for all three sessions in cash. On top of that money you will receive the money earned from your decisions. Additionally, you will receive a completion bonus of €20 after Session 3, provided you have shown up on time at each session, and have made all payments as agreed (more on this later). This completion bonus will be transferred to your bank account around one week after Session 3. For this payment, we will ask you for your bank details at the end of Session 3.

Please note, should you, arising through your own fault, fail to attend all sessions or fail to make any payments agreed upon you will be excluded from the remaining experiment and all payments associated with it. You will also be removed from the BEElab participant pool and thus not receive any further invitations for economic experiments.

3.E.2 Instructions throughout Session 1 (on screen before respective task + printout to reread)

3.E.2.1 Let’s go

The 90 (120 *in extension*) decision situations are separated into four parts. You will now receive the specific instructions for part 1.

3.E.2.2 Part 1

In this part, you will make a total of 10 decisions. In each decision, you can choose between receiving monetary amounts today, or in one month, at Session 2. For instance, one of these decisions could look like this:

| Option A | Option B | Your choice |
|-------------------|------------------------|---|
| Receive €18 today | Receive €20 in 4 weeks | <input type="checkbox"/> Option A <input type="checkbox"/> Option B |

If you choose Option A in this decision situation, you will receive €18 today. If you choose Option B, you will receive €20 in four weeks, at Session 2.

If you prefer to receive €18 today and nothing in four weeks, choose Option A.

If you prefer to receive €20 in four weeks and nothing today, choose Option B.

Please note that we guarantee the later payment, even if you cannot participate on the due date for any unforeseen reason. In that case, we will transfer the money to your bank account, or you can pick it up at the secretarial office of the department of economics (MPE) at the School of Business and Economics. At the end of today's session, you will receive a receipt containing the email address of the principal investigator, who you can contact should there be any issues with the payment process.

*Subsequently, participants made the 10 decisions of MPL1
(Appendix Table 3.10).*

3.E.2.3 Part 2

In the following part, you will make a total of 20 choices. All payments occur today but depend on the outcome of a coin flip. If a decision situation from this part has been randomly selected as the 'decision that counts', you will make this coin flip yourself after the experiment

today. The coin is fair. There is an equal chance of observing HEADS or TAILS.

For example, you might be asked to choose between the following options:

| Option A | | Option B | | Your choice | |
|---------------------|---------------------|---------------------|---------------------|-----------------------------------|-----------------------------------|
| Coin shows HEADS | Coin shows TAILS | Coin shows HEADS | Coin shows TAILS | | |
| €5 | €4 | €10 | €1 | <input type="checkbox"/> Option A | <input type="checkbox"/> Option B |

In this decision situation, if you choose Option A and the coin shows HEADS, you win €5; if the coin shows TAILS, you win €4. If you choose Option B and the coin shows HEADS, you win €10; if the coin shows TAILS, you win €1.

In some decision situations, one option will be a safe amount and in the other option, the amount depends on a coin flip. For instance, such a decision situation may look as follows:

| Option A | | Option B | | Your choice | |
|----------|--|---------------------|---------------------|-----------------------------------|-----------------------------------|
| safe | | Coin shows HEADS | Coin shows TAILS | | |
| €5 | | €10 | €1 | <input type="checkbox"/> Option A | <input type="checkbox"/> Option B |

In this decision situation, if you choose Option A you receive €5 for sure.

If you choose Option B and the coin shows HEADS, you win €10; if the coin shows TAILS, you win €1.

*Subsequently, participants made the 20 decisions of MPL2 and MPL3
(Appendix Tables 3.11 and 3.12).*

3.E.2.4 Part 3

In the following part, you will make a total of 30 (*45 in extension*) choices. This time, you will be offered a series of real savings contracts, that you can either accept or not accept. Savings contracts involve the payment of some monetary amount by you to the experimenter at an earlier date and the repayment of a monetary amount to you at a later date.

For example, consider the following contract:

| Savings Contract | | Your choice | |
|------------------|------------------------|---------------------------------|-------------------------------------|
| Pay €10 today | Receive €12 in 4 weeks | <input type="checkbox"/> Accept | <input type="checkbox"/> Not Accept |

Under such a contract, you pay the experimenter €10 today, and receive €12 in four weeks, at Session 2. Note that if you have accepted one of these contracts and in case it has been selected as the ‘decision that counts’, you may use your show-up fee of €15, to pay this amount today.

Please note that we guarantee the later payment, even if you cannot participate on the due date for any unforeseen reason. In that case, we will transfer the money to your bank account, or you can pick it up at the secretarial office of the department of economics (MPE) at the School of Business and Economics.

Some savings contracts are defined over dates in the future. This is an example of such a savings contract:

| Savings Contract | Your choice |
|--|---|
| Pay €10 in 4 weeks Receive €15 in 8 weeks | <input type="checkbox"/> Accept <input type="checkbox"/> Not Accept |

Under such a contract, you pay the experimenter €10 in four weeks, at Session 2; and receive €12 in eight weeks, at Session 3. Note that if you have accepted one of these contracts and in case it has been selected as the ‘decision that counts’, you need to bring the specified amount in cash at Session 2. In any case, you will receive a receipt today, specifying what payments you agreed to make at what session. Additionally, we will send a reminder email before the session at which your payment is due, specifying the amount you need to bring to the session.

The receipt you get also contains the email address of the principal investigator, who you can contact should there be any issues with the payment process.

Should you fail to pay the specified amount at the specified Session, you will be excluded from the experiment, and will not receive any further payments, including the completion bonus payment of €20.

Note that you always have the choice to not accept a savings contract! If you do not accept, you won’t pay any money at the earlier date, and won’t receive any money at the later date.

Subsequently, participants made the 30 (45 in extension) decisions of MPL4 and MPL5 (and MPL8 in extension) (Appendix Tables 3.13, 3.14 and 3.17).

3.E.2.5 Part 4

In the following task, you will make a total of 30 (*45 in extension*) choices. This time, you will be offered a series of real debt contracts, that you can either accept or not accept. Debt contracts involve the payment of some monetary amount by the experimenter to you at an earlier date and the repayment of a monetary amount by you to the experimenter at a later date.

For example, consider the following contract:

| Debt Contract | Your choice | |
|-------------------|--------------------|---|
| Receive €10 today | Pay €12 in 4 weeks | <input type="checkbox"/> Accept <input type="checkbox"/> Not Accept |

Under such a contract, the experimenter pays you €10 today, and you have to pay back €12 to the experimenter in four weeks, at Session 2. Note that similar to the Saving Contracts, some debt contracts are defined over dates in the future. Here is an example of such a debt contract:

| Debt Contract | Your choice | |
|------------------------|--------------------|---|
| Receive €10 in 4 weeks | Pay €12 in 8 weeks | <input type="checkbox"/> Accept <input type="checkbox"/> Not Accept |

Under such a contract, the experimenter pays you €10 in four weeks, at Session 2; and you have to pay back €12 to the experimenter in eight weeks, at Session 3.

Please note that should you accept a debt contract, we expect you to repay your debt in full, even if you cannot participate on the due date for any unforeseen reason. If you have accepted one of these contracts and in case it has been selected as the ‘decision that counts’, you need to bring the specified amount in cash at the respective session.¹⁸ In any

¹⁸Alternatively, you may also pay the specified amount via PayPal to the experimenter at the respective session.

case, you will receive a receipt today, specifying what payments you agreed to make at what session. Additionally, we will send a reminder email before the session at which your payment is due, specifying the amount you need to bring to the session.

Should you fail to pay the specified amount at the specified session, you will be excluded from the experiment, and will not receive any further payments, including the completion bonus payment of €20.

Note that you always have the choice to not accept a debt contract! If you do not accept, you won't receive any money at the earlier date, and won't have to pay back any money at the later date.

Subsequently, participants made the 30 (45 in extension) decisions of MPL6 and MPL7 (and MPL9 in extension) (Appendix Tables 3.15, 3.16 and 3.18).

3.E.2.6 Check-out questionnaire

You completed all decisions. Now you still need to fill out a questionnaire and then you are done with today's session.

In this part we ask you to answer some questions and rate some statements about yourself. Some of them you need to classify according to how much they resemble yourself, others need to be ranked according to how much you agree with them or you think society agrees with them, accordingly.

Subsequently, participants provided basic sociodemographic information and answered the collection of 54 items on debt behavior and attitudes as described in detail in Albrecht and Meissner, 2022.

3.E.2.7 Random draw

Please give a sign to the experimenter. The experimenter will then come to you in order to draw the decision that counts from the bingo cage and implement your choice. Afterwards today's payments will be completed and you can leave the lab.

3.E.3 Instructions at session 2 (on screen before respective task)

3.E.3.1 General intro

Today we ask you to solve some logical tasks and answer a set of questions. Additionally, at the end of the session, all monetary payments that are due today will be implemented.

3.E.3.2 CRT and numeracy

You will start by solving 19 logical tasks, afterwards, there will be a questionnaire. Before you start with the logical tasks you will see an example on the next screen.

In each task there will be a short text explaining an issue. Underneath you will find a box where you can type your answer. The answer may be in form of a number or text depending on the task. When appropriate the answer's unit of measurement is already given.

Note, once you typed your answer and hit the continue button you will proceed to the next task and not be able to return.

Subsequently, participants completed tasks on numeracy and cognitive reflection (see 3.B).

3.E.3.3 BFI

You finished all logical tasks. In the next section, we ask you to answer some questions about yourself.

All questions have the same structure: “I am someone who ...” followed by something like “is outgoing and sociable.” and need to be rated according to your level of agreement.

Subsequently, participants completed 30 items of the Big Five Inventory-2-S (see Appendix 3.B).

3.E.3.4 Preference module

There is one more section with questions about yourself to go.

Subsequently, participants completed 12 items of the Preference Survey Module (see Appendix 3.B).

3.E.4 Instructions at Session 3 (on-screen before respective task)

3.E.4.1 General intro

Today we ask you to solve some logical tasks and to answer a set of questions. Additionally, at the end of the session, all monetary payments that are due today will be implemented.

3.E.4.2 Raven

You will start by solving 36 logical tasks, afterwards, there will be a questionnaire. Before you start with the logical tasks you will see an example on the next screen.

In each task there will be a picture on the left side of the screen. In the upper half of the picture, you may see a puzzle with different pieces. Most pieces are shown while the space for the last piece is left blank. You need to choose from the suggestions in the lower half, which piece fits the blank in the puzzle best.

Note, once you typed your answer and hit the continue button you will proceed to the next task and not be able to return.

Subsequently, participants completed 36 Raven matrices (see Appendix 3.B).

3.E.4.3 Planned behavior

In the next section, we ask you to answer some questions about the likelihood that you will make certain purchases in the future and how you will finance them.

Subsequently, participants completed seven items on their likelihood to purchase certain things within the one year and the likelihood to loan-finance these purchases (see Appendix 3.B).

3.E.4.4 Financial literacy

In the next section, we ask you to answer 16 financial questions.

Please note, for your convenience, you may use the Windows built-in calculator. To start the calculator use the calculator button in the bottom right-hand corner of the screen. If you want to, you can try that now. Please note, you can also set the calculator to the scientific mode in case you want to do calculations involving exponents or the like.

Subsequently, participants completed 36 financial literacy items (see Appendix 3.B).

3.E.4.5 Hypothetical debt contracts

There is one more section to go. You will be asked how you would behave in a series of four different hypothetical situations.

Imagine your bank offered you a debt contract. Under this contract, you receive €100 from your bank today and have to pay back some amount in 6 months.

Please assume that in all these choices you must pay the full amount you owe to the bank on time.

Subsequently, participants completed the four-item staircase measure (see Albrecht and Meissner, 2022 for details).

3.E.4.6 Additional questions on honesty and trustworthiness of experimental environment

In the next section, we ask you to answer some questions about yourself and how you think about certain things.

Subsequently, answered eight items from the HEXACO-60 inventory in the honesty domain (see Appendix 3.B) and four further questions on the trustworthiness of the experimental environment.

3.F Contract form



Maastricht University

School of Business and Economics

Dr. Thomas Meissner
Assistant Professor

Department of Economics
Section MPE

direct line
+31 43 3883891

Maastricht
January 2020

Dear

based on today's experimental session you agree and confirm the following contractual details.

1. *Receipt of a show up fee of 15 € today.*
2. *Receipt of € decision-based payoff today.*
3. *Receipt of € decision-based payoff in four weeks on (, . . .2020).*
4. *Receipt of € decision-based payoff in eight weeks on (, . . .2020).*
5. *Receipt of 20 € completion bonus after Session 3, conditional on showing up at each session and making all payments as agreed.*
6. *Payment of € decision-based payment today.*
7. *Payment of € decision-based payment in four weeks on (, . . .2020).*
8. *Payment of € decision-based payment in eight weeks on (, . . .2020).*

For the payment of the completion bonus we will ask you for your bank details after Session 3.

Please also leave your mail address, so that we can send a reminder for future due payments.

Mail:

Date, Signature:



Visiting address
Tongersestraat 53
6211 LM Maastricht

T +31 43 3883891

www.maastrichtuniversity.nl
t.meissner@maastrichtuniversity.nl



Postal address
P.O. Box 616
6200 MD Maastricht
The Netherlands



4

The debt aversion survey module: An experimentally validated tool to measure individual debt aversion¹

Adapted from: David Albrecht and Thomas Meissner (2022). "The debt aversion survey module: An experimentally validated tool to measure individual debt aversion". In: *Working Paper*.

¹Corresponding author: david.albrecht@pm.me. Thomas Meissner acknowledges funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 795958. Replication files can be found via OSF: https://osf.io/xzhw4/?view_only=385a9542dd11478286ca43e79c2f891c

Abstract

We develop an experimentally validated, short and easy-to-use survey module for measuring individual debt aversion. To this end, we first estimate debt aversion on an individual level, using choice data from Meissner and Albrecht, 2022. This data also contains responses to a large set of debt aversion survey items, consisting of existing items from the literature and novel items developed for this study. Out of these, we identify a survey module comprising two qualitative survey items to best predict debt aversion in the incentivized experiment.

Keywords Debt Aversion · Preference Measurement · Survey Validation

JEL Classification C91 · D15 · D91 · G11

4.1 Introduction

In this paper, we develop an experimentally validated, short and easy-to-use survey module for measuring individual debt aversion. Debt aversion has been shown to affect financial decision-making in a variety of different contexts, such as financing (tertiary) education (Field, 2009; Caetano, Palacios, and Patrinos, 2019), private investments in real estate (Schleich, Faure, and Meissner, 2021) or entrepreneurial activity (Nguyen et al., 2020). In Meissner and Albrecht, 2022 we show that debt aversion has a strong impact on borrowing and saving decisions and that it captures a genuine preference that is distinct from other individual preferences for time, risk and losses. However, while debt aversion appears to have a strong influence on financial decision-making, it is notoriously difficult to measure, typically involving complex and resource-intensive experiments (see e.g. Meissner, 2016; Ahrens, Bosch-Rosa, and Meissner, 2022; Meissner and Albrecht, 2022). This complicates the study of debt aversion on a larger scale and impedes the understanding of debt aversion and its effects in representative populations.

In this paper, we develop an experimentally validated, short and easy-to-use survey module for measuring individual debt aversion. Using the choice data and theoretical framework of Meissner and Albrecht, 2022, we structurally estimate parameters of debt aversion on the individual level, using hierarchical maximum likelihood methods. Building on the methodology of the Global Preference Survey - GPS (Falk, A. Becker, Dohmen, Enke, et al., 2018; Falk, A. Becker, Dohmen, Huffman, et al., 2022) we then test the predictive power of a large set of debt aversion survey items, consisting of existing items from the literature and novel items developed for this study. Out of these, we select a smaller subset of survey items, that jointly best predict debt aversion in the incentivized experiment as our debt aversion survey module.

The module that provides the best trade-off between in- and out-of-sample fit, as well as brevity of implementation, consists of two Likert scale survey items, summarized in Table 4.1. This survey module is short, easy to implement and predicts debt aversion reasonably well.

It may therefore facilitate future work on debt aversion, in research where expensive and complex incentivized experiments are not feasible, such as large-scale representative surveys. Further, the survey module might also improve behavioral predictions and identification in research where debt aversion is not the main focus. In such contexts, it enables to easily control for individual preferences towards debt and thus may prevent inference on spurious relationships.

In the following, we first introduce the debt aversion survey module in section 4.2 and delineate its practical implementation in detail. Next, we summarize the design and procedures of the validation study in section 4.3 which provides experimentally measured individual preferences on risk, time and losses as well as debt aversion as the base of the survey module development. Section 4.4 outlines the details of identifying the debt aversion survey module. Finally, in section 4.5 we discuss the strengths and limitations of the identified survey module.

4.2 The debt aversion survey module

Balancing predictive accuracy and brevity of implementation, we propose the debt aversion survey module in Table 4.1. It consists of two qualitative survey items: $Q1$ and $Q2$. Table 4.1 provides the exact wording of questions and the corresponding response categories, which can be directly implemented in any questionnaire as part of an experiment or a survey.²

As in Meissner and Albrecht, 2022, we denote γ as the parameter of debt aversion. $\gamma = 1$ implies debt neutrality, $\gamma > 1$ implies debt aversion and $\gamma < 1$ implies debt affinity.³ To translate responses to the survey module into a quantitative measure of debt aversion, the coefficients in the last column of Table 4.1 can be used as weights to construct γ , as the weighted sum of the module's items and the constant. These weights

²Note that item $Q2$ asks about second-order beliefs and is thus introduced differently than $Q1$.

³For more details on the underlying theory, please see Appendix 4.B.

Table 4.1: Debt Aversion Survey Module

| Question/Item | Response | Coeff. |
|---|---|---------|
| Please rate the following statement: | | |
| <i>Q1</i> <i>Debt is an integral part of today's life.</i> | 1: strongly agree - 6: strongly disagree | 0.0045 |
| What do you think, how does the average participant in this survey/experiment rate the following statement: | | |
| <i>Q2</i> <i>There is no excuse for borrowing money.</i> | 1: strongly agree - 6: strongly disagree | -0.0067 |
| <i>Constant</i> | | 1.0694 |

represent estimated OLS coefficients stemming from the multivariate regression of the debt aversion parameter γ , the dependent variable, on the survey module's set of predictors as independent variables. Likert Scale ratings are coded in the range of 1 to 6 corresponding to answers from "strongly agree" to "strongly disagree". We include a hypothetical example in Appendix 4.A.

The survey module's quality can be quantified by several metrics. With respect to within-sample fit, the survey module and the parameter of debt aversion elicited through the incentivized experiment reach a correlation of 0.3073.⁴ This value is close to the lower end of in-sample correlations of the GPS survey module (Falk, A. Becker, Dohmen, Huffman, et al., 2022), which range from 0.37 to 0.67. Beyond in-sample fit, k-fold cross-validation yields an average mean absolute prediction error (MAE) of 0.0272.⁵ In other words, predicting the individual debt

⁴Preferences over risk, time and losses are not significantly correlated with debt aversion as measured by the survey module.

⁵The MAE is calculated based on 100 repeated random samples with k=5 and k=10 partitions of the data, respectively. The reported figure is the average mean absolute error over the two levels of k .

aversion parameter with the survey module is subject to an average imprecision of 0.0272 points. Evaluated in isolation, these quality metrics might understate the actual power of the survey module. In section 4.5, we set the metrics into perspective by comparing them to suitable benchmarks in particular and by discussing the virtues and limitations of the survey module in general.

The survey module's quality can be quantified by several metrics. With respect to within-sample fit, the survey module and the parameter of debt aversion elicited through the incentivized experiment reach a correlation of 0.3073.⁶ This value is close to the lower end of in-sample correlations of the GPS survey module (Falk, A. Becker, Dohmen, Huffman, et al., 2022), which range from 0.37 to 0.67. Beyond in-sample fit, k-fold cross-validation yields an average mean absolute prediction error (MAE) of 0.0272.⁷ In other words, predicting the individual debt aversion parameter with the survey module is subject to an average imprecision of 0.0272 points. Evaluated in isolation, these quality metrics might understate the actual power of the survey module. In section 4.5, we set the metrics into perspective by comparing them to suitable benchmarks in particular and by discussing the virtues and limitations of the survey module in general.

4.3 Validation study

To construct the debt aversion survey module we make use of the theoretical framework and experimental data from Meissner and Albrecht, 2022. Both taken together enable the quantification of debt aversion on the individual level. A summary of the theory of debt aversion can be found in Appendix 4.B. For a more detailed description of the theory

⁶Preferences over risk, time and losses are not significantly correlated with debt aversion as measured by the survey module.

⁷The MAE is calculated based on 100 repeated random samples with $k=5$ and $k=10$ partitions of the data, respectively. The reported figure is the average mean absolute error over the two levels of k .

and corresponding experiment, we refer the reader to the original paper. We use the experimental choice data to structurally estimate the parameter of debt aversion γ on the individual level. Using the individual estimates of γ , we can construct a debt aversion survey module that optimally predicts γ and thus the degree of debt aversion. The following paragraphs provide a summary of the experiment and detail the procedures to structurally estimate preference parameters on the individual level.

4.3.1 Preference elicitation and estimation

A total of 127 people completed a three-session sequence of experiments in the behavioral and experimental laboratory BEElab at Maastricht University conducted over the years 2019 to 2021. Important for this work, they decided on a total of 90 binary prospects defined over payments at different points in time, lotteries, as well as saving and debt contracts. These choices can be grouped in seven different multiple price lists (MPLs) which are presented in Appendix 4.C.

To construct the debt aversion survey module, we extend the estimation procedure of Meissner and Albrecht, 2022 to additionally estimate preference parameters on the individual level. In Meissner and Albrecht, 2022 we use the entirety of all 90 choices made by all participants to jointly estimate preference parameters of risk, loss and debt aversion as well as time discounting, using maximum likelihood procedures. We estimate preferences on the aggregate level, i.e. for the average decision-maker. Additionally, we estimate moments of the population distributions of preference parameters. Extending these procedures for this study, we employ hierarchical maximum likelihood estimation to retrieve individual preference parameters, loosely following Murphy and Brincke, 2018 and Farrell and Ludwig, 2008. This technique has been established to increase the reliability of individual preference parameter estimates which have to rely on far less data than aggregate estimations. Hierarchical maximum likelihood procedures estimate the set of individual preference parameters that is most likely to produce

the observed choices of the respective person, weighted by the probability of occurrence of such parameter estimates given the population distribution of parameter estimates.

Based on the random utility model outlined in Meissner and Albrecht, 2022, a decision maker with preference parameters $\omega = (\alpha, \beta, \gamma, \lambda)$ chooses option B if $U(X^B, \omega) + \varepsilon^B \geq U(X^A, \omega) + \varepsilon^A$. The probability of observing choice B can then be written as:

$$P^B(\theta) = F\left(\frac{U(X^B, \omega) - U(X^A, \omega)}{\mu}\right) = F(\Delta U(\theta)), \quad (4.1)$$

where F is the cumulative distribution function of $(\varepsilon^A - \varepsilon^B)$. We assume $(\varepsilon^A - \varepsilon^B)$ to follow a standard logistic distribution with distribution function $F(\xi) = (1 + e^{-\xi})^{-1}$, corresponding to the often termed Luce model (Luce, Suppes, et al., 1965; Holt and Laury, 2002) or Fechner error with logit link (Drichoutis and Lusk, 2014). $F(\Delta U(\theta))$ depends on the preference parameters for risk aversion, time discounting, debt aversion and loss aversion: α, δ, γ and λ , respectively and the error parameter μ . All five parameters, summarized by the vector $\theta = (\alpha, \beta, \gamma, \lambda, \mu)$, are identified on the individual level through hierarchical maximum likelihood estimation based on observed choices and the population distribution of preference parameters.⁸ The population distributions have already been retrieved in the first hierarchy level of estimation in Meissner and Albrecht, 2022. Thus, we here consider them as given by the normal density functions $d(\alpha), d(\delta), d(\gamma)$ and $d(\lambda)$. We denote the product of these normal densities as $d(\omega) = d(\alpha)d(\delta)d(\gamma)d(\lambda)$. The hierarchical likelihood function on the individual level then writes as

⁸In line with Murphy and Brincke, 2018 we do not consider the population distribution of the error parameter μ for hierarchical estimation, as it cannot be estimated and interpreted independently of other preference parameters.

$$\ln(L(\theta)) = \sum_j [\ln(F(\Delta U(\theta))w(d(\omega)))c_j + \ln(1 - F(\Delta U(\theta))w(d(\omega)))(1 - c_j)], \quad (4.2)$$

where $c_j = 0$ if Option A was chosen in choice j , $c_j = 1$ if Option B was chosen in choice j . Further, $w(d(\omega))$ denotes the product of preference parameters' density functions, given their distribution in the population, weighted by function $w(\cdot)$.

We introduce w to prevent a potential pitfall of hierarchical estimation. Scheibehenne and Pachur, 2015 argue that the relative influence of population distributions on individual preference parameters, as opposed to the influence of individual choices, may be too large. In other words, hierarchical estimation may produce excessive shrinkage of individual parameter estimates toward their population mean.⁹ Following a suggestion by Murphy and Brincke, 2018 we mitigate this effect by weighing $d(\omega)$. In particular, we introduce the weighting function

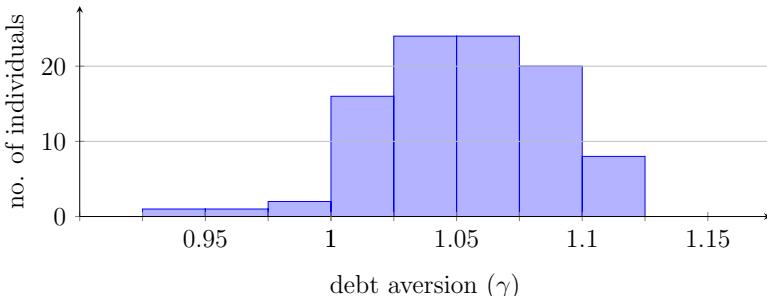
$$w(d(\omega)) = d(\omega)^s, \quad (4.3)$$

where s is a shrinkage factor determining the relative weight given to population distributions as opposed to individual choices when estimating individual preference parameters.¹⁰ If $s = 1$, then $w(d(\omega)) = d(\omega)$, for $s < 1$ the influence of population distributions on estimates decreases, and for $s > 1$ the influence increases.

⁹We find such excessive shrinkage to likely also apply to our setting, based on estimations for simulated decision makers, with hypothetical, known preference parameters.

¹⁰We employ the exponential form, as it preserves the range of $d(\omega)$, i.e. both $d(\omega)$ and $w(d(\omega))$ lie in the range of probabilities $[0, 1]$. Further, the exponential form introduces only one additional parameter to be estimated.

Figure 4.1: Individual level estimates of the debt aversion parameter γ .



We estimate the optimal shrinkage factor s based on simulated, hypothetical individuals with known preference parameters. To this end, we apply the outlined hierarchical ML procedure to estimate preference parameters from simulated choices. By comparing the estimated parameters to true, assumed parameters we can quantify the goodness-of-fit. Using grid-search, we find $s = 0.0139$ to minimize mean squared error (MSE) of estimated and true individual parameters of debt aversion γ .

By maximizing the hierarchical log-likelihood function (Equation 4.2) based on choice data and population distributions from Meissner and Albrecht, 2022 we derive point estimates for all preference parameters and the error parameter per individual.¹¹ In this way, we retrieve individual preference parameters, including γ , the coefficient of debt aversion, for 96 participants. Figure 4.1 summarizes the distribution of γ . Individual parameter estimates range from $\gamma = 0.9468$, corresponding

¹¹To increase the reliability of estimates we ensure that the same estimates can be retrieved using STATA's modified Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm and through maximization using the Nelder-Mead algorithm implemented in R. Individuals, whose parameters cannot be estimated consistently, i.e. estimates from R are more than 10% higher or lower than estimates from STATA, are discarded. For consistent estimates, we choose results from either STATA or R depending on which routine yielded a higher likelihood score for the given individual's parameter estimates.

to debt affinity, to $\gamma = 1.1171$, corresponding to debt aversion. Around, 96% of individuals are estimated as debt averse with a parameter $\gamma > 1$, the median level of debt aversion is $\gamma = 1.0523$.

4.3.2 Included survey items

Besides incentivized choices, we collected a battery of self-reported, qualitative measures of debt behavior and attitudes in Meissner and Albrecht, 2022. For this purpose, we composed a comprehensive set of 54 survey items, roughly spanning the clusters experience and usage of borrowing, appropriateness to be indebted, rules, norms and personal preferences on debt. The items are collected from previous studies using questionnaire-based measures of debt attitudes but also include novel items developed for this study. Items range from directly asking respondents to state “how much they (dis)like being in debt in general” to more indirect survey items around the topic of debt and money. The complete list of items can be found in Appendix, Table 4.D.1. In addition to the set of survey items, we implemented a hypothetical, multiple price list consisting of 15 debt contracts. To minimize the required response time, the hypothetical debt contracts were presented using the staircase method (Cornsweet, 1962). This mode of presentation allows to identify the switchpoint between accepting and not accepting each of the 15 contracts by only asking four successive questions. for detailed instructions see Appendix 4.D.2. We count the switchpoint in the hypothetical choice task as one additional quantitative survey item.

4.4 Identification of the debt aversion survey module

To identify an experimentally validated survey module for debt aversion we follow the procedures established for the GPS (Falk, A. Becker, Dohmen, Huffman, et al., 2022). To construct the survey module, we first consider the entirety of collected survey items, and all possible combinations thereof. We then identify the subset of items that yields

the best combination of predictive accuracy for individual debt aversion and brevity. We begin by considering a baseline pool of 55 potential best predictors including qualitative items collected from existing literature on debt attitudes, novel items developed for this study and the switchpoint in the hypothetical debt contract choice task. By stepwise reduction, we condense the amount of predictors to a debt aversion survey module of two items.

In the first step, we discard thirteen items that appear improper to construct an easy and widely applicable survey module. Six items have a specific focus on (debt) financing tertiary education, which makes them inappropriate for use beyond the university context. For further five items, we could not identify an intuitive directional hypothesis on the relation of the item and debt aversion. Lastly, two items exhibit a correlation with debt aversion, which goes against the direction of an intuitive hypothesis. A potential reason for this could be a misunderstanding of these items due to double negative wording.

As a second step, we consider a multitude of linear regressions, modelling the experimentally elicited debt aversion parameter γ per individual i as the dependent variable, and all possible subsets of the remaining 42 items as independent variables. In other words, we scrutinize all possible combinations of one item, two items, ..., n items as independent variables in a standard linear model, see Equation 4.4, and estimate the regression parameters β_0, \dots, β_p using ordinary least squares (OLS).

$$\gamma_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (4.4)$$

The number of potential combinations, increases rapidly in the number of considered items, and quickly becomes intractable with conventional computational resources.¹²

¹²42 combinations of a single predictor, i.e. each and every item itself. 861 combinations of two predictors, 11 480 combinations of three predictors, 111 930 combinations of four predictors, and so on, with a maximum of 538 257 874 440 combinations of 21

We therefore create a shortlist of predictor variables that appear in any of the ten best performing models, evaluated according to adjusted R^2 for models with one, two, three up to six predictors, respectively.¹³ This shortlist contains 16 items. Using the shortlist we run linear OLS regressions on all possible combinations of predictors, i.e. 524 288 regressions. In line with the GPS module (Falk, A. Becker, Dohmen, Huffman, et al., 2022), we use adjusted R^2 , a criterion of in-sample-fit, to identify the best subset for each number of predictors.

Third, to discriminate between models comprising different numbers of variables we additionally consider information criteria and estimates of out-of-sample predictive power based on cross-validation. We consider the Akaike Information Criterion (AIC) as introduced by Akaike, 1974 and the Bayesian or Bayes-Schwarz Information Criterion (BIC) as introduced by Schwarz, 1978. With respect to cross-validation, we implement k-fold cross-validations (Stone, 1974) splitting our sample into $k = 5$ and $k = 10$ data chunks with 100 random samples each to calculate mean squared prediction errors (MSE) for the parameter of debt aversion γ of the candidate models. The performance of the five best candidate models per number of predictors are summarized in Figure 4.2.

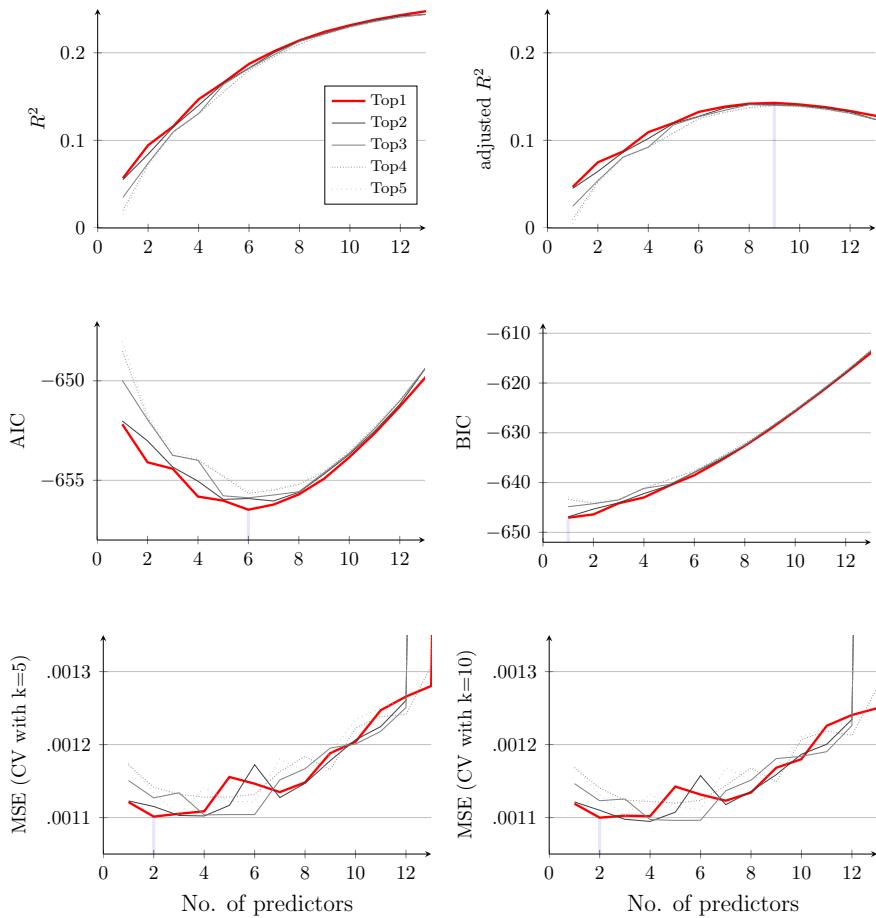
Bringing together the variety of performance measures calculated in the previous step, we identify a survey module containing two items as providing a good trade-off between brevity and predictive power. Figure 4.2 shows that adjusted R^2 is maximized for a model with nine predictors, while AIC favors a six and BIC a one predictor model. The discrepancy in terms of favored models does not come as a surprise, as the three performance criteria differ in the penalty they put on including additional predictors into the model. Adjusted R^2 incorporates the

predictors. As subsets/combinations of predictors are sampled disregarding the order of elements and without the possibility to include the same element more than once the number of combinations reaches a maximum for models including 21 variables. For larger subsets, the number of potential combinations decreases again.

There is only one subset containing all predictors from the pool.

¹³For the creation of this shortlist we run a total of 36 211 980 regressions.

Figure 4.2: Performance metrics of the candidate survey modules



Notes: **First row:** coefficients of determination; **Second row:** information criteria; **Third row:** mean squared error (MSE) derived through k-fold cross validation (CV) with $k = 5$ and $k = 10$ data chunks and 100 repetitions each

smallest and BIC the largest penalty. To complement in-sample fit based performance metrics, we scrutinize the cross-validation-based

MSE of the candidate models. Successively including more predictors, MSEs for $k = 5$ and $k = 10$ decrease for the Top1 models with up to two predictors, remain relatively flat up to four predictors and increases for more than four predictors. These favored numbers of predictors are largely corroborated by considering not only the best model (Top1) but also its close competitors (Top2 to Top5).

Just as in the GPS module (Falk, A. Becker, Dohmen, Huffman, et al., 2022) we value brevity of the survey module and thus favor the BIC. However, we do not want to fully disregard adjusted R^2 and AIC, as the discrepancy of suggested predictors is rather large. We reconcile the number of predictors by following the best model according to cross validation, comprising two predictors. This appears to be a reasonable compromise between BIC and AIC, which still puts a strong weight on brevity.

4.5 Discussion and conclusion

We propose a two-question survey module, that allows measuring debt aversion in large-scale surveys and experimental questionnaires with minimal effort and expenditure of time. The module is validated by use of an incentivized experiment, thus lending credibility to the resulting measure. In the following, we discuss the performance of the debt aversion survey module according to different metrics and in comparison to reasonable benchmarks. This way we provide guidance for researchers who face the trade-off between eliciting debt aversion via the short, but potentially less accurate survey module as opposed to incentivized lab experiments and structural estimations, which are expensive and involve considerable time and effort.¹⁴

As a starting point, we consider the in-sample fit of the debt aversion survey module as indicator of quality. The module reaches an R^2 of 0.0945, i.e. it accounts for around 10% of the variation of the parameter

¹⁴In Meissner and Albrecht, 2022, participants had to come to the lab on a total of three dates, with a total time of around 2.5 hours.

of debt aversion (from the incentivized experiment). This corresponds to a correlation of $\rho = 0.3073$. As discussed for the GPS module (Falk, A. Becker, Dohmen, Huffman, et al., 2022), this correlation cannot be evaluated against a benchmark of 1, unless measurement of preferences (both in the structural estimations and in the response to the survey items) was without measurement error.

To evaluate measurement error, repeated measurement on the same participants would be helpful. While this was not done by Meissner and Albrecht, 2022, the validation study for the GPS module finds a test-retest correlation of 0.5753 for risk aversion, and 0.8149 for time discounting when comparing their preference measure in one experiment to the same measure from an identical repeated experiment with the same participants (Falk, A. Becker, Dohmen, Huffman, et al., 2022). Evaluated against such a benchmark, our survey module appears to exhibit reasonable explanatory power for measuring debt aversion.

Beyond in-sample fit, the debt aversion survey module proves to be capable in terms of predictive power. K-fold cross validation yields a mean absolute prediction error (based on 100 repeated random samples) of 0.0272 for $k=5$ and $k=10$. In other words, predicting γ using the debt aversion survey module suffers an average error of ca. 0.0272.¹⁵ To set this into perspective we consider predicted choices for participants in the experiment of Meissner and Albrecht, 2022 in comparison to their actual choices. Specifically, we consider predictions based on the debt aversion parameter from the survey module competing against predictions based on the structurally estimated debt aversion parameter as a benchmark. Survey module based predictions are accurate in 89.76%, while structural estimation based predictions are accurate in 91.48% of the decision situations in the experiment. Thus, the survey module performs similarly well in predicting choices compared to the structurally estimated parameter of debt aversion.

Comparing our results to the module of the GPS (Falk, A. Becker, Dohmen, Enke, et al., 2018; Falk, A. Becker, Dohmen, Huffman, et al.,

¹⁵Calculated as the mean absolute prediction error over the two different ks .

2022), we see the debt aversion survey module as a useful addition to measure debt aversion alongside other key economic preferences in survey style investigations. A notable difference to the GPS module is that we do not identify the hypothetical version of incentivized choices, in our case the hypothetical debt contracts, as part of the set of the best-predictors of preferences. We see a likely explanation for this in the interdependence of debt aversion with other preference dimensions such as risk and loss aversion as well as discounting over time. As illustrated in Meissner and Albrecht, 2022, the switch point in a multiple price list consisting of debt contracts will not only depend on individual debt aversion, but potentially also on time preferences, risk preferences and loss aversion.

Summing up, we develop a short and easy-to-use experimentally validated survey module to measure debt aversion. We hope that our survey module will facilitate future research on debt aversion on a larger scale, where complex and incentivized experiments are often not feasible.

Appendix

4.A Calculation example

The following illustrates how answers to the debt aversion survey module are translated into $\hat{\gamma}$ the predicted structural estimate of debt aversion.

Figure 4.3: Debt aversion survey module answered by a hypothetical respondent

Please rate the following statement.

Debt is an integral part of today's life.

strongly agree strongly disagree

—

What do you think how does the average participant in this survey/experiment rate the following statement?

There is no excuse for borrowing money.

strongly agree strongly disagree

Let's consider the responses to the debt aversion survey module by a hypothetical person as depicted in Figure 4.3. In this particular case, the predicted debt aversion parameter, $\hat{\gamma}$, can be calculated as

$$\begin{aligned}\hat{\gamma} &= 1.0694 + \\ &\quad 0.0045 \times (5) - 0.0067 \times (2) \\ &= 1.0785\end{aligned}$$

where the numbers in brackets are the answers given to the survey items coded into numbers 1-6 from the left-most to the right-most answer increment, respectively.

We recommend implementing the survey module with the presented 6-point Likert scales, as this format is underlying all the validation procedures. However, should there be reasons to use Likert scales with more or less increments, the predicted debt aversion parameter $\hat{\gamma}$, can be calculated from Likert scales with minimum l and maximum h by transforming the answers x to the 6-point scale. To this end, one needs to replace the numbers in brackets by:

$$\left(\frac{x - l}{h - l} \times 5 + 1 \right) \quad (4.5)$$

4.B Theoretical framework

In Meissner and Albrecht, 2022, agents choose between intertemporal prospects that are defined over streams of monetary gains or losses in up to two periods. $\mathbf{x} = (x_t, x_T)$ denotes a stream of payments that offers x_t at time t , and x_T at time T , where $0 \leq t < T$. $X = (x_1, p_1; x_2, p_2; \dots; x_N, p_N)$ denotes an intertemporal prospect, that gives the payment stream \mathbf{x}_n with probability p_n . Intertemporal utility is written as:

$$U(X) = \mathbb{E} [\phi(t)v(x_t) + \phi(T)v(x_T) - \mathbb{1}_{debt}c(\mathbf{x})]$$

where $v(x_t)$ denotes atemporal utility of monetary gains and losses at time t . Agents discount future gains and losses with the discount function ϕ .

Saving contracts are payment streams characterized by $x_t < 0$ and $x_T > 0$. Inversely, *debt contracts* are payment streams characterized by $x_t > 0$ and $x_T < 0$. Cost of being in debt $c(\mathbf{x})$ is only incurred for debt contracts:

$$\mathbb{1}_{debt} = \begin{cases} 1 & \text{if } x_t > 0 \text{ and } x_T < 0 \\ 0 & \text{otherwise.} \end{cases}$$

The specific functional forms used to estimate preference parameters are identical to Meissner and Albrecht, 2022. Gains and losses of money are evaluated relative to a reference point ($x = 0$):

$$v(x) = \begin{cases} u(x) & \text{if } x \geq 0 \\ -\lambda u(-x) & \text{if } x < 0, \end{cases} \quad (4.6)$$

Utility in gains and losses is given by:

$$u(x) = \frac{(x + \varepsilon)^{1-\alpha} - \varepsilon^{1-\alpha}}{1 - \alpha} \quad (4.7)$$

The discount function is exponential:

$$\phi(\tau) = \frac{1}{(1 + \delta)^\tau} \quad (4.8)$$

Finally, the cost of being in debt is modelled as:

$$c(\mathbf{x}) = (1 - \gamma)\phi(T)v(x_T) \quad (4.9)$$

Here, γ is the parameter of debt aversion. A parameter of $\gamma = 1$ implies debt neutrality, $\gamma > 1$ implies debt aversion and $\gamma < 1$ implies debt affinity.

4.C Multiple price lists

Table 4.2: Multiple price list of intertemporal choices (MPL1)

| Choice | Option A | Option B |
|--------|----------------------------------|---------------------------------------|
| 1 | Receive an amount of €18.2 today | Receive an amount of €18.0 in 4 weeks |
| 2 | Receive an amount of €18.0 today | Receive an amount of €18.0 in 4 weeks |
| 3 | Receive an amount of €17.8 today | Receive an amount of €18.0 in 4 weeks |
| 4 | Receive an amount of €17.3 today | Receive an amount of €18.0 in 4 weeks |
| 5 | Receive an amount of €16.8 today | Receive an amount of €18.0 in 4 weeks |
| 6 | Receive an amount of €16.0 today | Receive an amount of €18.0 in 4 weeks |
| 7 | Receive an amount of €14.0 today | Receive an amount of €18.0 in 4 weeks |
| 8 | Receive an amount of €12.0 today | Receive an amount of €18.0 in 4 weeks |
| 9 | Receive an amount of €10.0 today | Receive an amount of €18.0 in 4 weeks |
| 10 | Receive an amount of €8.0 today | Receive an amount of €18.0 in 4 weeks |

Table 4.3: Multiple price list of certain payments vs. risky gambles (MPL2)

| Choice | Option A | | Option B | |
|--------|---------------------|---------------------|---------------------|---------------------|
| | Coin shows Heads | Coin shows Tails | Coin shows Heads | Coin shows Tails |
| 1 | €30 today | €30 today | €30 today | €1 today |
| 2 | €25 today | €25 today | €30 today | €1 today |
| 3 | €20 today | €20 today | €30 today | €1 today |
| 4 | €17 today | €17 today | €30 today | €1 today |
| 5 | €16 today | €16 today | €30 today | €1 today |
| 6 | €15 today | €15 today | €30 today | €1 today |
| 7 | €12 today | €12 today | €30 today | €1 today |
| 8 | €10 today | €10 today | €30 today | €1 today |
| 9 | €5 today | €5 today | €30 today | €1 today |
| 10 | €1 today | €1 today | €30 today | €1 today |

Table 4.4: Multiple price list of less risky vs. more risky gambles (MPL3)

| Choice | Option A | | Option B | |
|--------|---------------------|---------------------|---------------------|---------------------|
| | Coin shows Heads | Coin shows Tails | Coin shows Heads | Coin shows Tails |
| 1 | €14 today | €17 today | €17 today | €1 today |
| 2 | €14 today | €17 today | €20 today | €1 today |
| 3 | €14 today | €17 today | €25 today | €1 today |
| 4 | €14 today | €17 today | €28 today | €1 today |
| 5 | €14 today | €17 today | €29 today | €1 today |
| 6 | €14 today | €17 today | €30 today | €2 today |
| 7 | €14 today | €17 today | €30 today | €3 today |
| 8 | €14 today | €17 today | €32 today | €8 today |
| 9 | €14 today | €17 today | €32 today | €10 today |
| 10 | €14 today | €17 today | €32 today | €14 today |

Table 4.5: Multiple price list of 4-week saving contracts starting at Session 1 (MPL4)

| Choice | Early saving contracts | |
|--------|------------------------|---------------------------|
| | Session 1 (today) | Session 2 (in 4 weeks) |
| 1 | Pay an amount of €15 | Receive an amount of €45 |
| 2 | Pay an amount of €15 | Receive an amount of €40 |
| 3 | Pay an amount of €15 | Receive an amount of €36 |
| 4 | Pay an amount of €15 | Receive an amount of €34 |
| 5 | Pay an amount of €15 | Receive an amount of €32 |
| 6 | Pay an amount of €15 | Receive an amount of €30 |
| 7 | Pay an amount of €15 | Receive an amount of €28 |
| 8 | Pay an amount of €15 | Receive an amount of €26 |
| 9 | Pay an amount of €15 | Receive an amount of €24 |
| 10 | Pay an amount of €15 | Receive an amount of €22 |
| 11 | Pay an amount of €15 | Receive an amount of €20 |
| 12 | Pay an amount of €15 | Receive an amount of €18 |
| 13 | Pay an amount of €15 | Receive an amount of €16 |
| 14 | Pay an amount of €15 | Receive an amount of €14 |
| 15 | Pay an amount of €15 | Receive an amount of €12 |

Table 4.6: Multiple price list of 4-week saving contracts starting at Session 2 (MPL5)

| Choice | Late saving contracts | |
|--------|---------------------------|---------------------------|
| | Session 2 (in 4 weeks) | Session 3 (in 8 weeks) |
| 1 | Pay an amount of €15 | Receive an amount of €40 |
| 2 | Pay an amount of €15 | Receive an amount of €35 |
| 3 | Pay an amount of €15 | Receive an amount of €31 |
| 4 | Pay an amount of €15 | Receive an amount of €29 |
| 5 | Pay an amount of €15 | Receive an amount of €27 |
| 6 | Pay an amount of €15 | Receive an amount of €25 |
| 7 | Pay an amount of €15 | Receive an amount of €23 |
| 8 | Pay an amount of €15 | Receive an amount of €21 |
| 9 | Pay an amount of €15 | Receive an amount of €19 |
| 10 | Pay an amount of €15 | Receive an amount of €17 |
| 11 | Pay an amount of €15 | Receive an amount of €15 |
| 12 | Pay an amount of €15 | Receive an amount of €13 |
| 13 | Pay an amount of €15 | Receive an amount of €11 |
| 14 | Pay an amount of €15 | Receive an amount of €9 |
| 15 | Pay an amount of €15 | Receive an amount of €7 |

Table 4.7: Multiple price list of 4-week debt contracts starting at Session 1 (MPL6)

| Choice | Early debt contracts | |
|--------|--------------------------|---------------------------|
| | Session 1 (today) | Session 2 (in 4 weeks) |
| 1 | Receive an amount of €31 | Pay an amount of €15 |
| 2 | Receive an amount of €27 | Pay an amount of €15 |
| 3 | Receive an amount of €24 | Pay an amount of €15 |
| 4 | Receive an amount of €21 | Pay an amount of €15 |
| 5 | Receive an amount of €19 | Pay an amount of €15 |
| 6 | Receive an amount of €17 | Pay an amount of €15 |
| 7 | Receive an amount of €16 | Pay an amount of €15 |
| 8 | Receive an amount of €15 | Pay an amount of €15 |
| 9 | Receive an amount of €14 | Pay an amount of €15 |
| 10 | Receive an amount of €13 | Pay an amount of €15 |
| 11 | Receive an amount of €11 | Pay an amount of €15 |
| 12 | Receive an amount of €9 | Pay an amount of €15 |
| 13 | Receive an amount of €7 | Pay an amount of €15 |
| 14 | Receive an amount of €5 | Pay an amount of €15 |
| 15 | Receive an amount of €3 | Pay an amount of €15 |

Table 4.8: Multiple price list of 4-week debt contracts starting at Session 2 (MPL7)

| Choice | Late debt contracts | |
|--------|---------------------------|---------------------------|
| | Session 2 (in 4 weeks) | Session 3 (in 8 weeks) |
| 1 | Receive an amount of €33 | Pay an amount of €15 |
| 2 | Receive an amount of €30 | Pay an amount of €15 |
| 3 | Receive an amount of €27 | Pay an amount of €15 |
| 4 | Receive an amount of €24 | Pay an amount of €15 |
| 5 | Receive an amount of €22 | Pay an amount of €15 |
| 6 | Receive an amount of €20 | Pay an amount of €15 |
| 7 | Receive an amount of €18 | Pay an amount of €15 |
| 8 | Receive an amount of €16 | Pay an amount of €15 |
| 9 | Receive an amount of €15 | Pay an amount of €15 |
| 10 | Receive an amount of €14 | Pay an amount of €15 |
| 11 | Receive an amount of €12 | Pay an amount of €15 |
| 12 | Receive an amount of €10 | Pay an amount of €15 |
| 13 | Receive an amount of €8 | Pay an amount of €15 |
| 14 | Receive an amount of €6 | Pay an amount of €15 |
| 15 | Receive an amount of €3 | Pay an amount of €15 |

4.D Debt aversion survey module

4.D.1 Pool of debt survey items

Table 4.9: Pool of Debt Survey Items

| No. | Survey Item | Scale | Reference |
|--|--|-------------------------|-------------------------|
| <i>Usage</i> | | | |
| 1 | Did you ever use overdraft on your bank account? | yes/no | - |
| 2 | Do you use credit cards? | yes/no | (Eckel et al., 2016) |
| 3 | In total, how many credit cards with different accounts do you use? | categorical 0 to > 5 | - |
| 4 | If you have a credit card balance. Do you usually pay it off each month? | yes/no | (Eckel et al., 2016) |
| 5 | How would you categorize your access to loans/credits/capital? | Likert | - |
| 6 | Did you ever take out a loan at a bank? | yes/no | - |
| 7 | Do you owe money in student loans? | yes/no | (Eckel et al., 2016) |
| 8 | In total, what is your best guess of your outstanding debt as of today in €? (including informal loans, family, friends, etc.) | integer | - |
| 9 | How certain are you about your guess on your overall outstanding debt? | Likert | - |
| 10 | Does your current level of debt burden you? | Likert | (Eckel et al., 2016) |
| 11 | In total, what is your best guess of your savings as of today in €? | integer | - |
| 12 | How certain are you about your guess on your overall savings? | Likert | - |
| <i>Appropriateness: "Please rate the following statements"</i> | | | |
| 13 | It is okay to accrue debt for living the style you desire. | Likert | (Chudry et al., 2011) |
| 14 | It is okay to be in debt if you know you can pay it off. | Likert | (Haultain et al., 2010) |

(Continued on next page)

Chapter 4. The debt aversion survey module: An experimentally validated tool to measure individual debt aversion

| No. | Survey Item | Scale | Reference |
|-----|---|--------|---|
| 15 | It is ok to borrow money to pay for necessities (e.g. food, rent, utilities). | Likert | |
| 16 | It is ok to borrow money to pay for essential purchases (e.g. car, housing, appliances). | Likert | adapted based on: (Davies and Lea, 1995; Haultain et al., 2010; |
| 17 | It is ok to borrow money to finance investments (e.g. tertiary education, starting a business, solar panels). | Likert | Harrison and Agnew, 2016; George et al., 2018) |
| 18 | It is ok to borrow money to pay for luxuries (e.g. expensive holiday, status symbols). | Likert | |
| 19 | Students should take the maximum permissible student debt (loans/overdraft, etc.). | Likert | (Chudry et al., 2011) |
| 20 | Debt is an integral part of today's life. | Likert | (Davies and Lea, 1995; Haultain et al., 2010; Chudry et al., 2011) |
| 21 | Reducing/controlling debt leads to a better quality of life. | Likert | (Chudry et al., 2011) |
| 22 | Reducing/controlling debt leads to greater success. | Likert | (Chudry et al., 2011) |
| 23 | Reducing/controlling debt leads to feeling a sense of achievement. | Likert | (Chudry et al., 2011) |
| 24 | Reducing/controlling debt leads to a feeling that you are fitting in with friends. | Likert | (Chudry et al., 2011) |
| 25 | Reducing/controlling debt leads to being perceived as boring. | Likert | (Chudry et al., 2011) |
| 26 | Reducing/controlling debt leads to being perceived as tight. | Likert | (Chudry et al., 2011) |
| 27 | Reducing/controlling debt leads to enjoying yourself less. | Likert | (Chudry et al., 2011) |
| 28 | Once you are in debt it is very difficult to get out of it. | Likert | (Davies and Lea, 1995; Haultain et al., 2010) |
| 29 | Owing money is basically wrong. | Likert | (Haultain et al., 2010; Boatman et al., 2017) |
| 30 | You should always save up first before buying something. | Likert | (Davies and Lea, 1995; Haultain et al., 2010; Boatman et al., 2017) |

| No. | Survey Item | Scale | Reference |
|---|---|--------|---|
| 31 | There is no excuse for borrowing money. | Likert | (Davies and Lea, 1995; Haultain et al., 2010; Boatman et al., 2017) |
| 32 | Borrowing money for tertiary education is a good investment. | Likert | (Haultain et al., 2010) |
| 33 | You should rather restrict your lifestyle than go into debt. | Likert | - |
| <i>Personality: "Please rate the following statements"</i> | | | |
| 34 | I like to pay my debts as soon as possible. | Likert | (Walters et al., 2016) |
| 35 | I prefer to delay paying my debts if possible, even if it means paying more in total. | Likert | (Walters et al., 2016) |
| 36 | Having debts makes me feel uncomfortable. | Likert | (Walters et al., 2016) |
| 37 | Having debt doesn't bother me. | Likert | (Walters et al., 2016) |
| 38 | I dislike borrowing money. | Likert | (Schleich et al., 2021) |
| 39 | I feel OK borrowing money for 'essential' purchases e.g. cars, appliances, mortgage. | Likert | (Schleich et al., 2021) |
| 40 | I enjoy being able to borrow money to buy things I like, and to pay for things I cannot afford. | Likert | (Schleich et al., 2021) |
| 41 | I would rather be in debt than change my lifestyle. | Likert | (Haultain et al., 2010) |
| 42 | If I had to make an unexpected expenditure today of 500 € or more, I would use a credit card/borrow from a financial institution, family or friends. | Likert | (Eckel et al., 2016) |
| 43 | If I had to make an unexpected expenditure today of 5000 € or more, I would use a credit card/borrow from a financial institution, family or friends. | Likert | (Eckel et al., 2016) |
| 44 | I like saving money. | Likert | - |
| <i>Norms, i.e. second order beliefs (Krupka and Weber, 2013)¹³</i> | | | |
| 45 | It is okay to accrue debt for living the style you desire. | Likert | (Chudry et al., 2011) |
| 46 | Students should take the maximum permissible student debt (loans/overdraft, etc.). | Likert | (Chudry et al., 2011) |

(Continued on next page)

| No. | Survey Item | Scale | Reference |
|-----|--|--------|---|
| 47 | Debt is an integral part of today's life. | Likert | (Davies and Lea, 1995; Haultain et al., 2010; Chudry et al., 2011) |
| 48 | Once you are in debt it is very difficult to get out of it. | Likert | (Davies and Lea, 1995; Haultain et al., 2010) |
| 49 | Owing money is basically wrong. | Likert | (Haultain et al., 2010; Boatman et al., 2017) |
| 50 | You should always save up first before buying something. | Likert | (Davies and Lea, 1995; Haultain et al., 2010; Boatman et al., 2017) |
| 51 | There is no excuse for borrowing money. | Likert | (Davies and Lea, 1995; Haultain et al., 2010; Boatman et al., 2017) |
| 52 | It is okay to be in debt if you know you can pay it off. | Likert | (Haultain et al., 2010) |
| 53 | Borrowing money for tertiary education is a good investment. | Likert | (Haultain et al., 2010) |
| 54 | You should rather restrict your lifestyle than go into debt. | Likert | - |

Notes: All Likert scales follow a 6-point format, items without reference were created by the authors.

4.D.2 Hypothetical debt contracts

The multiple price list on hypothetical debt contracts contains a total of 15 decisions. Aiming at brevity of the debt aversion survey module, its implementation and validity has been tested using the staircase method. Thus respondents effectively only see and make four yes/no choices. Participants are asked whether they would, hypothetically, accept four financial debt contracts, under which they receive 100€ today, with the obligation to repay between 60 to 140€ in six months:

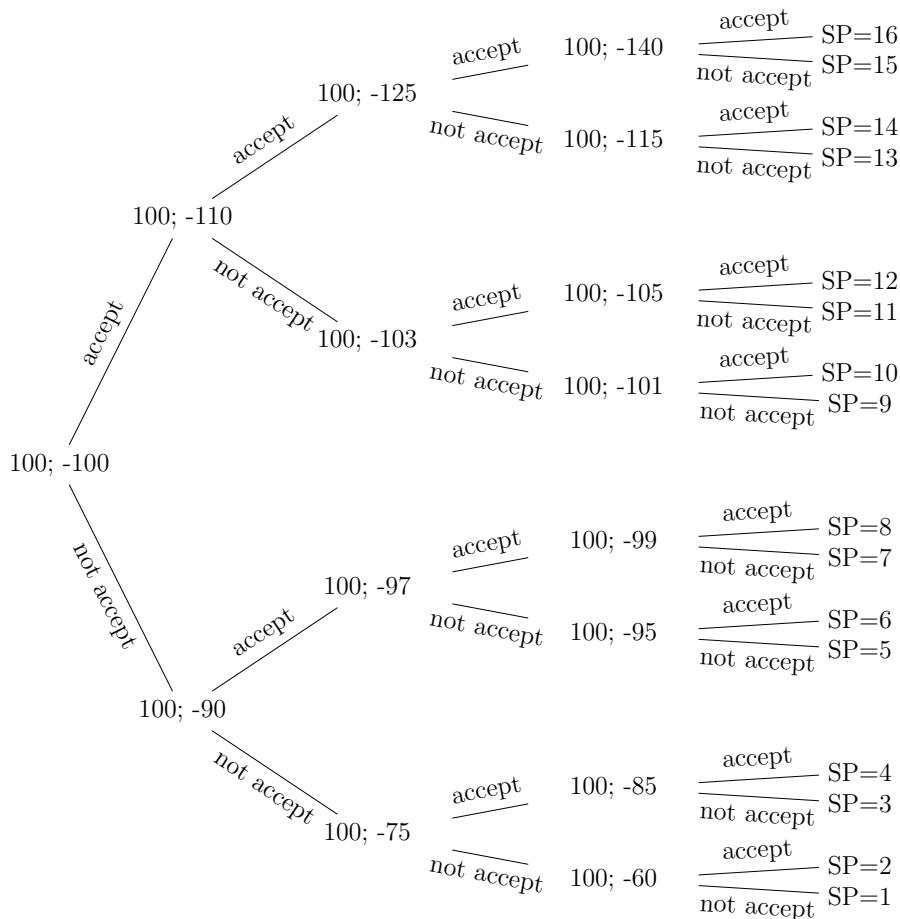
Imagine your bank offered you a debt contract. Under this contract you

*receive 100€ from your bank today and have to pay back XX€ in 6 months.
Please assume that you must pay the full amount you owe to the bank on time.*

Would you accept such a contract?

Figure 4.4 illustrates all 15 choices of the MPL. Nodes depict a hypothetical debt contract. Branches depict the available choices. Based on the choice in a specific node the path through the staircase is determined. The fixed, positive amounts of 100€ indicates the hypothetical amount to be received today and the node/contract-specific negative amounts indicate the respective hypothetical amount of repayment in six months (to be inserted in for XX in the above mentioned question text). Respondents start at the left-most node and work their way through four questions until reaching a end-point at the right side. The label in the end-point states the switchpoint (SP) associated with the given choice path.

Figure 4.4: Staircase elicitation of choices on hypothetical debt contracts



5

General Discussion

This thesis analyzes a selection of particular aspects of beliefs and preferences from different points of view.

In Chapter 2, I focus on a particular source of information that may influence beliefs: expert committee judgments. I elicit how the accuracy and perceived trustworthiness of such judgments are affected by the interaction format of committees. Moreover, I analyze the effects of committee members with hidden agendas who try to manipulate the group judgment to induce a consequent decision in their interest. Utilizing two complementary controlled lab experiments, I compare group judgments from unstructured face-to-face interaction, ubiquitous in real-world institutions, to group judgments from the scientifically endorsed, structured Delphi technique. Supporting its scientific endorsement, in scenarios without manipulation, the Delphi technique produces the most accurate judgments, which are indistinguishable from the Bayesian benchmark. With manipulation, however, face-to-face meetings produce more accurate results, which are also perceived as more trustworthy. Further, unlike Delphi meetings, I show that face-to-face meetings exhibit robustness towards manipulation, as the

accuracy of judgments from face-to-face meetings is indistinguishable in scenarios with and without manipulation. Pinpointing potential mechanisms behind accuracy differences, I find that hidden agendas lead to less information sharing and decrease the truthfulness of shared information for Delphi groups. In contrast, in FTF groups, the amount of shared information and of shared truthful information increases with hidden agendas.

In light of the broader scientific literature, my research contributes to analyses of how to best elicit the collective judgment of groups like expert committees. While previous research has focused on prediction markets (Choo, Kaplan, and Zultan, 2022) and structured face-to-face interaction following the nominal group technique (Maciejovsky and Budescu, 2020), I am first to scrutinize the effect of hidden agendas on accuracy and perceived trustworthiness of judgments from groups interacting according to the Delphi technique and in free-form face-to-face meetings.

From a policy perspective, the results can be taken as evidence that free-form face-to-face interaction should be used in settings with hidden agendas. Arguably, some forms of hidden agenda play a role in most real-world situations where a group of people jointly evaluate some decision-relevant criteria. As such, the prevalence of face-to-face meetings in practice might be ecologically rational. While my study uncovers the relative advantage of face-to-face meetings over the Delphi technique in situations with hidden agendas, further research may analyze whether they are the best among a wider set of interaction formats in situations with hidden agendas and whether the result also holds in different contexts.

In Chapter 3, Thomas Meissner and I analyze a specific domain of individual preferences: aversion toward debt. We use a series of financial decision scenarios in a controlled lab experiment to test for the existence and magnitude of debt aversion. The design of our experiment allows us to isolate the effect of genuine debt aversion from potentially confounding preferences, biases, beliefs, and constraints. We find that the

vast majority (89%) of participants exhibit debt aversion. We present a theoretical framework incorporating debt aversion as a standalone domain of economic preferences and estimate the model parameters based on observed choices in our experiments. As such, we develop a fundamental underpinning for research observing an intrinsic unwillingness to take on debt in a variety of contexts such as investment in tertiary education (Field, 2009; Caetano, Palacios, and Patrinos, 2019), energy-efficient technologies (Schleich, Faure, and Meissner, 2021), and entrepreneurial ventures (Nguyen et al., 2020). While we have made a first step to identifying debt aversion cleanly, many open questions remain. Further research could improve the knowledge on debt aversion by uncovering the exact mechanisms at play.

Understanding debt preferences is key to analyzing debt-related behavior. Some of the financially most significant choices people face in life potentially involve going into debt, e.g., to finance home ownership. Moreover, various policies use subsidized loans to spur socially desirable behavior, such as tertiary education. To enable short and easy elicitation of individual debt preferences, Chapter 4 presents a two-item survey module that approximates the behavior in the experiment from Chapter 3. In this way, we complement existing survey modules on economic preferences (Falk, A. Becker, Dohmen, Huffman, et al., 2022). This may prove useful to researchers and for policy advice on matters of debt-related decision-making.

Bibliography

- Abdellaoui, Mohammed, Han Bleichrodt, et al. (2013). "Is There One Unifying Concept of Utility? An Experimental Comparison of Utility Under Risk and Utility Over Time". In: *Management Science* 59.9, pp. 2153–2169. DOI: 10.1287/mnsc.1120.1690.
- Abdellaoui, Mohammed, Emmanuel Kemel, et al. (2019). "Measuring time and risk preferences in an integrated framework". In: *Games and Economic Behavior* 115, pp. 459–469. DOI: 10.1016/j.geb.2019.03.001.
- Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2019). "Preferences for Truth-Telling". In: *Econometrica* 87.4, pp. 1115–1153. DOI: <https://doi.org/10.3982/ECTA14673>.
- Ahrens, Steffen, Ciril Bosch-Rosa, and Thomas Meissner (2022). "Intertemporal consumption and debt aversion: a replication and extension". In: *Journal of the Economic Science Association*.
- Akaike, H. (1974). "A new look at the statistical model identification". In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723. DOI: 10.1109/TAC.1974.1100705.
- Albrecht, David (2023). *Mitigating manipulation in committees: Just let them talk!*
- Albrecht, David and Thomas Meissner (2022). "The debt aversion survey module: An experimentally validated tool to measure individual debt aversion". In: *Working Paper*.
- Almenberg, Johan et al. (2021). "Attitudes towards Debt and Debt Behavior*". In: *The Scandinavian Journal of Economics* 123.3, pp. 780–809. DOI: <https://doi.org/10.1111/sjoe.12419>.
- Amar, Moty et al. (2019). "Winning the Battle but Losing the War: The Psychology of Debt Management". In: *Journal of Marketing Research* 48.SPL, S38–S50. DOI: 10.1509/jmkr.48.SPL.S38.
- Amjahid, M. et al. (2017). "An Attack is Expected". In: *Die Zeit* Nr. 13/2017.

Bibliography

- Andersen, Steffen, Glenn W. Harrison, Arne Risa Hole, et al. (2012). "Non-linear mixed logit". In: *Theory and Decision* 73.1, pp. 77–96.
- Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, et al. (2008). "Eliciting Risk and Time Preferences". In: *Econometrica* 76.3, pp. 583–618. DOI: [10.1111/j.1468-0262.2008.00848.x](https://doi.org/10.1111/j.1468-0262.2008.00848.x).
- Andersson, Ola et al. (2020). "Robust inference in risk elicitation tasks". In: *Journal of Risk and Uncertainty*, pp. 1–15. DOI: [10.1007/s11166-020-09341-6](https://doi.org/10.1007/s11166-020-09341-6).
- Andreoni, James and Charles Sprenger (2012). "Risk preferences are not time preferences". In: *American Economic Review* 102.7, pp. 3357–76.
- Armstrong, J. S. (2006). "How to Make Better Forecasts and Decisions: Avoid Face-to-Face Meetings". In: *Foresight: The International Journal of Applied Forecasting* 5, pp. 3–15.
- Arrow, Kenneth J. et al. (2008). "The Promise of Prediction Markets". In: *Science* 320.5878, pp. 877–878. DOI: [10.1126/science.1157679](https://doi.org/10.1126/science.1157679).
- Atanasov, Pavel et al. (2017). "Distilling the Wisdom of Crowds: Prediction Markets vs. Prediction Polls". In: *Management Science* 63.3, pp. 691–706. DOI: [10.1287/mnsc.2015.2374](https://doi.org/10.1287/mnsc.2015.2374).
- Becker, Joshua Aaron, D. Brackbill, and D. Centola (2017). "Network dynamics of social influence in the wisdom of crowds". In: *Proc Natl Acad Sci U S A* 114.26, E5070–E5076. DOI: [10.1073/pnas.1615978114](https://doi.org/10.1073/pnas.1615978114).
- Becker, Joshua Aaron, Douglas Guilbeault, and Edward Bishop Smith (2021). "The Crowd Classification Problem: Social Dynamics of Binary-Choice Accuracy". In: *Management Science*. DOI: [10.1287/mnsc.2021.4127](https://doi.org/10.1287/mnsc.2021.4127).
- Belton, Ian, Alice MacDonald, et al. (2019). "Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process". In: *Technological Forecasting and Social Change* 147, pp. 72–82. DOI: [10.1016/j.techfore.2019.07.002](https://doi.org/10.1016/j.techfore.2019.07.002).
- Belton, Ian, George Wright, et al. (2021). "Delphi with feedback of rationales: How large can a Delphi group be such that participants are not overloaded, de-motivated, or disengaged?" In: *Technological*

-
- Forecasting and Social Change* 170. DOI: 10.1016/j.techfore.2021.120897.
- Berger, P.L. and T. Luckmann (2011). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Open Road Media.
- Besharat, Ali, Sajeev Varki, and Adam W. Craig (2015). "Keeping consumers in the red: Hedonic debt prioritization within multiple debt accounts". In: *Journal of Consumer Psychology* 25.2, pp. 311–316. DOI: 10.1016/j.jcps.2014.08.005.
- Best, Roger J. (1974). "An Experiment in Delphi Estimation in Marketing Decision Making". In: *Journal of Marketing Research* 11.4, pp. 448–452.
- Boatman, Angela, Brent J. Evans, and Adela Soliz (2017). "Understanding Loan Aversion in Education". In: *AERA Open* 3.1. DOI: 10.1177/2332858416683649.
- Bolger, Fergus, Andrew Stranieri, et al. (2011). "Does the Delphi process lead to increased accuracy in group-based judgmental forecasts or does it simply induce consensus amongst judgmental forecasters?" In: *Technological Forecasting and Social Change* 78.9, pp. 1671–1680. DOI: 10.1016/j.techfore.2011.06.002.
- Bolger, Fergus and George Wright (2011). "Improving the Delphi process: Lessons from social psychological research". In: *Technological Forecasting and Social Change* 78.9, pp. 1500–1513. DOI: 10.1016/j.techfore.2011.07.007.
- Bond C. F., Jr. and B. M. DePaulo (2006). "Accuracy of deception judgments". In: *Pers Soc Psychol Rev* 10.3, pp. 214–34. DOI: 10.1207/s15327957pspr1003_2.
- Brown, Alexander L. et al. (2022). "Meta-analysis of empirical estimates of loss aversion". eng. In: *Journal of Economic Literature*.
- Caetano, Gregorio, Miguel Palacios, and Harry A. Patrinos (2019). "Measuring Aversion to Debt: An Experiment Among Student Loan Candidates". In: *Journal of Family and Economic Issues* 40.1, pp. 117–131. DOI: 10.1007/s10834-018-9601-8.
- Callender, Claire and Jonathan Jackson (2005). "Does the Fear of Debt Deter Students from Higher Education?" In: *Journal of Social Policy* 34.4, pp. 509–540. DOI: 10.1017/s004727940500913x.

Bibliography

- Cameron, A Colin and Pravin K Trivedi (2005). *Microeometrics: methods and applications*. Cambridge university press.
- Chen, Daniel L., Martin Schonger, and Chris Wickens (2016). “oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments”. In: *SSRN Electronic Journal*.
- Cheung, Stephen L (2020). “Eliciting utility curvature in time preference”. In: *Experimental Economics* 23.2, pp. 493–525.
- Choo, Lawrence, Todd R. Kaplan, and Ro'i Zultan (2022). “Manipulation and (Mis)trust in Prediction Markets.” In: *Management Science*.
- Coller, Maribeth and Melonie B. Williams (1999). “Eliciting Individual Discount Rates”. In: *Experimental Economics* 2.2, pp. 107–127. DOI: 10.1023/a:1009986005690.
- Condorcet, Jean-Antoine-Nicolas de Caritat marquis de (1785). *Essai sur l'application de l'analyse a la probabilite des decisions rendues a la probabilite des voix*. 1st. Library of liberal arts ; LLA 159. Translated in 1976 to “Essay on the Application of Mathematics to the Theory of Decision-Making,” Indianapolis: Bobbs-Merrill.
- Conrads, Julian et al. (2013). “Lying and team incentives”. In: *Journal of Economic Psychology* 34, pp. 1–7. DOI: 10.1016/j.jeop.2012.10.011.
- Conte, Anna, John D. Hey, and Peter G. Moffatt (2011). “Mixture models of choice under risk”. In: *Journal of Econometrics* 162.1, pp. 79–88. DOI: 10.1016/j.jeconom.2009.10.011.
- Cornsweet, Tom N. (1962). “The Staircase-Method in Psychophysics”. In: *The American Journal of Psychology* 75.3. DOI: 10.2307/1419876.
- Cowgill, Bo and Eric Zitzewitz (2015). “Corporate Prediction Markets: Evidence from Google, Ford, and Firm X”. In: *The Review of Economic Studies* 82.4, pp. 1309–1341. DOI: 10.1093/restud/rdv014.
- Da, Zhi and Xing Huang (2020). “Harnessing the Wisdom of Crowds”. In: *Management Science* 66.5, pp. 1847–1867. DOI: 10.1287/mnsc.2019.3294.
- Daar, A. S. et al. (2002). “Top ten biotechnologies for improving health in developing countries”. In: *Nat Genet* 32.2, pp. 229–32. DOI: 10.1038/ng1002-229.

-
- Dalkey, Norman C (1975). "Toward a theory of group estimation". In: *The Delphi method: Techniques and applications*, pp. 236–261.
- Datta, S. and G. A. Satten (2008). "A signed-rank test for clustered data". In: *Biometrics* 64.2, pp. 501–7. DOI: 10.1111/j.1541-0420.2007.00923.x.
- Drichoutis, A. C. and J. L. Lusk (2014). "Judging statistical models of individual decision making under risk using in- and out-of-sample criteria". In: *PLoS One* 9.7, e102269. DOI: 10.1371/journal.pone.0102269.
- Duffy, John (2016). "Macroeconomics: A Survey of Laboratory Research". In: *The Handbook of Experimental Economics, Volume 2*. Ed. by John H. Kagel and Alvin E. Roth. Vol. 2. Princeton University Press, pp. 1–90.
- Duffy, John and Andreas Orland (2020). "Liquidity Constraints and Buffer Stock Savings: Theory and Experimental Evidence". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3672126.
- Eckel, Catherine C. et al. (2016). "Debt Aversion and the Demand for Loans for Postsecondary Education". In: *Public Finance Review* 35.2, pp. 233–262. DOI: 10.1177/1091142106292774.
- Ecken, Philipp, Tobias Gnatzy, and Heiko A. von der Gracht (2011). "Desirability bias in foresight: Consequences for decision quality based on Delphi results". In: *Technological Forecasting and Social Change* 78.9, pp. 1654–1670. DOI: 10.1016/j.techfore.2011.05.006.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, et al. (2018). "Global Evidence on Economic Preferences*". In: *The Quarterly Journal of Economics* 133.4, pp. 1645–1692. DOI: 10.1093/qje/qjy013.
- Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, et al. (2022). "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences". In: *Management Science*. DOI: 10.1287/mnsc.2022.4455.
- Farrell, S. and C. J. Ludwig (2008). "Bayesian and maximum likelihood estimation of hierarchical response time models". In: *Psychon Bull Rev* 15.6, pp. 1209–17. DOI: 10.3758/PBR.15.6.1209.

Bibliography

- Feddersen, Timothy and Wolfgang Pesendorfer (1998). "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting". In: *American Political Science Review* 92.1, pp. 23–35. DOI: 10.2307/2585926.
- Fehrler, Sebastian and Niall Hughes (2018). "How Transparency Kills Information Aggregation: Theory and Experiment". In: *American Economic Journal: Microeconomics* 10.1, pp. 181–209. DOI: 10.1257/mic.20160046.
- Felgenhauer, Mike and Hans Peter Grüner (2008). "Committees and Special Interests". In: *Journal of Public Economic Theory* 10.2, pp. 219–243. DOI: <https://doi.org/10.1111/j.1467-9779.2008.00359.x>.
- Field, Erica (2009). "Educational Debt Burden and Career Choice: Evidence from a Financial Aid Experiment at NYU Law School". In: *American Economic Journal: Applied Economics* 1.1, pp. 1–21. DOI: 10.1257/app.1.1.1.
- Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental Economics* 10.2, pp. 171–178. DOI: 10.1007/s10683-006-9159-4.
- Fischbacher, Urs and Franziska Föllmi-Heusi (2013). "Lies in Disguise – an Experimental Study on Cheating". In: *Journal of the European Economic Association* 11.3, pp. 525–547. DOI: 10.1111/jeea.12014.
- Friedman, Jeffrey A. and Richard Zeckhauser (2014). "Handling and Mishandling Estimative Probability: Likelihood, Confidence, and the Search for Bin Laden". In: *Intelligence and National Security* 30.1, pp. 77–99. DOI: 10.1080/02684527.2014.885202.
- Fries, Tilman et al. (2021). "Observability and lying". In: *Journal of Economic Behavior & Organization* 189, pp. 132–149. DOI: 10.1016/j.jebo.2021.06.038.
- Galton, Francis (1907). "Vox populi". In: *Nature* 75.
- Gaudecker, Hans-Martin von, Arthur van Soest, and Erik Wengström (2011). "Heterogeneity in Risky Choice Behavior in a Broad Population". In: *American Economic Review* 101.2, pp. 664–694. DOI: 10.1257/aer.101.2.664.

-
- Gimpel, Henner and Florian Teschner (June 2014). "Market-Based Collective Intelligence in Enterprise 2.0 Decision Making". In.
- Glimcher, P.W. and E. Fehr (2013). *Neuroeconomics: Decision Making and the Brain*. Elsevier Science.
- Gneezy, Uri (2005). "Deception: The Role of Consequences". In: *American Economic Review* 95.
- Gneezy, Uri, Agne Kajackaite, and Joel Sobel (2018). "Lying Aversion and the Size of the Lie". In: *American Economic Review* 108.2, pp. 419–453. DOI: 10.1257/aer.20161553.
- Goeree, Jacob K. and Leeat Yariv (2011). "An Experimental Study of Collective Deliberation". In: *Econometrica* 79.3, pp. 893–921. DOI: <https://doi.org/10.3982/ECTA8852>.
- Gopalan, Radhakrishnan et al. (2021). "Aversion to student debt? Evidence from low-wage workers". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3843961.
- Graefe, Andreas and J. S. Armstrong (2011). "Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task". In: *International Journal of Forecasting* 27.1, pp. 183–195. DOI: 10.1016/j.ijforecast.2010.05.004.
- Greiner, Ben (2015). "Subject pool recruitment procedures: organizing experiments with ORSEE". In: *Journal of the Economic Science Association* 1.1, pp. 114–125. DOI: 10.1007/s40881-015-0004-4.
- Grime, Megan M. and George Wright (2016). "Delphi Method". In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, pp. 1–6. DOI: 10.1002/9781118445112.stat07879.
- Griscom, Bronson W. et al. (2017). "Natural climate solutions". In: *Proceedings of the National Academy of Sciences* 114.44, pp. 11645–11650. DOI: 10.1073/pnas.1710465114.
- Hansen, Jan, Carsten Schmidt, and Martin Strobel (2004). "Manipulation in political stock markets - preconditions and evidence". In: *Applied Economics Letters* 11.7, pp. 459–463. DOI: 10.1080/1350485042000191700.
- Hanson, Robin and Ryan Oprea (2009). "A Manipulator Can Aid Prediction Market Accuracy". In: *Economica* 76.302, pp. 304–314. DOI: 10.1111/j.1468-0335.2008.00734.x.

Bibliography

- Hanson, Robin, Ryan Oprea, and David Porter (2006). "Information aggregation and manipulation in an experimental market". In: *Journal of Economic Behavior & Organization* 60.4, pp. 449–459. DOI: 10.1016/j.jebo.2004.09.011.
- Harrison, Glenn W., Morten I. Lau, and Melonie B. Williams (2002). "Estimating Individual Discount Rates in Denmark: A Field Experiment". In: *American Economic Review* 92.5, pp. 1606–1617. DOI: 10.1257/000282802762024674.
- Harrison, Glenn W. and E. Elisabet Rutström (2008). "Risk Aversion in the Laboratory". In: *Risk Aversion in Experiments*. Research in Experimental Economics. Emerald Group Publishing Limited, pp. 41–196. DOI: 10.1016/s0193-2306(08)0003-3.
- Haug, N. et al. (2020). "Ranking the effectiveness of worldwide COVID-19 government interventions". In: *Nat Hum Behav* 4.12, pp. 1303–1312. DOI: 10.1038/s41562-020-01009-0.
- Helka, Anna M. and Dominika Maison (2021). "Predictors of the Propensity to Incur Loans for Varying Purposes in the Future". In: *European Research Studies* 24.1, pp. 1114–1128.
- Hermann, Daniel and Mattheus Brenig (2022). "Dishonest online: A distinction between observable and unobservable lying". In: *Journal of Economic Psychology* 90, p. 102489. DOI: <https://doi.org/10.1016/j.joep.2022.102489>.
- Hermann, Daniel and Andreas Ostermaier (2018). "Be Close to Me and I Will Be Honest. How Social Distance Influences Honesty". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3131732.
- Holt, Charles A. and Susan K. Laury (2002). "Risk Aversion and Incentive Effects". In: *American Economic Review* 92.5, pp. 1644–1655. DOI: 10.1257/000282802762024700.
- Hossain, T. and R. Okui (2013). "The Binarized Scoring Rule". In: *The Review of Economic Studies* 80.3, pp. 984–1001. DOI: 10.1093/restud/rdt006.
- Hundtofte, Sean, Arna Olafsson, and Michaela Pagel (2019). "Credit smoothing". In: *Working Paper*.

-
- Ikeda, Shinsuke and Myong-Il Kang (2015). "Hyperbolic Discounting, Borrowing Aversion and Debt Holding". In: *Japanese Economic Review* 66.4, pp. 421–446. DOI: 10.1111/jere.12072.
- Imai, Taisuke, Tom A. Rutter, and Colin F. Camerer (2020). "Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets". In: *The Economic Journal*. DOI: 10.1093/ej/ueaa115.
- Jabotinsky, Hadar Yoana and Roee Sarel (2020). "Let it Flow: Information Exchange in Video Conferences versus Face-to-Face Meetings". In: *Working Paper*. DOI: <http://dx.doi.org/10.2139/ssrn.3589431>.
- Jiang, Yujing et al. (2020). "Wilcoxon Rank-Based Tests for Clustered Data with R Package clusrank". In: *Journal of Statistical Software* 96.6. DOI: 10.18637/jss.v096.i06.
- Jolson, Marvin A. and Gerald L. Rossow (1971). "The Delphi Process in Marketing Decision Making". In: *Journal of Marketing Research* 8.4, pp. 443–448.
- Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". In: *Econometrica* 47.2. DOI: 10.2307/1914185.
- KfW (2023). *Our promotion offer for private customers*. URL: <https://www.kfw.de/inlandsfoerderung/Private-customers/> (visited on 09/28/2023).
- Khalmetski, Kiryl, Bettina Rockenbach, and Peter Werner (2017). "Evasive lying in strategic communication". In: *Journal of Public Economics* 156, pp. 59–72. DOI: 10.1016/j.jpubeco.2017.10.002.
- Laibson, D. (1997). "Golden Eggs and Hyperbolic Discounting". In: *The Quarterly Journal of Economics* 112.2, pp. 443–478. DOI: 10.1162/003355397555253.
- Loewenstein, G. and D. Prelec (1992). "Anomalies in Intertemporal Choice: Evidence and an Interpretation". In: *The Quarterly Journal of Economics* 107.2, pp. 573–597. DOI: 10.2307/2118482.
- Lovallo, Dan et al. (2020). "Your company is too risk-averse". In: *Harvard Business Review* 98.2.

Bibliography

- Luce, Robert Duncan, Patrick Suppes, et al. (1965). "Preference, Utility, and Subjective Probability". In: *Handbook of Mathematical Psychology* 3, pp. 252–410.
- Maciejovsky, Boris and David V. Budescu (2013). "Markets as a structural solution to knowledge-sharing dilemmas". In: *Organizational Behavior and Human Decision Processes* 120.2, pp. 154–167. DOI: 10.1016/j.obhdp.2012.04.005.
- (2020). "Too Much Trust in Group Decisions: Uncovering Hidden Profiles by Groups and Markets". In: *Organization Science* 31.6, pp. 1497–1514. DOI: 10.1287/orsc.2020.1363.
- Manski, Charles F. (2004). "Measuring Expectations". In: *Econometrica* 72.5, pp. 1329–1376. DOI: <https://doi.org/10.1111/j.1468-0262.2004.00537.x>.
- Marett, Kent and Joey F. George (2012). "Barriers to Deceiving Other Group Members in Virtual Settings". In: *Group Decision and Negotiation* 22.1, pp. 89–115. DOI: 10.1007/s10726-012-9297-3.
- Martínez-Marquina, Alejandro and Mike Shi (2021). "The Burden of Household Debt". In: *Working Paper*.
- Matousek, Jindrich, Tomas Havranek, and Zuzana Irsova (2021). "Individual discount rates: a meta-analysis of experimental evidence". In: *Experimental Economics* 25.1, pp. 318–358. DOI: 10.1007/s10683-021-09716-9.
- Mattozzi, Andrea and Marcos Y Nakaguma (Nov. 2022). "Public Versus Secret Voting in Committees". In: *Journal of the European Economic Association* 21.3, pp. 907–940. DOI: 10.1093/jeea/jvac061.
- Mazar, Nina, On Amir, and Dan Ariely (2008). "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance". In: *Journal of Marketing Research* 45.6, pp. 633–644. DOI: 10.1509/jmkr.45.6.633.
- McDowell, M. and P. Jacobs (2017). "Meta-analysis of the effect of natural frequencies on Bayesian reasoning". In: *Psychol Bull* 143.12, pp. 1273–1312. DOI: 10.1037/bul0000126.
- Meijenfeldt, F. A. Bastiaan von, Paulien Hogeweg, and Bas E. Dutilh (2023). "A social niche breadth score reveals niche range strategies

-
- of generalists and specialists". In: *Nature Ecology & Evolution* 7.5, pp. 768–781. DOI: 10.1038/s41559-023-02027-7.
- Meissner, Thomas (2016). "Intertemporal consumption and debt aversion: an experimental study". In: *Experimental Economics* 19.2, pp. 281–298.
- Meissner, Thomas and David Albrecht (2022). "Debt Aversion: Theory and Measurement". In: *arXiv preprint arxiv.2207.07538*. DOI: 10.48550/ARXIV.2207.07538.
- Meissner, Thomas, Xavier Gassmann, et al. (2022). "Individual characteristics associated with risk and time preferences: A multi-country representative survey". In: *Working Paper*.
- Meissner, Thomas and Philipp Pfeiffer (2022). "Measuring preferences over the temporal resolution of consumption uncertainty". In: *Journal of Economic Theory* 200, p. 105379.
- Mellers, Barbara et al. (2014). "Psychological Strategies for Winning a Geopolitical Forecasting Tournament". In: *Psychological Science* 25.5, pp. 1106–1115.
- Murphy, Ryan O. and Robert H. W. ten Brincke (2018). "Hierarchical Maximum Likelihood Parameter Estimation for Cumulative Prospect Theory: Improving the Reliability of Individual Risk Parameter Estimates". In: *Management Science* 64.1, pp. 308–326. DOI: 10.1287/mnsc.2016.2591.
- Nelson, Bradley W. (1978). "Statistical manipulation of delphi statements: Its success and effects on convergence and stability". In: *Technological Forecasting and Social Change* 12.1, pp. 41–60. DOI: 10.1016/0040-1625(78)90034-3.
- Nguyen, Hang Thu et al. (2020). "Debt aversion, education, and credit self-rationing in SMEs". In: *Small Business Economics*. DOI: 10.1007/s11187-020-00329-9.
- Oosterbeek, Hessel and Anja van den Broek (2009). "An empirical analysis of borrowing behaviour of higher education students in the Netherlands". In: *Economics of Education Review* 28.2, pp. 170–177. DOI: 10.1016/j.econedurev.2008.01.005.
- Paaso, Mikael, Vesa Pursiainen, and Sami Torstila (2020). "Entrepreneur Debt Aversion and the Effectiveness of SME Support Programs: Ev-

Bibliography

- idence from the COVID-19 Pandemic". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3615299.
- Parenté, F.J. and J.K. Anderson-Parenté (1987). "Delphi inquiry systems". In: *Judgmental Forecasting*. Ed. by P. Ayton G. Wright. Chichester: Wiley, pp. 129–156.
- Pearsall, M. J. and V. Venkataramani (2015). "Overcoming asymmetric goals in teams: the interactive roles of team learning orientation and team identification". In: *J Appl Psychol* 100.3, pp. 735–48. DOI: 10.1037/a0038315.
- Peeters, R. and L. Wolk (2017). "Eliciting interval beliefs: An experimental study". In: *PLoS One* 12.4, e0175163. DOI: 10.1371/journal.pone.0175163.
- (2018). "Elicitation of expectations using Colonel Blotto". In: *Experimental Economics* 22.1, pp. 268–288. DOI: 10.1007/s10683-018-9596-x.
- Phelps, E. S. and R. A. Pollak (1968). "On Second-Best National Saving and Game-Equilibrium Growth". In: *The Review of Economic Studies* 35.2. DOI: 10.2307/2296547.
- Prelec, D. and G. Loewenstein (1998). "The Red and The Black: Mental Accounting of Savings and Debt". In: *Marketing Science* Vol. 17.Issue 1. DOI: 10.1287/mksc.17.1.4..
- Rosner, B., R. J. Glynn, and M. L. Lee (2003). "Incorporation of clustering effects for the Wilcoxon rank sum test: a large-sample approach". In: *Biometrics* 59.4, pp. 1089–98. DOI: 10.1111/j.0006-341x.2003.00125.x.
- Rothschild, David and Rajiv Sethi (2016). "Trading Strategies and Market Microstructure: Evidence from a Prediction Market". In: *The Journal of Prediction Markets* 10.1. DOI: <https://doi.org/10.5750/jpm.v10i1.1179>.
- Rowe, Gene and George Wright (1996). "The impact of task characteristics on the performance of structured group forecasting techniques". In: *International Journal of Forecasting* 12.1, pp. 73–89. DOI: 10.1016/0169-2070(95)00658-3.

-
- (1999). “The Delphi technique as a forecasting tool: issues and analysis”. In: *International Journal of Forecasting* 15.4, pp. 353–375. DOI: 10.1016/s0169-2070(99)00018-7.
 - (2001). “Expert Opinions in Forecasting: The Role of the Delphi Technique”. In: *Principles of Forecasting. International Series in Operations Research & Management Science*. Ed. by J. S. Armstrong. Boston: Springer. DOI: https://doi.org/10.1007/978-0-306-47630-3_7.
- Rowe, Gene, George Wright, and Fergus Bolger (1991). “Delphi: A reevaluation of research and theory”. In: *Technological Forecasting and Social Change* 39.3, pp. 235–251. DOI: 10.1016/0040-1625(91)90039-i.
- Rowe, Gene, George Wright, and Andy McColl (2005). “Judgment change during Delphi-like procedures: The role of majority influence, expertise, and confidence”. In: *Technological Forecasting and Social Change* 72.4, pp. 377–399. DOI: 10.1016/j.techfore.2004.03.004.
- Satopää, Ville, Marat Salikhov, and Elvira Moreno (2022). *BINtools*. Version 0.2.0.
- Satopää, Ville, Marat Salikhov, Philip E. Tetlock, et al. (2021). “Bias, Information, Noise: The BIN Model of Forecasting”. In: *Management Science*. DOI: 10.1287/mnsc.2020.3882.
- Scheibehenne, B. and T. Pachur (2015). “Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice”. In: *Psychon Bull Rev* 22.2, pp. 391–407. DOI: 10.3758/s13423-014-0684-4.
- Schlag, Karl H and Joël J van der Weele (2015). “A method to elicit beliefs as most likely intervals.” In: *Judgment & Decision Making* 10.5.
- Schleich, Joachim, Corinne Faure, and Thomas Meissner (2021). “Adoption of retrofit measures among homeowners in EU countries: The effects of access to capital and debt aversion”. In: *Energy Policy* 149. DOI: 10.1016/j.enpol.2020.112025.
- Schleich, Joachim, Xavier Gassmann, et al. (2019). “A large-scale test of the effects of time discounting, risk aversion, loss aversion, and

Bibliography

- present bias on household adoption of energy-efficient technologies". In: *Energy Economics* 80, pp. 377–393.
- Schwarz, Gideon (1978). "Estimating the Dimension of a Model". In: *The Annals of Statistics* 6.2, pp. 461–464. DOI: 10.1214/aos/1176344136.
- Serra-Garcia, Marta, Eric van Damme, and Jan Potters (2011). "Hiding an inconvenient truth: Lies and vagueness". In: *Games and Economic Behavior* 73.1, pp. 244–261. DOI: 10.1016/j.geb.2011.01.007.
- Serra-Garcia, Marta and Uri Gneezy (2021). "Mistakes, Overconfidence, and the Effect of Sharing on Detecting Lies". In: *American Economic Review* 111.10, pp. 3160–3183. DOI: 10.1257/aer.20191295.
- Smith, A. (1761). *The Theory of Moral Sentiments*. A. Millar.
- Stasser, Garold and William Titus (1985). "Pooling of Unshared Information in Group Decision Making: Biased Information Sampling During Discussion". In: *Journal of Personality and Social Psychology* 48.6, pp. 1467–1478.
- Stone, M. (1974). "Cross-Validatory Choice and Assessment of Statistical Predictions". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 36.2, pp. 111–147.
- Sunstein, Cass Robert (2005). "Group Judgments: Statistical Means, Deliberation, and Information Markets". In: *New York University Law Review* 80, p. 962. DOI: <https://doi.org/10.2139/ssrn.578301>.
- Sutter, Matthias (2009). "Deception Through Telling the Truth?! Experimental Evidence from Individuals and Teams". In: *The Economic Journal* 119.534, pp. 47–60. DOI: 10.1111/j.1468-0297.2008.02205.x.
- Teschner, Florian, David Rothschild, and Henner Gimpel (2017). "Manipulation in Conditional Decision Markets". In: *Group Decision and Negotiation* 26.5, pp. 953–971. DOI: 10.1007/s10726-017-9531-0.
- Tetlock, Philip E. (2005). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press.
- The 80,000 Hours team (2023a). *Improving decision making (especially in important institutions)*. URL: <https://80000hours.org/problem->

-
- profiles/improving-institutional-decision-making/ (visited on 09/29/2023).
- (2023b). *What are the most pressing world problems?* URL: <https://80000hours.org/problem-profiles/> (visited on 09/29/2023).
- Toma, C. and F. Butera (2009). "Hidden profiles and concealed information: strategic information sharing and use in group decision making". In: *Pers Soc Psychol Bull* 35.6, pp. 793–806. DOI: 10.1177/0146167209333176.
- Train, Kenneth E (2009). *Discrete choice methods with simulation*. Cambridge university press.
- Tversky, Amos and Daniel Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185.4157, pp. 1124–1131. DOI: 10.1126/science.185.4157.1124.
- United Nations (2015). *The 17 Goals*. URL: <https://sdgs.un.org/goals> (visited on 09/29/2023).
- Visser, Bauke and Otto H. Swank (2007). "On Committees of Experts". In: *The Quarterly Journal of Economics* 122.1, pp. 337–372.
- Von Neumann, John and Oskar Morgenstern (1944). "Theory of games and economic behavior Princeton". In: *Princeton University Press* 1947, p. 1953.
- Wakker, Peter P (2008). "Explaining the characteristics of the power (CRRA) utility family". In: *Health Economics* 17.12, pp. 1329–1344.
- Weimann, Joachim, Jeannette Brosig-Koch, et al. (2019). *Methods in experimental economics*. Springer.
- Wintle, B. et al. (2012). "The Intelligence Game: Assessing Delphi Groups and Structured Question Formats." In: DOI: <https://doi.org/10.4225/75/57a03816ac5cf>.
- Wittrock, Lars (2023). "Useful forecasting: belief elicitation for decision making". In: *Working Paper*.
- Wolfers, Justin and Eric Zitzewitz (2004). "Prediction Markets". In: *Journal of Economic Perspectives* 18, No. 2. DOI: 10.3386/w10504.
- Woudenberg, Fred (1991). "An evaluation of Delphi". In: *Technological Forecasting and Social Change* 40.2, pp. 131–150. DOI: 10.1016/0040-1625(91)90002-w.

Bibliography

Valorisation

This thesis may benefit society in two major domains: by improving institutional decision-making procedures and by informing policies involving debt-related choices.

Improving institutional decision-making is rated among one of the best ways to enhance prospects for the long-term future as evaluated by the non-governmental organization 80.000 Hours (The 80,000 Hours team, 2023a).¹ Good institutional decision-making is important to tackle the world's most pressing problems like risks from artificial intelligence, catastrophic pandemics, (nuclear) war, and climate change (The 80,000 Hours team, 2023b). Further, improved institutional decision-making cross-sectionally contributes to achieving the United Nations' Sustainable Development Goals (United Nations, 2015). Chapter 2 provides evidence how to mitigate manipulation in expert committees that inform and influence institutional decision-making. To achieve accurate and trustworthy expert committee judgments despite hidden agendas and manipulation, my research supports the usage of free-form face-to-face interaction. The Delphi technique is preferable in situations without hidden agendas where accuracy is the highest priority.

In various contexts, policies use subsidized loans to spur socially desirable investments and behaviors, e.g., tertiary education or sustainable buildings (KfW, 2023). Such policies may contribute to the United Nations' Sustainable Development Goals 4 - quality education, 7 - affordable and clean energy, 9 - industry, innovation and infrastructure, and 11 - sustainable cities and communities. For an effective design of such policies, understanding individual debt preferences is key. In the face of a largely debt-averse population, subsidized loans might not be the best policy tool. Moreover, if debt aversion correlates with

¹80,000 Hours is a non-profit organization analyzing ways to choose one's professional career (40 hours/week, 50 weeks/year, 40 years equals 80,000 hours) in a way that creates the most positive impact on the world.

individual characteristics, such as income or socioeconomic status, such policies could have unintended effects. For instance, loan-based policies to facilitate tertiary education for students from weak financial backgrounds might be particularly unattractive to these students if they are also more debt-averse. Chapter 3 provides insights into debt aversion's existence, prevalence, and magnitude. Chapter 4 presents a survey module suited to elicit debt aversion and thus enable better informed, evidence-based design of debt-related policies.

Summary

This thesis analyzes a selection of particular aspects of beliefs and preferences. Both are important dimensions influencing (individual) decision-making. Beliefs refer to people's expectations about the true state of the world. Preferences define how much people like one option compared to other options.

In Chapter 2, I investigate a particular source of information that may influence beliefs: expert committee judgments. Many institutional decisions build on the collective judgment of small groups, like expert committees or advisory boards. Such group judgments should be accurate and perceived as trustworthy to inform decisions optimally. Accuracy and trustworthiness are potentially at risk as some group members may have a hidden agenda and manipulate the judgment to induce a consequent decision in their own interest. In my work, I use a group decision-making experiment to quantify the effect of hidden agendas on accuracy and trustworthiness. Moreover, I analyze whether different interaction formats of the committee may mitigate the negative effects of manipulation. I find that just letting people talk in a free-form face-to-face meeting leads to group judgments whose accuracy with manipulation is indistinguishable from scenarios without manipulation. As a comparison, I consider the scientifically endorsed Delphi technique. In this format, the committee interacts by exchanging anonymous written messages according to a structured protocol. Supporting the scientific endorsement, Delphi judgments are highly accurate in situations without manipulation. However, unlike face-to-face meetings, the Delphi technique is significantly less accurate with manipulation. Moreover, judgments from face-to-face meetings are generally perceived as more trustworthy than Delphi judgments. In situations with manipulation, judgments from face-to-face meetings are simultaneously more accurate and perceived as more trustworthy than Delphi judgments. This can be taken as evidence that face-to-face

meetings should be used in institutional decision-making in practice whenever hidden agendas might play a role.

In Chapter 3, Thomas Meissner and I uncover a particular domain of preferences: aversion toward debt. Some of the financially most significant choices people face in life potentially involve going into debt, e.g., to finance home ownership or the take-up of tertiary education. Yet evidence from various contexts suggests that people exhibit an intrinsic unwillingness to take on debt. We reproduce this finding in a controlled experiment and isolate genuine debt aversion from potentially confounding preferences, biases, beliefs, and constraints. We find that most participants (89%) exhibit debt aversion. To quantify the magnitude of debt aversion, we estimate the “borrowing premium” the average participant would require to accept getting into debt as compared to a debt-neutral person. This premium amounts to ca. 16% of the principal. In other words, if a debt-neutral person would only accept a loan of at least €100 at certain repayment terms, the debt-averse average person would only accept a loan of at least €116 at the same repayment terms, i.e. the same absolute amount to be repaid at the same time.

In Chapter 4, Thomas Meissner and I present a short survey module to elicit individual debt aversion. Understanding debt aversion is key to studying debt-related decision-making and informing policy-making that uses subsidized loans to spur socially desirable behavior, such as tertiary education. We identify two simple questions that predict behavior in the experiment of Chapter 3. First, how much do you agree that “Debt is an integral part of today’s life.”? Second, how much do you think the average survey respondent agrees, “There is no excuse for borrowing money.”? The answers to these questions can be used approximately quantify the respondent’s degree of debt aversion without going through the incentivized experiment presented in Chapter 3.

Summary

Samenvatting

In dit proefschrift analyseer ik een selectie van bepaalde aspecten van overtuigingen en voorkeuren. Beide zijn belangrijke dimensies die (individuele) besluitvorming beïnvloeden. Overtuigingen beschrijven hoe waarschijnlijk iemand denkt dat een bepaalde staat de werkelijke staat van de wereld is of zal zijn. Voorkeuren geven aan in hoeverre iemand een optie verkiest boven andere opties.

In hoofdstuk 2 onderzoek ik één specifieke informatiebron die overtuigingen kan beïnvloeden: beoordelingen of evaluaties van deskundigencommissies. Veel institutionele beslissingen zijn gebaseerd op collectieve beoordelingen van beslissingsrelevante criteria door kleine groepen zoals commissies en adviesorganen. Dergelijke beoordelingen moeten accuraat zijn en als betrouwbaar worden beschouwd, om een optimale basis voor besluitvorming te bieden. De nauwkeurigheid en betrouwbaarheid kunnen echter in gevaar komen wanneer een of meerdere groepsleden een eigen verborgen agenda hebben en de beoordeling manipuleren om beslissingen in hun eigen belang te zwaaien. In mijn werk gebruik ik een experiment om de invloed van zulke verborgen agenda's op de nauwkeurigheid en betrouwbaarheid van groepsbeoordelingen te kunnen kwantificeren. Ik onderzoek ook of verschillende vormen van groepsinteractie de negatieve effecten van verborgen agenda's kunnen verminderen.

Mijn resultaten laten zien dat ongestructureerde face-to-face meetings leiden tot groepsbeoordelingen waarvan de nauwkeurigheid in situaties met en zonder verborgen agenda's niet te onderscheiden is. Ter vergelijking gebruik ik het wetenschappelijk geprezen Delphi interactieformat. In dit format interageert de groep door anonieme schriftelijke berichten uit te wisselen volgens een gestructureerd protocol. Evaluaties van groepen die interacteren volgens het Delphi formaat zijn zeer accuraat in situaties zonder verborgen agenda's. In tegenstelling tot

ongestructureerde face-to-face meetings, is de Delphi techniek echter aanzienlijk minder nauwkeurig in situaties met verborgen agenda's. Bovendien worden beoordelingen van persoonlijke ontmoetingen over het algemeen als betrouwbaarder ervaren dan Delphi beoordelingen. In situaties met verborgen agenda's zijn beoordelingen van face-to-face meetings tegelijkertijd nauwkeuriger en als betrouwbaarder beschouwd dan Delphi beoordelingen. Dit is een argument om in de praktijk ongestructureerde face-to-face meetings te gebruiken bij institutionele besluitvorming wanneer verborgen agenda's een rol kunnen spelen.

In hoofdstuk 3 onderzoeken Thomas Meissner en ik een speciale dimensie van voorkeuren: schuldaversie. Sommige van de belangrijkste financiële beslissingen die mensen in hun leven nemen kunnen gepaard gaan met het aangaan van schulden, zoals bijvoorbeeld de beslissing om een eigen huis of studies te financieren. Er zijn echter aanwijzingen uit verschillende contexten dat mensen terughoudend zijn om schulden aan te gaan. Wij herhalen deze bevinding in een gecontroleerd experiment en isoleren het effect van puur schuldaversie van mogelijk andere voorkeursdimensies, systematische biases, overtuigingen en beperkingen. We vinden dat de meeste deelnemers (89%) een aversie hebben tegen schulden. Om de mate van schuldaversie te kwantificeren schatten we de 'schuldpremie' die gemiddelde deelnemers nodig hebben om schulden te accepteren in vergelijking met personen die niet schuldavers zijn. Deze premie bedraagt ongeveer 16% van het geleende bedrag. Met andere woorden, als een schuld-neutrale persoon een lening van ten minste 100 euro accepteert tegen bepaalde aflossingsvoorwaarden, zou de gemiddelde schuld-averse persoon slechts een lening van ten minste 116 euro accepteren tegen dezelfde aflossingsvoorwaarden, d.w.z. hetzelfde absolute bedrag dat op hetzelfde moment moet worden terugbetaald.

In hoofdstuk 4 presenteren Thomas Meissner en ik een korte enquête-module om individuele schuldaversie te meten. Inzicht in schuldaversie is essentieel voor het bestuderen van schuldgerelateerde beslissingen en voor het ontwerpen van empirisch onderbouwde beleidsinstrumenten die gesubsidieerde leningen gebruiken om sociaal wense-

lijk gedrag, zoals hoger onderwijs, te bevorderen. Op basis van het experiment uit hoofdstuk 3 hebben we twee eenvoudige vragen geïdentificeerd die gedrag in het experiment voorspellen. Ten eerste, in hoeverre bent u het ermee eens dat “schulden vandaag de dag een integraal onderdeel van het leven vormen”? Ten tweede, in hoeverre denkt u dat de gemiddelde enquêteerdeelnemer het eens is met de stelling: “Er is geen excuus voor het lenen van geld”? Op basis van de antwoorden op deze vragen kan de mate van schuldaversie van de respondenten ongeveer even goed gekwantificeerd worden als door het veel uitbreidere experiment in hoofdstuk 3.

Zusammenfassung

In dieser Dissertation analysiere ich eine Auswahl bestimmter Aspekte von Überzeugungen und Präferenzen. Beides sind wichtige Dimensionen, welche die (individuelle) Entscheidungsfindung beeinflussen. Überzeugungen beschreiben, für wie wahrscheinlich jemand es hält, dass ein bestimmter Zustand, der wahre Zustand der Welt ist bzw. sein wird. Präferenzen geben an, wie jemand eine Option im Vergleich zu anderen Optionen mag.

In Kapitel 2 untersuche ich eine bestimmte Informationsquelle, die Überzeugungen beeinflussen kann: Urteile bzw. Bewertungen von Expertenkommittees. Viele institutionelle Entscheidungen beruhen auf der kollektiven Bewertung eines Sachverhalts durch kleine Gruppen, wie Kommittees und beratenden Gremien. Solche Bewertungen sollten genau sein und als vertrauenswürdig wahrgenommen werden, um eine optimale Entscheidungsgrundlage zu bieten. Genaugigkeit und Vertrauenswürdigkeit sind jedoch potenziell gefährdet, wenn einige Gruppenmitglieder eine versteckte, eigene Agenda haben und die Bewertung manipulieren könnten, um Entscheidungen in ihrem eigenen Interesse herbeizuführen. In meiner Arbeit verwende ich ein Experiment, um die Auswirkungen von solchen versteckten Agenden auf die Genaugigkeit und Vertrauenswürdigkeit von Gruppenbewertungen zu quantifizieren. Außerdem untersuche ich, ob verschiedene Interaktionsformate der Gruppe die negativen Auswirkungen von versteckten Agenden abschwächen können.

Ich zeige auf, dass unstrukturierte face-to-face meetings zu Bewertungen durch die Gruppe führen, deren Genaugigkeit in Situationen mit Gruppenmitgliedern mit versteckten Agenden nicht von Szenarien ohne versteckte Agenden zu unterscheiden ist. Als Vergleich ziehe ich das wissenschaftlich angepriesene Delphi Interaktionsformat heran. Bei diesem Format interagiert die Gruppe durch den Austausch anonymer, schriftlicher Nachrichten nach einem strukturierten Protokoll.

Bewertungen von Gruppen die nach dem Delphi Format interagieren sind in Situationen ohne versteckte Agenden sehr genau. Im Gegensatz zu unstrukturierten face-to-face Meetings ist die Delphi Technik in Situationen mit versteckten Agenden jedoch deutlich ungenauer. Darüber hinaus werden Bewertungen aus face-to-face Meetings im Allgemeinen als vertrauenswürdiger wahrgenommen als Delphi Bewertungen. In Situationen mit versteckten Agenden sind Bewertungen aus face-to-face Meetings gleichzeitig genauer und werden als vertrauenswürdiger wahrgenommen als Delphi Bewertungen. Dies ist ein wissenschaftliches Argument dafür, dass unstrukturierte face-to-face meetings bei institutionellen Entscheidungen in der Praxis eingesetzt werden sollten, wenn versteckte Agenden eine Rolle spielen könnten.

In Kapitel 3 decken Thomas Meissner und ich eine besondere Dimension von Präferenzen auf: die Abneigung bzw. Aversion gegen Schulden. Einige der finanziell bedeutsamsten Entscheidungen, die Menschen im Leben treffen, bedingen möglicherweise die Aufnahme von Schulden, z. B. zur Finanzierung von Wohneigentum oder eines Studiums. Es gibt jedoch Hinweise aus verschiedenen Kontexten, die darauf hindeuten, dass Menschen eine intrinsische Abneigung haben, sich zu verschulden. Wir reproduzieren diesen Befund in einem kontrollierten Experiment und isolieren den Effekt eigentlicher Schuldenaversion von potenziell anderen Präferenzdimensionen, systematischen Verzerrungen im Entscheidungsverhalten, Überzeugungen und Restriktionen. Wir stellen fest, dass die meisten Teilnehmenden (89%) schuldenavers sind. Um das Ausmaß von Schuldenaversion zu quantifizieren, schätzen wir die "Verschuldungsprämie", die durchschnittliche Teilnehmende im Vergleich zu einer schuldenneutralen Personen benötigen, um eine Verschuldung zu akzeptieren. Diese Prämie beläuft sich auf ca. 16% der Kreditsumme. Mit anderen Worten, wenn eine schuldenneutrale Person einen Kredit von mindestens 100 Euro zu bestimmten Rückzahlungsbedingungen akzeptierte, würde die durchschnittliche, schuldenaverse Person nur einen Kredit von mindestens 116 Euro zu den gleichen Rückzahlungsbedingungen, d. h. dem gleichen absoluten Rückzahlungsbetrag, der zum gleichen Zeitpunkt zurückzuzahlen ist,

akzeptieren.

In Kapitel 4 stellen Thomas Meissner und ich ein kurzes Umfragemodul vor, um individuelle Schuldenaversion zu ermitteln. Das Verständnis von Schuldenaversion ist essenziell zur Untersuchung schuldenbezogener Entscheidungen und zum Design von evidenzbasierten Politkinstrumenten, die subventionierte Kredite einsetzen, um sozial erwünschtes Verhalten, wie z. B. Hochschulbildung, zu fördern. Auf Basis des Experiments in Kapitel 3 haben wir zwei einfache Fragen ermittelt, die das Verhalten im Experiment aus Kapitel 3 vorhersagen. Erstens: Wie sehr stimmen Sie zu, dass "Schulden ein integraler Bestandteil des heutigen Lebens sind"? Zweitens: Wie sehr, glauben Sie, stimmt der durchschnittliche Umfrageteilnehmende der Aussage zu: "Es gibt keine Entschuldigung dafür, Geld zu leihen."? Anhand der Antworten auf diese Fragen lässt sich der Grad der Schuldenaversion der Befragten annähernd so gut quantifizieren, wie durch das deutlich aufwändigeren Experiment in Kapitel 3.

Zusammenfassung

Zusammenfassung

About the author

David Albrecht was born in Germany in 1992. He received his Bachelor's degree in Economics and Sociology from Georg August University Göttingen (2014) and his Master's degree in Economics and Management Science from Humboldt University Berlin (2017). After graduation, he worked as a management consultant in the German public sector. Motivated to find scientific, evidence-based solutions to some of the challenges he experienced as a consultant, he joined Maastricht University to pursue his PhD in Human Decisions and Policy Design. The results of his PhD research are published in this issue.

About the author

About the author
