Essays in Applied Microeconomics

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum (Dr. rer. pol.)

eingereicht an der Wirtschaftswissenschaftlichen Fakultät
der Heinrich-Heine Universität Düsseldorf
von

Vasilisa Petrishcheva, M.Sc.
geb. am 19. Januar 1994 in Moskau

Erste Gutachterin: Prof. Dr. Hannah Schildberg-Hörisch
Zweiter Gutachter: Prof. Dr. Alexander Rasch

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Hannah Schildberg-Hörisch, for her continuous guidance, valuable feedback, and support throughout my time as a doctoral researcher. I want to especially thank Hannah for her support during my job market and the terrible events of 2021. I am grateful to my second supervisor, Alexander Rasch, for helping me develop new ideas and supporting my academic curiosity with his feedback.

I thank my co-authors, Gerhard Riener, Christian Waibel, Chi Trieu, and Matthias Hunold, for sharing their experience and knowledge and exchanging ideas. I also want to thank Paul Heidhues and Mats Köster. My dissertation benefited tremendously from their advice and our insightful conversations.

I am infinitely grateful to my parents, Marina Petrishcheva and Andrey Petrishchev, for their love and support and for believing in me every step of the way. Without their countless sacrifices, even starting (let alone finishing) this dissertation would not be possible. I am grateful to my friends from all around the world, Margarita Vardanyan, Melina Anagnostopoulou Ribeiro, Nesma Ali, Amy Fotheringham, Svetlana Korchagina, and Judith Schwartz, for always having my back. Last but not least, I thank my partner, Tobias Werner, for helping me keep my sanity through this challenging journey and for cheering on every little success with me.

Introduction

This dissertation consists of four chapters that contribute to applied microeconomics. My co-authors and I combine theoretical and experimental methods to study behavioral paradoxes and market inefficiencies. Chapters 1 and 2 contribute to the understanding of image concerns. They study the dynamic implications of how individuals perceive themselves and are perceived by others. Chapter 3 focuses on markets where sellers have the informational advantage and how mitigating consumers' limited attention can partially correct the market inefficiencies arising due to asymmetric information. Chapter 4 shows that profit-shifting within partially vertically integrated entities can facilitate or hinder input and customer foreclosure depending on the restrictions the minority shareholder protection can offer.

Chapter 1 explores whether loss aversion applies to social image concerns. In a simple model, we combine loss aversion in social image concerns and attitudes towards lying. We then test its predictions in a laboratory experiment. Subjects are first ranked publicly in a social image relevant domain, intelligence. This initial rank serves as a within-subject reference point. After subjects have experienced a change in rank over time, they are offered scope for lying to improve their final, publicly reported rank. We find evidence for loss aversion in social image concerns. Subjects who face a loss in social image lie more than those experiencing gains if they care about social image. Individual-level analyses document a discontinuity in lying behavior when moving from rank losses to gains, indicating a kink in the value function for the social image.

Chapter 2 studies how individuals update beliefs about their self-image in case of positive and negative shocks to their self-image, and how these updated beliefs translate into

willingness to acquire self-image-relevant information. The experimental design allows testing whether self-image concerns are reference-dependent and loss aversion applies to self-image concerns. In the experiment, subjects work on an IQ test, a self-image-relevant task, before they face a gain or a loss in self-image induced by an exogenous shift in the task complexity. I find evidence for overly optimistic belief updating for subjects who experience a loss in self-image and overly pessimistic belief updating for those with a gain in self-image. Then, I elicit their willingness to pay to acquire feedback and analyze whether individuals who experience a loss in self-image are more likely to want feedback than those with a gain in this domain. They are, on average, willing to pay to avoid information. Larger changes in beliefs lead to an increase in the willingness to pay for self-image-relevant feedback. I find no evidence supporting loss aversion in self-image concerns, i.e., subjects with marginal positive and negative belief differences do not have significantly different willingness to acquire information. Furthermore, I propose a simple stylized theoretical framework that offers a possible explanation for the patterns in belief updating and information avoidance in the experimental data.

Chapter 3 analyzes how consumers' limited attention affects outcomes in a monopolistic market of credence goods (such as healthcare, repair services, legal advice). Our study is motivated by discrepancies between theoretical predictions and empirical evidence on market outcomes when customers can verify the type of quality they receive, as well as recent calls for more transparency in sellers' costs in some real-world markets. Whereas theory predicts market efficiency with equal markups for different qualities and sufficient quality provision, observations from laboratory experiments yield contradicting evidence of inefficiency. Our study presents both theoretical arguments and experimental evidence that customers' limited attention to sellers' costs can explain these differences. In our experiment, we find that when costs are made salient to customers, the market becomes more efficient. Sellers are more likely to provide sufficient quality, and prices are significantly closer to equal markups. Furthermore, we find that social preferences appear to be important for market outcomes.

Chapter 4 focuses on the incentives of firms that hold partial vertical ownership to

foreclose rivals. Compared to a full vertical merger, with partial ownership, a firm may obtain only part of the target's profit but nevertheless be able to influence the target's strategy significantly. Levy et al. (2018) argue that it makes foreclosure more likely than a full merger. The target may be either a supplier or a customer, which opens the scope for either input foreclosure or customer foreclosure. We show that the incentives to foreclose can be higher, equal, or even lower with partial ownership than with a vertical merger, depending on how the protection of minority shareholders and transfer price regulations are specified.

Chapter 1

Loss Aversion in Social Image Concerns

Co-authored with Gerhard Riener and Hannah Schildberg-Hörisch

1.1 Introduction

Humans care how they are perceived by their fellow humans and go a great length to build up a positive image of themselves (e.g., Andreoni and Bernheim, 2009; Ariely et al., 2009; Bénabou and Jean Tirole, 2006; Bursztyn and Jensen, 2017; Ewers and Zimmermann, 2015; Soetevent, 2011). These carefully crafted images are at stake in everyday interaction, and reputation can decline rapidly. Casual observations suggest that when social image is at risk of being lost people engage in lies and denial to maintain it in many domains of economic life. Managers who do not reach expected targets may engage in fraudulent behavior—as happened recently in the manipulation of car emission tests (Aurand et al., 2018). A person losing her job may leave the house everyday pretending to her family that she is still employed. However, the reference point for status loss does not necessarily have to come from own achievements or calamities, it may also be transmitted through generations as a sense of class entitlement (Alsop, 2008). In the 2019 college admission scandal, affluent parents criminally conspired to influence admission decisions of prestigious colleges (Halleck, 2019; Lovett, 2020). While the special role of losses has been extensively documented in the monetary domain (Barberis, 2013; Camerer, 1998; Kahneman and Tversky, 1979; Wakker, 2010), the effect of losses on moral behavior deserves a closer look.

Does trying to shield oneself from a loss in social image generally lead to more morally deviant behavior than striving for a gain in social image? Or is it a particular behavior of those people who are more inclined to immoral decisions that can lead to tragic fall in the first place? Measuring losses of social image is hard to imagine in the field and the extent of lying difficult to observe. Hence, we design a parsimonious laboratory experiment to test for the presence of loss aversion in social image concerns.

In the experiment, subjects either experience a potential loss or gain in their social image over time, while keeping average social image constant. We then offer subjects scope for improving their social image by lying about their true type. This allows us to test whether—on average—subjects lie more (and are thus willing to incur higher lying costs) when they experience losses than when they experience gains in their social image.

Our results provide evidence for loss aversion in social image concerns. We find that subjects who sufficiently care about their social image—as measured by an independent survey instrument—lie more when experiencing losses as opposed to similar-sized gains in social image over time. Further individual-level analyses document that the extent of lying decreases discontinuously when moving from small losses to small gains in social image. This pattern in lying behavior is compatible with loss aversion in social image concerns but not a simple concave utility function for changes in social image.

Our main contribution is thus documenting loss aversion in social image concerns. Importantly, our findings imply that loss aversion can also play a role in the non-material domain of social image. So far, loss aversion is widely documented for money (e.g., Booij and Van de Kuilen, 2009; Pennings and Smidts, 2003) and material goods (e.g., Kahneman, Knetsch, et al., 1990)[1], but evidence on whether humans have the same inclination when it comes to social image utility is lacking.

Image concerns expand over various domains:[2] People care about being perceived smart and skillful (e.g., Burks et al., 2013; Ewers and Zimmermann, 2015), prosocial and altruistic (e.g., Carpenter and Myers, 2010), pro-environmental (e.g., S. E. Sexton and A. L. Sexton, 2014) and supportive of fair trade (e.g., Friedrichsen and Engelmann, 2018), trustworthy (Abeler et al., 2019), promise-keeping (Grubiak, 2019), or wealthy (Leibenstein, 1950).

In our experiment, we induce social image concerns by letting subjects perform an IQ test and reporting its results publicly. However, signaling skillfulness can be a two-sided sword as Austen-Smith and Fryer Jr (2005) show in a two-audience signaling model. For example, high ability students may under-invest in education because such investments lead to rejection by their peer group.[3] So it is important to establish that an IQ test is indeed suitable to induce social image that is worth striving for in our university student sample. This is underlined by Ewers and Zimmermann (2015) who document that, in

[1] See Bleichrodt et al. (2001) for an application to health outcomes.

[2] Bursztyn and Jensen (2017) present a detailed overview of the recent literature on social image concerns.

[3] Bursztyn, Egorov, et al. (2019) show that students are less likely to sign up for an SAT preparation course and to take an SAT exam itself, if their choices are observable. They therefore forgo educational investment due to possible social stigma.

a student sample similar to the one used in this study, subjects misreport their private information on ability in a laboratory context in order to appear more skillful even when strong monetary incentives are given to tell the truth.

While there is plenty of evidence that many people care about social image, recent, both theoretical and empirical work stresses that there is heterogeneity in the extent to which people care about social image and whether they do so at all. For example, Bursztyn and Jensen (2017) expand the model of Bénabou and Jean Tirole (2006) to explicitly account for heterogeneity in social image concerns.[4] Friedrichsen and Engelmann (2018) empirically reject the hypothesis of homogeneous image concerns and show that individuals react differently to image-building opportunities. In our experiment, we will therefore measure each subject's individual extent of social image concerns.

On top of addressing image concerns, this study also contributes to the growing literature on lying behavior, extensively summarized in Abeler et al. (2019).[5] Based on a comprehensive meta-analysis, Abeler et al. (2019) identify two main channels why people prefer to tell the truth, namely, lying costs that increase in the size of a lie and image concerns for being perceived as an honest person. Our theoretical framework and experiment design build on their work. First, our experimental design ensures that lying cannot be detected such that image concerns for being seen as an honest person by others cannot play a role in the context of our experiment. Second, in order to avoid possible interactions between loss aversion in the monetary and social image domain, our design offers subjects a flat payment and uses the extent of lying, i.e., the lying costs subjects are willing to incur, to quantify how much they suffer from losing or gain from improving their social image. Therefore, our finding that subjects who care about their social image report more dishonestly than others speaks to situations in which honest reporting of private information is key but not incentive-compatible. Since lying in the laboratory is a predictor of dishonesty and rule violations in real life (Dai et al., 2018; Hanna and Wang,

[4]Their theoretical framework distinguishes conformists who experience social pressure to act in a socially desirable way, contrarians who feel pressured to act differently from what is socially desirable, and those who are not subject to social image concerns at all.

[5]Abeler et al. (2019) provide a web interface where they present a detailed overview on recent experiments on lying.

2017), our findings suggest that monitoring efforts should be targeted at individuals who strongly care about their reputation.

We also relate to the literature which links the concept of loss aversion to lying behavior. Grolleau et al. (2016) and Schindler and Pfattheicher (2017) compare the extent of lying for individuals who face monetary losses and gains. They find that participants misreport more to avoid a monetary loss than they do to increase their monetary gain. Garbarino et al. (2019) show that the less likely a low monetary payoff is, the more likely individuals lie to avoid it. In a series of experiments involving deception, Pettit et al. (2016) show that subjects threatened by status loss cheat more.

The paper proceeds as follows: Section 1.2 describes the experiment design and procedures, before we outline our hypotheses in Section 2.3. Results are presented in Section 1.4 and Section 1.5 concludes.

1.2 Experiment design

General setup Our experiment consists of two stages. Stage 1 is designed to establish a personal reference point for social image utility—a publicly reported rank in an intelligence test—against which subjects can fall short of or improve their image in Stage 2. In the second stage, we induce a change of the rank. Subjects are then informed about their true rank and offered scope to manipulate the reporting of their rank to their peers. We test whether subjects whose average rank deteriorates—who experience a loss in social image—misreport their rank more strongly than those who experience an improvement in their rank. We pay special attention towards analyzing misreporting behavior around the reference point in social image in order to identify a possible discontinuity in misreporting as predicted by loss aversion.

We create social image concerns through reporting a subject's ranking in a standardized test of fluid intelligence—Raven's Progressive Matrices test (1983)—to two randomly selected peers. Fluid intelligence encompasses logical reasoning and abstract thinking and constitutes an image providing trait for university students.[6] Public reporting of results

[6]Our approach is similar to Falk and Szech (2020), Ewers and Zimmermann (2015), Zimmermann

shall hence create social image utility. In order to strengthen this link we explicitly mention in the instructions that the matrices (labeled as picture puzzles) are designed to measure fluid intelligence, that fluid IQ is an important part of an individual's overall IQ, and that such or related tasks are often employed in recruitment processes.

At the beginning of each session, two subjects per session are randomly assigned the role of peer observers. We randomly draw one observer from all male subjects and the other from all female subjects. This avoids possible gender-specific observer effects. After the observers have been determined, they stand up in front of the other subjects and announce "I am one of the two observers". The other subjects are randomly assigned to one of two sequences that vary the order of the quizzes over the two stages of the experiment. In sequence *HardEasy* subjects work on a *Hard* quiz in Stage 1 and an *Easy* quiz in Stage 2 and in *EasyHard* on an *Easy* quiz in Stage 1 and a *Hard* quiz in Stage 2. At the end of the experiment, all subjects in both sequences have worked on the exactly same 48 matrices. All subjects—including the observers—received the same instructions. Then subjects performed two quizzes (consisting of 24 matrices each) and after each quiz report their relative performance (rank) to the observers. In the second stage, subjects have the possibility to lie in order to improve their rank before reporting it. Figure 1.1 illustrates the timeline of the experiment that we explain in detail below.

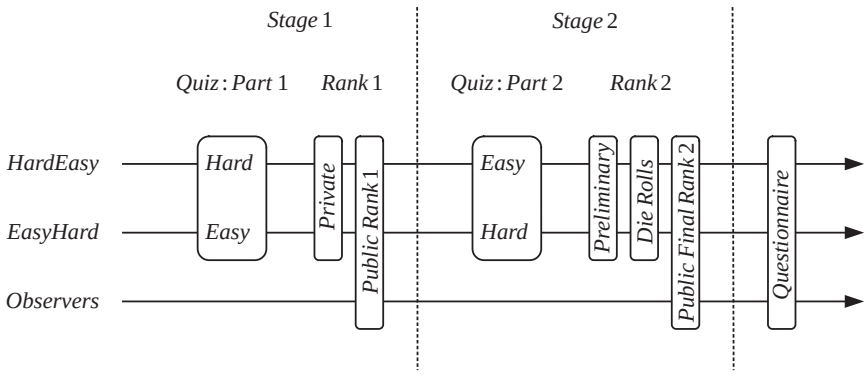


Figure 1.1: Timeline

(2020) and Burks et al. (2013) who also use reporting of the performance in IQ or knowledge tests to induce image concerns.

Matrices task and sequences The original Raven's Progressive Matrices test (RPM) consists of 60 matrices that are divided into 5 equally sized sets (A to E) which increase in difficulty. Figure 1.2 provides an example of a Raven's Progressive Matrix. Subjects have to choose that box below the picture puzzle which is the best logical fit to the empty box within the picture. Progressive means that the matrices are increasing in difficulty. In our design, we do not use the 12 matrices of the easiest set A since we expect our student subjects to solve them all correctly. We split the remaining 48 matrices in two parts consisting of 24 matrices each that we will use for the quizzes. One quiz is easier (*Easy*), while the other is harder (*Hard*). We calibrated the two sets such that *Hard* has a higher likelihood to contain matrices that have been solved by fewer subjects in a reference sample. The reference sample includes 413 observations (students) from a previous experiment which took place at the same lab in 2014. Subjects of the reference group solved exactly the same overall 48 matrices as our subjects.[7] In both quizzes, the difficulty of the tasks is gradually increasing over time. Importantly, both quizzes contain tasks from sets B (easy) to E (difficult) to ensure that subjects do not perceive the difference in difficulty across quizzes as major. Matrices in quiz *Easy* and *Hard* do not repeat or overlap.

Subjects have 30 seconds to work on each matrix. The time limit ensures that performance is comparable across subjects: both within our experiment and with respect to the reference sample we use, in which subjects also had 30 seconds to work on each matrix. On average, it took subjects 11.5 seconds to answer a matrix. 2.7% of answers were provided in the last five seconds and in only 0.7% of cases subjects ran out of time, which suggests that the time limit was not restrictively binding. For each correctly solved matrix, subjects get one point. Wrong answers or no answer within the 30 seconds time limit do not give any points.

[7]The *Easy* quiz consists of the following matrices: B1, B5, B6, B7, B8, B9, B10, B11, B12, C1, C2, C3, C7, C8, C9, C10, C12, D2, D3, D5, D7, E2, E6, and E11. The *Hard* quiz contains the following matrices: B2, B3, B4, C4, C5, C6, C11, D1, D4, D6, D8, D9, D10, D11, D12, E1, E3, E4, E5, E7, E8, E9, E10, and E12.

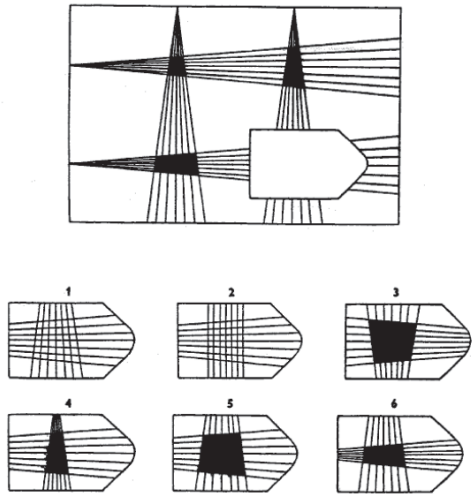


Figure 1.2: Example of a Raven's progressive matrix

Stage 1 After completing the sequence specific Raven's Matrices, subjects received private feedback on their relative performance (i.e., Rank 1) on their screen telling them that "X% of the participants of the reference group have a higher rank than you in Quiz 1". A lower X (lower rank) implies better relative performance. The instructions provide several examples how individual rank is calculated and how to interpret it.[8]

To determine the rank, we compare the share of correctly solved matrices among the first 24 matrices to the distribution of the share of correctly solved matrices among all 48 matrices of the reference sample. Our calibration of the matrix distribution between *Easy* and *Hard* ensures that subjects in sequence *EasyHard* will on average rank better than subjects in sequence *HardEasy* in Quiz 1 since both groups are compared to the same reference sample but the first 24 matrices are easier for subjects in sequence *EasyHard* than in *HardEasy*.

Subjects report their rank in the first stage to the observers. This establishes the individual Rank 1 as a personal reference point for social image concerns. Since subjects are randomized into sequences, their initial reference points before the feedback on Rank 1

[8]We explicitly explain in instructions:

> "For example, the statement "9% of participants of the reference group have a higher rank than you in part 1" implies that "9% performed better than you (i.e., they solved a higher share of the overall 48 matrices from part 1 and 2 correctly than you) and 90% worse (i.e., they solved a lower share of the matrices correctly than you). That means you belong to the 10% of best performers in solving the matrices that were designed to measure fluid IQ."

are the same on average (given skill, ability, etc.). We give both subjects and observers detailed instructions on the reporting procedure to control the reporting process using the same protocol for all sessions. We instruct subjects to fill in report sheets named "Rank 1" and "Rank 2" in Stages 1 and 2, respectively, and to present these sheets to observers who verify the report. No further verbal communication between subjects and observers is allowed, i.e., the entire reporting procedure happens in silence. Report sheets contain two pieces of information: a 4-digit individual code and a rank. After each Stage, observers see a table on their computer screen in which each individual code corresponds to a rank, and thus can compare the report sheet to the true information from the table. If the reported rank matches the true rank, observers stamp the report sheet to verify it.[9] We organized our laboratory setup in a way that subjects cannot see observers' computer screens while reporting their rank. Additionally, to assure anonymity, we use 4-digit individual codes instead of cubicle numbers which, in the unlikely case of a subject seeing the table on the observer's screen, makes it uninformative.

Stage 2 Subjects work on the remaining 24 matrices. For subjects in sequence *Easy-Hard*, Stage 2 is more complicated than Stage 1. In expectation, they rank worse than in Stage 1. For subjects in sequence *HardEasy*, rank improves in expectation. We construct a Preliminary Rank 2 by comparing the overall individual correctly solved number of matrices to their distribution in the reference group. After completing the task in this stage, both Rank 1 and the Preliminary Rank 2 are displayed privately to each subject, so that subjects can compare their ranking in the two stages. While average Preliminary Rank 2 (that is calculated based on the performance on the same 48 matrices for all subjects) does not differ systematically across sequences, subjects' average reference point (Rank 1) will be better in sequence *EasyHard* than *HardEasy*. The purpose of the two sequences is thus twofold: first, to add an element of variation to subjects' reference points (Rank 1) in Stage 1 and second, to ensure a balanced data set in which about 50% of subjects will experience losses and gains in social image when moving from Stage 1 to 2.

[9]Examples of filled in and verified report sheets (in German) as well as their translations to English are shown in Appendix Figures 1.A.4 and 1.A.5 for Ranks 1 and 2, respectively.

Die reports After learning about their ranks, subjects are asked to throw a die twice and report the rolled numbers. The first reported number is then added to the number of correctly solved matrices in the reference group. The second reported number is added to a subject's own number of correctly solved matrices, giving the subjects two ways of cheating on the final reported rank that bear exactly the same consequences for their social image.

We use a modified version of the die roll task by Fischbacher and Föllmi-Heusi (2013).[10] Each subject rolls the die in private in the cubical so that no one, including the experimenters, can observe the actually rolled numbers.[11] Lying cannot be detected at the individual level in the die roll task. However, the underlying distribution of true die roll outcomes is known such that it can be observed whether and how much subjects lie on average as a group. Hence, we will conduct our main analysis on the group level, e.g., comparing subjects who experience gains and losses in social image.

Building on the work of Abeler et al. (2019), we use total lying costs which increase in the size of the lie to quantify utility changes due to changes of social image. Importantly, this approach enables us to isolate loss aversion in social image concerns. If subjects could pay to improve their final reported rank, paying money would induce a loss in the monetary domain and a gain in social image at the same time. Using lying costs to quantify utility changes due to changes of social image instead avoids the additional monetary domain of loss aversion and possible interaction effects with loss aversion in the social image domain that would make it impossible to isolate loss aversion in social image concerns.

Including two die rolls instead of only one has the advantage that subjects are not forced to over-report their Rank 2. With just one die roll, any reported rolled number would result in a better Final Rank 2 than Preliminary Rank 2. With two die rolls,

[10] In Fischbacher and Föllmi-Heusi (2013), subjects roll a die once, report on the rolled number (which does not necessarily need to be the truly rolled number), and are paid according to the reported number (i.e., higher numbers give a higher payoff except for 6, which pays zero). We build on the original die roll task but adjust it for our purposes in two aspects. First, instead of using monetary payoffs, we reward subjects with additional points which add up to the number of correctly solved matrices. Thus, lying enables subjects to improve their rank. Second, our subjects are told to throw the die twice.

[11] According to Gneezy, Kajackaite, et al. (2018), the fact that the experimenter cannot observe participants' true outcomes facilitates lying.

however, a subject's Final Rank 2 can either be better or worse than or equal to the Preliminary Rank 2, depending on whether subjects report a lower, higher or equal number to be added to the own score compared to the number to be added to the reference group's score.

In order to avoid that subjects' lying behavior depends on their beliefs on others' lying and to be able to interpret lying as a reflection of image concerns independent of individual beliefs, it is important to construct a ranking system which compares subjects to a predetermined reference group one by one. In contrast, if we based the ranking system on comparing subjects only within the current experiment (for example, ranking them from best to worst score), there would be an incentive to add a higher number to the own score if subjects expect others to add a high number to their score.

Further remarks Introducing observers instead of allowing subjects to report their rank to each other has two major advantages. First, our subjects do not get feedback on others' rank which could affect their perception of their own social image. Second, observers only know about the existence of a "further task" on top of the second quiz in Stage 2 and that the score in this task will feed into a subject's Final Rank 2. Observers are not informed about the exact nature of the die roll task, do not know how and to which extent the further task influences final ranks, and this is common knowledge to all subjects.[12] Consequently, subjects do not risk losing social image because of possible reputation cost of being seen as a liar. The remaining subjects receive the instructions regarding the die roll task on their computer screen after they have worked on Part 2 of the quiz.

Once the reported die rolls have been added and Final Rank 2 calculated, subjects go to observers again and report their Final Rank 2. After Stage 2, observers' information tables include, for each subject, the individual code, Final Rank 2, Rank 1 and the difference between Final Rank 2 and Rank 1. This is common knowledge for all subjects. Reporting procedures are the same as in Stage 1.

[12]The role of observers is passive: They are not allowed to communicate with subjects.

Procedural details and implementation Our experiment design and hypotheses are preregistered on AEA RCT Registry.[13] We conducted our experiment using zTree (Fischbacher, 2007b). After two pilot sessions as a prerequisite for power calculations, we run 19 main sessions in the DICE Lab, University of Düsseldorf between November 2018 and November 2019. 383 subjects participated, 38 as observers. Our sample mainly consists of a student population and was recruited using ORSEE (Greiner, 2015). 142 subjects were male, 203 were female. Age varied between 18 and 63 years with a median age of 23 years and 95% of subjects being younger than 33 years. No particular exclusion criteria applied. Subjects were randomized to sequences within each session.

All participants received a flat payment of 12 Euro, but no additional performance-contingent payment for correctly solving the matrices, which was clearly communicated to the subjects. Subjects' behavior thus indicates image concerns as a possible motive for exerting effort on solving the matrices correctly, even if this does not increase their monetary reward. On average, subjects earned €12.65, which includes the €12 flat payment plus one lottery outcome (as described below). In total, the experiment lasted about 90 minutes (including payment).

Post-experimental questionnaire The questionnaire provides information on socio-economic and demographic characteristics (age, gender, high school GPA, last math grade at school, student status and field of study, previous participation in experiments). It also assessed subjects' general willingness to take risks, based on a question from the German Socio-Economic Panel (GSOEP) questionnaire as well as the importance of social image, using the following question (similar to the one used by Ewers and Zimmermann (2015)): "How important is the opinion that others hold about you to you?". Additionally, following Gächter, E. J. Johnson, et al. (2021) and Fehr and Goette (2007), we measure loss aversion in the monetary domain using a set of incentivized lotteries which subjects can choose to accept or decline. Appendix 1.E provides the exact wording of the entire questionnaire.

[13] Petrishcheva, Vasilisa, Gerhard Riener, and Hannah Schildberg-Hörisch. 2019. "Loss Aversion in Social Image Concerns." AEA RCT Registry. April 09. https://doi.org/10.1257/rct.3422-5.0.

1.3 Hypotheses

We derive our hypotheses based on theoretical predictions described in Appendix 1.A. Our model integrates three key psychological features that—up to now have been treated separately—into individual utility: (1) agents gain positive utility from social image, (2) agents experience loss aversion in the social image domain, i.e., losses of social image loom larger than gains of the same size, and (3) agents dislike lying, i.e., they experience costs of misreporting the true state of the world. First, we show that individuals with social image concerns will not under-report, leading to Hypothesis 1.

Hypothesis 1. (Social-image relevance of task)

On average, subjects will weakly over-report their score.

In our experiment design, over-reporting implies that subjects report higher die rolls for themselves than for the reference group to be able to report a better Final Rank 2 to the observers. Since subjects have already been informed about their own Preliminary Rank 2 before their decision which die rolls to report, it seems plausible to assume that subjects can only misreport their rank to the observers but not lie to themselves. Over-reporting then establishes the relevance of social as opposed to self-image concerns for our subjects as a whole.

Hypothesis 2. (Loss aversion in social image concerns)

(a) Losses versus Gains: On average, subjects with sufficiently strong social image concerns over-report more if they experience a loss than a gain in social image.

(b) Discontinuity: There is a discontinuity in the extent of over-reporting at the reference point, i.e., when moving from losses to gains in social image.

We derive theoretical predictions for Hypothesis 2 in Appendix 1.A. Our model integrates three key psychological features that — up to now have been treated separately — into individual utility: (1) agents gain positive utility from social image, (2) agents experience loss aversion in the social image domain, i.e., losses of social image loom larger than gains of the same size, and (3) agents dislike lying, i.e., they experience costs of

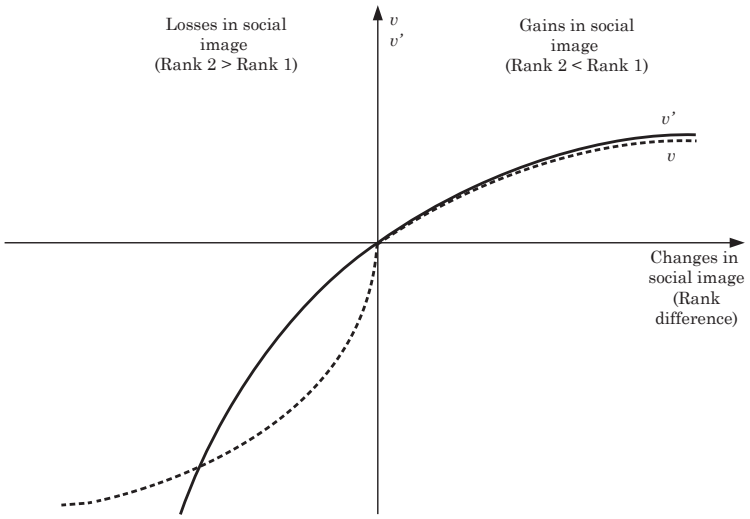
misreporting the true state of the world. We assume that the three components are additively separable. The value function of changes in social image satisfies the standard assumptions of prospect theory (Kahneman and Tversky, 1979): It is concave for gains, convex for losses and has a kink at the reference point. For simplicity, we assume that social image utility and lying costs are linear. Our predictions do not require linearity but make the model tractable. Hypothesis 2 follows directly from Proposition 1.

According to Hypothesis 2(a), we expect subjects who experience a loss in social image (i.e., Rank 2>Rank 1) to over-report more than subjects who experience a gain in social image (Rank 2<Rank 1). Over-reporting is reflected in the difference in die roll reports, i.e., the reported number to be added to own performance minus the reported number to be added to the reference group's performance. If, on average, this difference is higher for subjects experiencing losses than gains, this provides first evidence in line with loss aversion in social image concerns: on average, subjects who risk losing social image are ready to lie more than those with social image gains.

However, an alternative explanation for such a pattern is a simple concave utility function for changes in social image, which also implies that losses in social image induce stronger changes in utility than equally sized gains in social image.[14] For an illustration of a standard concave utility function, see the solid, black line in Figure 1.3. Losses in social image realize if Rank 2 is larger than Rank 1 that marks the intersection of both axes in Figure 1.3. In contrast, the dashed line depicts a value function that is compatible with the assumption of loss aversion.

Hypothesis 2(b) serves the purpose to differentiate between these two competing explanations for evidence in line with Hypothesis 2(a). The hypothesis is derived from a particularity in the shape of the value function as postulated in prospect theory. Figure 1.3 illustrates the existence of a kink in the value function of social image at a rank difference of zero when Rank 2 coincides with Rank 1. This kink implies a discontinuity in the first derivative of the value function when subjects move from the loss to the gain domain. We thus expect to observe a discontinuity in the extent of over-reporting as well

[14]See, e.g., Butera et al. (2022) who experimentally show that the social image utility is overall concave.



Note: We illustrate possible functions for changes in social image utility as realized at the end of Stage 2: the solid, black line a standard concave utility function v', the dashed line a value function v that is compatible with the assumption of loss aversion. The horizontal axis measures changes in social image. We define Rank 1 and Rank 2 as values between 0 and 100, with lower values corresponding to better performance. Negative values on the horizontal axis are hence realized if Rank 2>Rank 1 and stand for losses in social image, positive values on the horizontal axis are realized if Rank 2<Rank 1 such that subjects experience gains in social image.

Figure 1.3: Illustration of potential utility functions for changes in social image

when subjects move from losses to gains in social image. Since the value function's first derivative is higher for losses than gains close to the reference point, over-reporting should decrease. Evidence in line with Hypothesis 2(b) is compatible with loss aversion in social image concerns, but not a concave utility function for changes in social image.

1.4 Results

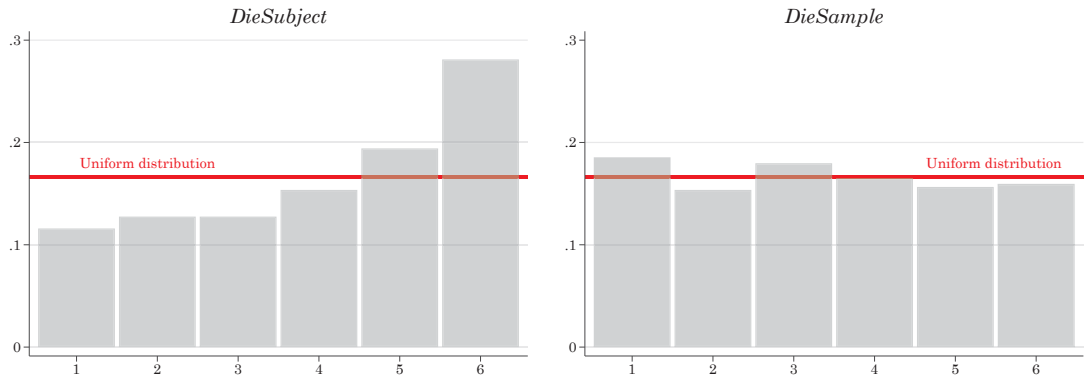
We will first establish that the matrices task is a source of social image-concerns, before we analyze how subjects react to losses as opposed to gains in social image.

1.4.1 Social image relevance of the matrices task

Subjects exerted substantial effort on the quizzes. They solved on average 38.8 out of all 48 matrices correctly. No subject solved less than 20 matrices, and more than 90% of subjects gave 34 or more correct answers. Since correct answers are not incentivized monetarily, substantial effort provision suggests image concerns as one of the driving

forces behind solving the matrices along with a potential intrinsic motivation for solving this type of tasks.

Subjects reported two values about their die rolls: The variable *DieSubject* which is added to their own score and the variable *DieSample* which is added to the scores of all subjects in the reference sample. In the absence of lying, die roll reports for each of the variables should follow a discrete uniform distribution with the support $\{1, \ldots, 6\}$ and an average of 3.5. Figure 1.4 displays histograms of *DieSubject* (left) and *DieSample* (right) as well as the probability density function of the uniform distribution (red line). The average of *DieSubject* is 4.03 and we reject the null hypothesis for the point prediction (t-test, H_0: $DieSubject = 3.5$, $p < 0.0001$).[15] The distribution of *DieSubject* is also highly significantly different from the discrete uniform distribution (Pearson's χ^2-test, $p < 0.0001$) and left-skewed. In contrast, the average of *DieSample* is 3.43 which is not significantly different from 3.5 (t-test, $p = 0.4614$). Moreover, the distribution of *DieSample* does not differ significantly from the discrete uniform distribution (Pearson's χ^2-test, $p = 0.881$).



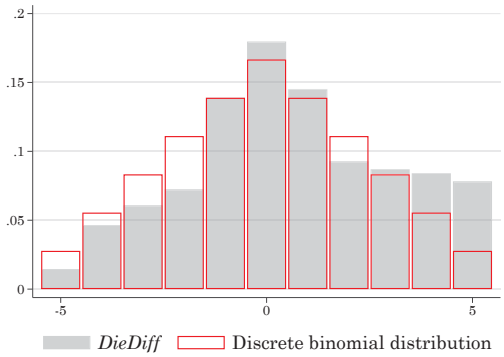
Note: Figures illustrate histograms of *DieSubject* (left) and *DieSample* (right). Horizontal axis indicates reported die rolls (from 1 to 6). Vertical axis indicates the fraction of subjects who reported the respective die rolls. Absent misreporting, die rolls should follow uniform distributions (red lines).

Figure 1.4: Distributions of *DieSubject* and *DieSample* by treatment

Subtracting *DieSample* from *DieSubject* results in the die roll difference, $DieDiff$, which indicates whether subjects improve or worsen their Final Rank 2 through reporting. The higher $DieDiff$, the better becomes Final Rank 2. In principle, $DieDiff$ can vary

[15] Throughout the paper, we report two-sided tests and refer to results as (weakly/highly) significant if the two-tailed test's p-value is smaller than 0.05 (0.10/0.01).

between -5 and 5, and, in the absence of lying, follows a discrete binomial distribution with zero mean. Our subjects report an average die roll difference of 0.59 which is highly significantly different from zero (t-test, $p < 0.0001$). As illustrated in Figure 1.5, the values of 4 and 5 are significantly over-reported (binomial probability tests, two-sided $p = 0.0253$ and $p < 0.0001$ for the values of 4 and 5, respectively). Thus, subjects lie both fully (maximal over-reporting) and partially (less than maximal over-reporting) which is in line with our theoretical predictions in Appendix 1.A.1 and experimental evidence of Gneezy, Kajackaite, et al. (2018) and Fischbacher and Föllmi-Heusi (2013). Over-reporting high values of $DieDiff$ provides further evidence that subjects perceive our matrices task as image-relevant and additionally shows that *social* image concerns matter: as all subjects know their Preliminary Rank 2, over-reporting their own score is unlikely to improve their self-image.[16]



Note: Figure illustrates histogram of $DieDiff$. Horizontal axis indicates a reported die roll difference (from -5 to 5, higher $DieDiff$ means adding more to one's own score). Vertical axis indicates the fraction of subjects who reported the respective die roll difference. Absent misreporting, die roll difference should follow the discrete binomial distribution (red outlines).

Figure 1.5: Reported die roll difference

Result 1. *Subjects report higher die rolls to be added to their own score than expected by rolling a fair die.*

This first set of results suggests that, on average, public reporting of own performance in the Raven's matrices induces social image concerns and that subjects engage in lying in order to report better ranks to the observers.

[16]Similarly, Burks et al. (2013) conclude that individuals' overstatement of own abilities is more likely induced by social as opposed to self-image concerns.

1.4.2 Gains and losses in social image

We now turn to the role of gains and losses in social image for reporting behavior. Obviously, loss aversion in social image can only be observed for those subjects who indeed care about their social image and do so sufficiently to bear the lying costs involved. While we have shown above that many of our subjects do over-report, it is also well documented that people are heterogeneous in the degree of social image concerns (see Bursztyn and Jensen, 2017; Friedrichsen and Engelmann, 2018) and lying costs (Abeler et al., 2019). This is also true in our sample, as Figure 1.A.6 in Section 1.B shows.

We are particularly interested in testing whether subjects with social image concerns are loss averse in social image. We therefore present three sets of results: (a) evidence from subjects with social image concerns, (b) evidence from subjects without social image concerns and (c) evidence for our sample as a whole. We classify subjects based on a median sample split on social image concerns as measured at the individual level through our survey instrument: "How important is the opinion that others hold about you to you?" (11-point Likert scale, social image concerns if answer 6 or higher).[17]

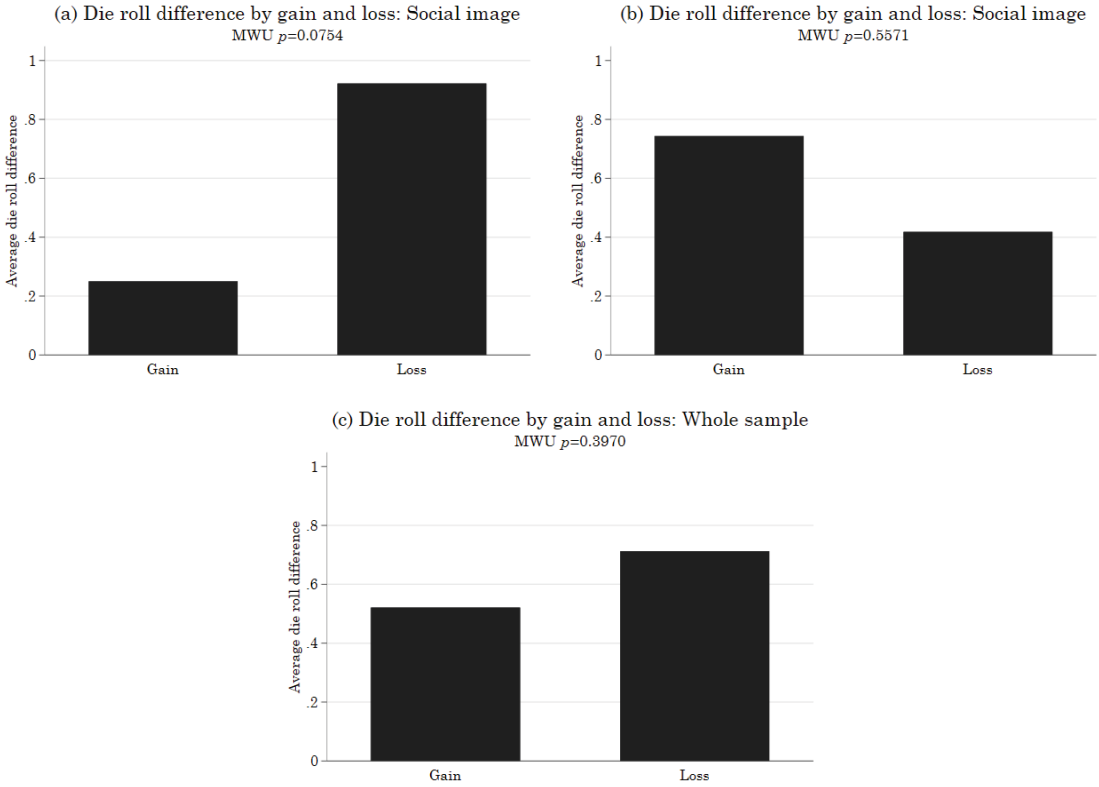
We push subjects into the gain or loss domain by randomly varying the sequence in which subjects performed the tasks. In 87.8 percent of the cases, we were successful in inducing losses and gains as intended by the respective sequence.[18] In the following, we will provide evidence based on whether subjects are in the loss or gain domain of social image, our subject of interest.

The gain-loss border In Figure 1.6, we compare reported die roll differences for subjects who experience gains and losses in social image. Positive rank differences are labeled as "Gain" indicating better performance in Part 2 than in Part 1, and negative

[17]Social image concerns do not differ significantly between *HardEasy* and *EasyHard* (MWU test, $p = 0.151$). Social image concerns are also not significantly correlated with Rank 1 and Preliminary Rank 2 ($p = 0.327$ and $p = 0.997$, respectively).

[18]There are 10 subjects with a rank difference of zero and 3 subjects with a negative rank difference in *HardEasy*. In *EasyHard*, 20 subjects have a zero rank difference and 19 subjects a positive rank difference. By introducing Gain and Loss, we reassign those overall 42 of 345 individuals to the intended category. Overall, 38.3% of subjects have a negative rank difference, 8.7% have a rank difference of zero and the remaining 53% have a positive rank difference. Subjects with a rank difference of zero are assigned to the Gain category. Hence, there are 213 out of 345 subjects in the Gain category and 132 out of 345 subjects in the Loss category.

rank differences as "Loss". As illustrated in Figure 1.6(a), subjects with image concerns who experience a loss in social image misreport more than those who experience a gain (MWU test, $p = 0.0754$), which is in line with Hypothesis 2(a). We see a similar, however statistically insignificant, pattern for the sample as a whole in Figure 1.6(b), i.e., irrespective of whether subjects care about their social image or not.[19]



Note: This Figure illustrates reported die roll differences for subjects who experience gains versus losses in social image. Vertical axis indicates average die roll difference (from -5 to 5, higher $DieDiff$ means adding more to one's own score). Absent misreporting, average die roll difference should be zero. Figure (a) shows differences for subjects with social image concerns, Figure (b) shows differences for subjects without social image concerns, and Figure (c) reports differences in the sample as a whole. Above each figure, we report MWU test results comparing distributions of $DieDiff$ for the respective groups. For Figure (a), the diffence in reported die roll difference between subjects with social image concerns in *Gain* and *Loss* is robust to the variation in social image concerns threshold (MWU test, $p = 0.086$ for 82 subjects who reported the importance of social image to be 8 or above, $p = 0.005$ for 146 subjects who reported the importance of social image to be 7 or above, $p = 0.075$ for 173 subjects who reported the importance of social image to be 6 or above, $p = 0.193$ for 209 subjects who reported the importance of social image to be 5 or above, and $p = 0.317$ for 238 subjects who reported the importance of social image to be 4 or above). The difference remains significant for subjects with strong image concerns and fades as we include additional subjects with weaker social image concerns.

Figure 1.6: Reported die roll difference by gains and losses in social image

Result 2. *On average, subjects with social image concerns over-report more if they experience a loss than a gain in social image.*

[19]We further present a robustness check that controls for ability in Table 1.A.1.

Assuming loss aversion in social image concerns and the standard shape of the value function, it is not surprising that differences in misreporting are not that large when comparing rank losses and gains of all sizes. As the value function depicted in Figure 1.3 illustrates, a further implication of the standard assumptions regarding the value function is that small rank losses and gains will induce the largest marginal changes in social image utility. We thus expect to observe larger differences in misreporting when comparing small losses and gains in rank, but only small differences for larger losses and gains in rank.

In order to differentiate between the two possible explanations of misreporting behavior—a concave utility function for changes in social image concerns versus loss aversion in social image concerns—, we proceed by taking a look at the behavior of subjects close to the gain-loss border. We present results from a regression discontinuity design in Table 1.1. The regression discontinuity specification maps the first derivative of the value function v which is commonly assumed to be larger for losses than for gains around zero and discontinuous at zero. Allowing for a discontinuity (RD) at a rank difference of zero (i.e., at the origin in Figure 1.3), we explore whether subjects report systematically different die roll differences when moving from the loss to the gain domain in social image. If we find such a significant discontinuity in the derivative of the value function at the rank difference of zero, the empirical approximation of the value function has a kink—as is generally assumed in prospect theory. In contrast, such a kink is not compatible with a standard concave utility function v' for changes in social image.

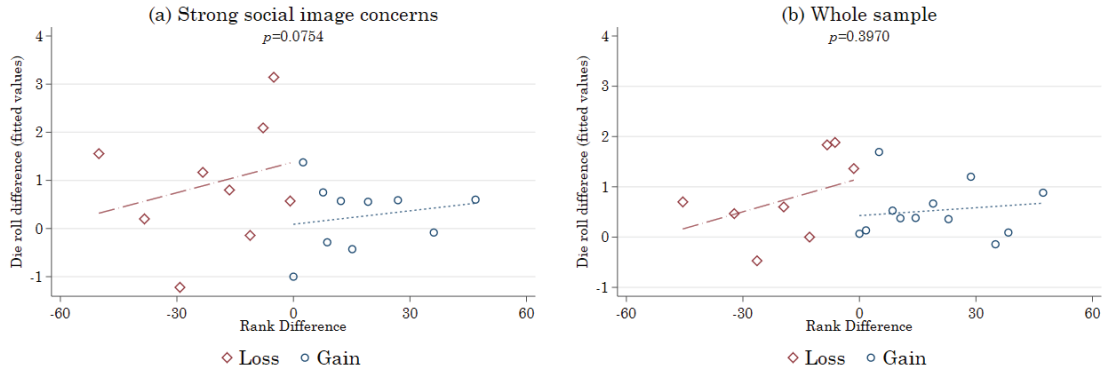
Table 1.1 indeed documents a significant discontinuity at the rank difference of zero, both for subjects with social image concerns and for the sample as a whole. Findings are similar in two different specifications: (i) an RD tobit specification focusing on subjects with rank differences between -10 and 10 in columns (1) and (3) and (ii) the robust procedure of Calonico et al. (2014) (CCT), employing the MSE-optimal bandwidth selection criterion in columns (2) and (4). On average, subjects below the threshold who experience a small loss in social image report 1.2–1.6 higher die roll differences than those above who experience a small gain in social image, see columns (3) and (4) in Panel A. For subjects with social image concerns, this discontinuity is even more pronounced: those below

Table 1.1: Regression discontinuity design

	Social image concerns		Whole sample	
	Tobit	CCT	Tobit	CCT
	(1)	(2)	(3)	(4)
Panel A: Without individual characteristics				
RD estimates	1.856***	1.975**	1.205**	1.599**
Clustered Std. Err.	(0.658)	(0.866)	(0.562)	(0.782)
Conventional p-value	0.006	0.023	0.034	0.041
Robust p-value		0.037		0.075
Number of obs.	66	94	123	190
Panel B: With individual characteristics				
RD estimates	2.235***	2.004***	1.228**	1.912**
Clustered Std. Err.	(0.514)	(0.698)	(0.517)	(0.818)
Conventional p-value	0.000	0.004	0.019	0.019
Robust p-value		0.005		0.035
Number of obs.	66	81	123	176

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ based on conventional p-values. Standard errors clustered at the session level in parentheses. Reported estimations in columns (1) and (2) refer to subjects who reported the importance of social image concerns to be 6 or higher and to the whole sample in columns (3) and (4). In columns (1) and (3), we estimate a two-limit tobit model focusing on subjects with rank differences between -10 and 10. In columns (2) and (4), we use local-linear estimators around a rank difference of zero with Epanechnikov kernels, the MSE-optimal bandwidth selection criterion (Calonico et al. (2014), CCT). For CCT, number of observations indicates the number of effective observations for an optimal bandwidth in a given regression. For model (2), total number of observations is 173; for model (4), 345. Individual characteristics include gender, age, squared age, field of study (indicators for economics, psychology as opposed to other), high school GPA, IQ (measured by Preliminary Rank 2) and measures for loss aversion in the monetary domain, intensity of social image concerns, and risk aversion. We present robustness checks of (1) and (2) for different thresholds of social image concerns in Figure 1.A.7.

the threshold report on average 1.9–2.0 higher die roll differences than those above, see columns (1) and (2) in Panel A. We illustrate this discontinuity in Figure 1.7 by plotting how the rank difference is associated with the die roll difference.



Note: Figures (a) and (b) illustrate the dynamics of die roll differences gain-loss size (rank difference) for different samples: (a) shows subjects with social image concerns and (b) the whole sample. Both Figures display differences between subjects who experience losses versus gains in social image. Each diamond represents the mean of the die roll difference and the rank difference withing each of the 20 equal-sized bins. Dotted and dashed lines represent a linear fit based on all the data from the respective sample for subjects with gains and losses in social image, respectively.

Figure 1.7: Die roll difference by Rank 1

The results from the RD design can be interpreted in a causal manner under the assumption that subjects just below and above the threshold (with rank differences of $[-10, 0]$ compared to $(0, 10]$) do not differ systematically in other dimensions than the one that defines the threshold. Using the comprehensive data from our post-experimental questionnaire[20], we establish in Table 1.2 that subjects do not differ significantly with respect to their extent of social image concerns, loss aversion in the monetary domain, risk aversion, field of study, final GPA at school, and fluid IQ (measured by Preliminary Rank 2).[21] This is true for both subjects with social image concerns and the sample as a whole. Differences in age are significant for the sample with social image concerns only. However, according to the results presented in the online appendix of the meta-analysis of Abeler et al. (2019), age is not a significant predictor of misreporting behavior when controlling for age and age squared as we do in our specifications. There are less female than male participants with rank differences of $[-10, 0]$ compared to $(0, 10]$. If we

[20]Exact variable definitions are provided in Appendix 1.E.

[21]The absence of significant differences in Preliminary Rank 2 implies that differences in rank differences are driven by differences in Rank 1. This is exactly what we intended by the design of the two matrix sequences.

Table 1.2: Individual characteristics around the gain-loss border

Rank difference	Social image [-10, 10]	Whole sample [-10, 10]
	(1)	(2)
Social image concerns	0.645	0.404
Loss aversion	0.472	0.186
Risk aversion	0.389	0.398
High school GPA	0.515	0.814
Fluid IQ (Preliminary Rank 2)	0.643	0.508
Field of study: Economics	0.419	0.506
Field of study: Psychology	0.673	0.727
Field of study: Other	0.391	0.376
Gender (1 if female)	0.202	0.020
Age	0.004	0.114

Note: We compare individual characteristics of subjects we use in the RDD (see column (1) of Table 1.1, i.e., those subjects with rank differences of [-10, 0) who experience a small loss to those with rank differences of [0, 10] who experience a small gain. For gender (1 if female, 0 else) and field of study (1 if economics or psychology or other, respectively, 0 else), we report p-values of Fisher's exact tests and of MWU tests for all other variables. Column (1) refers to subjects who report the importance of social image concerns to be 6 or higher, column (2) to the sample as a whole.

only consider those subjects with image concerns, the most relevant group under study, the difference in gender composition is no longer significant. Moreover, we do not find significant differences in misreporting by gender in our data.

Finally, Panel B of Table 1.1 that displays the regression discontinuity results when including all control variables confirms the significant discontinuity at the rank difference of zero; estimated coefficients remain rather stable. Subjects who experience a small loss in social image report 1.2–1.9 higher die roll differences than those who experience a small gain in social image; these numbers increase to 2.0–2.2 for subjects with social image concerns. In sum, we thus feel confident that a causal interpretation of our RD estimates is warranted.

Result 3. *We observe a significant discontinuity in over-reporting at the reference point, indicating a kink in the value function for social image as predicted by loss aversion.*

1.5 Conclusion

Does loss aversion apply to social image concerns? We observe that individuals who care about their reputation lie more if they are threatened by a loss than when facing a gain in social image. Taking a closer look at subjects' behavior when moving from losses to gains in social image, we find a sharp decrease in lying—providing evidence for loss aversion in social image irrespective of the individual extent of social image concerns.

More generally, our findings underline that loss aversion can also play a role in the non-material domain. While loss aversion is a well-established phenomenon for money and material goods (Kahneman, Knetsch, et al., 1991), our findings take a first step in a new line of research investigating the relevance of loss aversion to non-material sources of utility such as various drivers of reputation or self-image.

Since our experimental paradigm quantifies utility changes due to changes in social image by the amount of lying that individuals are willing to engage in, our findings also speak to the manifold situations in which honest reporting of private information is of great importance but not necessarily incentive-compatible. Dai et al. (2018) have shown that dishonesty in the lab can predict fraud and rule violation in real life. Our results reveal that individuals who care about their social image tend to report more dishonestly than others when their reputation is at stake. Monitoring efforts should thus be targeted at those individuals. One could also try to make it harder to lie while keeping a good reputation, e.g., via transparency, naming-and-shaming, or reputation systems (see also Abeler et al., 2019).

Finally, we find that the way social image evolves over time affects behavior. While making a decision, this reference-dependence implies that individuals may not only take present or discounted future reputation into consideration, but also account for the history of their social image. Two otherwise identical individuals may thus take opposite actions only due to differences in their social image in the past.

Appendix

1.A Theoretical Framework

Consider a two-period decision-making environment where $t \in \{1, 2\}$ indicates the period. In the first period, the agent receives a signal of her type s_1 that is communicated to herself and her peers. One can think of it as her social image relevant performance, and we assume that this signal establishes a reference point concerning her true type.

In the second period, she learns about her true type $\tilde{s}_2$, while peers are only going to see a signal of the true type s_2. This signal can be actively misrepresented in an unverifiable manner by the agent. In each period t, she derives $u(s_t)$ from the signal of her social image, where $u(\cdot)$ is assumed to be linear in s.

We model the cost of misrepresenting the true state following Abeler et al. (2019) and Khalmetski and Sliwka (2019). The true state of the world is $\omega \in [-\overline{\omega}, \overline{\omega}]$. The agent's report of the true state is $r \in [-\overline{r}, \overline{r}]$ with $\overline{\omega} = \overline{r} > 0$. In period $t = 2$, her final public signal s_2 consists of her actual performance $\tilde{s}_2$ plus her report of the true state r ($s_2 = \tilde{s}_2 + r$). The agent dislikes misreporting the true state and experiences lying costs $c(\omega, r)$. Lying costs are zero if the state is reported truthfully, i.e., $c(\omega, \omega) = 0$, and positive otherwise. Lying costs depend on the size of misreporting and are symmetric around ω, i.e., $c(\omega, \omega + a) = c(\omega, \omega - a)$ for all $a \in \mathbb{R}$. Moreover, we assume that lying costs are linear.[22] So we can write $c(\omega, r) = |\omega - r|$.[23] As the agent only makes a choice in period 2, we limit our attention to the utility in the second period:

$$\phi_2(r) = \theta^{social}[u(\tilde{s}_2 + r) + v(\tilde{s}_2 + r - s_1)] - \theta^{lying}|\omega - r|,$$

which she maximizes with respect to her report r. θ^{social} represents the sensitivity to social image that may differ across agents (Bursztyn and Jensen, 2017). θ^{lying} represents the

[22] Abeler et al. (2019) assume lying costs to be two-part, including a fixed cost of lying and a cost that is linear in the probability that an agent lied. Our analyses investigate the first-order derivatives and we can omit the fixed cost of lying by focusing on interior solutions, i.e., we consider cases where fixed costs of lying are not high enough to observe only truthful reporting.

[23] Note that in our model, agents follow a teleological moral theory that can be seen as a form of act consequentialism. In contrast, agents who adhere to a deontological normative moral reasoning would never engage in lying as it is considered a moral wrong, independent of the cost structure and the other parameters of the model. Also, in contrast to Abeler et al. (2019), we do not model social image concerns of being seen as a liar as we explicitly rule them out in our experimental design.

agent's sensitivity to lying (Gibson et al., 2013). We assume that social image utility is linear. The signals of ability, s_1 and $\tilde{s}_2$ are parameters,[24] hence utility in period 1 is fixed ($\phi_1 = \theta^{social} u(s_1)$), and we just consider the utility function in period 2 for maximization. We assume the following value function for changes in social image, which has a first derivative that is discontinuous at zero but differentiable otherwise:

$$v(s): \quad v(\Delta) < -v(-\Delta).$$

The value function satisfies the standard assumptions of prospect theory (Kahneman and Tversky, 1979). Negative deviations from the reference point s_1 have a larger absolute impact on utility than equally sized positive deviations, i.e., $v'(\Delta) < v'(-\Delta)$. Additionally, the value function is concave for gains ($v''(\Delta) < 0$ for $\Delta > 0$) and convex for losses ($v''(\Delta) > 0$ for $\Delta < 0$).

The first observation follows directly: An agent without social image concerns never misreports the true state. If $\theta^{social} = 0$, agent's utility in period 2 is reduced to $\phi_2 = -\theta^{lying}|\omega - r|$, which reaches its maximum when lying costs are minimized, i.e., in the absence of lying. Hence, an agent who does not care about her social image will always report truthfully: $r = \omega$.

From now on, we focus on agents with social image concerns, i.e., $\theta^{social} > 0$. The utility derived from social image is weakly increasing when the agent's report increases ($\partial u(\tilde{s}_2 + r)/\partial r \geq 0$) because the agent obtains positive marginal utility when the signal improves. $\partial v(\tilde{s}_2 + r - s_1)/\partial r > 0$ is independent of whether an agent is in the loss or gain domain (or shifts from the loss to the gain domain) with regard to social image. Lying costs are positive whenever the true state is misreported and $\partial c(\omega, r)/\partial r > 0$ if $\omega < r$ and $\partial c(\omega, r)/\partial r < 0$ if $\omega > r$.

The following observation is straightforward: An agent never underreports the true state. Given the true state ω, an agent always strictly prefers to report $r = \omega$ to any $\tilde{\omega} < \omega$

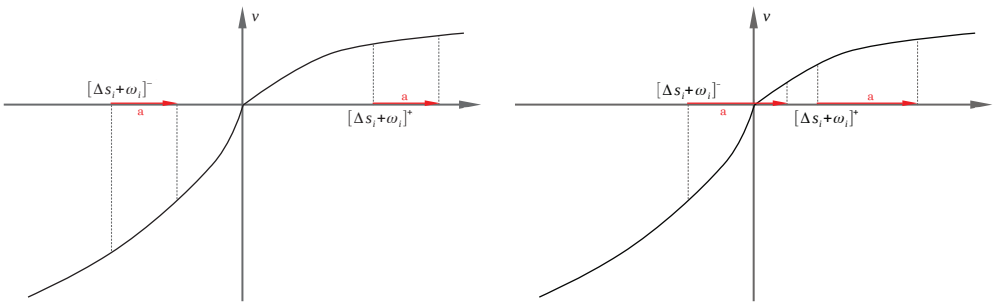
[24] In laboratory experiments, subjects tend to exert close to maximal effort in real-effort tasks (Araujo et al., 2016; Corgnet et al., 2015; Gächter, Huang, et al., 2016; Goerg et al., 2019). The same is true in IQ tests like the Raven's Progressive Matrices that we will use (Eckartz et al., 2012). Hence, we assume s_1 and $\tilde{s}_2$ to be parameters, not variables.

because under-reporting lowers utility due to three factors. First, an individual obtains weakly lower utility derived from social image: $u(\tilde{s}_2+\tilde{\omega}) \leq u(\tilde{s}_2+\omega)$. Second, the level of value function is lower at $\tilde{\omega}$ than at ω for any value of Δ, i.e., $v(\Delta+\tilde{\omega}) < v(\Delta+\omega)$, because $\partial v(\tilde{s}_2 + r - s_1)/\partial r > 0$. Third, reporting $r = \omega$ yields zero lying costs while reporting $\tilde{\omega}$ misreports the true state, which is costly, i.e., $c(\omega, \tilde{\omega}) > c(\omega, \omega)$. Additionally, if $\omega = \overline{\omega}$ and an agent does not under-report, it directly follows that the agent will report truthfully (i.e., $r = \omega$).

Comparing Gains and Losses We derive the level of optimal misreporting behavior for agents who experience gains and losses in social image in Section 1.A.1. Our main interest, however, concerns behavior closely around the reference point.

Proposition 1. *There is more incentive to lie if an agent experiences a loss rather than a gain in social image of the same size. There is a discontinuity in lying at the reference point.*

Proof. We compare cases denoted $(\Delta+\omega)^+$ and $(\Delta+\omega)^-$ in which $(\Delta+\omega)^+ = -(\Delta+\omega)^-$. Those cases are driven by changes in s_1 or ω, i.e., holding $\tilde{s}_2$ constant, and they both imply zero lying costs and symmetry. We assume that an agent makes a lying decision after observing the true state ω. We illustrate the proof in Figure 1.A.1. Since agents will not lie downwards, we only consider the case of $r \geq \omega$. We know that for $r = \omega$ the following holds:



Note: We display a value function that is in line with standard assumptions of prospect theory (Kahneman and Tversky, 1979) to illustrate the intuition of the proof of Proposition 1. In the top figure, we show the case of sufficiently small a, such that an agent in the loss domain remains in the loss domain after reporting $r = \omega + a$. In the bottom figure, we present a case of a sufficiently large a: In that case, an agent in the loss domain who reports $r = \omega + a$ switches to the gain domain.

Figure 1.A.1: Illustration of a value function

$$v'\left((\Delta+\omega)^{+}\right) < v'\left((\Delta+\omega)^{-}\right). \tag{1.1}$$

Moreover, the value function is convex for losses, i.e., for any $a > 0$ it is true that

$$v'\left((\Delta+\omega)^{-}\right) < v'\left((\Delta+\omega+a)^{-}\right),$$

and concave for gains, such that

$$v'\left((\Delta+\omega)^{+}\right) > v'\left((\Delta+\omega+a)^{+}\right).$$

Then Condition 1.1 also holds for $r = \omega + a$:

$$v'\left((\Delta+r)^{+}\right) < v'\left((\Delta+r)^{-}\right) \tag{1.2}$$

and therefore reporting $r = \omega + a > \omega$ is more attractive if an agent is in the loss domain than the gain domain. Note that if a is sufficiently large, $v\left((\Delta+\omega)^{-}\right) < 0$ but $v\left((\Delta+r)^{-}\right) > 0$ which means that the agent has been in the loss domain before reporting but has entered the gain domain by over-reporting. Condition 1.2 still holds in this case. Additionally, as $(\Delta+r) \to 0$, i.e., the change in social image becomes marginal, Condition 1.2 still holds strictly. Therefore, we observe a discontinuity in over-reporting at the reference point. □

To summarize, our model predicts that an agent with social image concerns never under-reports the true state. If an agent cares about her social image and the true state is not the best possible, she might engage in misreporting. Importantly, an agent has more incentives to misreport her true state if she experiences a loss in social image than a gain in social image of the same size and we expect to observe a discontinuity in over-reporting at the reference point for social image.

1.A.1 Optimal misreporting

In the following, we analyze behavior with respect to gains and losses in social image. We assume that subjects care about social image — $\theta^{social} > 0$ — and define $\theta = {}^{\theta^{lying}}/_{\theta^{social}}$ which expresses the agent's relative sensitivity to lying. We restrict the following analysis to over-reporting since we have already established that under-reporting will not occur. Following Fischbacher and Föllmi-Heusi (2013), we study the conditions under which an agent engages in full and partial lying. In the following, we refer to "full lying" whenever an agent reports $r = \overline{\omega} > \omega$ and to "partial lying" whenever an agent reports $\omega < r < \overline{\omega}$.

An agent makes a decision about r by comparing marginal benefits $(\partial u/\partial r + \partial v/\partial r)$ and marginal costs θ of misreporting. For $r > \omega$, θ and $u' = \partial u/\partial r$ are constant, so the agent's trade-off boils down to comparing a constant $C = \theta - u'$ to $\partial v/\partial r$, which varies with r.

We discuss all the cases for gains in social image in Section 1.A.1 and illustrate them Figure 1.A.2. In Section 1.A.1, we consider all the cases for losses in social image and illustrate them in Figure 1.A.3. We display C and $\partial v/\partial r$ on the vertical axis and $r - \omega$ on the horizontal axis. The horizontal axis shows normalized reports: zero on the horizontal axis corresponds to truthful reporting and positive values on the horizontal axis correspond to over-reporting. We denote an interior solution for reporting $\tilde{r}$.

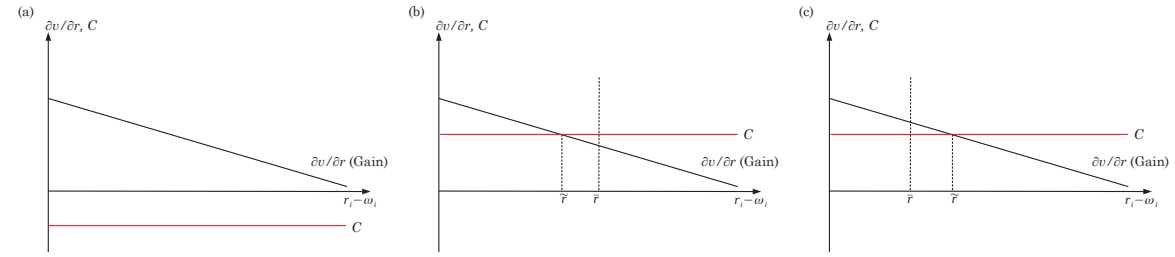
Gain in social image

What happens if the agent finds herself in the gain domain after learning about her true type $\tilde{s_2}$? Positive misreporting may only further increase the gain in social image, as $\partial v/\partial r$ is positive and $\partial^2 v/\partial^2 r$ is negative. We show that there exist threshold levels of θ that determine the extent of lying. We derive the following proposition:

Proposition 2. *An agent who experiences a gain in social image*

- *reports truthfully if* $\theta \geq \theta^{true}_{gain}$,
- *lies fully if* $\theta \leq \theta^{full}_{gain}$, *and*
- *lies partially if* $\theta \in (\theta^{full}_{gain}; \theta^{true}_{gain})$.

We provide a case-by-case proof of Proposition 2 in Lemmas 1, 2 and 3.



Note: C and $\partial v/\partial r$ are on the vertical axis and the normalized reports $(r-\omega)$ are on the horizontal axis. Zero on the horizontal axis corresponds to truthful reporting and positive values on the horizontal axis correspond to over-reporting. In Figures (a) and (c) we illustrate cases where full over-reporting $(r=\bar{r})$ is optimal. In Figure (b) we illustrate the case where partial over-reporting $(r=\tilde{r}<\bar{r})$ is optimal.

Figure 1.A.2: Gain in social image concerns: Partial versus full lying

Lemma 1. *An agent who experiences a gain in social image lies fully if* $\theta \leq \theta_{gain}^{full}$.

Proof. The agent chooses to lie fully if reporting $r=\overline{r}=\overline{\omega}$ yields marginal costs that are the same or lower than the marginal benefits of lying:

$$\underbrace{\theta - u'}_{C} \leq \underbrace{\frac{\partial v(\tilde{s}_2+\overline{\omega}-s_1)}{\partial r}}_{\partial v/\partial r}.$$

By rearranging with respect to θ we get

$$\theta \leq \theta_{gain}^{full} = \left(\frac{\partial v(\tilde{s}_2+r-s_1)}{\partial r}\bigg|_{r=\overline{\omega}}^{-} + u' \right).$$

Since $\partial v/\partial r$ is strictly decreasing for agents who experience a gain in social image and C remains constant, the maximization problem is always concave. Therefore, if the agent is sufficiently insensitive to lying, i.e., $\theta \leq \theta_{gain}^{full}$, she will lie fully $(r=\overline{\omega})$. □

Lemma 2. *An agent who experiences a gain in social image reports truthfully if* $\theta \geq \theta_{gain}^{true}$.

Proof. An agent with a gain in social image will engage in misreporting if

$$\theta < \theta_{gain}^{true} = \left(\frac{\partial v(\tilde{s}_2+r-s_1)}{\partial r}\bigg|_{r=\omega} + u' \right).$$

θ_{gain}^{true} indicates a threshold lying sensitivity: An agent with $\theta \geq \theta_{gain}^{true}$ prefers truthful reporting because costs of lying outweigh the benefits of reporting the true state $r=\omega$.

□

Additionally, if the agent's sensitivity to lying is not low enough to lie fully but not high enough to report truthfully, she will engage in partial misreporting.

Lemma 3. *An agent who experiences a gain in social image lies partially if* $\theta \in (\theta_{gain}^{full}; \theta_{gain}^{true})$.

Results from Lemmas 1, 2 and 3 lead to Proposition 2.

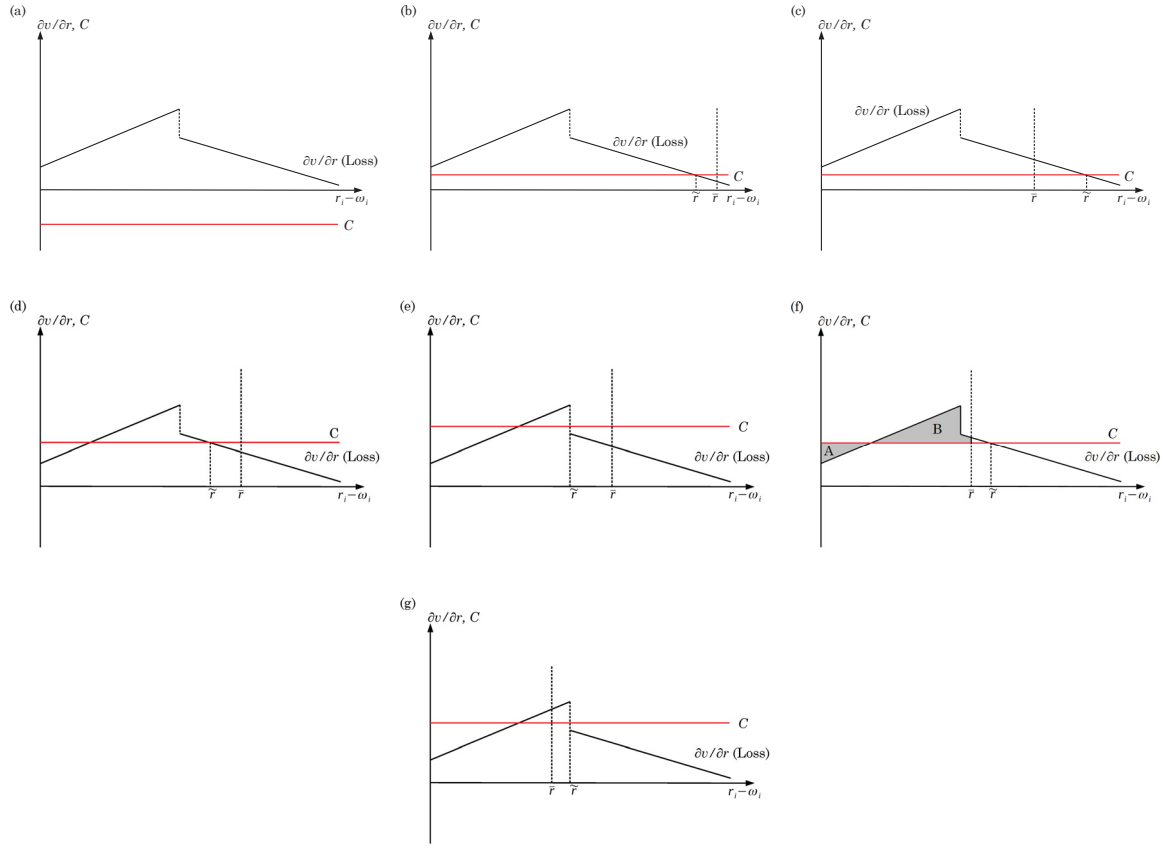
Loss in social image

In this subsection, we consider an agent who experiences a loss in social image. Importantly, positive misreporting may lead to various consequences in case of a loss in social image, namely, it may (a) decrease an existing loss in social image, (b) fully eliminate an existing loss in social image and (c) fully eliminate an existing loss and induce a gain in social image. Hence, $\partial v/\partial r$ is discontinuous and consists of two pieces as shown in Figure 1.A.3. The value function is convex in losses. Therefore, for small $(r - \omega)$, $\partial v/\partial r$ is positive and increasing which indicates that the agent moves from a larger loss to a smaller loss in social image. The discontinuity point corresponds to the social image loss being fully eliminated by overreporting such that the agent does neither experience a loss nor a gain in the social image domain. However, increasing $(r - \omega)$ even more puts the agent in the gain domain in social image. The value function is concave for gains in social image, hence, $\partial v/\partial r$ remains positive but becomes decreasing. In the gain domain, lying becomes instantly relatively less attractive due to discontinuity of the value function at the reference point. We derive the following conditions for lying:

Proposition 3. *An agent who experiences a loss in social image*

- *reports truthfully if* $\theta \geq \theta_{loss}^{true}$ *or* $\phi_2(\omega) \geq \phi_2(\tilde{r})$,
- *lies fully if* $\theta < \theta_{loss}^{full}$ *or* $\phi_2(\omega) < \phi_2(\bar{r})$, *and*
- *lies partially otherwise.*

We discuss agents' incentives to lie case by case. First, we consider a case with $C \leq 0$ as shown in Figure 1.A.3(a).



Note: C and $\partial v/\partial r$ are on the vertical axis and the normalized reports $r-\omega$ are on the horizontal axis. Zero on the horizontal axis corresponds to truthful reporting and positive values on the horizontal axis correspond to over-reporting. In Figure (a) we illustrate the case where $C \leq 0$ and full over-reporting is optimal. Figure (b) shows the case with $C \in (0, \partial v/\partial r(r=\omega)]$ and $\bar{r} \geq \tilde{r}$, where partial over-reporting is optimal. Figure (c) shows the case with $C \in (0, \partial v/\partial r(r=\omega)]$ and $\bar{r} < \tilde{r}$, where full over-reporting is optimal. In Figures (d)-(g), we present cases with $C \in (\partial v/\partial r(r=\omega), \partial v/\partial r(r=s_1-\tilde{s}_2)]$. We highlight areas A and B to demonstrate the intuition for identifying the global maximum, i.e., the comparison of $\phi_2(\omega)$ and $\phi_2(\tilde{r})$. In Figure (d), it is optimal to over-report partially because $\tilde{r} < \bar{r}$ and $\phi_2(\omega) < \phi_2(\tilde{r})$ (area A is smaller than area B). In Figure (e), it is optimal to report truthfully because $\tilde{r} < \bar{r}$ and $\phi_2(\omega) \geq \phi_2(\tilde{r})$ (area A is larger than area B). In Figure (f), agents' optimal strategy is to over-report fully because $\tilde{r} \geq \bar{r}$ and $\phi_2(\omega) < \phi_2(\bar{r})$ (area A is smaller than area B). Finally, in Figure (g), the optimal strategy is to report truthfully because $\tilde{r} \geq \bar{r}$ and $\phi_2(\omega) \geq \phi_2(\bar{r})$ (area A is larger than area B).

Figure 1.A.3: Loss in social image concerns: Partial versus full lying

Lemma 4. *If $C \leq 0$, an agent who experiences a loss in social image lies fully.*

Proof. C is non-positive whenever the marginal utility gain from misreporting u' outweighs the marginal lying costs θ even without taking into consideration additional marginal benefits from the value function $\partial v/\partial r$. Therefore, lying comes at relatively low costs and the agent chooses to misreport maximally. Her optimal report is then $r = \bar{\omega}$. □

Next, we consider a case shown in Figures 1.A.3(b) and 1.A.3(c), namely, $C(r=\omega) \leq$

$\partial v/\partial r(r=\omega)$.

Lemma 5. *If $C \in (0, \partial v/\partial r(r=\omega)]$, an agent who experiences a loss in social image lies fully if $\bar{r} \leq \tilde{r}$, and lies partially otherwise.*

Proof. For any $r < \tilde{r}$, $\partial v/\partial r$ is always larger than C and hence $r = \tilde{r}$ is a global maximum. If $\tilde{r} \leq \bar{r}$ as in Figure 1.A.3(b), the agent reports $r = \tilde{r}$. Otherwise, she chooses the highest possible report $r = \bar{r}$ as in Figure 1.A.3(c). □

We proceed to cases shown in Figures 1.A.3(d)-1.A.3(g). The results are summarized in the following Lemma:

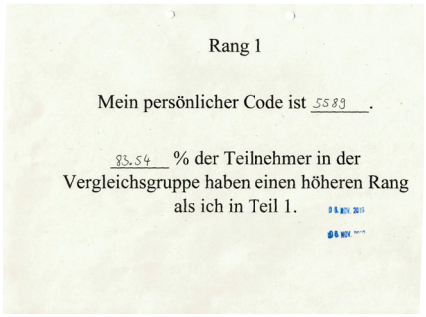
Lemma 6. *If $C \in (\partial v/\partial r(r=\omega), \partial v/\partial r(r=s_1-\tilde{s}_2)]$, an agent who experiences a loss in social image reports truthfully if $\phi_2(\omega) \geq \min(\phi_2(\tilde{r}); \phi_2(\bar{r}))$, lies fully if $\phi_2(\omega) < \phi_2(\bar{r})$ and $\tilde{r} \geq \bar{r}$, and lies partially otherwise.*

Proof. For $C \in (\partial v/\partial r(r=\omega), \partial v/\partial r(r=s_1-\tilde{s}_2)]$, we can no longer be sure that reporting $\tilde{r}$ or $\bar{r}$ yields the global maximum because $\partial v/\partial r$ is no longer larger than C for any $r < \tilde{r}$. In contrast, the agent strictly prefers truthful reporting to ϵ-misreporting. Hence, we should (a) consider whether the agent is able to report $\tilde{r}$ and can only report at most $\bar{r}$, and (b) additionally focus on whether she chooses the optimal misreporting or no misreporting at all.

If $\tilde{r} \leq \bar{r}$, the agent faces a trade-off between reporting truthfully ($r = \omega$) and lying partially ($r = \tilde{r}$). She chooses to report truthfully if $\phi_2(\omega) \geq \phi_2(\tilde{r})$. If $\tilde{r} > \bar{r}$, the agent has a choice between full overreporting ($r = \bar{r}$) and truthful reporting ($r = \omega$). Analogously, she reports truthfully if $\phi_2(\omega) \geq \phi_2(\bar{r})$. □

Finally, taking all the cases from Lemmas 4, 5 and 6 into consideration, we formulate Proposition 3. The thresholds θ^{true}_{loss} and θ^{full}_{loss} are derived analogously to θ^{true}_{gain} and θ^{full}_{gain}. However, since these thresholds depend on changes in social image, we mark them "gain" and "loss" to indicate that this difference.

1.B Additional Figures


Rang 1

Mein persönlicher Code ist 5589.

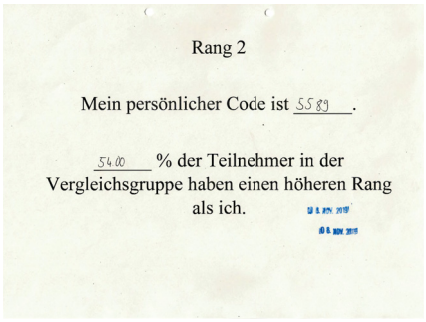
83,54 % der Teilnehmer in der Vergleichsgruppe haben einen höheren Rang als ich in Teil 1.

Rank 1

My individual code is _____.

_____ % of the participants of the reference group have a higher rank than me in part 1.

Figure 1.A.4: Rank 1 report sheet (original in German and translated to English)


Rang 2

Mein persönlicher Code ist 5589.

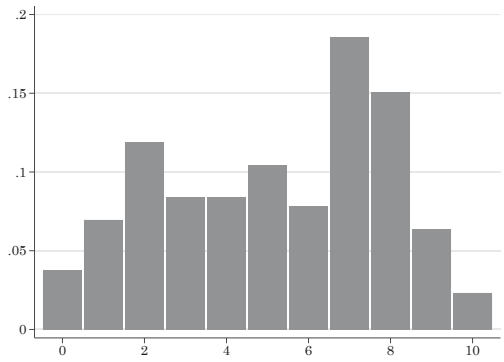
54,00 % der Teilnehmer in der Vergleichsgruppe haben einen höheren Rang als ich.

Rank 2

My individual code is _____.

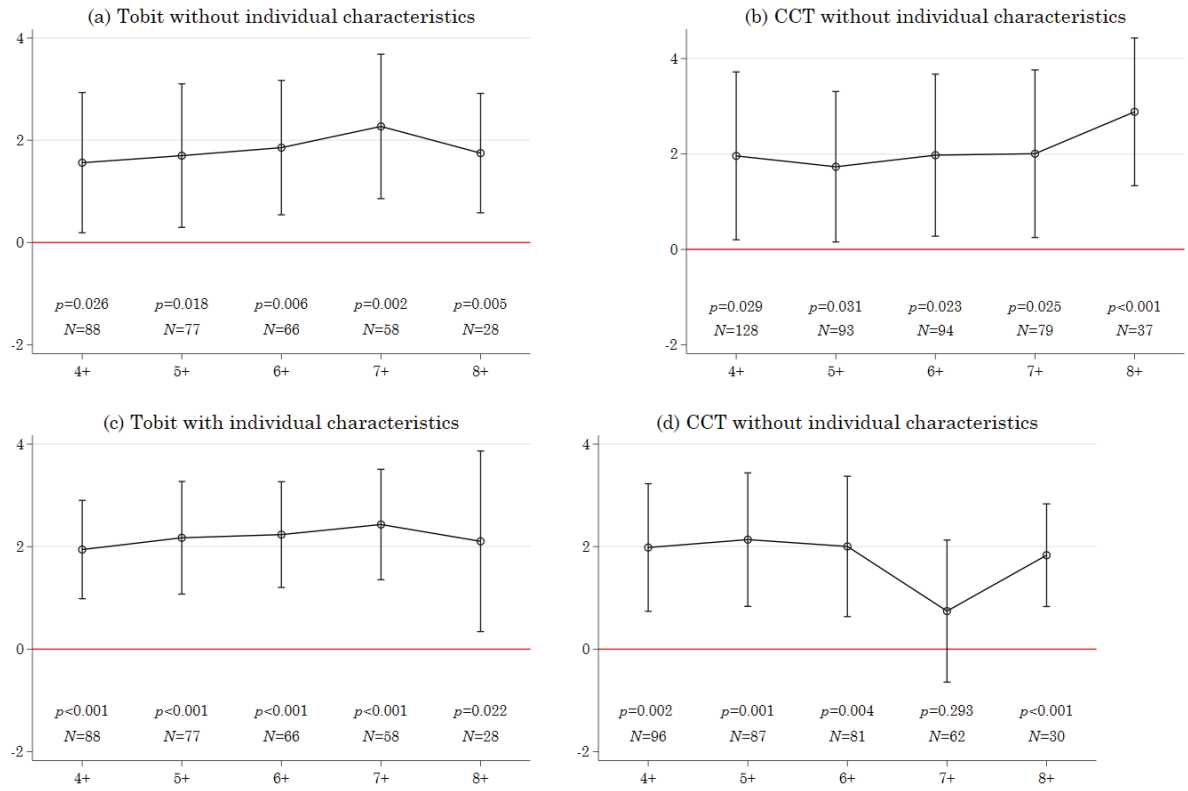
_____ % of the participants of the reference group have a higher rank than me.

Figure 1.A.5: Rank 2 report sheet (original in German and translated to English)



Note: Importance of social image concerns is measured on a 11-point Likert scale based on the question "How important is the opinion that others hold about you to you?".

Figure 1.A.6: Self-reported importance of social image



Note: Figures (a) and (c) display the robustness checks for (1) in Panels A and B of Table 1.1, respectively. Figures (b) and (d) display the robustness checks for (2) in Panels A and B of Table 1.1, respectively. The RD estimates are shown on the vertical axis. Error bars correspond to the 95% confidence intervals of the estimates. In Figures (b) and (d), we report conventional CCT estimates and p-values. The horizontal axis indicates various subsamples based on the measure of social image concerns. "4+" indicates the subsample of subjects who reported the importance of social image concerns to be 4 or above, "5+" indicates 5 or above, and so on. "6+" corresponds to the original median split results reported in Table 1.1. In columns (1) and (3), we estimate a two-limit tobit model focusing on subjects with rank differences between -10 and 10. In Figures (b) and (d), we use local-linear estimators around a rank difference of zero with Epanechnikov kernels, the MSE-optimal bandwidth selection criterion (CCT). For CCT, number of observations indicates the number of effective observations for an optimal bandwidth in a given regression. Individual characteristics include gender, age, squared age, field of study (indicators for economics, psychology as opposed to other), high school GPA, IQ (measured by Preliminary Rank 2) and measures for loss aversion in the monetary domain, intensity of social image concerns, and risk aversion.

Figure 1.A.7: Robustness checks: Social image threshold

1.C Additional Tables

Table 1.A.1: Robustness check: Proxy for ability

Panel A: All rank differences		
	Social image concerns	Whole sample
	(1)	(2)
Loss	1.610**	0.892
	(0.772)	(0.553)
Rank 2	0.013	0.008
	(0.009)	(0.007)
Loss × Rank 2	-0.020	-0.016
	(0.014)	(0.011)
Constant	-0.254	0.265
	(0.426)	(0.315)
Number of obs.	173	345
Panel B: Rank differences between -10 and 10		
	Social image concerns	Whole sample
	(3)	(4)
Loss	2.112**	1.726**
	(0.955)	(0.697)
Rank 2	0.003	0.012
	(0.011)	(0.009)
Loss × Rank 2	-0.007	-0.014
	(0.016)	(0.013)
Constant	0.127	0.190
	(0.517)	(0.422)
Number of obs.	66	123

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust errors are in parentheses. We include Rank 2 as a proxy for subjects' ability because it corresponds to their performance in the IQ test. Reported estimations in columns (1) and (3) refer to subjects who reported the importance of social image concerns to be 6 or higher and to the whole sample in columns (2) and (4). In columns (1) and (2), we estimate a two-limit tobit model focusing on subjects with all rank differences. In columns (3) and (4), we estimate a two-limit tobit model focusing on subjects with rank differences between -10 and 10 (small gains and losses in social image).

1.D Instructions of the Experiment

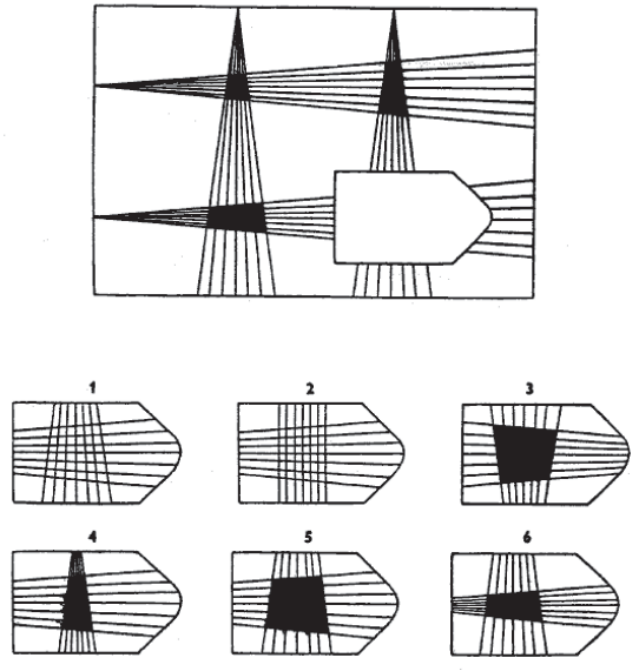
1.D.1 English

General Instructions

We warmly welcome you to this economic experiment. Please read the following instructions carefully! If you have any questions, please raise your hand from the cubicle—we will then come to your seat. It is not allowed to talk to other participants of the experiment, use mobile phones or start other programs on the computer during the experiment. Non-compliance with these rules will result in exclusion from the experiment and all payments. You will receive a fixed payment of €12 for participating in this experiment, which will be paid in cash at the end of the experiment. On the following pages we describe the exact procedure of the experiment.

Part 1 of the Experiment

Parts 1 and 2 consist each of 24 tasks, which are often used to measure so-called fluid intelligence of a person. The fluid intelligence is an important part of the general intelligence of humans. These or similar tasks are also often used by companies in the context of recruitment procedures. Each task corresponds to a picture puzzle. Here you can see an example:



Each picture puzzle shows in its upper part a pattern in a box, in which a "piece of the puzzle" in the lower right corner is left out. Your task is to select one of the puzzle pieces listed below the box, which will logically fill the blank lower right corner of the pattern in the box. Please enter the number of the puzzle piece that you think fits best on the screen. The number of a puzzle piece is stated above each puzzle piece. There is always exactly one piece that fits best.

You have 30 seconds to complete each picture puzzle. For each correctly completed picture puzzle you receive one point. As commonly done with intelligence tests, correct answers are not paid extra. You will receive 0 points for each wrongly answered picture puzzle or if you do not enter the best fitting piece of the puzzle within 30 seconds.

After you have completed all 24 picture puzzles in Part 1, you will first receive a private feedback on your rank on the computer screen, indicating how well you performed in solving the picture puzzles. The feedback has the following form: "X % of the participants of the reference group have a higher rank than you in Part 1". The reference group consists of 413 participants of a previous laboratory experiment conducted in 2014 here at the DICE Lab of the University of Düsseldorf, who have worked on the same picture puzzles as you do in the course of this experiment. So the feedback "9% of the participants of the reference group have a higher rank than you in Part 1" means that 9% performed better than you (i.e., solved a higher percentage of the total 48 picture puzzles from Parts 1 and 2 correctly than you) and 90% performed worse (i.e., solved a lower percentage of picture puzzles correctly than you). So you belong to the 10% of the best at answering the picture puzzles designed to measure individual fluid intelligence. The feedback "83% of the participants of the reference group have a higher rank than you in Part 1" means that 83% performed better and 16% worse than you. So you are among the 17% of the worst in answering the picture puzzles.

Before Part 2 of the experiment starts, you have to inform two so-called "Observers" about your performance in the experiment. Please use the report sheet available in your cubicle. Your cubicle number is already entered. Please enter legibly the number, which you received as feedback on the computer screen, in the sentence "___ % of the participants

of the reference group have a higher rank than me in Part 1" in the report sheet "Rank 1". Please enter your personal code, which is also displayed on the screen, in the free field next to it: "My personal code is ___". Observers sit in the cubicles number 1 and 2 in the laboratory (directly in front of the entrance door). Please go there with the completed report sheet and show it silently to Observers as soon as your cubicle number is called by the experimenter. This ensures that each participant informs Observers individually without other participants knowing her/his rank. A two-column table will be displayed on the Observers' computer screens, assigning each personal code the corresponding rank in Part 1. Each Observer will silently compare your report sheet with the information in the table and stamp it. Afterwards, please return to your cubicle in silence. Part 2 of the experiment will begin as soon as all participants have informed Observers of their rank.

The Different Participants in the Experiment

At the beginning of the experiment, each participant randomly drew a chip with a number indicating his cubicle number. The cubicle numbers have the following additional meaning: The participants who have randomly drawn cubicle numbers 1 and 2 have the role of "Observers" described above. Since the chips with even numbers were reserved for female participants and the chips with odd numbers for male participants, there is always one male Observer and one female Observer. These will introduce themselves to you shortly before the actual experiment begins by standing up and saying "I am one of the two Observers". Observers—just like all other participants—will receive this printed explanation of the rules of the experiment, which you are reading, for information about the experiment.

All other participants in the experiment with cabin numbers 3 or higher solve the picture puzzles described above. Each participant is randomly assigned to one of two groups: Group A or Group B. Throughout the whole experiment, all participants of both groups will solve exactly the same 48 picture puzzles, 24 in Part 1 and 24 in Part 2. The further task in part 2 of the experiment is also exactly the same for both groups. Only the order in which the picture puzzles are processed differs between group A and B. The group membership has no further meaning. In Parts 1 and 2 you belong to the same

group.

Part 2 of the Experiment

Part 2 of the experiment is very similar to Part 1. First you work on 24 more picture puzzles following the same rules (30 seconds time per puzzle, 1 point for correct answers, 0 points otherwise, etc.). After you have completed remaining 24 picture puzzles in Part 2, you will receive a private feedback on your preliminary rank in Part 2 on the computer screen, indicating how well you have done in the 48 picture puzzles in Parts 1 and 2. The feedback again has the following form: "X% of the participants of the reference group have a higher rank than you". The reference group is again the 413 participants of a previous lab experiment here in the DICE Lab of the HHU from 2014, who have solved the same 48 picture puzzles as you. In addition, the rank you had in Part 1 of the experiment is displayed as a reminder.

The only difference to Part 1 is that you have one more task, which is also used to calculate your final rank in Part 2. You will then receive a private feedback on your final rank in Part 2, which is calculated based on the 48 picture puzzles in Parts 1 and 2 and your score in the further task in Part 2. Details of the further task and how exactly it is included in the calculation of the final rank in Part 2 will be explained on the computer screen during the course of the experiment. For calculation of your final rank the same reference group is used again as for your rank in Part 1 and the preliminary rank in Part 2. The detailed explanations of the further task in Part 2 are given only to the participants, but not to the two Observers.

Just like at the end of Part 1, you still have to inform the two Observers about your performance, i.e., your final rank, in Part 2. Please use the report sheet which is available in your cubicle. In addition, under "Rank 2", please enter legibly in the sentence "___ % of the participants of the reference group have a higher rank than me", which you have received as feedback on your final rank on the computer screen. Please go to two Observers with the completed report sheet and show it to them in silence as soon as your cubicle number is called up by an experimenter. This again ensures that each participant informs Observers individually without the other participants knowing her/his rank. A

table with four columns is now displayed to Observers on your computer screen, which assigns to each personal code the corresponding rank in Part 1, the final rank and the difference in rank between the rank in Part 1 and the final rank.

The observers will, also in silence, compare your report sheet with the information in the table and stamp it. Afterwards, please return to your cabin in silence.

End and Payment of the Experiment

After Part 2 of today's experiment, there will be some more screens with questions before we proceed to the payment of €12. We will call you individually by cubicle number for payment. If you have any questions now, please raise your hand out of the cubicle. Experiment supervisor will then come to your seat to answer your questions. Do not ask questions out loud!

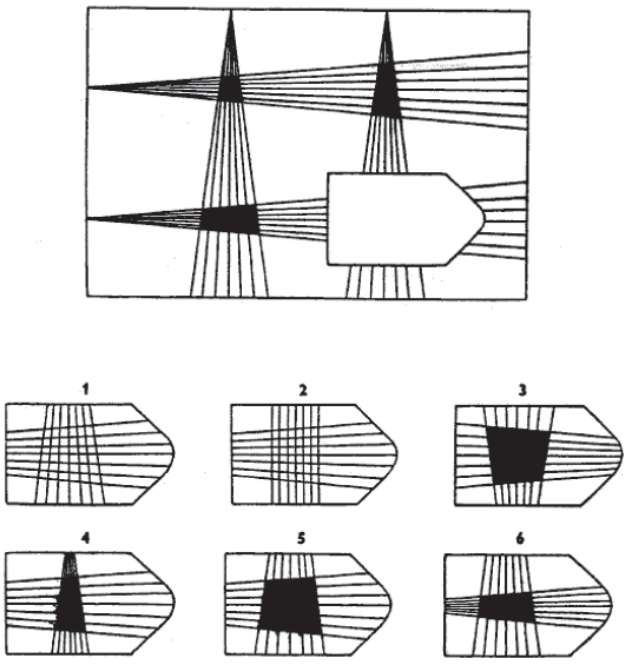
1.D.2 German (original)

Allgemeine Erklärungen

Wir begrüßen Sie herzlich zu diesem wirtschaftswissenschaftlichen Experiment. Lesen Sie die folgenden Erklärungen bitte gründlich durch! Wenn Sie Fragen haben, strecken Sie bitte Ihre Hand aus der Kabine – wir kommen dann zu Ihrem Platz. Während des Experiments ist es nicht erlaubt, mit den anderen Experimentteilnehmern zu sprechen, Mobiltelefone zu benutzen oder andere Programme auf dem Computer zu starten. Die Nichtbeachtung dieser Regeln führt zum Ausschluss aus dem Experiment und von allen Zahlungen. Für die Teilnahme an diesem Experiment erhalten Sie pauschal 12 Euro, die Sie am Ende dieses Experiments bar ausbezahlt bekommen. Auf den nächsten Seiten beschreiben wir den genauen Ablauf des Experiments.

Teil 1 des Experiments

In Teil 1 und 2 bearbeiten Sie jeweils 24 Aufgaben, die oft verwendet werden, um die sogenannte fluide Intelligenz eines Menschen zu bestimmen. Die fluide Intelligenz ist ein wichtiger Bestandteil der allgemeinen Intelligenz des Menschen. Oft werden solche oder ähnliche Aufgaben auch im Rahmen von Einstellungsverfahren von Unternehmen verwendet. Jede Aufgabe entspricht einem Bilderrätsel. Hier sehen Sie ein Beispiel:



Jedes Bilderrätsel zeigt in seinem oberen Teil ein Muster in einem Kasten, in dem unten rechts ein "Puzzlestück" ausgelassen ist. Ihre Aufgabe ist es, eines der unterhalb des Kastens aufgeführten Puzzlestücke auszuwählen, das die leere, untere rechte Ecke des Musters im Kasten logisch passend füllt. Bitte geben Sie dazu die Nummer des Puzzlestücks, das Ihrer Meinung nach am besten passt, auf dem Bildschirm ein. Die Nummer eines Puzzlestücks steht oberhalb jedes Puzzlestücks. Es gibt immer genau ein am besten passendes Puzzlestück.

Für die Bearbeitung eines Bilderrätsels haben Sie jeweils 30 Sekunden Zeit. Für jedes richtig beantwortete Bilderrätsel erhalten Sie einen Punkt. Wie dies bei Intelligenztests üblich ist, werden richtige Antworten nicht extra bezahlt. Sie erhalten 0 Punkte für jedes falsch beantwortete Bilderrätsel oder falls Sie innerhalb der 30 Sekunden keine Eingabe zum Ihrer Meinung nach am besten passenden Puzzlestück machen. Nachdem Sie alle 24 Bilderrätsel in Teil 1 bearbeitet haben, erhalten Sie auf dem Computerbildschirm zunächst ein privates Feedback zu Ihrem Rang, der angibt, wie gut Sie bei den Bilderrätseln abgeschnitten haben. Das Feedback hat die folgende Form: "X % der Teilnehmer in der Vergleichsgruppe haben einen höheren Rang als Sie in Teil 1". Die Vergleichsgruppe sind dabei 413 Teilnehmer an einem vorherigen Laborexperiment hier im DICE Lab der HHU aus dem Jahr 2014, die dieselben Bilderrätsel bearbeitet haben, wie Sie es im Laufe

dieses Experiments tun. Das Feedback "9 % der Teilnehmer in der Vergleichsgruppe haben einen höheren Rang als Sie in Teil 1" bedeutet also, dass 9 % besser abschneiden als Sie (d.h. einen höheren Anteil der gesamten 48 Bilderrätsel aus Teil 1 und 2 korrekt gelöst haben als Sie) und 90 % schlechter (d.h. einen niedrigen Anteil an Bilderrätseln korrekt gelöst haben als Sie). Sie gehören also zu den 10 % der Besten beim Beantworten der Bilderrätsel, die konzipiert wurden, um die individuelle fluide Intelligenz zu messen. Das Feedback "83 % der Teilnehmer in der Vergleichsgruppe haben einen höheren Rang als Sie in Teil 1" bedeutet, dass 83 % besser abschneiden als Sie und 16 % schlechter. Sie gehören also zu den 17 % der Schlechtesten beim Beantworten der Bilderrätsel.

Bevor Teil 2 des Experiments beginnt, müssen Sie noch zwei sogenannte "Beobachter" über Ihr Abschneiden im Experiment informieren. Bitte verwenden Sie dazu das DIN-A4-Blatt, das in Ihrer Kabine bereitliegt. Ihre Kabinennummer ist bereits eingetragen. Bitte tragen Sie unter "Rang 1" gut leserlich die Zahl in den Satz ein "___ % der Teilnehmer in der Vergleichsgruppe haben einen höheren Rang als ich in Teil 1", die Sie als Feedback auf dem Computerbildschirm erhalten haben. Tragen Sie bitte Ihren persönlichen Code, der ebenfalls auf dem Bildschirm angezeigt wird, daneben in das freie Feld ein: "Mein persönlicher Code ist ___". Die Beobachter sitzen in den Kabinen mit Nummer 1 und 2 im Labor (direkt gegenüber der Eingangstür). Bitte gehen Sie mit dem ausgefüllten DIN-A4-Blatt dorthin und zeigen es schweigend den Beobachtern, sobald Ihre Kabinennummer vom Experimentator aufgerufen wird. So wird sichergestellt, dass jeder Teilnehmer die Beobachter einzeln informiert, ohne dass die anderen Teilnehmer seinen Rang erfahren. Den Beobachtern wird auf ihrem Computerbildschirm eine Tabelle mit zwei Spalten angezeigt, die jedem persönlichen Code den entsprechenden Rang in Teil 1 zuordnet. Beide Beobachter werden, ebenfalls schweigend, Ihr DIN-A4-Blatt mit den Angaben in ihrer Tabelle vergleichen und jeweils abstempeln. Bitte begeben Sie sich dann schweigend wieder zurück in Ihre Kabine. Teil 2 des Experiments beginnt, sobald alle Teilnehmer die Beobachter über ihren Rang informiert haben.

Die verschiedenen Teilnehmer am Experiment

Zu Beginn des Experiments hat jeder Teilnehmer zufällig einen Chip mit einer Zahl

gezogen, die seine Kabinennummer angibt. Die Kabinennummern haben folgende weitere Bedeutung: Die Teilnehmer, die zufällig die Kabinennummern 1 und 2 gezogen haben, haben die Rolle der oben beschriebenen "Beobachter". Da die Chips mit den geraden Zahlen für die Frauen und die Chips mit den ungeraden Zahlen für die Männer reserviert waren, gibt es immer jeweils einen männlichen Beobachter und eine weibliche Beobachterin. Diese werden sich vor Beginn des eigentlichen Experiments kurz bei Ihnen vorstellen, in dem sie aufstehen und sagen "Ich bin eine/r der beiden Beobachter". Die Beobachter erhalten—genau wie die anderen Teilnehmer—diese ausgedruckte Erklärung der Regeln des Experiments, die Sie gerade lesen, zur Information über das Experiment.

Alle anderen Teilnehmer am Experiment mit den Kabinennummern 3 oder höher lösen die oben beschriebenen Bilderrätsel. Dabei wird jeder Teilnehmer zufällig einer von zwei Gruppen zugelost: Gruppe A oder Gruppe B. Im Laufe des gesamten Experiments bearbeiten alle Teilnehmer beider Gruppen exakt dieselben 48 Bilderrätsel, jeweils 24 in Teil 1 und 24 in Teil 2. Auch die weitere Aufgabe in Teil 2 des Experiments ist exakt dieselbe für beide Gruppen. Nur die Reihenfolge, in der die Bilderrätsel bearbeitet werden, unterscheidet sich zwischen Gruppe A und B. Eine weitere Bedeutung hat die Gruppenzugehörigkeit nicht. In Teil 1 und 2 gehören Sie zu derselben Gruppe.

Teil 2 des Experiments

Teil 2 des Experiments ist Teil 1 sehr ähnlich. Zunächst bearbeiten Sie 24 weitere Bilderrätsel nach denselben Regeln (30 Sekunden Zeit pro Rätsel, 1 Punkt für richtige Antworten, 0 Punkte sonst etc.). Nachdem Sie die weiteren 24 Bilderrätsel in Teil 2 bearbeitet haben, erhalten Sie auf dem Computerbildschirm zunächst ein privates Feedback zu Ihrem vorläufigen Rang in Teil 2, der angibt, wie gut Sie bei den insgesamt 48 Bilderrätseln in Teil 1 und 2 abgeschnitten haben. Das Feedback hat wieder die folgende Form: "X % der Teilnehmer in der Vergleichsgruppe haben einen höheren Rang als Sie." Die Vergleichsgruppe sind dabei wieder die 413 Teilnehmer an einem vorherigen Laborexperiment hier im DICE Lab der HHU aus dem Jahr 2014, die dieselben 48 Bilderrätsel bearbeitet haben wie Sie. Außerdem wird zur Erinnerung angezeigt, welchen Rang Sie in Teil 1 des Experiments hatten.

Der einzige Unterschied zu Teil 1 ist, dass Sie eine weitere Aufgabe haben, die auch in die Berechnung Ihres finalen Rangs in Teil 2 einfließt. Anschließend erhalten Sie ein privates Feedback zu Ihrem finalen Rang in Teil 2, der auf Grundlage der 48 Bilderrätsel in Teil 1 und 2 und Ihrem Abschneiden in der weiteren Aufgabe in Teil 2 berechnet wird. Details zur weiteren Aufgabe und wie genau sie in die Berechnung des finalen Rangs in Teil 2 einfließt, werden im Verlauf des Experiments auf dem Computerbildschirm erklärt. Zur Berechnung Ihres finalen Rangs wird wieder dieselbe Vergleichsgruppe herangezogen wie für Ihren Rang in Teil 1 und den vorläufigen Rang in Teil 2. Die detaillierten Erklärungen zur weiteren Aufgabe in Teil 2 erhalten nur die Teilnehmer, aber nicht die beiden Beobachter.

Genau wie zum Abschluss von Teil 1 müssen Sie noch die zwei Beobachter über Ihr Abschneiden, also Ihren finalen Rang, in Teil 2 informieren. Bitte verwenden Sie dazu wieder das DIN-A4-Blatt, das in Ihrer Kabine bereitliegt. Bitte tragen Sie nun zusätzlich unter "Rang 2" gut leserlich die Zahl in den Satz ein "___ % der Teilnehmer in der Vergleichsgruppe haben einen höheren Rang als ich", die Sie als Feedback über Ihren finalen Rang auf dem Computerbildschirm erhalten haben. Bitte gehen Sie mit dem ausgefüllten DIN-A4-Blatt zu den beiden Beobachtern und zeigen es ihnen schweigend, sobald Ihre Kabinennummer von einem Experimentator aufgerufen wird. So wird wieder sichergestellt, dass jeder Teilnehmer die Beobachter einzeln informiert, ohne dass die anderen Teilnehmer seinen Rang erfahren. Den Beobachtern wird auf ihrem Computerbildschirm nun eine Tabelle mit vier Spalten angezeigt, die jedem persönlichen Code den entsprechenden Rang in Teil 1, den finalen Rang sowie die Rangdifferenz zwischen Rang in Teil 1 und dem finalen Rang zuordnet.

Die Beobachter werden, ebenfalls schweigend, Ihr DIN-A4-Blatt mit den Angaben in ihrer Tabelle vergleichen und abstempeln. Bitte begeben Sie sich dann schweigend wieder zurück in Ihre Kabine.

Ende und Auszahlung des Experiments

Nach Teil 2 des heutigen Experiments folgen dann noch einige Bildschirme mit Fragen u. Ä., bevor wir zur Auszahlung der 12 Euro kommen. Wir werden Sie einzeln nach

Kabinennummer zur Auszahlung aufrufen.

Falls Sie jetzt Fragen haben, halten Sie bitte die Hand aus der Kabine. Ein Leiter des Experiments wird dann an Ihren Platz kommen, um Ihre Fragen zu beantworten. Stellen Sie Fragen keinesfalls laut!

1.D.3 Additional Instructions on the Computer Screen: Die Roll Task

English

There is a die in your cubicle. Please roll the die twice in your cubicle.

Please enter the numbers between 1 and 6 which you rolled on the first and second die rolls on the computer screen. The first number you rolled is added to the number of correctly solved picture puzzles of each participant of the reference group. The second die roll is added to your own number of correctly solved picture puzzles. Your total output is equal to the number of correctly solved picture puzzles in Parts 1 and 2 of the 48 picture puzzles plus the number of points you entered for the second die roll. Your total output is used to calculate your final rank. Your total output is compared with the total output of the peer group. The total output of a participant in the comparison group is equal to the number of correctly solved picture puzzles out of the 48 picture puzzles plus the number of points you entered for the first die roll. Your final rank will be shown to Observers and you will report it to the Observers at the end.

You may, of course, roll the die more often, for example to check that the die is working properly. If you have thrown more than twice, the other throws after the first two do not have any special meaning.

German (original)

In Ihrer Kabine liegt ein Würfel bereit. Bitte würfeln Sie zwei Mal in Ihrer Kabine.

Bitte geben Sie dann auf dem Computerbildschirm ein, welche Augenzahl zwischen 1 und 6 Sie beim ersten und zweiten Wurf gewürfelt haben. Die erste gewürfelte Augenzahl wird zur Anzahl der korrekt gelösten Bilderrätsel jedes Teilnehmers in der Vergleichsgruppe dazu gezählt. Die zweite gewürfelte Augenzahl wird zur Anzahl der von Ihnen korrekt gelösten Bilderrätsel dazu gezählt. Ihre entstehende Gesamtleistung entspricht also der Anzahl der von Ihnen korrekt gelösten Bilderrätsel in Teil 1 und 2 von den insgesamt 48 Bilderrätseln plus der von Ihnen eingegebenen Augenzahl vom zweiten Würfel-

wurf. Ihre Gesamtleistung wird verwendet, um Ihren finalen Rang zu berechnen. Dabei wird Ihre Gesamtleistung mit der Gesamtleistung der Vergleichsgruppe verglichen. Die Gesamtleistung eines Teilnehmers der Vergleichsgruppe entspricht der Anzahl der von ihm / ihr korrekt gelösten Bilderrätsel von den insgesamt 48 Bilderrätseln plus die von Ihnen eingegebene Augenzahl vom ersten Würfelwurf. Ihr finaler Rang wird den Beobachtern angezeigt und Sie werden ihn den Beobachtern abschließend berichten.

Natürlich können Sie gerne auch häufiger würfeln, z.B. um festzustellen, dass der Würfel richtig funktioniert. Falls Sie häufiger als zwei Mal gewürfelt haben, haben die weiteren Würfe nach den ersten beiden keine besondere Bedeutung.

1.E Questionnaire

1.E.1 English

Please fill out the following questionnaire now before we proceed to the payment. Please enter the following personal data. If you want to enter decimal numbers, please use a dot (.) instead of a comma (,).

- Age
- Gender (male/female)
- Final grade point average at high school (Abiturnote) (1.0–6.0)
- Last math grade (1.0–6.0)
- Last German grade (1.0–6.0)
- Field of study/job
- How much money do you have available each month (after deducting fixed costs such as rent, insurance, etc.)?
- How much money do you spend each month (after deducting fixed costs such as rent, insurance, etc.)?
- In how many economic science experiments have you (approximately) already participated?
- On a scale of 0 to 10, how would you rate your willingness to take risks? 0 means not willing to take risks at all and 10 means completely willing to take risks.
- How important is the opinion that others hold about you to you? Please answer on a scale of 0 to 10, where 0 is not important at all and 10 is extremely important.
- Have you ever solved similar tasks as the picture puzzles before? (Yes/No)

- If so, how long ago approximately? Please indicate the approximate number of months.

Below, please answer a few more questions about lotteries in which you can earn or lose money in addition to the €12 if you decide to accept the lotteries.

Listed below are 6 different lotteries. For each of the 6 lotteries you can choose whether to accept or decline the lottery. If you choose to decline a lottery, your payout will not change. If you accept a lottery, you will realize either an additional gain or an additional loss based on the €12.

At the end of the experiment, one of the 6 lotteries is randomly selected. So you should make each lottery decision as if it was your only decision. The selected lottery is then drawn to determine whether the additional gain or loss will be realized.

Lottery 1: With 50% probability you lose €2 and with 50% probability you win €6. (accept / reject)
Lottery 2: With 50% probability you lose €3 and with 50% probability you win €6. (accept / reject)
Lottery 3: With 50% probability you lose €4 and with 50% probability you win €6. (accept / reject)
Lottery 4: With 50% probability you lose €5 and with 50% probability you win €6. (accept / reject)
Lottery 5: With 50% probability you lose €6 and with 50% probability you win €6. (accept / reject)
Lottery 6: With 50% probability you lose €7 and with 50% probability you win €6. (accept / reject)

1.E.2 German (original)

Füllen Sie nun bitte die folgenden Fragen aus, bevor wir zur Auszahlung kommen. Bitte geben Sie die folgenden Daten zu Ihrer Person an. Wenn Sie Kommazahlen eingeben möchten, nutzen Sie bitte einen Punkt (.) statt eines Kommas (,).

- Alter
- Geschlecht (männlich/weiblich)
- Abiturdurchschnittsnote (1.0-6.0)
- Letzte Mathenote (1.0-6.0)
- Letzte Deutschnote (1.0-6.0)
- Studienfach/Tätigkeit
- Wie viel Geld haben Sie monatlich (nach Abzug von Fixkosten wie Miete, Versicherungen etc.) zur Verfügung?
- Wie viel Geld geben Sie monatlich aus (nach Abzug von Fixkosten wie Miete, Versicherungen etc.)?
- An wie vielen wirtschaftswissenschaftlichen Experimenten haben Sie (ungefähr) bereits teilgenommen?
- Wie schätzen Sie Ihre Risikobereitschaft auf einer Skala von 0 bis 10 ein? Dabei bedeutet 0 überhaupt nicht risikobereit und 10 vollkommen risikofreudig.
- Wie wichtig ist Ihnen die Meinung, die andere über Sie haben? Bitte antworten Sie auf einer Skala 0 bis 10. Dabei ist 0 überhaupt nicht wichtig und 10 extrem wichtig.
- Haben Sie schon einmal ähnliche Aufgaben wie die Bilderrätsel gelöst? (Ja/Nein)
- Falls ja, wie lange ist das ungefähr her? Bitte geben Sie die ungefähre Zahl der Monate an.

Im Folgenden beantworten Sie bitte noch ein paar Fragen zu Lotterien, bei denen Sie noch einmal zusätzlich zu den €12 Geld verdienen oder auch verlieren können, falls Sie sich entscheiden, die Lotterien zu akzeptieren.

Unten sind 6 verschiedene Lotterien aufgelistet. Sie können für jede der 6 Lotterien wählen, ob Sie die Lotterie akzeptieren oder ablehnen möchten. Falls Sie eine Lotterie

ablehnen, bleibt Ihre Auszahlung unverändert. Falls Sie eine Lotterie akzeptieren, werden Sie ausgehend von den €12 entweder einen zusätzlichen Gewinn oder einen zusätzlichen Verlust realisieren.

Am Ende des Experiments wird zufällig eine der 6 Lotterien ausgewählt. Sie sollten also jede Lotterieentscheidung so fallen, als wäre es Ihre einzige Entscheidung. Die ausgewählte Lotterie wird anschließend ausgelost, damit feststeht, ob sich der zusätzliche Gewinn oder Verlust realisiert.

Lotterie 1: Mit 50% Wahrscheinlichkeit verlieren Sie €2 und mit 50% Wahrscheinlichkeit gewinnen Sie €6. (akzeptieren / ablehnen)
Lotterie 2: Mit 50% Wahrscheinlichkeit verlieren Sie €3 und mit 50% Wahrscheinlichkeit gewinnen Sie €6. (akzeptieren / ablehnen)
Lotterie 3: Mit 50% Wahrscheinlichkeit verlieren Sie €4 und mit 50% Wahrscheinlichkeit gewinnen Sie €6. (akzeptieren / ablehnen)
Lotterie 4: Mit 50% Wahrscheinlichkeit verlieren Sie €5 und mit 50% Wahrscheinlichkeit gewinnen Sie €6. (akzeptieren / ablehnen)
Lotterie 5: Mit 50% Wahrscheinlichkeit verlieren Sie €6 und mit 50% Wahrscheinlichkeit gewinnen Sie €6. (akzeptieren / ablehnen)
Lotterie 6: Mit 50% Wahrscheinlichkeit verlieren Sie €7 und mit 50% Wahrscheinlichkeit gewinnen Sie €6. (akzeptieren / ablehnen)

Chapter 2

Willful Ignorance and Reference Dependence of Self-Image Concerns

2.1 Introduction

Information is a key element in most economic decisions. Individuals tend to seek certainty and avoid ambiguity. Yet, in many situations, people prefer to deliberately avoid information and remain willfully ignorant.[1] Examples of information avoidance range from everyday interactions to high-stake decisions with long-term effects. For instance, individuals may not want to learn that the holiday season made them put on some weight or that there is a better deal for a recent purchase (Sweeny et al., 2010). In a health context, people tend to avoid learning about their genetic risks for cancer or the Huntington's disease (Oster et al., 2013; Ropka et al., 2006) and their HIV status even when offered monetary incentives to do so (Thornton, 2008). In a finance context, investors tend to monitor their portfolios and balances closely on "good days", e.g., on paycheck days or when the market goes up, and avoid logging into their accounts on "bad days" (Karlsson et al., 2009; Olafsson et al., 2018). In a workplace, managers often forego helpful feedback to avoid learning that their earlier decisions were incorrect because they want to maintain their professional self-image (Deshpande and Kohli, 1989; Schulz-Hardt et al., 2000; Zaltman, 1983). In the case of prosocial behavior, individuals often prefer to remain uninformed about the actual effectiveness of their altruistic actions or charitable donations and carry on a feeling of warm glow due to the fact of their deed but not to the impact on its recipient (Niehaus, 2014). Similarly, people tend to avoid learning about the potential harm their actions may yield for others (Dana et al., 2007; Grossman and Van Der Weele, 2017; Serra-Garcia and Szech, 2021).

This paper studies implications of changes in self-image for the demand for feedback as well as the evolution of the self-image itself. I conduct a laboratory experiment to analyze individuals' willingness to avoid self-image-relevant feedback after having them work on more difficult or easier tasks in the first part of the experiment. The key novelty of the paper is two-fold. First, varying the complexity of the tasks allows inducing exogenous shock to subjects' performance measurable on an individual level. Second, by

[1] For a comprehensive multidisciplinary literature overview of information avoidance, see Golman et al. (2017).

complementing this approach with multiple elicitations of individuals' beliefs about their performance, I observe the impact of exogenous positive and negative shocks to self-image on an individual level as well. I investigate whether subjects who expect positive feedback are more likely to acquire information than those who expect negative feedback. I also test for reference-dependence of self-image concerns as well as for loss aversion in the self-image domain.

The experimental data provide strong evidence of information avoidance independently of whether the expected feedback is positive or negative. Individuals tend to have a stronger willingness to avoid feedback if they expect it to be negative. In line with expectations, subjects update beliefs about their performance upwards if they work on easier tasks in the first part of the experiment, which translates into an improvement in their self-image. Subjects who work on harder tasks first update beliefs downwards, indicating the deterioration of their self-image. Moreover, subjects update beliefs about their self-image only slightly after the easier task, while subjects who have done the harder task first update much stronger. At the end of the experiment, after individuals worked on both hard and easy tasks, their beliefs about their performance in the intelligence test go back to the pre-experiment levels. This result indicates that subjects did not find the overall complexity of the IQ test surprising.

I propose a stylized theoretical framework that offers a simple explanation to the experimental findings, in particular, the surprising patterns in belief updating. The framework captures the idea of disappointment aversion (Gul, 1991). It follows closely the setup of Gollier and Muermann (2010) and models the trade-off between the ex-ante feelings and the risk of ex-post disappointment. In this framework, agents, who derive utility from self-image, first manage expectations and choose an optimal degree of optimism. Then, they decide whether they want to acquire self-image-relevant information. In the context of my experiment, subjects may choose to be optimistic about their performance and derive utility ex-ante at the cost of a possible disappointment at the end of the experiment. Alternatively, participants may stay pessimistic in their beliefs throughout the experiment and likely be positively surprised at the end.

My experimental setup uses intelligence as a self-image relevant domain and lets subjects work on an IQ test.[2] To induce exogenous gains and losses in their self-image, I randomize whether subjects first complete a more difficult or easier part of a standard IQ test. This design feature allows to induce a sharp symmetric heterogeneous shock in subjects' performance in the first part that I observe on an individual level. After working on the easier part, subjects on average perceive their performance positively, and thus expect the good feedback. On the contrary, when initially working on the harder part perceived performance is on average worse, as is the feedback they expect. I employ a continuous willingness-to-pay (WTP) measure to elicit subjects' exact willingness to acquire feedback.

In order to be able to study perceived gains and losses in self-image, I elicit subjects' beliefs about their performance at three points over the course of the experiment. Prior belief elicitation takes place before subjects work on the IQ test. After inducing a gain or a loss in self-image by letting them work on easier or harder tasks, respectively, I elicit their beliefs again. The second belief elicitation allows to observe whether they update their beliefs, and whether they do so differently when they expect more positive and negative feedback. Furthermore, after the whole IQ test is complete and all subjects worked on the exactly same tasks, I elicit beliefs to analyze whether they recovered from the exogenous shock in self-image.

Multiple belief elicitations are an important feature of my design. While in standard economic theory the ultimate purpose of beliefs is to assist in the decision making processes, many recent studies have shown, both theoretically and experimentally, that individuals tend to hold motivated beliefs and argued that beliefs can be a choice variable (Bénabou and Jean Tirole, 2002; Kőszegi, 2006). Experimental evidence shows that people dislike updating their beliefs negatively and react to noisy negative signals much weaker than to the positive ones (Coutts, 2019; Eil and Rao, 2011; Golman et al., 2017; Zimmermann, 2020). In other words, a gain in self-image might be internalized stronger than a loss of the same magnitude. In contrast, individuals react stronger to losses than

[2]Intelligence, or IQ, is a commonly used self-image relevant domain. See, e.g., Fein and Spencer (1997), Santos-Pinto and Sobel (2005) and Castagnetti and Schmacker (2022).

to gains in monetary and material domains (Kahneman, Knetsch, et al., 1990; Kahneman and Tversky, 1979) as well as with respect to health outcomes (Bleichrodt et al., 2001) and social image (Petrishcheva et al., 2020). It is important to observe not only actual differences in one's performance but also perceived ones. This paper focuses on disentangling these effects by looking at the willingness to acquire feedback of individuals who experience measurable perceived gains and losses in self-image. When analyzing subjects' willingness to acquire feedback, I take into consideration how they updated their beliefs.

This paper is organized as follows. Section 2.2 discusses the experimental design, implementation and technical details. In Section 2.3, I formulate the hypothesis. I present the results in Section 2.4. I discuss my results and propose a stylized theoretical framework in Section 2.5. Section 2.6 concludes.

2.2 Experimental design

My experimental setup includes three stages as shown in Figure 2.1. In Stage 1, I elicit subjects' prior beliefs about their performance in the upcoming IQ test. I treat prior beliefs as a within-subject reference point in intelligence, a self-image-relevant domain.[3] In Stage 2, I induce an exogenous shift in self-image. By introducing treatments *Loss* and *Gain*, I put subjects' self-image at either loss or gain by varying the task complexity. I then ask subjects whether they are willing to acquire feedback about their performance and elicit their willingness-to-pay to do so and their beliefs about their performance. The second belief elicitation is necessary to see whether the treatment variation worked, i.e., whether subjects indeed expect losses and gains when I assume they do. In Stage 3, I let subjects work on the remaining tasks of the IQ test and elicit their performance beliefs upon completion.

First, I analyze belief updating for those with gains and losses in self-image. I focus on the two main aspects: Is subjects' belief updating (a) going in the direction suggested by the treatment and (b) symmetric for gains and losses of self-image?

[3]See, e.g., Fein and Spencer (1997), Santos-Pinto and Sobel (2005) and Castagnetti and Schmacker (2022).

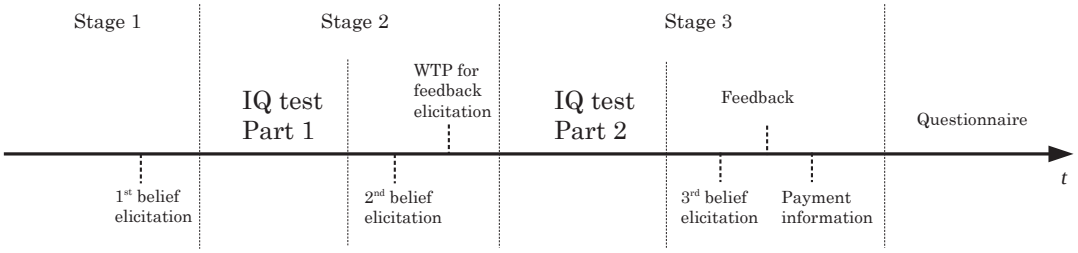


Figure 2.1: Timeline of the experiment

Additionally, this design allows to analyze subjects' willingness to pay to acquire self-image-relevant feedback both unconditionally and conditionally on belief updating. Varying task complexity allows to induce an *objective* performance shift. Since subjects do not receive any signals about their performance except their *subjective* perception of it before they report the willingness to pay to acquire feedback, it is crucial to control for their beliefs when analyzing their WTP for feedback. I test whether subjects who care about their self-image avoid ego-relevant feedback. Then, I analyze whether those who experience a loss in self-image are more willing to acquire feedback than those who experience gain. I also test whether subjects with marginal self-image losses have a disproportionately higher willingness to acquire feedback than those with marginal gains in self-image.

IQ test In this experiment, subjects work on Raven's Progressive Matrices (RPMs; Raven, 1983), which are designed to measure fluid intelligence and often used in economic experiments to induce image concerns (e.g., Zimmermann, 2020, and Ewers and Zimmermann, 2015). In Figure 2.2, there are two examples of RPMs. They are picture puzzles with a missing piece. Among the available answers, subjects should choose the best logical fit to fill in the blank space. RPM tests commonly consist of five sets of matrices (A to E), with 12 matrices in each set. These sets progress in difficulty. Set A includes the easiest matrices; Set B is slightly harder, and so on. Set E contains the 12 hardest matrices. In Figure 2.2, the left matrix is one of the easier matrices from the set B (B3), and the right one is among the most complicated tasks from the set E (E10). Based on the reference data,[4] I expect student subjects to solve all the matrices in set A correctly. Hence, I do

[4]The reference sample includes 413 observations (students) from a previous experiment that took place at the same lab in 2014 who worked on the full set (all 60) of the same RPM matrices.

not use the 12 easiest matrices in this experiment but the 48 matrices from sets B to E.

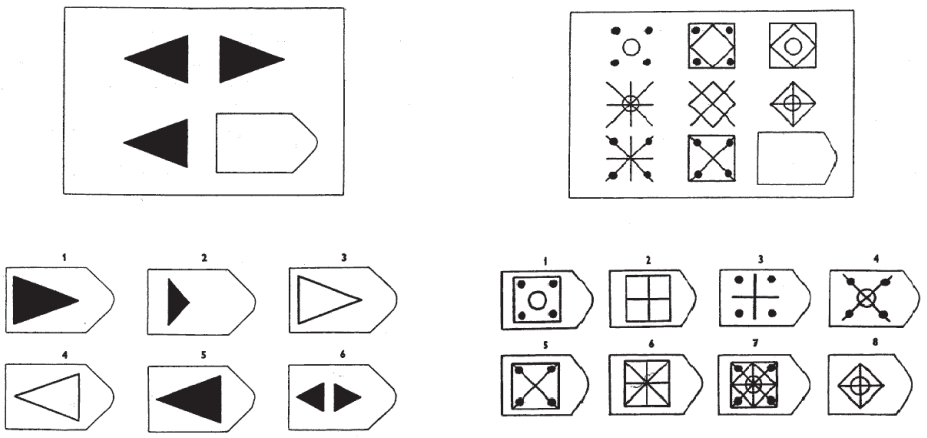


Figure 2.2: Examples of Raven's progressive matrices

I split 48 matrices into two parts: *Easy* and *Hard*. Matrices from sets B and C belong to the *Easy* part. Matrices from sets D and E form the *Hard* part. Both parts are progressive, i.e., they start with easy tasks and get more complicated over time. Matrices in parts *Easy* and *Hard* do not repeat or overlap. Subjects get one point if they solve a matrix correctly and get zero points otherwise. Subjects have a time limit of 30 seconds per matrix, which ensures that their performance is comparable within the experiment and to the reference sample, where the same time limit was imposed.

Stage 1 After reading general instructions and answering control questions,[5] subjects proceed to the first belief elicitation. I elicit their prior beliefs about their overall performance, i.e., the number of correctly solved matrices in both parts.

Belief elicitation procedure In the belief elicitation screen, subjects get the following information:[6]

- A summary about the performance of the reference sample. I tell subjects that in 2014, 413 individuals worked on the same picture puzzles in the DICE Lab.

[5]Original and translated versions of instructions and control questions (in German and English, respectively) are available in Appendix 2.C.

[6]See a complete belief elicitation screen in Appendix 2.C.

Additionally, I give them a short description of the data, namely: (a) no previous participant solved all 48 matrices, (b) the average participant solved 39 matrices, and (c) all previous participants solved at least 20 matrices or more.

- A figure with the performance of the reference sample. I show a histogram with scores displayed on the horizontal axis and the frequency (i.e., percent of the participants) on the vertical axis.

- A disclaimer saying "Carefully and honestly answering the question is in your best interest". Following Danz et al. (2020), I do not explain the exact monetary incentive structure in advance to reduce errors in belief elicitation. Instead, I tell them that the precise payment rule details are available by request at the end of the experiment.

- A slider with values between 0 and 48 and no default value where subjects should indicate how many matrices they think they will solve correctly.

- A phrase "I think I will solve X out of 48 picture puzzles correctly. It means that I think I will perform better than Y% of previous participants", which completes automatically when they choose or adjust the slider.

I incentivize the decision using the binarized scoring rule (Danz et al., 2020; Hossain and Okui, 2013). According to the binarized scoring rule, an individual may earn a fixed payment. The probability of receiving it increases the closer is her guess to the true outcome. In the context of my experimental design, participants can earn one euro in each belief elicitation task. Throughout the experiment, I used experimental currency units (ECU). The exchange rate was 1 euro = 20 ECU.[7] If their belief is correct, i.e., their perceived number of correctly solved matrices corresponds to their actual number of correctly solved matrices, they get a bonus of 20 ECU with a probability of one. Importantly, with the binarized scoring rule, subjects still have a small probability to get paid for the belief elicitation task, even if their guess and their actual performance differ a lot. Hence, their payoffs are not (directly) indicative of their performance.

[7] In the instructions, I refer to ECU as thalers (Taler) which is a commonly used ECU in the DICE Lab.

Subjects' prior beliefs about their performance in the IQ test serve as a within-subject reference point in intelligence. The procedure of belief elicitations is always the same. I always ask subjects about their beliefs about their overall performance. Payoffs of multiple belief elicitations are independent.

Stage 2 Subjects work on Part 1 of the test. In treatment *Gain*, Part 1 is *Easy*, such that subjects, on average, solve more matrices than they expected and hence can expect positive feedback about their performance. In treatment *Loss*, on the contrary, subjects work on *Hard* tasks, so they, on average, perform worse than expected. After participants completed 24 tasks in Part 1, I elicit their beliefs following the same procedure as described above.

After the second belief elicitation, I ask subjects about their willingness-to-pay to get feedback using the Becker-DeGroot-Marschak mechanism (G. M. Becker et al., 1964; BDM; see the screen in Appendix 2.C). On a scale from -100 to 100 ECU (-5 to 5 euro), they report how much they would like to pay for feedback. Subjects are aware that WTP of -100 ECU guarantees that they will not receive information about their performance. WTP of 100 ECU means that they will certainly get feedback, and WTP of zero yields a 50% chance of receiving feedback about the number of matrices they solved correctly. I draw a random price for feedback from a uniform distribution with a support on the interval $[-100; 100]$. If the random price for feedback is smaller than or equal to the participants' WTP, they pay the price and receive feedback. If the random price for feedback is higher than their WTP, they do not pay the price and do not receive the feedback.

Stage 3 Subjects work on the remaining 24 RPM tasks. It means that subjects from treatment *Gain* work now on the *Hard* part, while those from treatment *Loss* work on the *Easy* part. After Stage 3, all subjects have worked on the same 48 picture puzzles described above. Once subjects complete the task, I elicit beliefs about their performance again before they receive (or not) their feedback. I display their feedback in the same format as belief elicitation, i.e., it says "You solved X out of 48 picture puzzles correctly.

This means that you performed better than Y% of previous participants".

Questionnaire After the main experiment is complete, subjects fill out a questionnaire. It contains the main sociodemographic characteristics such as age, gender, the field of study, occupation, current GPA (or last degree GPA), high school GPA as well as average monthly budget and spending. Additionally, I ask them about their experience in the lab, and collect independent measures of loss aversion in the monetary domain (Fehr and Goette, 2007; Gächter, E. J. Johnson, et al., 2021), risk aversion, and time preferences (Falk, A. Becker, T. J. Dohmen, et al., 2016). Furthermore, subjects report the intensity of their social image concerns by answering the question "How important is the opinion that others hold about you to you?" following Ewers and Zimmermann (2015). I measure their overconfidence by letting them work on real-effort slider tasks and eliciting their beliefs about their performance (similar to S. Chen and Schildberg-Hörisch, 2019). Additionally, I elicit the intensity of self-image concerns following the approach of Aquino and Reed II (2002) and Grossman and Van Der Weele (2017). Subjects get a list of six statements about the importance of being kind, generous, and fair to their sense of self. They can choose whether they agree or disagree with those statements on a six-point Likert scale (from 0 indicating "strongly disagree" to 5 indicating "strongly agree"). Following Grossman and Van Der Weele (2017), I sum the points from evaluating all six statements to generate a measure of self-image concerns. The exact wording of each question is in Appendices 2.C.7 and 2.C.8. The independent measure of loss aversion in monetary domains is a set of incentivized lotteries. There are six lotteries and subjects can decide whether they accept or reject participation in each of them. One of the lotteries is paid out randomly at the end of the experiment. Each lottery yields a 50% chance of winning 12 ECU and a 50% chance of losing 4, 6, 8, 10, 12, or 14 ECU. Subjects do not earn any additional money if they rejected a lottery.

The independent measure of overconfidence is incentivized as well. There are 11 slider tasks, and subjects should position each slider in the middle (between 49 and 51 on a 0-100 scale). For each correctly solved slider task, subjects received 2 ECU. Furthermore, subjects could receive additional 10 ECU if their guess about how many sliders they solved

correctly was sufficiently accurate according to the binarized scoring rule (Hossain and Okui, 2013).

Payment structure Total earnings are only paid out upon completion of the experiment to prevent subjects from potentially dropping out. Subjects received a show-up fee of 3.70 euro as well as a 5 euro endowment at the beginning of the experiment, which might be used to pay for the feedback about their performance. The 5 euro endowment assures that, to ensure (not) getting feedback, the stakes are relatively high. However, subjects cannot make an absolute loss after their decision is realized. Additionally, subjects face three rounds of belief elicitations (before the experiment, after Part 1, and after Part 2) which pay 1 euro each with a probability that depends on the correctness of their belief. On top of that, loss aversion and overconfidence measures were monetarily incentivized.

Technical details and procedure This experiment was conducted online with subjects from the DICE Lab, University of Düsseldorf, in June 2021. For each session, all subjects took part in a web-conference call where they could ask clarifying questions or receive technical support if needed.[8] The experiment is preregistered on AEA RCT Registry[9], received an IRB approval[10] and was programmed using oTree (D. L. Chen et al., 2016). Subjects were recruited via Orsee (Greiner, 2015). Original instructions (in German) and the translated version of the instructions (in English) are in Appendix 2.C. Subjects earned 13.3 euro on average for the experiment, which lasted approximately 45 minutes.[11] No subjects dropped out of the experiment. During the experiment, participants could not communicate with or see each other.

I conducted six online sessions with 20-24 participants each. In total, 132 subjects participated in the experiment: 67 of them were assigned to treatment *Gain* and the

[8]Li et al. (2021) find that using web-conference calls in online experiments leads to outcomes comparable to those the laboratory experiments for various economic games.

[9]Petrishcheva, Vasilisa. 2021. "Willful Ignorance and Reference-Dependence of Self-Image Concerns." AEA RCT Registry. June 09.

[10]IRB Approval No. 49nWIXIa

[11]Subjects earned at least 9.7 and at most 17.7 euro in this experiment. In addition to a show-up fee of 3.7 euro and an endowment of 5 euro, subjects' earnings depended on numerous decisions, namely, belief elicitations, willingness-to-pay for feedback, loss aversion lotteries, and performance in the overconfidence tasks. Subjects were not able to make an absolute loss in this experiment.

remaining 65 to treatment *Loss*. As reported in Table 2.A.1, the sample is well-balanced with respect to individual characteristics between treatments, such that no exclusion criteria apply.

2.3 Hypotheses

In this section, I formulate four pre-registered hypotheses regarding belief updating and information avoidance in my experiment. First, I hypothesize that the share of subjects with negative willingness-to-pay to acquire feedback relevant to their self-image in the IQ domain will be non-negligible.

Hypothesis 1. (Willful ignorance)

Individuals who care about their self-image may avoid feedback relevant for their self-image.

This experimental design induces changes in subjects' performance in an IQ test, a self-image-relevant domain. Acquiring or avoiding feedback may influence subjects' utility derived from their self-image. Hence, following the literature on information avoidance and willful ignorance (e.g., Golman et al., 2017; Kőszegi, 2006), I expect subjects may avoid information relevant for their self-image. Next, I formulate a hypothesis about how subjects update beliefs about their performance in the IQ test when I introduce positive and negative shocks to their performance.

Hypothesis 2. (Asymmetric belief updating)

Individuals who care about their self-image may update their beliefs stronger if they experience a gain in a self-image-relevant domain compared to a loss in a self-image-relevant domain of the same size.

In line with motivated beliefs literature (Coutts, 2019; Eil and Rao, 2011; Golman et al., 2017; Zimmermann, 2020), I hypothesize that the absolute difference between prior beliefs and the first posterior beliefs will be larger for subjects in *Gain* than in *Loss*. It implies that subjects who on average experience gains in their self-image update their

beliefs stronger than those who experience losses of the same size in their self-image. In presence of a rather strong but very noisy signal about their performance (their own perception of their performance), I expect subjects who observe a negative signal to be more hesitant to update their beliefs about their IQ compared to those who observed a positive signal.

Next, I formulate the following hypothesis for *perceived* gains and losses of the same size:

Hypothesis 3. (Reference-dependence)

On average, individuals who care about their self-image and expect a loss in their self-image are more willing to acquire self-image-relevant information than those who expect a gain in their self-image of the same size.

I expect that individuals with a perceived loss will be more willing to acquire feedback about their performance than those with a perceived gain in self-image. The key novelty of this paper is analyzing the reference dependence of self-image concerns. More specifically, I test whether subjects who expect a loss in self-image have a higher willingness to pay to acquire feedback than those who expect a gain in self-image. If an individual expects a loss in self-image, positive feedback may serve as a tool to avoid this loss. Moreover, this paper focuses on individuals who experience marginal gains and losses. According to prospect theory (Kahneman and Tversky, 1979), there is a kink in the value function for changes in self-image which results in a kink in incentives to acquire self-image-relevant feedback. I hence formulate Hypothesis 4:

Hypothesis 4. (Loss aversion)

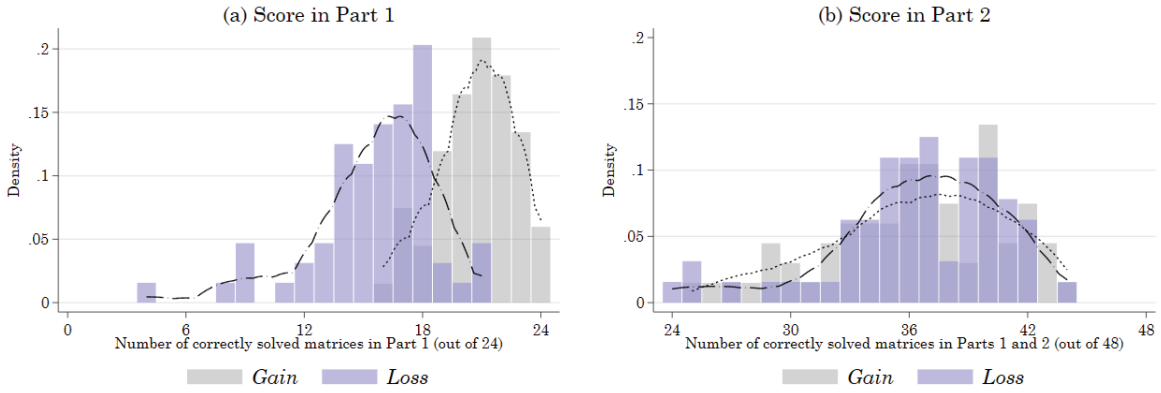
Individuals who care about their self-image and expect a marginal loss in their self-image are more willing to acquire information than those who expect a marginal gain in their self-image.

2.4 Results

This section is organized as follows. First, I discuss results related to subjects' performance in the IQ test in Parts 1 and 2 in Section 2.4.1. Next, I analyze subjects' beliefs about their intelligence in Section 2.4.2. In Section 2.4.3, I discuss their willingness-to-pay to receive self-image-relevant feedback.

2.4.1 IQ

Despite not being monetarily incentivized, subjects exerted substantial effort on solving the Raven's Progressive Matrices. On average, they solved 36.4 matrices correctly. Out of 48 matrices that subjects have worked on, they gave at least 24 and at most 44 correct answers.[12]



Note: Figures (a) and (b) display score distributions by treatment for Parts 1 and 2, respectively. In Figure (a), the horizontal axis shows the total number of correctly solved matrices after subjects completed Part 1. In Figure (b), the horizontal axis displays values between 24 and 48 because no subject solved less than 24 matrices correctly. In Figure (b), the horizontal axis shows the total number of correctly solved matrices after subjects completed Part 2. The vertical axis shows density. I show the histograms of score distributions and the kernel density estimates for treatments *Gain* and *Loss* in each figure. I estimate density using Epanechnikov kernels with an optimal bandwidth.

Figure 2.3: Score distributions by treatment

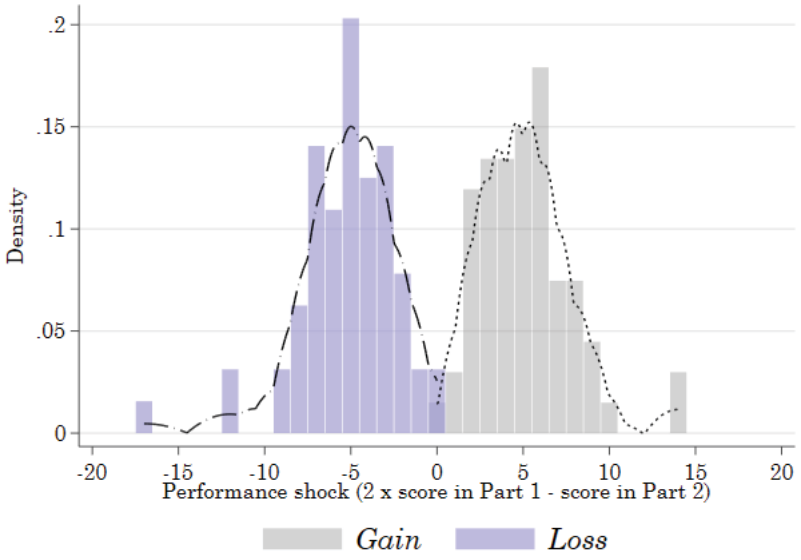
As intended by the experimental design, there are no significant differences in the distributions of the overall performance of subjects in treatments *Gain* and *Loss* (p=0.937).[13] I display the distributions of the score in Part 2 (overall performance) by treatment in

[12]Only one participant did not solve any matrices correctly by letting the 30-second timers run out. I exclude this subject from further analysis.

[13]In my analyses, I report two-sided Mann-Whitney U test results unless specified otherwise. I refer to results as (highly/weakly) statistically significant if the respective p-values are below 0.05 (0.01/0.1).

Figure 2.3(b). The average number of correct answers is 36.3 and 36.4 in *Gain* and *Loss*, respectively. Working on part *Easy* first and on part *Hard* second (treatment *Gain*) leads to similar overall scores as working on part *Hard* first and on part *Easy* second (treatment *Loss*). Hence, there is no evidence for order effects in my experiment.

After subjects worked on Part 1 of the IQ test (the first 24 tasks), I document a substantial difference in performance between treatments *Gain* and *Loss*. The average number of correctly solved matrices is 20.7 in treatment *Gain* and 15.6 in treatment *Loss*. The difference in performance between treatments is highly statistically significant. In Figure 2.3(a), I illustrate the distributions of performance in Part 1 by treatment.



Note: This Figure displays histograms of the performance shock (2 × score in Part 1 - score in Part 2) by treatment. The dashed line corresponds to the density estimates with Epanechnikov kernels and an optimal bandwidth.

Figure 2.4: Performance shock (by treatment)

My experiment introduces a shock to subjects' self-image by affecting their score in Part 1. I define a shock by comparing subjects' total number of correctly solved matrices (score in Part 2) and an extrapolated number of correctly solved matrices, i.e., the number of matrices they would have correctly solved if they carried on the same performance (2 × score in Part 1). The distributions of the performance shock are shown in Figure 2.4. The difference in shock distributions is highly statistically significant ($p<0.001$). Moreover, in absolute terms, performance shocks in *Gain* and *Loss* do not differ significantly

(p=0.904) which indicates their symmetry for treatments *Gain* and *Loss*. Additionally, the performance shock I introduce in Part 1 aligns with the treatment assignment. As shown in Figure 2.4, there are no overlapping values of shock for treatments *Gain* and *Loss* except for zeros which account for two observations in treatment *Loss* and only one observation in treatment *Gain*. Hence, the score in Part 1 can act as a precise continuous individual-level measure of treatment that I will rely on in my analyses.

2.4.2 Beliefs about IQ

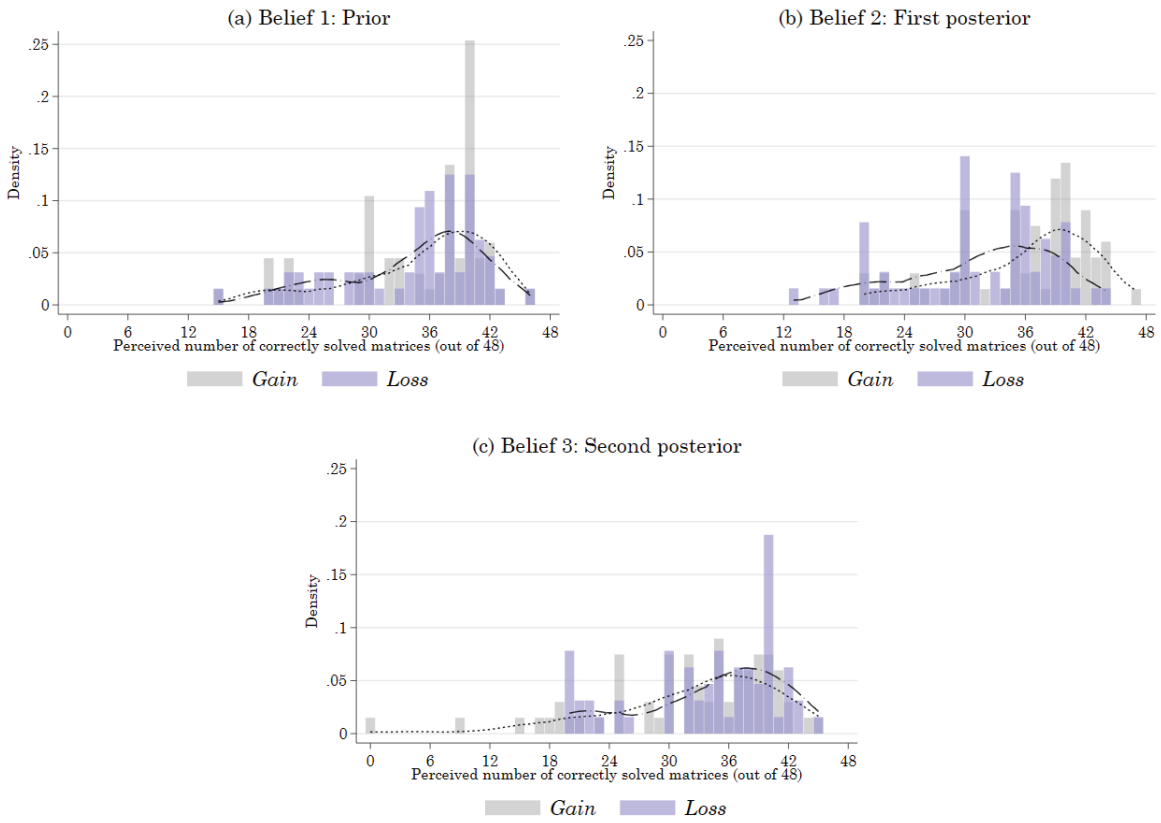
There are three belief elicitations in this experiment. I denote them Beliefs 1, 2, and 3, respectively. Belief 1 corresponds to the participants' prior belief about their performance which I elicit before they start working on the IQ test. Belief 2 is a subjects' first posterior belief which I elicit in the middle of the IQ test, namely after they worked on the first 24 out of 48 matrices and after the treatment variation took place. Belief 3 is a second posterior belief. Its elicitation takes place after subjects worked on all 48 matrices. I present summary statistics of subjects' beliefs in Table 2.1 and distributions of beliefs in Figure 2.4.2.

Table 2.1: Summary statistics: Beliefs (by treatment)

	Loss	*Gain*	Difference
Belief 1	34.30	35.03	p=0.325
Belief 2	31.58	36.07	p<0.001
Belief 3	34.05	32.27	p=0.224
N	64	67	131

Note: I show mean values of Beliefs 1, 2 and 3 for treatments *Loss* and *Gain*. Beliefs 1, 2 and 3 indicate subjects' guesses about their number of correctly solved matrices (0 to 48). I compare distributions of Beliefs 1, 2 and 3 between treatments and report two-sided MWU test p-values.

I measure Belief 1 before the treatment variation affects the course of the experiment, hence creating a belief baseline for my analysis. Unsurprisingly, participants of treatments *Loss* and *Gain* do not differ in their prior beliefs about performance in the IQ test (p=0.325). On average, subjects believe they will solve 34.3 and 35.0 Raven's Progressive Matrices correctly in treatments *Loss* and *Gain*, respectively.



Note: Figures (a)-(c) display distributions of Beliefs 1, 2, and 3 by treatment, respectively. The horizontal axis shows the total number of matrices that subjects expect to solve correctly (out of 48). The vertical axis shows density. I show the histograms of score distributions and the kernel density estimates for treatments *Gain* and *Loss* in each figure. I estimate density using Epanechnikov kernels with an optimal bandwidth.

Figure 2.5: Belief distributions by treatment

My treatment manipulation is designed to affect Belief 2. I shift participants' beliefs in the positive direction in treatment *Gain*, such that their Belief 2 is more optimistic than their Belief 1. In treatment *Loss*, on the contrary, participants update their beliefs negatively, i.e., Belief 2 is less optimistic than Belief 1. I find a highly significant difference in Belief 2 between subjects from *Gain* and *Loss* ($p<0.001$).

I define the belief difference as the difference between the first posterior beliefs about the IQ and the prior beliefs about the IQ: (Belief 2 - Belief 1). Hence, positive belief difference implies updating beliefs positively, i.e., subjects thinking they will solve more matrices than they initially assumed. Negative belief difference, vice versa, refers to updating beliefs negatively. Subjects expect to solve fewer matrices correctly than they thought before.

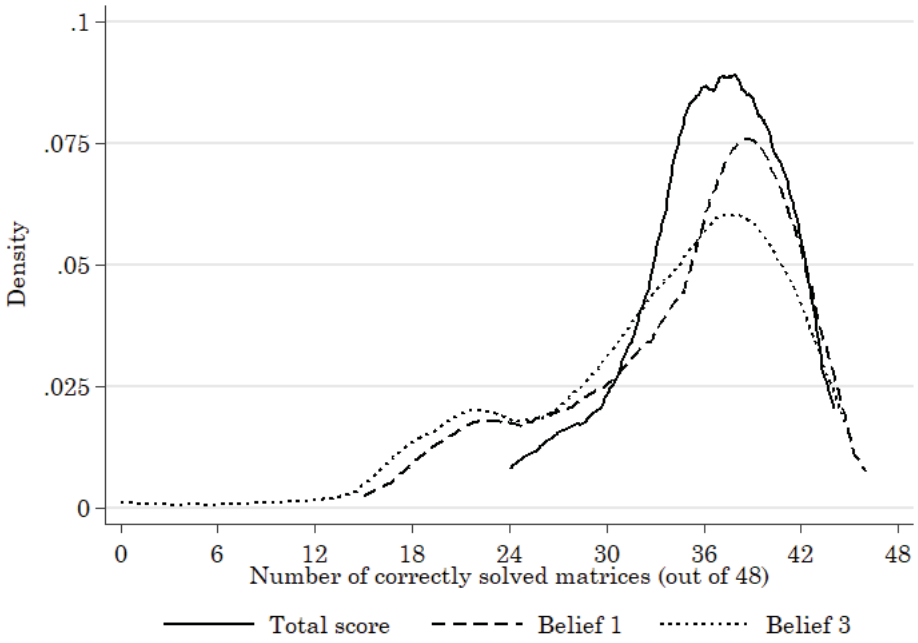
Subjects' average belief difference is -2.72 in treatment *Loss* and 1.04 in treatment *Gain* ($p<0.001$). Thus, subjects (a) update beliefs about their performance in the IQ test according to their treatment assignment and (b) update their beliefs stronger if they experience a loss in the self-image domain. The latter result is statistically significant as well and provides evidence for asymmetric belief updating ($p=0.018$). Moreover, I find that these results hold on an individual level. Subjects in both treatments update their beliefs weaker than the performance shock they experience. Subjects' belief difference is 4.07 matrices lower in treatment Gain and 2.47 higher in treatment Loss than the performance shock they experience.

According to previous findings, individuals hold motivated beliefs, dislike updating their beliefs negatively and react to noisy negative signals much weaker than to the positive ones (Coutts, 2019; Eil and Rao, 2011; Golman et al., 2017; Zimmermann, 2020). In contrast to those findings, subjects in my experiment update stronger in absolute terms when facing a negative shock to their self-image than a positive one. This result is in line with subjects' inclination to avoid a possible disappointment at the end of the experiment. I discuss the mechanism which could lead to these patterns in belief updating in Section 2.5.

Under-confidence about IQ I compare subjects' prior beliefs about their IQ and their actual performance in the IQ test. Contrary to the consensus in economic and psychological literature,[14] I detect significant under-confidence using the Wilcoxon matched-pairs signed-ranks test ($p=0.044$). On average, participants believe they will solve 1.7 matrices fewer than they do. The degree of under-confidence in the IQ domain does not vary between treatments ($p=0.721$). Furthermore, subjects remain under-confident after they have completed the task, i.e., all 48 matrices. While the average performance results in 36.4 correct answers, the average Belief 3 is only 33.1 correct answers, and the difference is highly statistically significant (Wilcoxon matched-pairs signed-ranks test, $p<0.001$). The degree of under-confidence does not differ significantly between treatments ($p=0.258$). Crucially, this under-confidence is intelligence-specific. The survey measure of confidence,

[14] See, e.g., Burks et al. (2013) and Heck et al. (2018).

based on 11 real-effort slider tasks, shows that subjects are significantly overconfident (Wilcoxon matched-pairs signed-ranks test, $p<0.001$) and expect to solve 1.52 tasks more correctly than they do.



Note: This Figure displays distributions of total scores (actual performance in the IQ test) as well as prior and second posterior beliefs about performance in the IQ test. Horizontal axis shows the total number of correctly solved matrices. Vertical axis shows the kernel density estimates using Epanechnikov kernels with an optimal bandwidth.

Figure 2.6: Performance in the IQ test and beliefs about it

In Figure 2.6, I display kernel density estimates for subjects' total performance along with their prior and second posterior beliefs about it. Prior beliefs are unaffected by treatment assignment by design. Second posterior beliefs (Belief 3) are elicited at the end of the experiment, i.e., after subjects observed and worked on all matrices.[15] In Figure 2.6, I observe that belief distributions are more left-skewed than the distribution of total scores.

In belief elicitation instructions, I gave subjects an overview of the performance of the reference sample, where, among other information,[16] I included the following statements: (a) no previous participant solved all 48 matrices, (b) the average participant solved 39 matrices, and (c) all previous participants solved at least 20 matrices or more. Subjects

[15]I do not compare their total performance and Belief 2 because Belief 2 is directly affected by treatment, which leads to positive or negative belief shocks in treatments *Gain* and *Loss*, respectively.

[16]See detailed screenshots in Appendix 2.C.

could see these statements in all belief elicitations. Interestingly, 1.5% and 5.3% of subjects still reported their perceived number of correct answers to be less than 20 in Beliefs 1 and 3, respectively. Moreover, 63.4% and 68.7% of them thought they would perform worse than an average participant of the reference sample (i.e., solve less than 39 matrices) in Beliefs 1 and 3, respectively.

In Result 1, I summarize the findings of belief updating.

Result 1. ***(Belief updating)***

(a) Subjects have on average a negative belief difference in treatment Loss and a positive belief difference in treatment Gain.

(b) Subjects in treatment Loss update their beliefs stronger in absolute terms than subjects in treatment Gain.

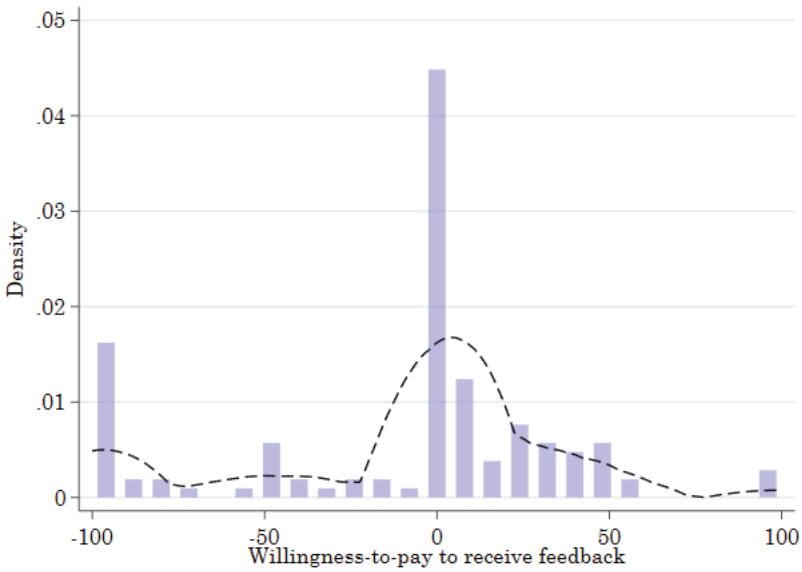
(c) Subjects are on average under-confident in their beliefs about their performance. The degree of under-confidence does not differ between treatments.

I find that subjects update their beliefs asymmetrically indicating that subjects hold motivated beliefs in the intelligence domain. This effect is, however, in the opposite direction as postulated in Hypothesis 2. Subjects in treatment *Loss* update their beliefs stronger than subjects in treatment *Gain*. While this result contrasts previous findings in the literature, I offer a simple possible explanation that is also in line with under-confidence and reference-dependence in the intelligence domain in Section 2.5.

2.4.3 Willful ignorance

In this subsection, I analyze subjects' willingness to pay to acquire feedback. The WTP measure varies between -100 ECU and 100 ECU, where -100 means that an individual certainly wants no feedback, 100 implies that an individual definitely wants feedback, and 0 corresponds to a 50% chance of getting feedback.

On average, subjects reported a negative willingness to pay of -9.5 ECU. I show the distribution of the WTP in Figure 2.7. 42.0% of subjects reported a positive willingness-to-pay for feedback, implying they were ready to forego monetary benefits to increase



Note: This Figure displays a histogram of the willingness-to-pay to receive feedback. The dashed line corresponds to the density estimates with Epanechnikov kernels and an optimal bandwidth.

Figure 2.7: Willingness-to-pay for feedback

their chances of acquiring feedback. However, only 2.3% of all participants had a WTP of 100. In total, 28.2% of subjects reported a negative willingness-to-pay to receive feedback. Moreover, 10.7% of all participants had a willingness-to-pay of -100 that guarantees no feedback about their performance in the IQ test.

Result 2. ***(Willful ignorance)***

(a) On average, subjects report a negative willingness-to-pay for self-image-relevant feedback.

(b) 28.2% of subjects report a negative willingness-to-pay to acquire feedback.

(c) 10.7% of subjects report a willingness-to-pay of -100 ECU that guarantees no feedback about their performance in the IQ test.

In line with Hypothesis 1, a non-negligible share of participants has a negative WTP for feedback relevant to their self-image. Moreover, approximately one in ten participants chooses to avoid feedback with certainty.

2.4.4 Reference dependence of self-image concerns

Participants were on average willing to pay -7.7 ECU in treatment *Gain* and -11.3 ECU in treatment *Loss* but the difference is not statistically significant (p=0.814). I report summary statistics of the WTP by treatment in Table 2.2. The shares of subjects who reported a positive, zero, and negative WTP is similar in treatments *Gain* and *Loss*. My sample size can detect a difference in willingness-to-pay of 9.3 ECU with 80% power and a significance level of 5%. With a scale from -100 to 100 ECU, the minimal detectable size of 9.3 ECU accounts for only 4.65% of the maximal shift and thus represents a minimal economically significant effect size.

Table 2.2: Summary statistics: Willingness-to-pay for feedback (by treatment)

WTP	*Loss*	*Gain*	Difference
Negative	0.297	0.269	p=0.846
Zero	0.281	0.313	p=0.707
Positive	0.422	0.418	p=1.000
N	64	67	131

Note: This table shows shares of subjects whose reported WTP to receive feedback is negative, zero and positive for treatments *Loss* and *Gain*. I compare these shares between treatments and report two-sided Fisher's exact test p-values.

To analyze reference dependence of self-image concerns, I account for how subjects update their beliefs when analyzing their WTP. I design treatments *Gain* and *Loss* to shift subjects' first posterior beliefs about their performance in the IQ test (Belief 2) by influencing their performance in Part 1. Hence, subjects endogenously update their beliefs taking into account their exogenous prior beliefs and an exogenous shock to their score in Part 1. In Table 2.3, I conduct 2SLS regressions to analyze the impact of beliefs on willingness-to-pay for feedback. I estimate the following regression:

$$\text{WTP}_i = \alpha + \beta(\text{Belief difference})_i + \gamma(\text{Belief 1})_i + \varepsilon_i,$$

which is equivalent to

$$\text{WTP}_i = \alpha + (\gamma - \beta)(\text{Belief 1})_i + \beta(\text{Belief 2})_i + \varepsilon_i.$$

I estimate the effect of belief difference on the willingness to pay for feedback on an individual level (i). The prior belief about the performance in the IQ test is exogenous. It is a proxy for the subjects' ability, and I elicit it before the treatment variation happens. The endogeneity concern arises with respect to Belief 2. Since I find evidence for motivated beliefs in my experimental data, I expect subjects to make decisions to update from Belief 1 to Belief 2 endogenously. Arguably, there might be unobservable individual effects that influence subjects' belief difference through Belief 2 and could be correlated with the error term. One likely candidate is the degree of optimism about the performance in Part 1 that can be potentially associated with belief updating and willingness to pay for feedback. Therefore, I instrument belief difference with the score in Part 1. Additionally, I include Belief 1 in both stages to account for differences in WTP for subjects with different levels of perceived ability.[17]

Relevance In the first stage, Belief 2 forms under the influence of two main criteria: previous beliefs about IQ and an exogenous shock introduced by the treatment. As I discussed in Section 2.4.1, the treatment shock affects subjects not only by treatment but also individually. Hence, using the score in Part 1 as an instrument provides me greater precision on an individual level.

Belief difference is strongly correlated with the score in Part 1. The correlation coefficient is 0.41 and highly statistically significant ($p<0.001$). This correlation emerges through the correlation between the score in Part 1 and Belief 2 (corr=0.47, $p<0.001$) but not between the score in Part 1 and Belief 1 (corr=0.13, p=0.134).

Exogeneity For the IV approach to be valid, the instrumental variable should be exogenous. In the discussed setup, the score in Part 1 should influence willingness-to-pay for feedback only through Belief 2.

The score in Part 1 is unobservable to subjects. They observe the complexity of the tasks in Part 1 and receive no additional signal about their performance. Hence, the only available information they have about the score in Part 1 *is* their perceived performance

[17] My results are robust to excluding Belief 1 as presented in Table 2.A.2.

in Part 1. Since subjects' perceived performance in Part 1 is fully reflected in Belief 2, there is no other channel through which score in Part 1 influences WTP for feedback except Belief 2. I rely on the assumption of the maximal effort provision in the IQ test that, as discussed in Section 2.4.1, is consistent with the observed performance.

Table 2.3: Instrumental variable approach: Willingness-to-pay for feedback

(a) First stage

Dependent variable:	Belief difference
Score in Part 1	0.782***
	(0.124)
Belief 1	-0.384***
	(0.078)
Constant	-1.734
	(3.252)
N	131
F statistic	39.72

(b) Second stage

Dependent variable:	WTP
Belief difference	2.966**
	(1.375)
Belief 1	0.773
	(0.676)
Constant	-33.942
	(22.952)
N	131

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are in parentheses. In these tables, I present the instrumental variable regression estimated via 2SLS. I report results of the first stage in Table (a) and results of the second stage in Table (b).

Interpretation On average, both the score in Part 1 and Belief 1 have a strong and highly significant impact on belief difference. The F statistic of 39.72 indicates that the score in Part 1 is a strong instrument. According to the first-stage results presented in Table 2.3(a), one additional correctly solved matrix increases belief difference by approximately 0.8 matrices. The results are highly statistically significant.

The second stage shows the impact of the prior beliefs and the belief difference on the willingness to pay to acquire self-image-relevant feedback. I document that one standard deviation increase in belief difference leads to a statistically significant increase in willingness-to-pay for feedback by 18.1 ECU. Prior beliefs about subjects' ability do not affect willingness-to-pay for self-image-relevant feedback significantly.

Result 3. ***(Reference dependence)***

(a) A standard deviation increase in belief difference leads to a statistically significant increase in willingness-to-pay for feedback by on average 18.1 ECU.

(b) Prior beliefs about subjects' own ability do not affect willingness-to-pay for self-image-relevant feedback significantly.

(c) The difference in willingness-to-pay between participants of treatments Gain and Loss is not statistically significant (p=0.814).

In line with Hypothesis 3, I find that participants' belief difference influences willingness-to-pay for feedback. However, contrary to Hypothesis 3, higher belief difference leads to higher willingness-to-pay. It indicates that participants who expect a gain in self-image are, on average, more willing to acquire information than those who, on average, expect a loss in their self-image. Indeed, subjects who expect bad news are more likely to avoid information than those who expect good news.

Importantly, this finding is belief-driven. A fixed performance shock introduced by the treatment assignment has no significant impact on subjects' average willingness to receive feedback. However, treatments affect subjects' beliefs about their performance in the IQ test asymmetrically. As discussed in Section 2.4.2, subjects update beliefs weaker when they expect a gain in self-image than when they expect a loss in self-image. Belief differences then lead to differences in willingness-to-pay on an individual level resulting in the higher WTP to receive feedback the larger the belief difference becomes.

2.4.5 Loss aversion in self-image concerns

To analyze whether loss aversion applies to self-image concerns, I focus on subjects with marginal perceived gains and losses. I apply a regression discontinuity design (RDD) to estimate local average treatment effects. Specifically, I use kink RDD. I aim at capturing the effect of small belief differences on willingness-to-pay to acquire feedback. Loss aversion implies a kink in incentives to receive feedback. Hence, I adjust my design to capture a kink, not a discontinuity. I present the results in Table 2.4.

Table 2.4 provides several specifications. Column (1) presents a kink RDD without additional control variables. Since belief difference is highly but not perfectly correlated with the treatment assignment, I include treatment as a control variable and present covariate-adjusted estimates in column (2). I further control for observable individual

Table 2.4: Regression discontinuity design: Willingness-to-pay for feedback

	(1)	(2)	(3)
RDD estimates	-0.949	-1.894	-4.381
	(12.299)	(12.462)	(10.650)
Covariates	none	treatment	treatment and individual characteristics
N	86	86	86

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are in parentheses. In this table, I present kink RDD local linear estimates with Epanechnikov kernels and a bandwidth of 5. In column (1), there are no additional control variables. In column (2), I control for treatment assignment. In column (3), I control for treatment assignment and individual characteristics. Individual characteristics include age, gender, occupation, field of study, monthly budget and spending, experience in laboratory experiments, number of correctly answered control questions, current GPA, high school GPA and IQ test results, measures of risk aversion, time preferences, overconfidence and intensity of social and self-image concerns. The reported number of observations indicates how many observations were actually used given a particular bandwidth selection criterion. Estimations are based on all 131 observations.

characteristics in column (3). These characteristics include age, gender, occupation, the field of study, monthly budget and spending, experience in laboratory experiments, number of correctly answered control questions, proxies for ability,[18] measures of risk aversion, time preferences, overconfidence, and intensity of social and self-image concerns.[19]

I find no evidence for loss aversion in self-image concerns. Subjects around the cut-off, i.e., those with belief differences close to zero, do not differ significantly in their willingness to pay to acquire self-image-relevant feedback. This finding is robust for specifications presented in Table 2.4 and specifications with shorter and longer bandwidths reported in Table 2.A.3.

Result 4. ***(Loss aversion)***

I find no significant difference in the effect of belief difference on willingness-to-pay for subjects with marginal gains and losses in self-image concerns.

In this experiment, I find no evidence that supports Hypothesis 4. I observe no significant difference in the effect of belief difference on willingness-to-pay for feedback between subjects with small positive and negative belief differences.

[18] Proxies for ability include current GPA, high school GPA, and IQ test results.

[19] See Section 2.2 and Table 2.A.1 for detailed explanation and summary statistics of all individual characteristics.

2.5 Mechanism

The belief formation I observe in my experiment is in contrast to the pre-registered hypotheses. In the following, I discuss a theoretical framework that offers an ex-post rationalization of those findings. The mechanism is in line with the idea of disappointment aversion (Gul, 1991) and follows closely the setup of Gollier and Muermann (2010). Gollier and Muermann (2010) propose that the decision-maker faces a trade-off between the ex-ante feelings and the risk of ex-post disappointment and chooses an optimal degree of optimism. In the context of my experiment, subjects may decide to be optimistic about their performance and derive utility ex-ante at the cost of a possible disappointment at the end of the experiment. Alternatively, participants may stay pessimistic in their beliefs throughout the experiment and likely be positively surprised at the end.

When reporting their Belief 2, subjects are aware that they only worked on Part 1 of the test, and there are 24 more matrices to solve. Arguably, subjects want to avoid a loss of self-image at the end of the experiment. Then, updating their beliefs weaker if participants are in gain can be optimal to avoid any possible disappointment at the end of the experiment.In other words, there might exist reference dependence not only within actions (willful ignorance) but also within the reported beliefs of the participants.

I observe that subjects who reported the WTP of -100 that guarantees that they do not receive feedback, were initially significantly more overconfident in the intelligence domain than others (p=0.024) and update their beliefs negatively (Belief 2 worse than Belief 1) in both treatments. The performance of these subjects in Parts 1 and 2 is not significantly different from other participants in the respective treatments (p=0.309 and p=0.287 in Parts 1 and 2, respectively). However, after they decide that they certainly do not want to receive feedback, their beliefs recover (Belief 3 better than Belief 2) in both treatments. Those findings might indicate that, at first, these subjects try to avoid disappointment in their performance by lowering the expectations, i.e., by adjusting their beliefs downwards. Yet, after they learn that they can avoid feedback altogether, they recover their beliefs accordingly.

I consider the following stylized framework to examine the mechanism of belief updat-

ing and incentives to acquire information that I observe in my experimental data.

Setup There are two stages denoted $t \in \{1, 2\}$. In $t = 1$, agents hold a prior belief about their type based on a self-image-relevant characteristic. In my experiment, this self-image-relevant characteristic is the number of correctly solved matrices in the IQ test. In $t = 2$, they face an exogenous shock to this characteristic, update their beliefs in response to the shock and choose whether to acquire or avoid information that affects their self-image.

I consider dual-self agents who derive reference-dependent utility from self-image. The concept of dual selves distinguishes the "rule chooser" and the "rule user", or a rational and an emotional self, for each agent (Bénabou and Jean Tirole, 2002; Eil and Rao, 2011; Fudenberg and Levine, 2006; Greiff, 2019). In my setup, the dual-self agent consists of two decision-makers: the rational self (R) and the emotional self (E). In $t = 2$, the emotional self shapes motivated beliefs, and the rational self takes beliefs as given and decides whether to acquire information.

An agent holds a prior belief about her type $n_1 = n \in [0, N]$ in period $t = 1$. She derives utility

$$\phi_1(n_1) = u(n_1),$$

where $u(\cdot)$ is an increasing and differentiable utility function from self-image. In $t = 2$, agents experience an exogenous self-image shock $s \in (0, \bar{s})$ with $\bar{s} < 1/2$ which can influence their perceived type either positively or negatively. Agents perceive this shock with a degree of optimism $\alpha \in [-a, a]$ with $a < 1$. Their motivated posterior beliefs are $n_{2m}^{Gain} = [1 + (1 + \alpha)s]n > n_1$ if they are exposed to a positive shock, and $n_{2m}^{Loss} = [1 - (1 - \alpha)s]n < n_1$ if they are exposed to a negative one. Essentially, the agent's beliefs are influenced by a shock s and the degree of optimism α determines the agent's sensitivity to this shock. I call agents "optimistic" when $\alpha > 0$ because it corresponds to overestimating the positive shock and underestimating the negative one. I refer to agents as "neutral" if $\alpha = 0$ and "pessimistic" if $\alpha < 0$.

Belief updating In my experiment, subjects do not know about the possibility of acquiring or avoiding feedback before they arrive at the respective decision screen. Hence, when reporting Belief 2, I assume that their status quo is that they *will* receive feedback about their performance in the IQ test. It is plausible to assume that, without any additional indication, individuals who work on an IQ test would expect to receive results upon completing the test.

The emotional self E endogenously chooses an optimal degree of optimism, while the rational self R takes it as given. E knows that R holds motivated beliefs n_{2m} and that the agent will receive information about her ability and will have to update to n_{2u} (unmotivated beliefs. Hence, her posterior beliefs will become $n_{2u}^{Gain} = (1+s)n$ if she is exposed to a positive shock, and $n_{2u}^{Loss} = (1-s)n$ if she is exposed to a negative one. Furthermore, E knows that R will experience losses whenever $n_{2m} < n_1$ or $n_{2u} < n_{2m}$, and gains otherwise. Therefore, E's objective is to choose α in the best interest of R. On the one hand, E wants to maximize the gain or minimize the loss when R updates her beliefs from n_1 to n_{2m}. On the other hand, E takes into account maximizing gains or minimizing losses from R updating from n_{2m} to n_{2u}. First, the emotional self maximizes the following utility function with respect to α:

$$\phi_{2E}(n_1, n_{2m}, n_{2u}|\text{information}) = u(n_{2u}) + I_{Loss}(n_{2u} - n_{2m}) + I_{Loss}(n_{2m} - n_1),$$

where I_{Loss} is an index function which equals $\lambda > 1$ whenever its argument is negative and one otherwise. I assume that agents' reference point is their prior belief about their type n_1. Hence, negative deviations from n_1 have a larger absolute impact on utility than equally-sized positive deviations.

Proposition 1. *Agents who are exposed to the positive self-image shock are non-optimistic* ($\alpha \in [-1, 0]$).

Proof. See Appendix 2.B.1 for the proof. □

Essentially, the agent with a positive shock to her self-image will experience a gain

when she updates from n_1 to n_{2m}. Depending on the degree of optimism the emotional self chooses, this gain may be relatively small if the agent's beliefs are pessimistic and relatively large if her beliefs are optimistic. Additionally, she may experience a gain or a loss when she updates from n_{2m} to n_{2u}. Her emotional self wants her to avoid this potential loss and hence keeps her motivated beliefs non-optimistic to avoid the disappointment when updating from n_{2m} to n_{2u}.

Then, conditional on receiving information about their performance, agents' utility with the optimal degree of optimism is

$$\phi_2^{Gain}(n_1, n_{2u}|\text{information}, \alpha^*) = u(n_{2u}) + (n_{2u} - n_1).$$

Next, I examine how agents with a negative self-image shock update their beliefs. I summarize my findings in Proposition 2.

Proposition 2. *Agents who are exposed to the negative self-image shock are non-pessimistic ($\alpha \in (0, 1]$).*

Proof. See Appendix 2.B.2 for the proof. □

The intuition behind this finding is as follows. The agent with a negative shock to her self-image will experience a loss when she updates from n_1 to n_{2m}. If her beliefs are pessimistic, this perceived loss will be relatively large, i.e., larger than the shock s. Conversely, if her beliefs are optimistic, her perceived loss is relatively small. Additionally, she may experience a gain or a loss when she updates from n_{2m} to n_{2u}. Since losses have a stronger negative impact on the agent's utility than gains of the same size, the agent's emotional self wants her to avoid the "excessive" potential utility loss when updating from n_1 to n_{2m}. The additional utility of having the gain when updating from n_{2m} to n_{2u} cannot compensate this loss. Therefore, the agent's motivated beliefs are non-pessimistic.

Then, conditional on receiving information about their performance, agents' utility with the optimal degree of optimism is

$$\phi_2^{Loss}(n_1, n_{2u}|\text{information}, \alpha^*) = u(n_{2u}) + \lambda(n_{2u} - n_1).$$

Incentives to acquire information The rational self R learns about the possibility of choosing whether to acquire or avoid information. She chooses to acquire information whenever her expected utility from acquiring information is higher than from avoiding it, conditional on an optimal degree of optimism. If the agent decides to acquire information, she has to forego the utility from her optimism α and perceive the performance shock objectively. If the agent acquires the information, her utility in is $\phi_2(n_1, n_{2u}|\text{information}, \alpha^*)$. If she chooses to avoid information, she holds her motivated beliefs n_{2m} and has the following utility:

$$\phi_2(n_1, n_{2m}|\text{no information}, \alpha^*) = u(n_{2m}) + I_{Loss}(n_{2m} - n_1).$$

Agents choose to acquire information whenever

$$\phi_2(n_1, n_{2u}|\text{information}, \alpha^*) \geq \phi_2(n_1, n_{2m}|\text{no information}, \alpha^*).$$

I analyze the optimal information avoidance for agents in *Gain* and *Loss* separately. I summarize my findings of the incentives to acquire information for agents who experience a positive shock to their self-image in Proposition 3.

Proposition 3. *Agents who are exposed to the positive self-image shock acquire information conditional on their optimal degree of optimism ($\alpha^* \in [-1, 0]$).*

Proof. See Appendix 2.B.3 for the proof. □

Agents who are exposed to the positive self-image shock update their beliefs non-optimistically. Therefore, acquiring feedback improves their utility from self-image and yields a gain due to shifting beliefs from n_{2m} to $n_{2u} \geq n_{2m}$.

I proceed to analyze the incentives to acquire information for agents with a negative shock to their self-image. I show my findings in Proposition 4.

Proposition 4. *Agents who are exposed to the negative self-image shock avoid information conditional on their optimal degree of optimism ($\alpha^* \in [0, 1]$).*

Proof. See Appendix 2.B.4 for the proof. □

Agents who are experiencing the negative self-image shock update their beliefs non-pessimistically. Hence, acquiring feedback would deteriorate their utility from self-image and yields a loss due to shifting beliefs from n_{2m} to $n_{2u} \leq n_{2m}$. Therefore, agents with a negative self-image shock optimally avoid information relevant to their self-image.

I analyze this mechanism in a stylized framework where dual-self agents are optimally non-optimistic if they experience the positive shock in their self-image and optimally non-pessimistic when they experience a negative shock to their self-image. These patterns in belief updating are in line with my experimental data. I observed that subjects in treatment *Gain* update their beliefs by 4.07 matrices weaker than the positive performance shock they experience. In other words, if subjects in treatment *Gain* were neutral agents with $\alpha = 0$, their belief difference would have been 4.07 matrices larger. It indicates that subjects in treatment *Gain* are indeed pessimistic in their belief updating. Conversely, subjects in treatment *Loss* are optimistic. They update their beliefs by 2.47 matrices weaker than the negative performance shock they experience. If subjects in treatment *Loss* were neutral agents, their belief difference would have been 2.47 matrices lower.

In the proposed mechanism, the agents who experience a positive shock choose to acquire information, and the agents who experience a negative shock prefer to avoid it. This result is driven by the fact that the shock influences the optimal degree of optimism which in turn drives the updating process. My experimental data shows that an increase in the difference between the first posterior belief (Belief 2) and the prior belief (Belief 1) indeed leads to a higher willingness to pay for information.

2.6 Conclusion

This paper sheds light on the complexity and the dynamic nature of self-image concerns. Individual perception of oneself is naturally belief-driven. Thus, understanding the motivation behind updating beliefs in this domain and the channels through which beliefs shape one's self-image is crucial for all decisions where self-image plays a role.

In this paper, I analyze individuals' willingness to avoid self-image-relevant information after I expose them to positive or negative shocks in their self-image. I complement

this approach with multiple elicitations of beliefs about their self-image. They allow me to observe the impact of positive and negative shocks on an individual level. In my experiment, I induce an exogenous shift in self-image by introducing treatments *Loss* and *Gain*. Then, I ask subjects whether they are willing to acquire feedback about their performance and elicit their willingness-to-pay to do so and their beliefs about their performance.

As intended by the experimental design, individuals assigned to treatment *Gain* have a positive change in beliefs driven by a positive shock to their performance. Individuals in treatment *Loss*, on the contrary, update their beliefs negatively in line with a negative exogenous performance shock they experience. Interestingly, subjects in treatment *Loss* update their beliefs stronger than subjects in treatment *Gain*. Moreover, subjects in both treatments are, on average, under-confident and pessimistic in their beliefs about their intelligence. I propose a stylized theoretical framework to analyze the underlying mechanism. A possible explanation for this pessimism in beliefs is disappointment aversion (Gollier and Muermann, 2010; Kőszegi and Rabin, 2007).

On average, subjects report a negative willingness to pay for feedback relevant to their self-image. Almost one-third of participants reported a negative willingness-to-pay to acquire feedback. Moreover, about one in ten subjects had the lowest possible willingness-to-pay that guarantees no feedback about their performance in the IQ test.

I document causal evidence for reference dependence of self-image concerns. I find that an increasing change in beliefs about the performance in the IQ test leads to a statistically significant increase in willingness to pay for feedback. Furthermore, prior beliefs about subjects' ability do not affect willingness-to-pay for self-image-relevant feedback significantly. Hence, the difference in willingness-to-pay between participants of treatments *Gain* and *Loss* being not statistically significant is driven by asymmetric belief updating. Moreover, I find no significant difference in the effect of belief difference on willingness-to-pay for subjects with marginal gains and losses in self-image concerns.

Generally, this paper studies the implications of self-image for the demand for relevant feedback and the evolution of their self-image itself. My findings may have broad implica-

tions in various domains like health, finance, labor, prosocial and altruistic behavior, etc. While avoiding information may maximize the short-term utility of an individual, it may yield severe welfare losses in the long run or negatively affect individuals themselves as well as those around them. For example, managers may avoid helpful feedback to maintain their self-image as a professional. It hinders them from becoming better managers and potentially affects the performance of their entire team. Curating effective feedback systems can therefore be vital for the well-being of the firms. Charitable donors who often prefer to remain uninformed about the actual effectiveness of donations experience a short-term warm glow from their actions. However, making a more informed choice could lead to more effective use of their resources. Another prominent and recent example comes from the rising necessity of lesson and lecture recordings. Many teachers may be reluctant to watch them back, despite apparent benefits for improving their teaching style, to protect their ego. Detrimental effects of losses in self-image may be even more pronounced if individuals do not hold a strong prior in a particular domain. For example, new employees or students may be particularly vulnerable groups. Hence, the task allocators in the workplace and the designer of educational programs may regard the self-image effects and their possible consequences for feedback avoidance. Careful consideration of whether individuals experience gains or losses in self-image is crucial, as they can hinder individuals from acquiring relevant information.

My findings offer several avenues for future research. First, motivated belief updating relies strongly on the subjects' status quo in a given environment. Therefore, influencing subjects' perception of the status quo may shed more light on the formation of motivated beliefs. Furthermore, individuals tend to internalize negative feedback weaker than positive one (Zimmermann, 2020). Combining this finding with the evidence from my experimental data on stronger belief updating in the presence of a negative signal but *without feedback* may be insightful. Moreover, I focus on intelligence as a self-image-relevant domain in this paper. However, many studies have previously documented that individuals derive self-image utility from a wide range of characteristics, e.g., beauty (Eil and Rao, 2011) or morality (Gneezy, Saccardo, et al., 2020). Investigating whether in-

dividual behavior in case of gains and losses in other self-image-relevant domains follows similar patterns might be the next step towards a deeper understanding of how individuals perceive themselves.

Appendix

2.A Additional tables

Table 2.A.1: Differences in individual characteristics in treatments *Gain* and *Loss*

Individual characteristics	Variable type	Min	Max	*Loss*	*Gain*	Difference
Age	Continuous	18	49	25.844	24.985	0.362
Gender: 1 if female	Binary	0	1	0.656	0.522	0.156
Gender: 1 if diverse	Binary	0	1	0.000	0.015	1.000
Occupation: 1 if student	Binary	0	1	0.891	0.925	0.555
Field of study: 1 if economics	Binary	0	1	0.344	0.343	1.000
Field of study: 1 if psychology	Binary	0	1	0.016	0.030	1.000
Lab experience	Continuous	1	500	18.313	7.627	0.353
Current GPA	Continuous	1	4	2.207	2.230	0.969
High school GPA	Continuous	1	3.7	2.097	2.260	0.210
Monthly budget	Continuous	0	4000	532.359	505.299	0.434
Monthly spending	Continuous	0	1500	328.297	299.179	0.467
Control questions (# correct)	Continuous	1	3	2.781	2.896	0.155
Risk aversion	Continuous	1	10	4.969	5.448	0.309
Overconfidence	Continuous	-6	10	2.063	1.000	0.066
Time preferences	Continuous	1	10	7.250	7.090	0.757
Social image concerns	Continuous	0	10	4.938	5.254	0.524
Self-image concerns	Continuous	0	60	38.000	38.612	0.967
Loss aversion	Continuous	0	6	3.531	3.493	0.861
N				64	67	131

Note: I show summary statistics for subjects' individual characteristics in treatments *Loss* and *Gain*. I report the mean, minimal and maximal values of each variable. I also display p-values for treatment comparison for each corresponding variable. I compare the distributions of the variables marked "Continuous" using two-sided MWU tests. I compare the distributions of the variables marked "Binary" using two-sided Fisher's exact tests. Gender is a categorical variable (m/f/d). I test the differences between treatments by category. A detailed description of how I measure all individual characteristics is provided in Appendix 2.C.7 and 2.C.8 in English and German (original), respectively. Subjects' occupation was originally elicited as binary and indicates if an individual is a student. Field of study is a categorical variable and contains multiple fields, namely, mathematics or science, medicine, psychology, law or social sciences, economics, other and "I do not study". Following Abeler et al., 2019, I focus on economics and psychology. Lab experience indicates a self-reported number of economic experiments the subject has participated in. Please note that, despite the maximum of 500, 95% of subjects participated in 30 or fewer experiments. 79% of all subjects participated in 10 or fewer experiments. Current GPA and high school GPA reflect the standardized German grading system, with 1.0 corresponding to the best possible grade and 4.0 to the worst passing grade. Monthly budget and spending are measured in Euro, with fixed costs like rent already subtracted. Variable "Control questions" indicates the number of correctly answered control questions about the instructions of the current experiment (out of 3). Risk aversion, time preferences, and social image concerns are measured on an 11-point Likert scale (0-10). Larger reported values correspond to having a higher willingness to take risks, being more patient, and having stronger social image concerns, respectively. Overconfidence may vary between -11 and 11. Negative values of overconfidence correspond to under-confidence; Larger values imply stronger overconfidence. Self-image concerns is a measure that varies between -30 and 60 and indicates the intensity of self-image concerns, with larger values indicating stronger self-image concerns. Loss aversion may vary between zero and 6, and larger values mean stronger loss aversion.

Table 2.A.2: Instrumental variable approach robustness check: Willingness-to-pay for feedback

(a) First stage

Dependent variable: Belief difference	
Score in Part 1	0.686***
	(0.116)
Constant	-13.303***
	(2.263)
N	131
F statistic	34.73

(b) Second stage

Dependent variable: WTP	
Belief difference	3.245**
	(1.511)
Constant	-6.920
	(4.272)
N	131

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are in parentheses. In these tables, I present the instrumental variable regression estimated via 2SLS. I report results of the first stage in Table (a) and results of the second stage in Table (b).

Table 2.A.3: Regression discontinuity design robustness check: Willingness-to-pay for feedback

	Bandwidth = 4			Bandwidth = 6		
	(1)	(2)	(3)	(4)	(5)	(6)
RDD estimates	3.872	3.827	1.079	2.255	1.679	4.946
	(16.048)	(16.300)	(14.447)	(7.220)	(7.199)	(6.771)
Covariates	none	treatment	treatment and individual characteristics	none	treatment and individual characteristics	treatment
N	73	73	73	92	92	92

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are in parentheses. In this table, I present kink RDD local linear estimates with Epanechnikov kernels. In columns (1)-(3), I use a bandwidth of 4. In columns (4)-(6), I use a bandwidth of 6. In column (1), there are no additional control variables. In column (2), I control for treatment assignment. In column (3), I control for treatment assignment and individual characteristics. Individual characteristics include age, gender, occupation, field of study, monthly budget and spending, experience in laboratory experiments, number of correctly answered control questions, current GPA, high school GPA and IQ test results, measures of risk aversion, time preferences, overconfidence and intensity of social and self-image concerns. The reported number of observations indicates how many observations were actually used given a particular bandwidth selection criterion. Estimations are based on all 131 observations.

2.B Proofs

2.B.1 Proof of Proposition 1

For an agent exposed to a positive performance shock (*Gain*), the emotional self maximizes the following utility with respect to the degree of optimism α:

$$\phi_{2E}^{Gain}(n_1, n_{2m}, n_{2u}|\text{information}) = u\Big((1+s)n\Big)+I_{Loss}\Big((1+s)n-[1+(1+\alpha)s]n\Big)+\Big([1+(1+\alpha)s]n-n\Big).$$

Importantly, $\phi_{2E}^{Gain}(n_1, n_{2m}, n_{2u}|\text{information})$ is non-differentiable at $(1+s)n-[1+(1+\alpha)s]n=0$ or $\alpha=0$. It is because at $\alpha=0$, I_{Loss} switches between one and λ, thus creating a kink in the utility function. Hence, I consider cases of $\alpha>0$ and $\alpha\leq 0$ separately.[20]

Case $\alpha > 0$ The utility can be simplified and becomes

$$\phi_{2E}^{Gain}(n_1, n_{2m}, n_{2u}|\text{information}) = u\Big((1+s)n\Big)+\lambda\Big((1+s)n-[1+(1+\alpha)s]n\Big)+\Big([1+(1+\alpha)s]n-n\Big).$$

Notably,

$$\left.\frac{\partial(\phi_{2E}^{Gain}(n_1, n_{2m}, n_{2u}|\text{information}))}{\partial\alpha}\right|_{\alpha>0} = sn(1-\lambda) < 0.$$

Hence, the agents utility decreases in α for $\alpha > 0$. Positive α generates a relatively large gain when the agent updates from n_1 to n_{2m}. However, it also leads to a loss while updating from n_{2m} to n_{2u}. Since the agent dislikes losses more than she appreciates gains of the same size, larger α affects the agent's utility negatively.

Case $\alpha \leq 0$ The utility can be simplified and becomes

$$\phi_{2E}^{Gain}(n_1, n_{2m}, n_{2u}|\text{information}) = u\Big((1+s)n\Big) + \Big((1+s)n-n\Big).$$

for $\alpha\leq 0$. The agents' utility does not depend on the degree of optimism α. The agent is in gain while updating from n_1 to n_{2m} and from n_{2m} to n_{2u}. Therefore, any $\alpha\in[-1,0]$ is optimal.

[20] I examine the cases of $\alpha=0$ and $\alpha<0$ together because both for neutral and pessimistic agents I_{Loss} equals one, while for the optimistic agents $I_{Loss}=\lambda$.

2.B.2 Proof of Proposition 2

For an agent exposed to a negative performance shock (*Loss*), the emotional self maximizes the following utility with respect to the degree of optimism α:

$$\phi_{2E}^{Loss}(n_1, n_{2m}, n_{2u}|\text{information}) = u\Big((1-s)n\Big)+I_{Loss}\Big((1-s)n-[1-(1-\alpha)s]n\Big)+\Big([1-(1-\alpha)s]n-n\Big).$$

$\phi_{2E}^{Loss}(n_1, n_{2m}, n_{2u}|\text{information})$ is non-differentiable if $(1-s)n-[1-(1-\alpha)s]n = 0$ or $\alpha = 0$. Therefore, I consider cases of $\alpha \geq 0$ and $\alpha < 0$ separately.

Case $\alpha < 0$ The expected utility can be simplified and becomes

$$\phi_{2E}^{Loss}(n_1, n_{2m}, n_{2u}|\text{information}) = u\Big((1-s)n\Big)+\Big((1-s)n-[1-(1-\alpha)s]n\Big)+\lambda\Big([1-(1-\alpha)s]n-n\Big).$$

The agent's utility increases in α for $\alpha < 0$, since

$$\left.\frac{\partial(\phi_{2E}^{Loss}(n_1, n_{2m}, n_{2u}|\text{information}))}{\partial\alpha}\right|_{\alpha<0} = sn(\lambda-1) > 0.$$

Hence, negative α generates a relatively large loss when the agent updates from n_1 to n_{2m}. However, it also leads to a gain while updating from n_{2m} to n_{2u}. Since the agent dislikes losses more than she appreciates gains of the same size, smaller α affects the agent's utility negatively.

Case $\alpha \geq 0$ The utility can be simplified and becomes

$$\phi_{2E}^{Loss}(n_1, n_{2m}, n_{2u}|\text{information}) = u\Big((1-s)n\Big) + \Big((1-s)n-n\Big).$$

for $\alpha \geq 0$. The agents' utility does not depend on alpha. The agent is in loss while updating from n_1 to n_{2m} and from n_{2m} to n_{2u}. Therefore, any $\alpha \in [0,1]$ is optimal. Therefore, an agent with a negative performance shock is optimally non-pessimistic.

2.B.3 Proof of Proposition 3

The agent with a positive self-image shock acquires information about her performance whenever

$$u\Big((1+s)n\Big) + \Big((1+s)n - n\Big) \geq u\Big([1+(1+\alpha)s]n\Big) + \Big([1+(1+\alpha)s]n - n\Big). \quad (2.1)$$

As established in Proposition 1, agents who experience a positive shock are optimally non optimistic, i.e., have $\alpha^* \in [-1, 0]$. For any $\alpha^* \in [-1, 0]$, $u\Big((1+s)n\Big) \geq u\Big([1+(1+\alpha)s]n\Big)$ and $(1+s)n \geq [1+(1+\alpha)s]n$. Therefore, Condition 2.1 always holds. Hence, agents who are exposed to the positive self-image shock acquire information conditional on their optimal degree of optimism ($\alpha^* \in [-1, 0]$).

2.B.4 Proof of Proposition 4

The agent with a negative self-image shock acquires information about her performance whenever

$$u\Big((1-s)n\Big) + \lambda\Big((1-s)n - n\Big) \geq u\Big([1-(1-\alpha)s]n\Big) + \lambda\Big([1-(1-\alpha)s]n - n\Big). \quad (2.2)$$

As established in Proposition 2, agents who experience a positive shock are optimally non-pessimistic, i.e., have $\alpha^* \in [0, 1]$. For any $\alpha^* \in [0, 1]$, $u\Big((1-s)n\Big) \leq u\Big([1-(1-\alpha)s]n\Big)$ and $(1-s)n \leq [1-(1-\alpha)s]n$. Therefore, Condition 2.2 never holds. Hence, agents who are exposed to the negative self-image shock avoid information conditional on their optimal degree of optimism ($\alpha^* \in [0, 1]$).

2.C Instructions of the Experiment

2.C.1 General instructions: English

Please read the following instructions carefully! The amount of money you earn in this experiment strongly depends on your decisions. If you have any questions, please write a message to the experimenters in the chat. We will reply as soon as we can. During the experiment, it is not allowed to talk to other participants of the experiment or other people, use mobile phones or start other programs on the computer. Non-compliance with these rules will result in exclusion from the experiment and all payments. On the following pages we describe the exact procedure of the experiment.

In this experiment, we calculate your earnings using experimental currency units (talers). At the end of this experiment, all your earnings will be converted from talers to euro using the following exchange rate:

1 taler = 5 cents.

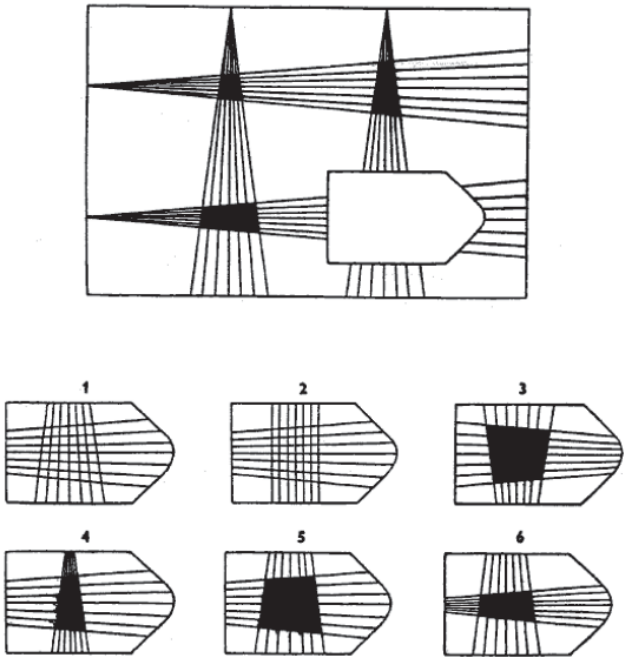
You will receive **a fixed payment of 74 talers** for participating in this experiment, which will be paid at the end of the experiment independent of your decisions in the experiment. Additionally, you receive **an endowment of 100 talers** which you might use in the course of the experiment. Please note that you receive your payments only upon completion of the entire experiment. In the following, there is a description of the exact experimental procedure.

Overview of the Experiment

This experiment consists of **48 tasks** (24 tasks in Part 1 and 24 tasks in Part 2), which are often used **to measure so-called fluid intelligence of a person**. The fluid intelligence is an important part of the general intelligence of humans. These or similar tasks are also often used by companies in the context of recruitment procedures. **Each task corresponds to a picture puzzle.**

Each picture puzzle shows in its upper part a pattern in a box, in which a "piece of the puzzle" in the lower right corner is left out. Your task is to select one of the puzzle pieces

Example for a picture puzzle:



listed below the box, which will logically fill the blank lower right corner of the pattern in the box. **Please enter the number of the puzzle piece that you think fits best on the screen.** The number of a puzzle piece is stated above each puzzle piece. There is always exactly one piece that fits best.

You have **30 seconds** to complete each picture puzzle. For each correctly completed picture puzzle you receive one point. **As commonly done with intelligence tests, correct answers are not paid extra.** You will receive 0 points for each wrongly answered picture puzzle or if you do not enter the best fitting piece of the puzzle within 30 seconds.

All participants in the experiment **work on exactly the same 48 picture puzzles** described above. Each participant is randomly assigned to one of two groups: Group A or Group B. Throughout the whole experiment, all participants of both groups will solve exactly the same 48 picture puzzles, 24 in Part 1 and 24 in Part 2. Only the order in which the picture puzzles are processed differs between group A and B, which has an influence on the relative complexity of the parts. The group membership has no further meaning. In Parts 1 and 2 you belong to the same group.

Part 1 of the Experiment

Before you start working on the picture puzzles, there will be some screens with questions. Then, you work on 24 picture puzzles following the rules described above (30 seconds time per puzzle, 1 point for correct answers, 0 points otherwise, etc.). After you have completed all 24 picture puzzles in Part 1, there will be some screens with questions before we proceed to Part 2.

Part 2 of the Experiment

Part 2 of the experiment is very similar to Part 1. You work on 24 more picture puzzles following the same rules (30 seconds time per puzzle, 1 point for correct answers, 0 points otherwise, etc.).

End and Payment of the Experiment

After Part 2 of today's experiment, there will be some more screens with information and questions before we proceed to the payment.

If you have any questions now,
please write a message to the experimenters in the chat.
We will reply as soon as we can.

Control questions

1. According to which rule will your earnings be converted from the experimental currency units (talers) to euro? (correct answer - c)

   (a) 1 taler = 1 cent

   (b) 1 taler = 3 cents

   (c) 1 taler = 5 cents

   (d) 1 taler = 10 cents

2. How many tasks are you going to work on? (correct answer - c)

   (a) 24

   (b) 30

(c) 48

(d) 60

3. How much time do you have to work on each picture puzzle? (correct answer - b)

(a) 15 seconds

(b) 30 seconds

(c) 45 seconds

(d) 60 seconds

2.C.2 General instructions: German (original)

Bitte lesen Sie die folgenden Instruktionen sorgfältig durch! Die Höhe Ihres Gewinns bei diesem Experiment hängt wesentlich von Ihren Entscheidungen ab. Wenn Sie Fragen haben, schreiben Sie bitte eine Nachricht an die ExperimentatorInnen im Chat. Wir werden so schnell wie möglich antworten. Während des Experiments ist es nicht erlaubt, mit anderen Teilnehmenden des Experiments oder anderen Personen zu sprechen, Handys zu benutzen oder andere Programme auf dem Computer zu starten. Die Nichteinhaltung dieser Regeln führt zum Ausschluss vom Experiment und sämtlicher Zahlungen. Auf den folgenden Seiten beschreiben wir den genauen Ablauf des Experiments.

In diesem Experiment berechnen wir Ihren Gewinn in Form von experimentellen Währungseinheiten (Taler). Am Ende des Experiments werden alle Ihre Gewinne unter Verwendung des folgenden Wechselkurses von Taler in Euro umgerechnet:

1 Taler = 5 Cent.

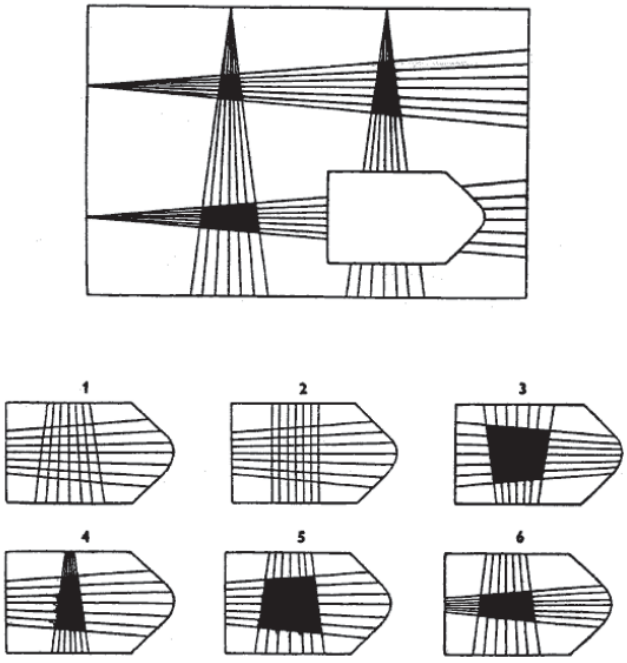
Sie erhalten **eine feste Zahlung von 74 Taler** für die Teilnahme an diesem Experiment, die am Ende des Experiments unabhängig von Ihren Entscheidungen im Experiment ausgezahlt wird. Zusätzlich erhalten Sie **eine Anfangsausstattung von 100 Taler**, die Sie im Laufe des Experiments verwenden können. Bitte beachten Sie, dass Sie Ihre Zahlungen erst nach Abschluss des gesamten Experiments erhalten. Im Folgenden finden Sie eine Beschreibung des genauen Versuchsablaufs.

Überblick über das Experiment

Dieses Experiment besteht aus **48 Aufgaben** (24 Aufgaben in Teil 1 und 24 Aufgaben in Teil 2), **die häufig zur Messung der** sogenannten **fluiden Intelligenz einer Person verwendet werden**. Die fluide Intelligenz ist ein wichtiger Teil der allgemeinen Intelligenz des Menschen. Diese oder ähnliche Aufgaben werden auch oft von Unternehmen im Rahmen von Einstellungsverfahren eingesetzt. **Jede Aufgabe entspricht einem Bilderrätsel.**

Jedes Bilderrätsel zeigt im oberen Teil ein Muster in einem Kasten, bei dem ein "Puzzleteil" in der unteren rechten Ecke ausgelassen ist. Ihre Aufgabe ist es, eines der unter

Beispiel für ein Bilderrätsel:



dem Kasten aufgeführten Puzzleteile auszuwählen, das die leere untere rechte Ecke des Musters im Kasten logisch ausfüllt. **Bitte geben Sie die Nummer des Puzzleteils ein, das Ihrer Meinung nach am besten in den Rahmen passt.** Die Nummer eines Puzzleteils ist über jedem Puzzleteil angegeben. Es gibt immer genau ein Teil, das am besten passt.

Sie haben **30 Sekunden Zeit**, um die einzelnen Bilderrätsel zu lösen. Für jedes richtig ausgefüllte Bilderrätsel erhalten Sie einen Punkt. **Wie bei Intelligenztests üblich, werden richtige Antworten nicht zusätzlich vergütet.** Sie erhalten 0 Punkte für jedes falsch beantwortete Bilderrätsel oder wenn Sie nicht innerhalb von 30 Sekunden das am besten passende Teil des Rätsels auswählen.

Alle Teilnehmenden des Experiments **arbeiten an genau den gleichen 48 Bilderrätseln**, die oben beschrieben wurden. Die Teilnehmenden werden zufällig einer von zwei Gruppen zugewiesen: Gruppe A oder Gruppe B. Während des gesamten Experiments lösen alle Teilnehmenden beider Gruppen genau die gleichen 48 Bilderrätsel, 24 in Teil 1 und 24 in Teil 2. Nur die Reihenfolge, in der die Bilderrätsel bearbeitet werden, unterscheidet sich zwischen Gruppe A und B, was einen Einfluss auf die relative Komplexität der Teile hat. Die Gruppenzugehörigkeit hat keine weitere Bedeutung. Sie gehören in

Teil 1 und 2 der gleichen Gruppe an.

Teil 1 des Experiments

Bevor Sie mit der Bearbeitung der Bilderrätsel beginnen, werden mehrere Seiten mit Fragen angezeigt. Dann bearbeiten Sie 24 Bilderrätsel nach den oben beschriebenen Regeln (30 Sekunden Zeit pro Rätsel, 1 Punkt für richtige Antworten, ansonsten 0 Punkte, usw.). Nachdem Sie alle 24 Bilderrätsel in Teil 1 gelöst haben, werden erneut ein paar Seiten mit Fragen gezeigt, bevor wir zu Teil 2 übergehen.

Teil 2 des Experiments

Teil 2 des Experiments ist sehr ähnlich zu Teil 1. Sie bearbeiten 24 weitere Bilderrätsel nach den gleichen Regeln (30 Sekunden Zeit pro Rätsel, 1 Punkt für richtige Antworten, ansonsten 0 Punkte, usw.).

Ende und Bezahlung des Experiments

Nach Teil 2 des heutigen Experiments werden noch einige Seiten mit Informationen und Fragen angezeigt, bevor wir zur Bezahlung übergehen.

Wenn Sie jetzt noch Fragen haben,

schreiben Sie bitte eine Nachricht an die ExperimentatorInnen im Chat.

Wir werden so schnell wie möglich antworten.

Kontrollfragen

1. Nach welcher Regel wird Ihr Gewinn von der experimentellen Währungseinheit (Taler) in Euro umgerechnet? (richtige Antwort - c)

   (a) 1 Taler = 1 Cent

   (b) 1 Taler = 3 Cent

   (c) 1 Taler = 5 Cent

   (d) 1 Taler = 10 Cent

2. Wie viele Aufgaben werden Sie bearbeiten? (richtige Antwort - c)

   (a) 24

(b) 30

(c) 48

(d) 60

3. Wie viel Zeit haben Sie für die Bearbeitung der einzelnen Bilderrätsel? (richtige Antwort - b)

(a) 15 Sekunden

(b) 30 Sekunden

(c) 45 Sekunden

(d) 60 Sekunden

2.C.3 Belief elicitations: English

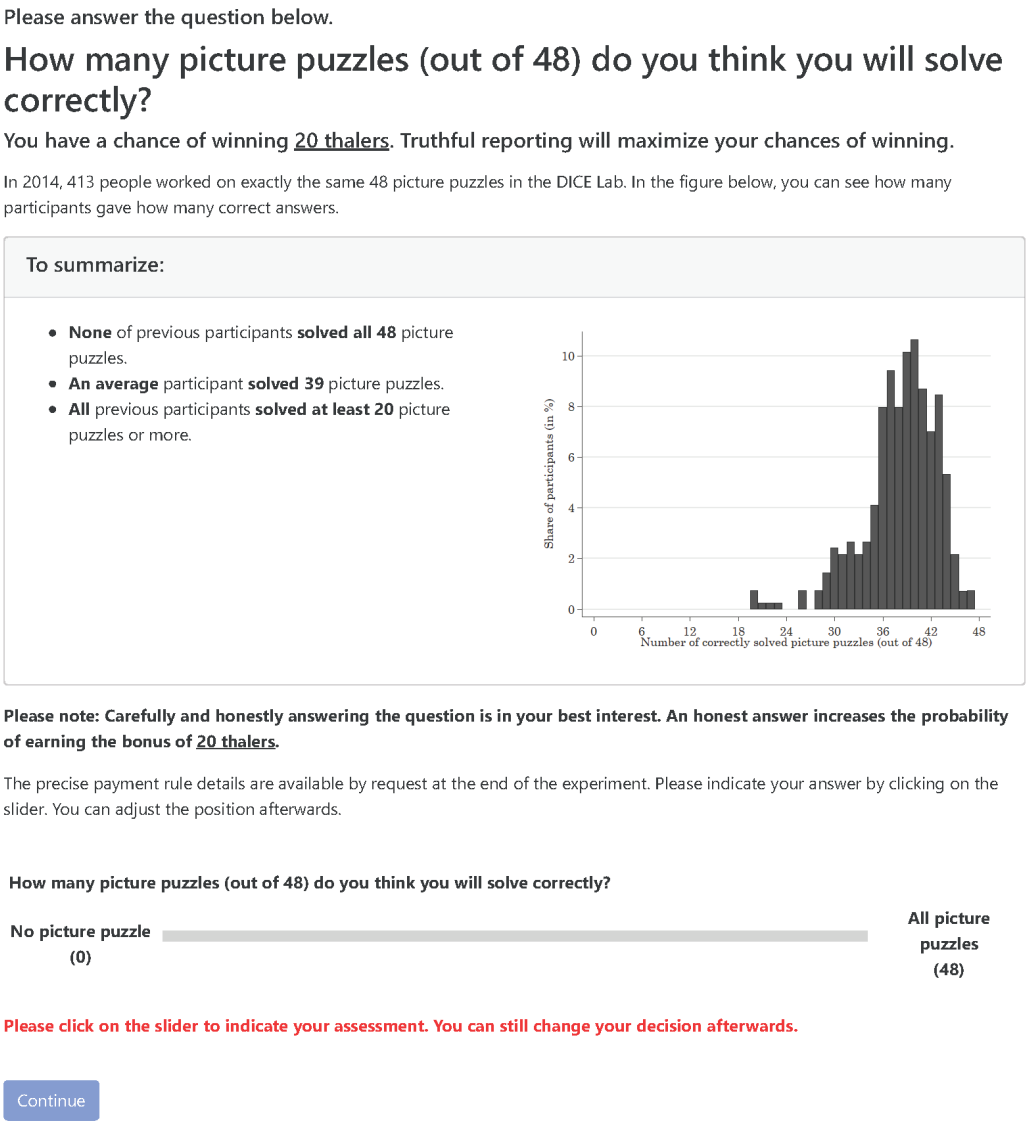


Figure 2.A.1: Belief elicitation screen

Please answer the question below.

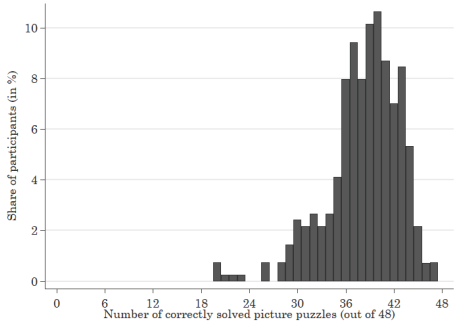
How many picture puzzles (out of 48) do you think you will solve correctly?

You have a chance of winning 20 thalers. Truthful reporting will maximize your chances of winning.

In 2014, 413 people worked on exactly the same 48 picture puzzles in the DICE Lab. In the figure below, you can see how many participants gave how many correct answers.

To summarize:

- **None** of previous participants **solved all 48** picture puzzles.
- **An average** participant **solved 39** picture puzzles.
- **All** previous participants **solved at least 20** picture puzzles or more.



Please note: Carefully and honestly answering the question is in your best interest. An honest answer increases the probability of earning the bonus of 20 thalers.

The precise payment rule details are available by request at the end of the experiment. Please indicate your answer by clicking on the slider. You can adjust the position afterwards.

How many picture puzzles (out of 48) do you think you will solve correctly?

No picture puzzle (0) — **All picture puzzles (48)**

I think, I will solve 38 out of 48 picture puzzles correctly. This means that I think I will perform better than 38.01% of previous participants.

Continue

Figure 2.A.2: Belief elicitation screen (answered)

2.C.4 Belief elicitations: German (original)

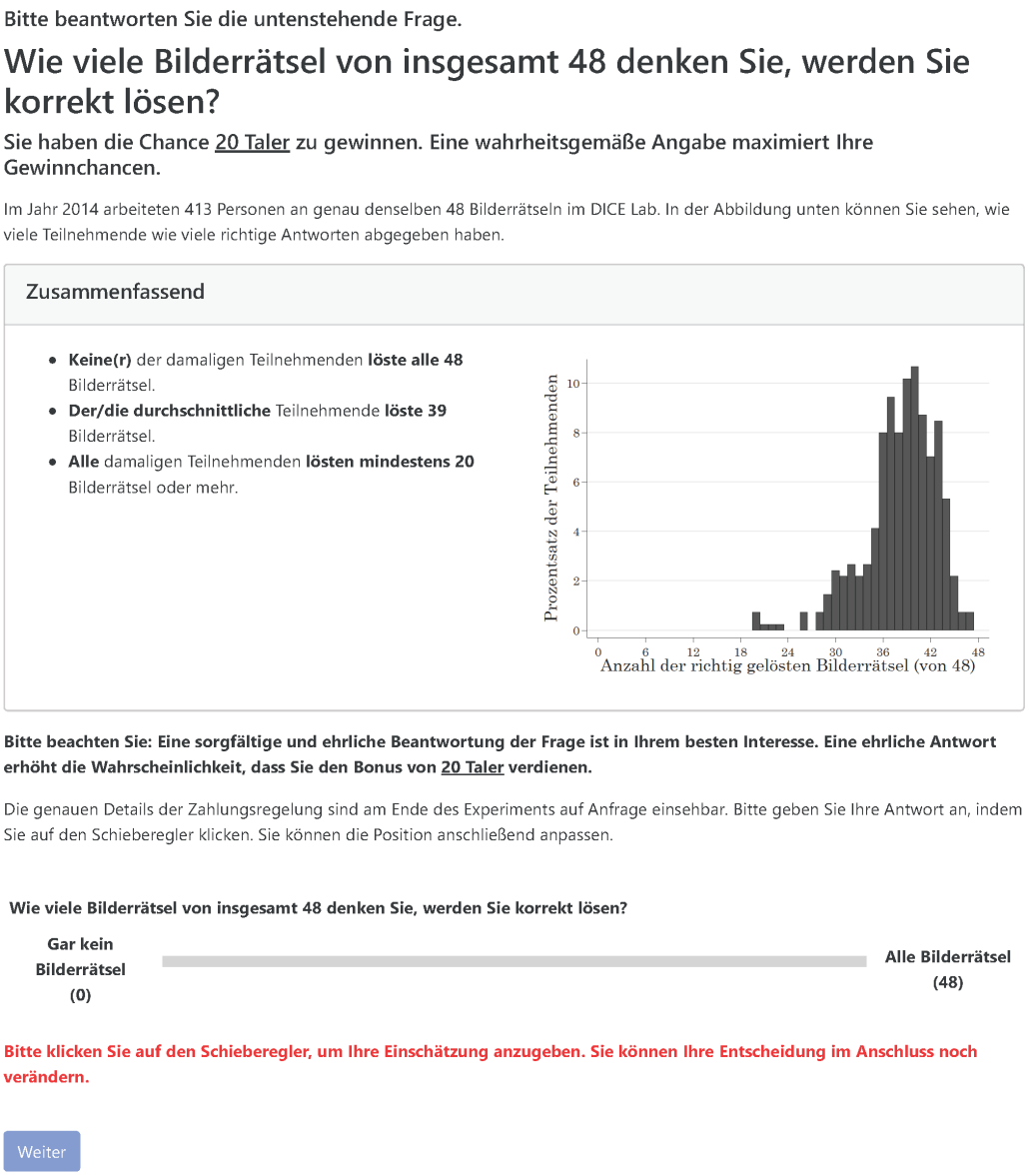


Figure 2.A.3: Belief elicitation screen

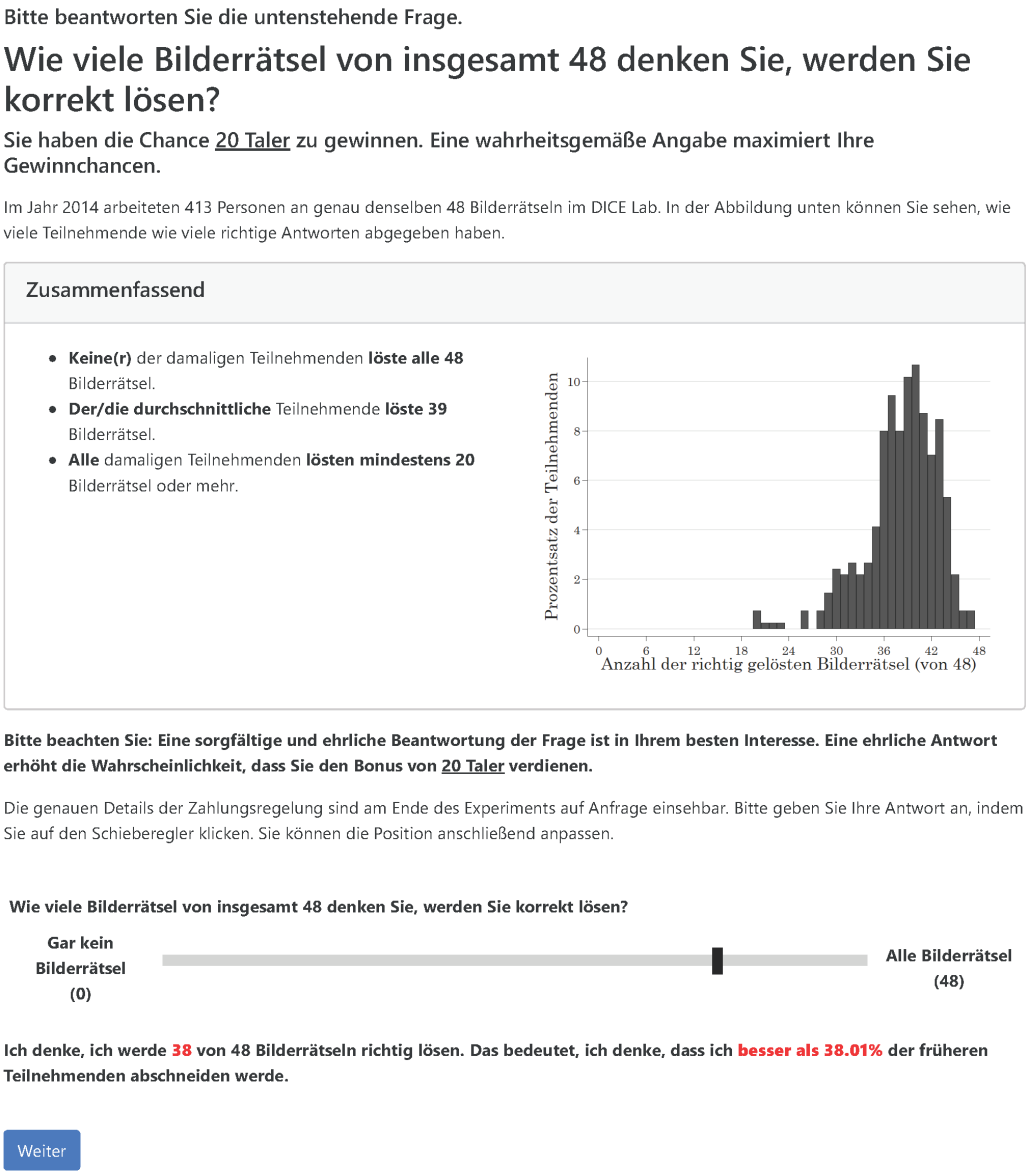


Figure 2.A.4: Belief elicitation screen (answered)

2.C.5 Willingness-to-pay for feedback: English

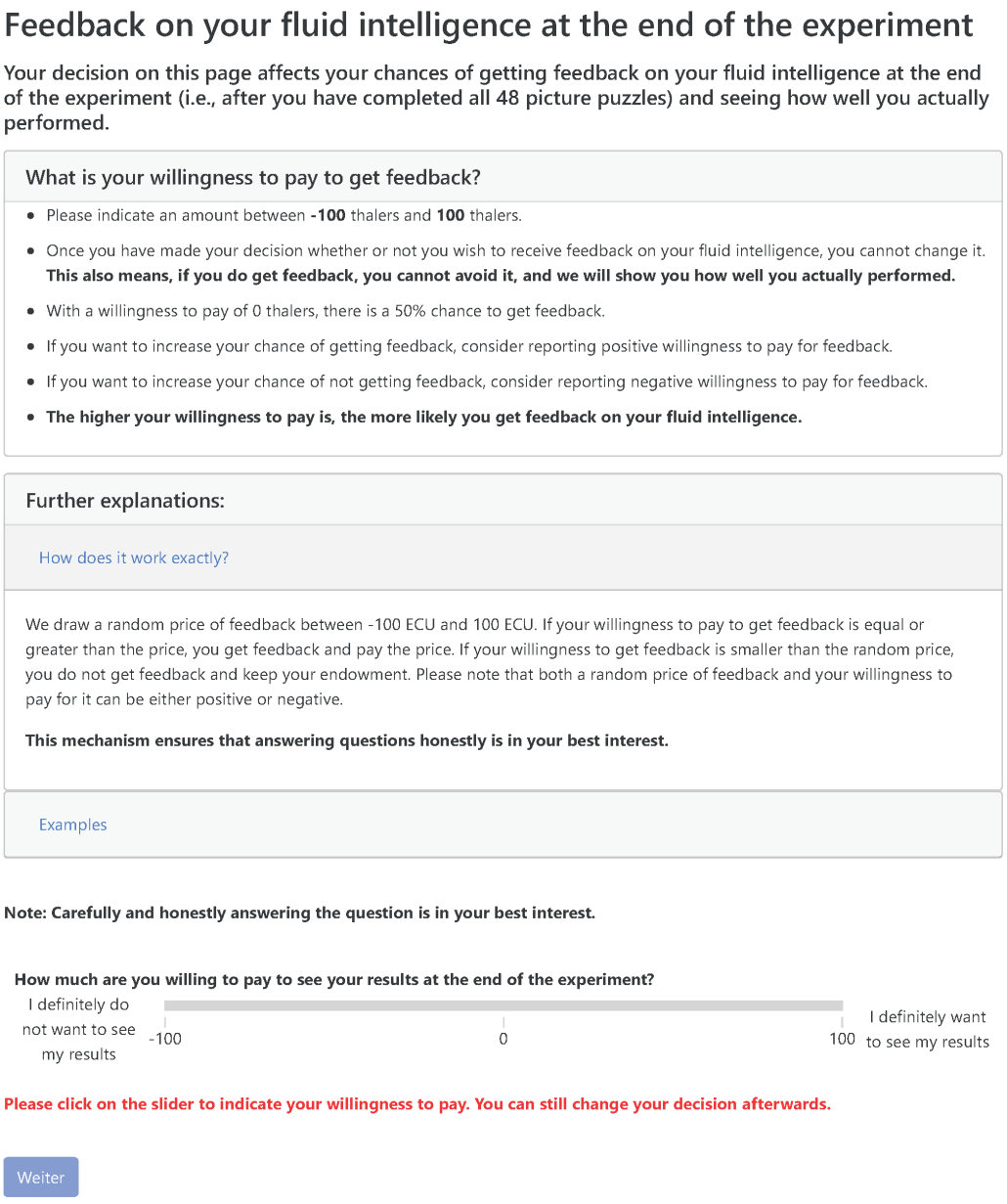


Figure 2.A.5: Willingness-to-pay for feedback screen

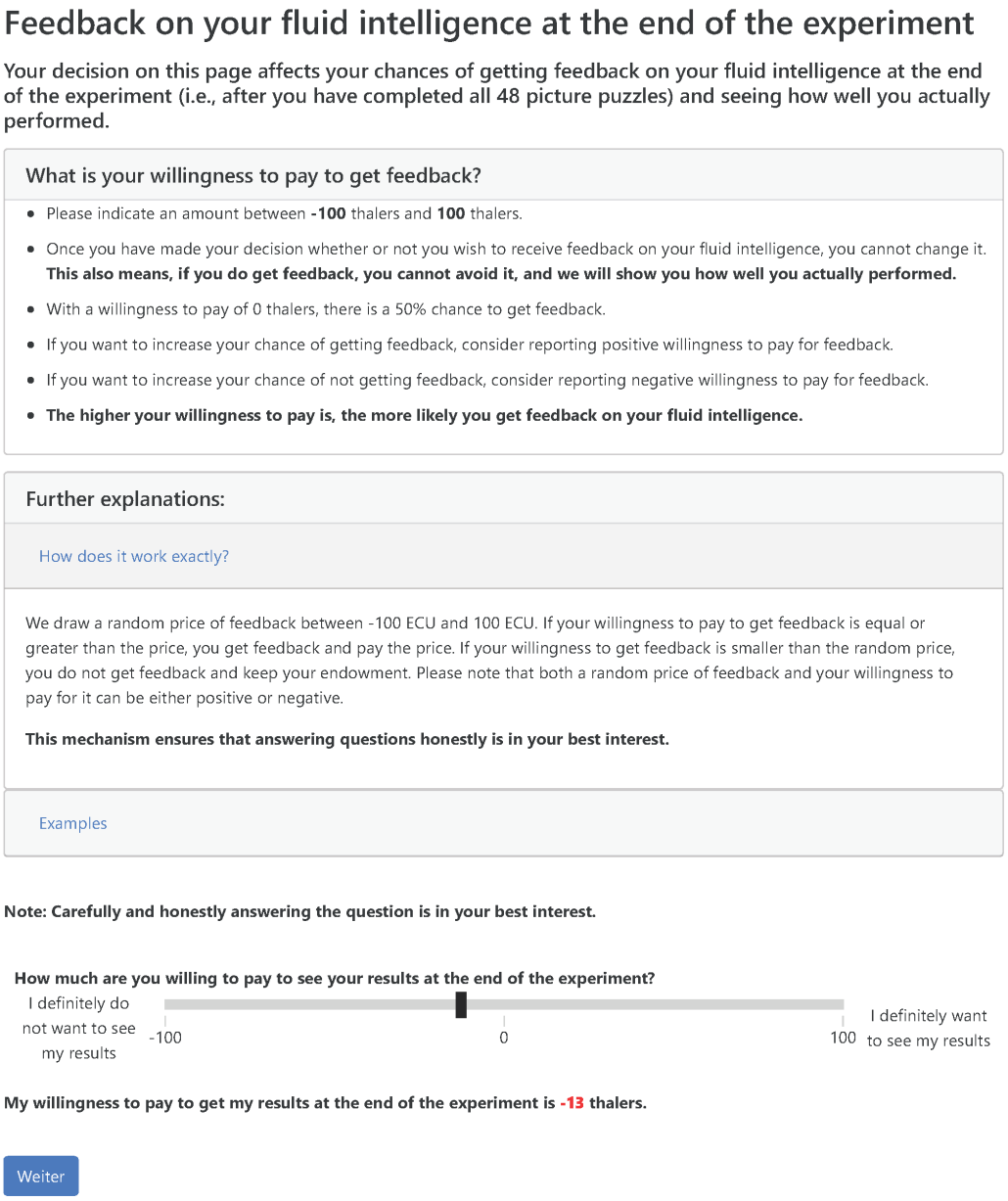
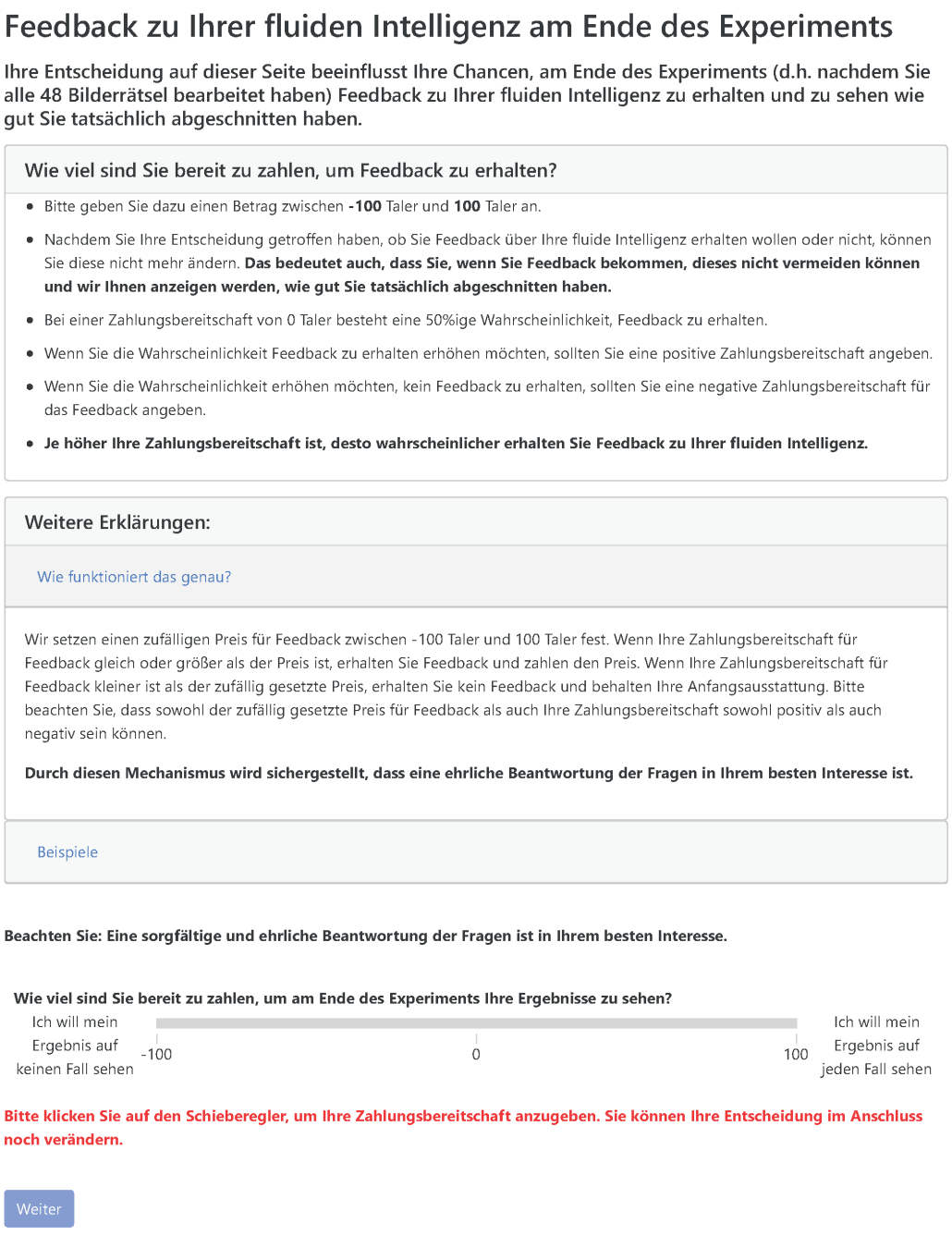


Figure 2.A.6: Willingness-to-pay for feedback screen (answered)

2.C.6 Willingness-to-pay for feedback: German (original)



Feedback zu Ihrer fluiden Intelligenz am Ende des Experiments

Ihre Entscheidung auf dieser Seite beeinflusst Ihre Chancen, am Ende des Experiments (d.h. nachdem Sie alle 48 Bilderrätsel bearbeitet haben) Feedback zu Ihrer fluiden Intelligenz zu erhalten und zu sehen wie gut Sie tatsächlich abgeschnitten haben.

Wie viel sind Sie bereit zu zahlen, um Feedback zu erhalten?

- Bitte geben Sie dazu einen Betrag zwischen **-100** Taler und **100** Taler an.
- Nachdem Sie Ihre Entscheidung getroffen haben, ob Sie Feedback über Ihre fluide Intelligenz erhalten wollen oder nicht, können Sie diese nicht mehr ändern. **Das bedeutet auch, dass Sie, wenn Sie Feedback bekommen, dieses nicht vermeiden können und wir Ihnen anzeigen werden, wie gut Sie tatsächlich abgeschnitten haben.**
- Bei einer Zahlungsbereitschaft von 0 Taler besteht eine 50%ige Wahrscheinlichkeit, Feedback zu erhalten.
- Wenn Sie die Wahrscheinlichkeit Feedback zu erhalten erhöhen möchten, sollten Sie eine positive Zahlungsbereitschaft angeben.
- Wenn Sie die Wahrscheinlichkeit erhöhen möchten, kein Feedback zu erhalten, sollten Sie eine negative Zahlungsbereitschaft für das Feedback angeben.
- **Je höher Ihre Zahlungsbereitschaft ist, desto wahrscheinlicher erhalten Sie Feedback zu Ihrer fluiden Intelligenz.**

Weitere Erklärungen:

Wie funktioniert das genau?

Wir setzen einen zufälligen Preis für Feedback zwischen -100 Taler und 100 Taler fest. Wenn Ihre Zahlungsbereitschaft für Feedback gleich oder größer als der Preis ist, erhalten Sie Feedback und zahlen den Preis. Wenn Ihre Zahlungsbereitschaft für Feedback kleiner ist als der zufällig gesetzte Preis, erhalten Sie kein Feedback und behalten Ihre Anfangsausstattung. Bitte beachten Sie, dass sowohl der zufällig gesetzte Preis für Feedback als auch Ihre Zahlungsbereitschaft sowohl positiv als auch negativ sein können.

Durch diesen Mechanismus wird sichergestellt, dass eine ehrliche Beantwortung der Fragen in Ihrem besten Interesse ist.

Beispiele

Beachten Sie: Eine sorgfältige und ehrliche Beantwortung der Fragen ist in Ihrem besten Interesse.

Wie viel sind Sie bereit zu zahlen, um am Ende des Experiments Ihre Ergebnisse zu sehen?

Ich will mein Ergebnis auf keinen Fall sehen | -100 | 0 | 100 | Ich will mein Ergebnis auf jeden Fall sehen

Bitte klicken Sie auf den Schieberegler, um Ihre Zahlungsbereitschaft anzugeben. Sie können Ihre Entscheidung im Anschluss noch verändern.

Weiter

Figure 2.A.7: Willingness-to-pay for feedback screen

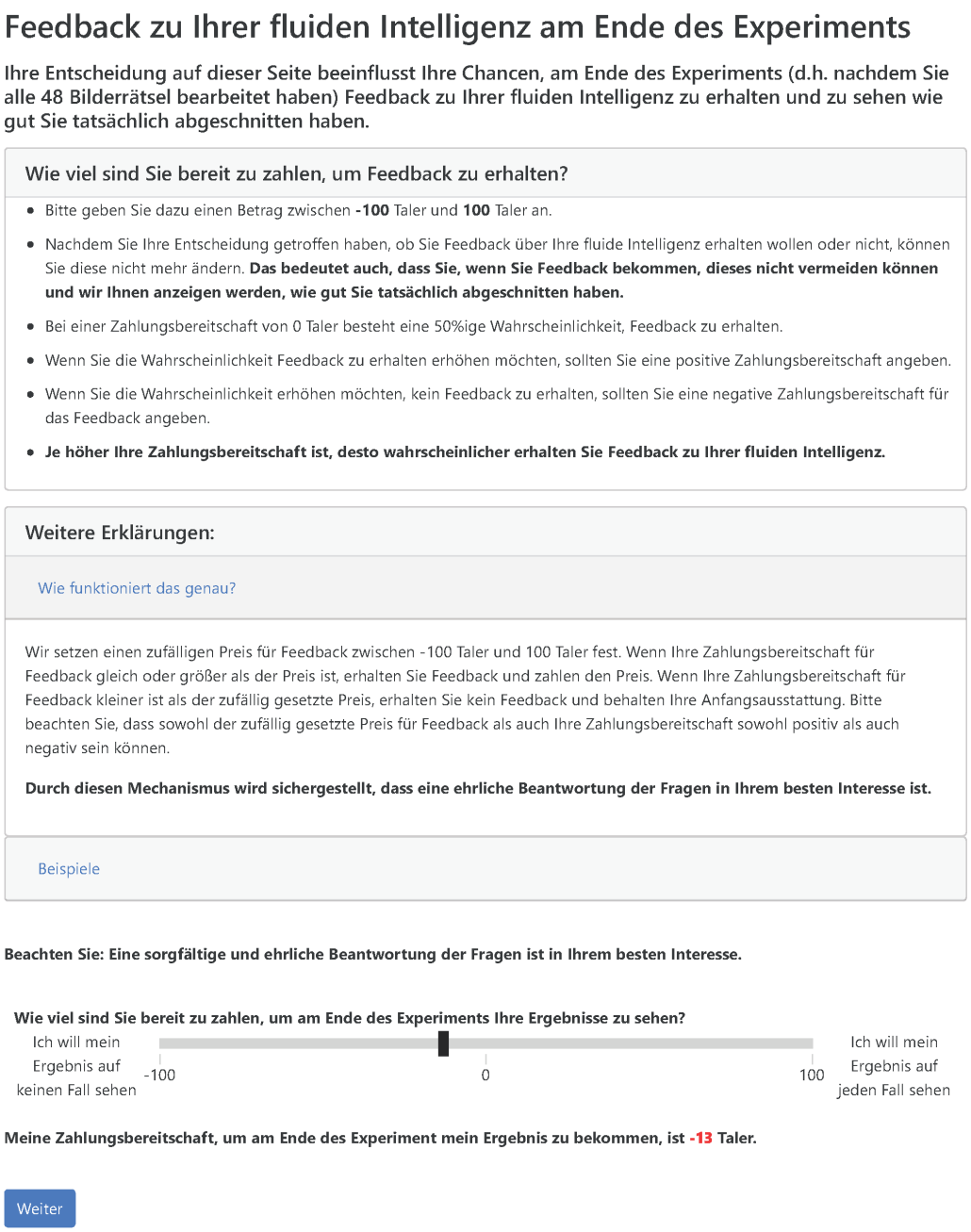


Figure 2.A.8: Willingness-to-pay for feedback screen (answered)

2.C.7 Questionnaire: English

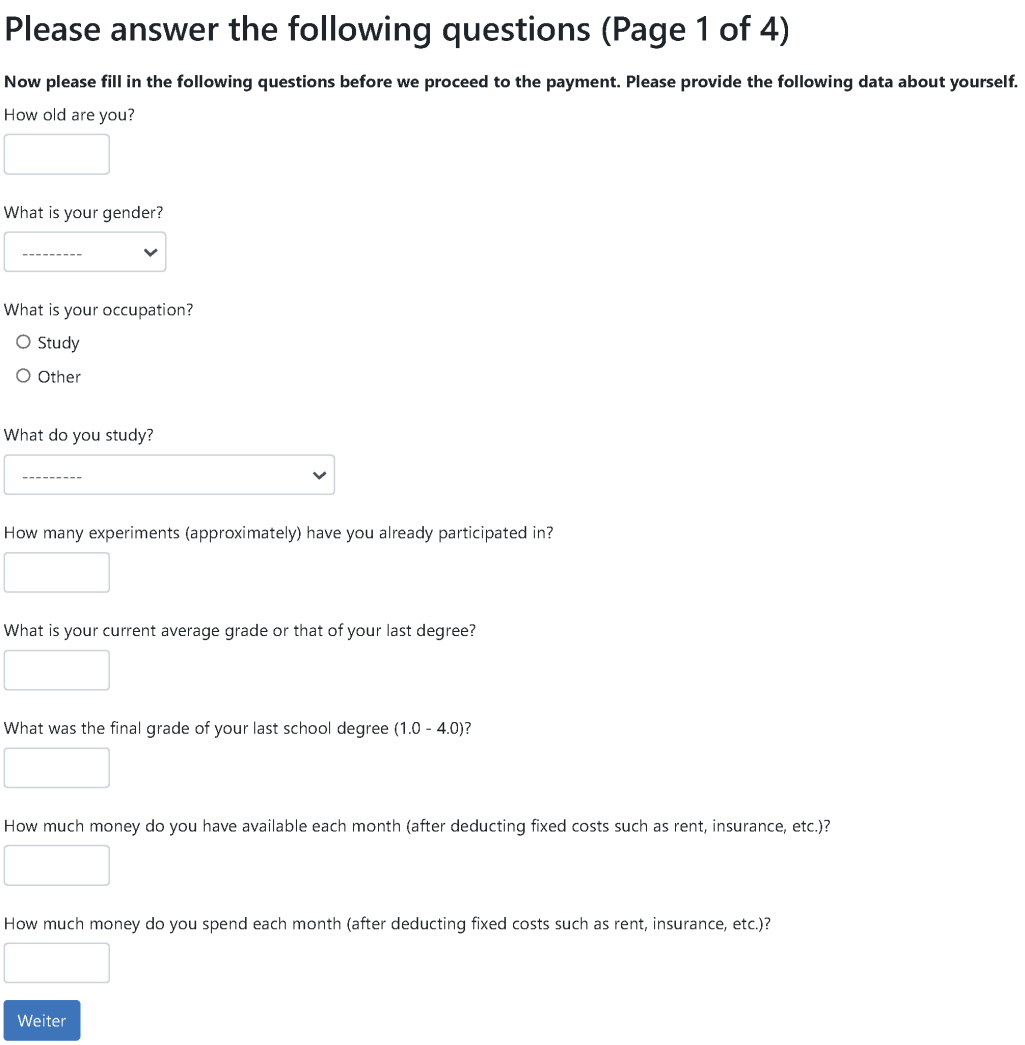


Figure 2.A.9: Questionnaire: Page 1 of 4

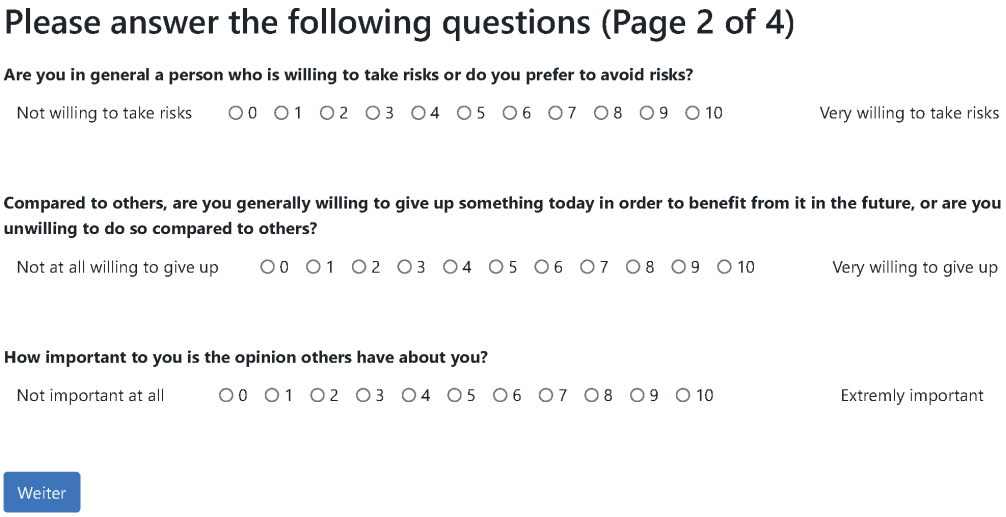


Figure 2.A.10: Questionnaire: Page 2 of 4

Please answer the following questions (Page 3 of 4)

How strongly do you agree with the following statements? Please answer on a scale of 0 to 5, where 0 means you do not agree with the statement at all and 5 means you agree very strongly.

This is about the characteristic "fairness"

1. I would feel good if I was a person who had this quality.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

2. Being someone who has this quality is an important part of who I am.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3. A big part of my emotional well-being is tied to having this quality.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4. I would be ashamed to be a person who has this characteristic.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

5. Having this characteristic is not really important to me.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

6. Having this quality is an important part of my self-image.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

This is about the characteristic "generosity"

1. I would feel good if I was a person who had this quality.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

2. Being someone who has this quality is an important part of who I am.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3. A big part of my emotional well-being is tied to having this quality.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4. I would be ashamed to be a person who has this characteristic.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

5. Having this characteristic is not really important to me.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

6. Having this quality is an important part of my self-image.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

This is about the characteristic "kindness"

1. I would feel good if I was a person who had this quality.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Figure 2.A.11: Questionnaire: Page 3 of 4

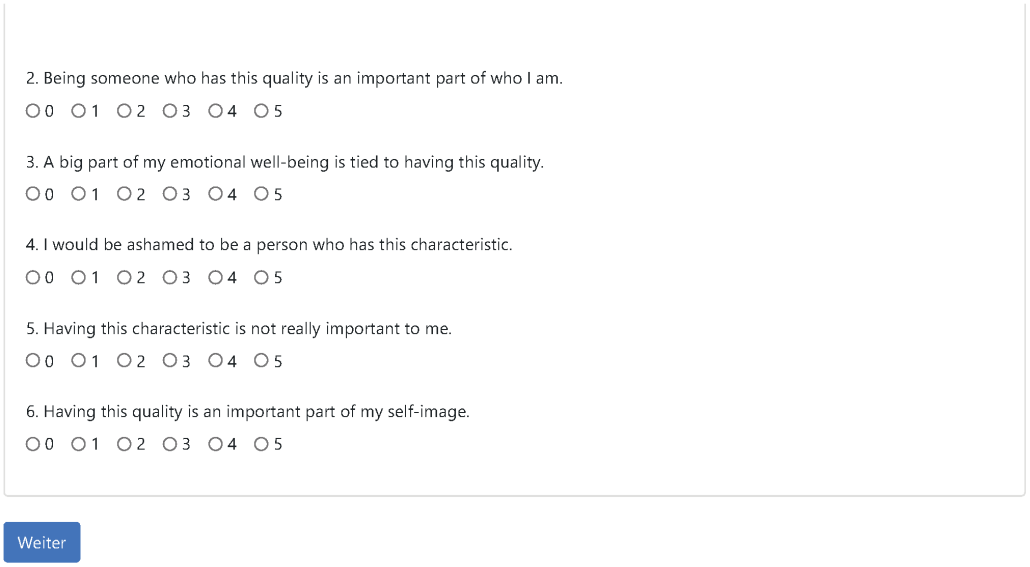


Figure 2.A.12: Questionnaire: Page 3 of 4

Lottery decisions (Page 4 of 4)

Please answer a few more questions about lotteries where you can earn or lose money once again if you decide to accept them.

Below are listed **6 different lotteries**:

- For each of the 6 lotteries, you can choose to accept or reject the lottery.
- If you reject a lottery, your payment will remain unchanged. If you accept a lottery, you will realize an additional profit or an additional loss.
- At the end of the experiment, one of the 6 lotteries is randomly selected.
- So, you should make each lottery decision as if it was your only decision. The selected lottery is then drawn to determine whether the additional profit or loss will be realized.

Lottery	Your decision			
With a probability of 50 % you **lose** 4 thalers. With a probability of 50 % you **win** 12 thalers.	**Accept**	○	○	**Reject**
With a probability of 50 % you **lose** 6 thalers. With a probability of 50 % you **win** 12 thalers.	**Accept**	○	○	**Accept**
With a probability of 50 % you **lose** 8 thalers. With a probability of 50 % you **win** 12 thalers.	**Accpet**	○	○	**Reject**
With a probability of 50 % you **lose** 10 thalers. With a probability of 50 % you **win** 12 thalers.	**Accept**	○	○	**Reject**
With a probability of 50 % you **lose** 12 thalers. With a probability of 50 % you **win** 12 thalers.	**Accept**	○	○	**Reject**
With a probability of 50 % you **lose** 14 thalers. With a probability of 50 % you **win** 12 thalers.	**Accept**	○	○	**Reject**

Weiter

Figure 2.A.13: Questionnaire: Page 4 of 4

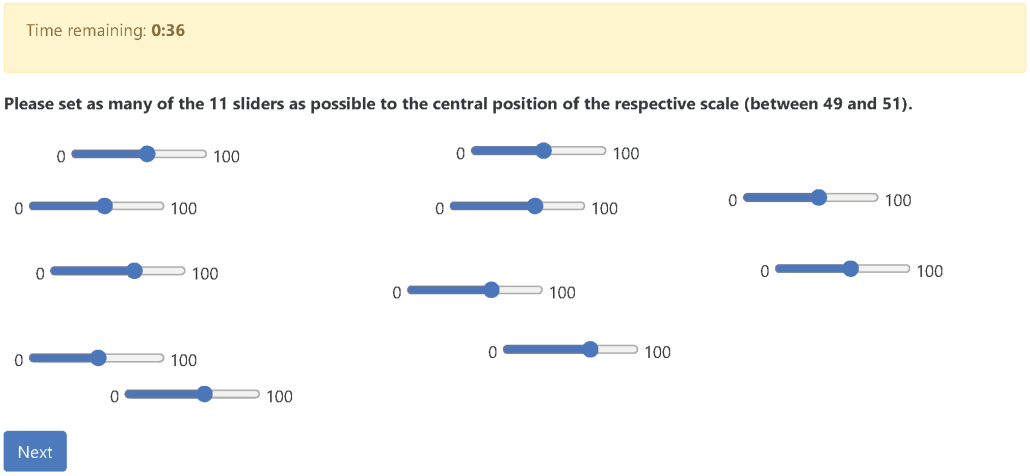


Figure 2.A.14: Overconfidence elicitation: Slider task

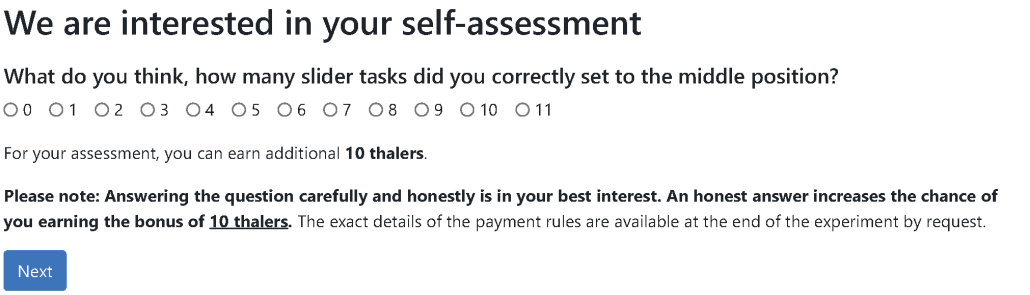


Figure 2.A.15: Overconfidence elicitation: Self-assessment

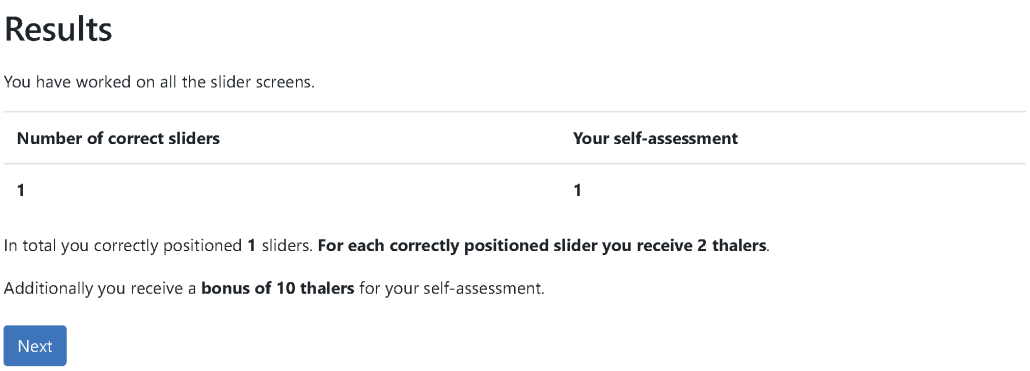
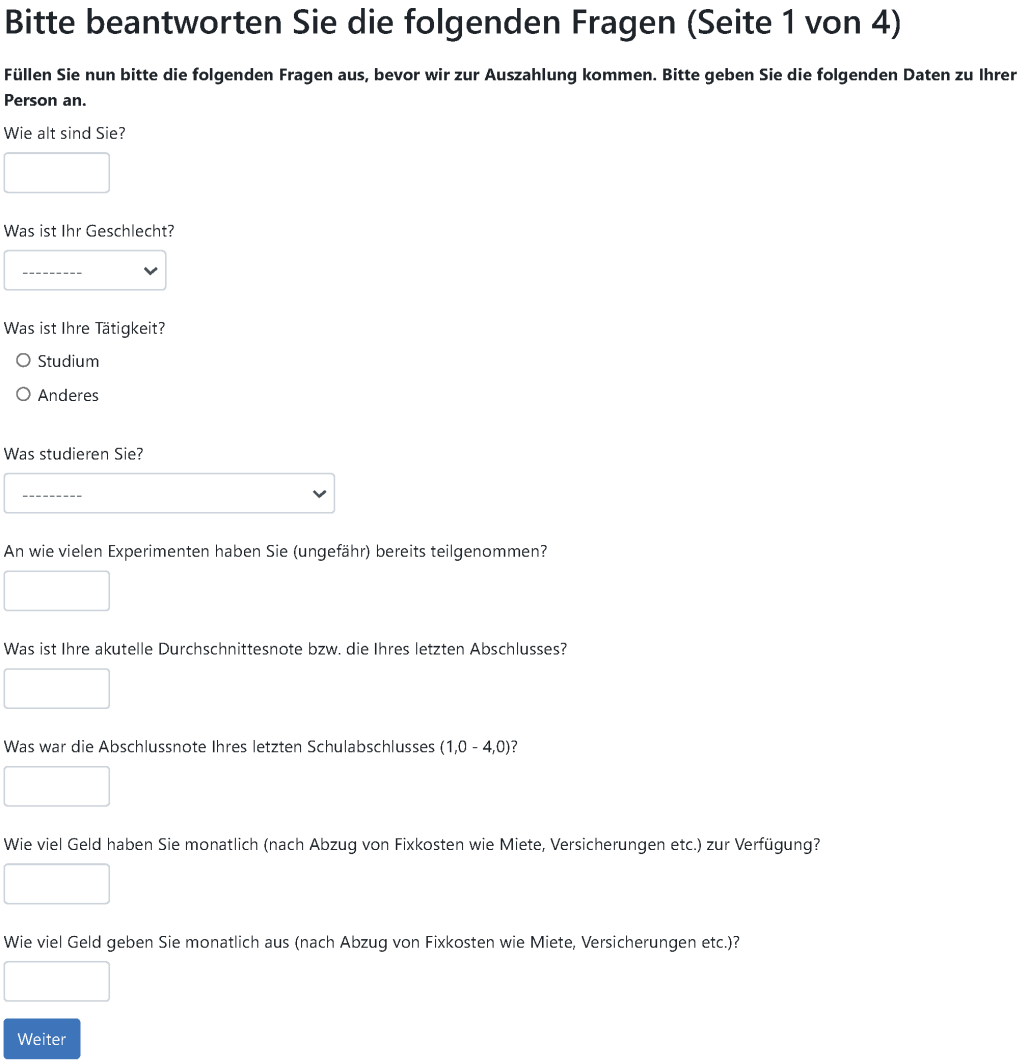


Figure 2.A.16: Overconfidence elicitation: Feedback

2.C.8 Questionnaire: German (original)



Bitte beantworten Sie die folgenden Fragen (Seite 1 von 4)

Füllen Sie nun bitte die folgenden Fragen aus, bevor wir zur Auszahlung kommen. Bitte geben Sie die folgenden Daten zu Ihrer Person an.

Wie alt sind Sie?

Was ist Ihr Geschlecht?

Was ist Ihre Tätigkeit?

- Studium
- Anderes

Was studieren Sie?

An wie vielen Experimenten haben Sie (ungefähr) bereits teilgenommen?

Was ist Ihre akutelle Durchschnittesnote bzw. die Ihres letzten Abschlusses?

Was war die Abschlussnote Ihres letzten Schulabschlusses (1,0 - 4,0)?

Wie viel Geld haben Sie monatlich (nach Abzug von Fixkosten wie Miete, Versicherungen etc.) zur Verfügung?

Wie viel Geld geben Sie monatlich aus (nach Abzug von Fixkosten wie Miete, Versicherungen etc.)?

Weiter

Figure 2.A.17: Questionnaire: Page 1 of 4

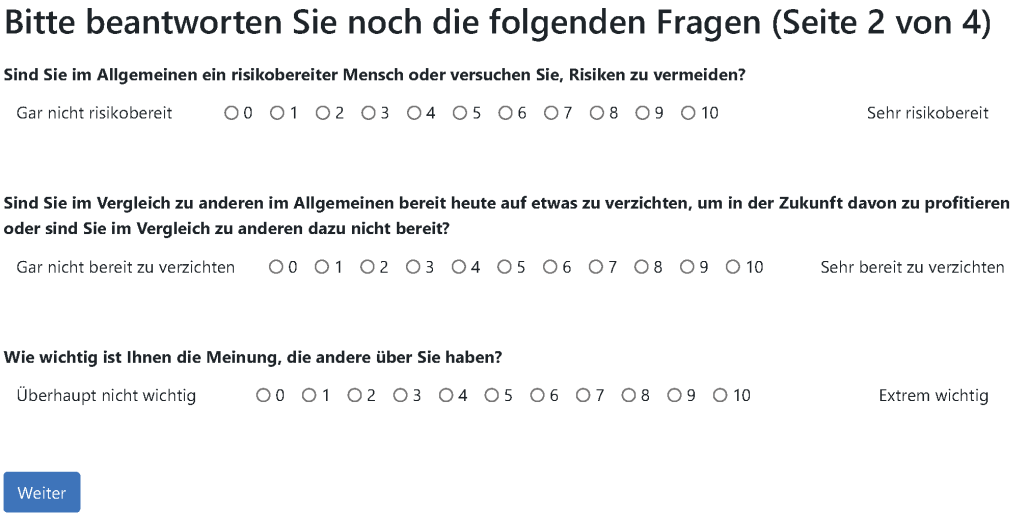


Figure 2.A.18: Questionnaire: Page 2 of 4

Bitte beantworten Sie noch die folgenden Fragen (Seite 3 von 4)

Wie sehr stimmen Sie den folgenden Aussagen zu? Bitte antworten Sie auf einer Skala von 0 bis 5. Dabei bedeutet 0, dass Sie der Aussage gar nicht zustimmen und 5, dass Sie sehr stark zustimmen.

Es geht um die Eigenschaft "Fairness"

1. Ich würde mich gut fühlen, wenn ich eine Person wäre, die diese Eigenschaft hat.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

2. Jemand zu sein, der diese Eigenschaft hat, ist ein wichtiger Teil von dem, was ich bin.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3. Ein großer Teil meines emotionalen Wohlbefindens ist damit verbunden, diese Eigenschaft zu haben.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4. Ich würde mich schämen, eine Person zu sein, die diese Eigenschaft hat.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

5. Diese Eigenschaft zu haben, ist für mich nicht wirklich wichtig.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

6. Diese Eigenschaft zu haben, ist ein wichtiger Teil meines Selbstverständnisses.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Es geht um die Eigenschaft "Großzügigkeit"

1. Ich würde mich gut fühlen, wenn ich eine Person wäre, die diese Eigenschaft hat.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

2. Jemand zu sein, der diese Eigenschaft hat, ist ein wichtiger Teil von dem, was ich bin.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3. Ein großer Teil meines emotionalen Wohlbefindens ist damit verbunden, diese Eigenschaft zu haben.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4. Ich würde mich schämen, eine Person zu sein, die diese Eigenschaft hat.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

5. Diese Eigenschaft zu haben, ist für mich nicht wirklich wichtig.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

6. Diese Eigenschaft zu haben, ist ein wichtiger Teil meines Selbstverständnisses.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Es geht um die Eigenschaft "Freundlichkeit"

1. Ich würde mich gut fühlen, wenn ich eine Person wäre, die diese Eigenschaft hat.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Figure 2.A.19: Questionnaire: Page 3 of 4

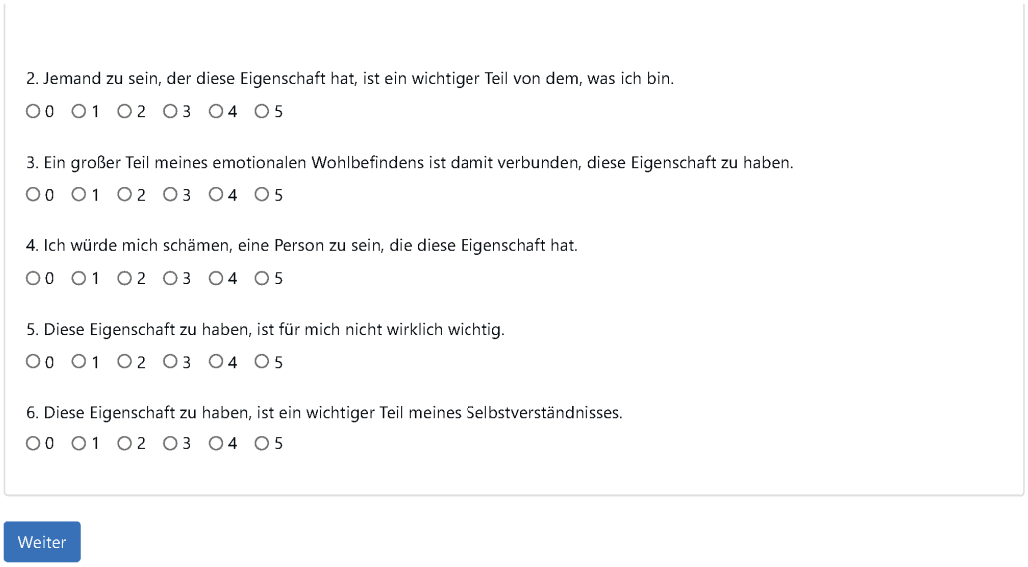


Figure 2.A.20: Questionnaire: Page 3 of 4

Lotterieentscheidungen (Seite 4 von 4)

Bitte beantworten Sie im Folgenden noch ein paar Fragen zu Lotterien, bei denen Sie noch einmal Geld verdienen oder auch verlieren können, falls Sie sich entscheiden, die Lotterien zu akzeptieren.

Unten sind **6 verschiedene Lotterien** aufgelistet:

- Sie können für jede der 6 Lotterien wählen, ob Sie die Lotterie akzeptieren oder ablehnen möchten.
- Falls Sie eine Lotterie ablehnen, bleibt Ihre Auszahlung unverändert. Falls Sie eine Lotterie akzeptieren, werden Sie einen zusätzlichen Gewinn oder einen zusätzlichen Verlust realisieren.
- Am Ende des Experiments wird zufällig eine der 6 Lotterien ausgewählt.
- Sie sollten also jede Lotterieentscheidung so treffen, als wäre es Ihre einzige Entscheidung. Die ausgewählte Lotterie wird anschließend ausgelost, um festzustellen, ob sich der zusätzliche Gewinn oder Verlust realisiert.

Lotterie	Ihre Entscheidung			
Mit einer Wahrscheinlichkeit von 50 % **verlieren** Sie: 4 Taler. Mit einer Wahrscheinlichkeit von 50 % **gewinnen** Sie 12 Taler.	**Akzeptieren**	○	○	**Ablehnen**
Mit einer Wahrscheinlichkeit von 50 % **verlieren** Sie: 6 Taler. Mit einer Wahrscheinlichkeit von 50 % **gewinnen** Sie 12 Taler.	**Akzeptieren**	○	○	**Ablehnen**
Mit einer Wahrscheinlichkeit von 50 % **verlieren** Sie: 8 Taler. Mit einer Wahrscheinlichkeit von 50 % **gewinnen** Sie 12 Taler.	**Akzeptieren**	○	○	**Ablehnen**
Mit einer Wahrscheinlichkeit von 50 % **verlieren** Sie: 10 Taler. Mit einer Wahrscheinlichkeit von 50 % **gewinnen** Sie 12 Taler.	**Akzeptieren**	○	○	**Ablehnen**
Mit einer Wahrscheinlichkeit von 50 % **verlieren** Sie: 12 Taler. Mit einer Wahrscheinlichkeit von 50 % **gewinnen** Sie 12 Taler.	**Akzeptieren**	○	○	**Ablehnen**
Mit einer Wahrscheinlichkeit von 50 % **verlieren** Sie: 14 Taler. Mit einer Wahrscheinlichkeit von 50 % **gewinnen** Sie 12 Taler.	**Akzeptieren**	○	○	**Ablehnen**

Weiter

Figure 2.A.21: Questionnaire: Page 4 of 4

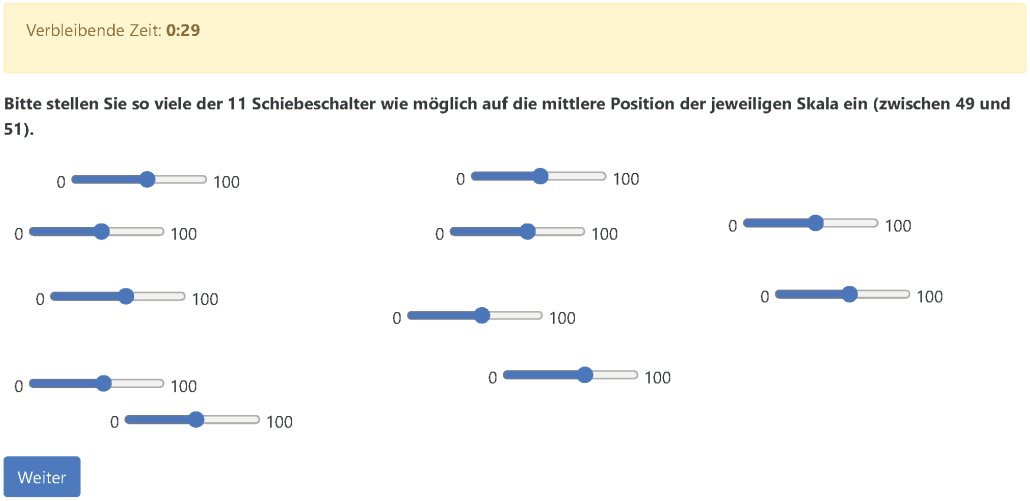


Figure 2.A.22: Overconfidence elicitation: Slider task

Figure 2.A.23: Overconfidence elicitation: Self-assessment

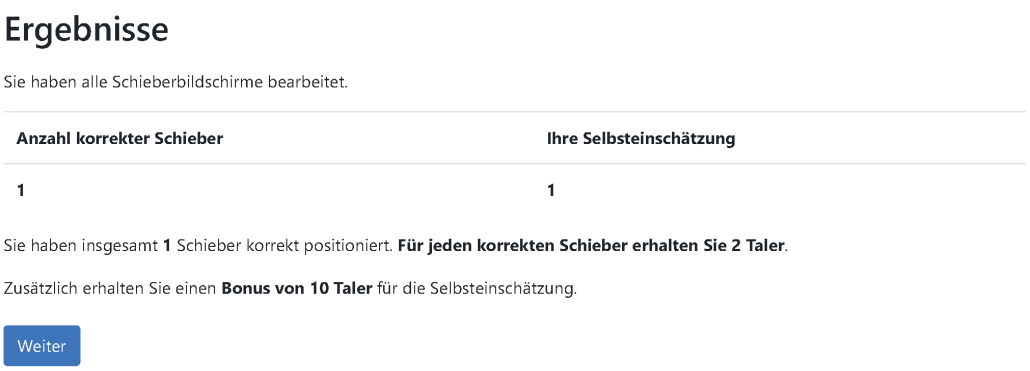


Figure 2.A.24: Overconfidence elicitation: Feedback

Chapter 3

Limited Attention in Credence Goods Markets

Co-authored with Alexander Rasch, Chi Trieu and Christian Waibel

3.1 Introduction

The distinct feature of credence goods markets is informational asymmetry. Sellers are experts and have an informational advantage over customers. More precisely, sellers know which types of services customers need, whereas customers do not Darby and Karni (1973). Customers have to trust experts that the experts provide the correct service. Experts may exploit their informational advantage by providing more or more expensive services than necessary. Prime examples are markets for repair services and healthcare.

One of the key theoretical predictions is that (liable) experts should have no incentives to provide an inappropriate amount of service whenever customers can verify the type of service (Dulleck and Kerschbamer, 2006). In equilibrium, experts post prices with equal markups for the different types of services. By posting equal-markup prices, experts credibly signal to perform the type of service that the customer needs. Because customers anticipate that experts provide necessary services under equal markups, customers' willingness to pay for a service is maximal. A monopolistic expert sets these equal markup prices in a way to fully extract customer rent. In a competitive credence goods market, prices cover experts' marginal costs of providing a service.

In real markets, however, these predictions appear to contradict observations. The FBI estimates that up to 10% of the 3.3 trillion US dollars of yearly health expenditures in the United States are due to fraud (Federal Bureau of Investigation, 2011).[1] Gottschalk et al. (2020) show that 28% of dentists' treatment recommendations involve overtreatment recommendations. In car repair services, Taylor (1995), Schneider (2012), and Rasch and Waibel (2018) report fraudulent behavior by garages. Kerschbamer, Neururer, et al. (2016) document fraud in computer repair services. Balafoutas, Beck, et al. (2013) and Balafoutas, Kerschbamer, et al. (2015) identify fraud in the market for taxi rides.

So far, the literature has offered different reasons to explain such discrepancies between the theoretical results and real-life observations. Explanations include expert heterogeneity (see, for example, Dulleck and Kerschbamer, 2009, Frankel and Schwarz, 2014, and Hilger, 2016), the coexistence of selfish and conscientious experts (see, for instance, Liu,

[1] For an overview of the phenomenon of so-called physician-induced demand (PID), see McGuire (2000)

2011 and Fong et al., 2014), and a lack or ban of price discrimination (see, for example, Dulleck and Kerschbamer, 2006). In this paper, we offer an alternative explanation: customers' limited attention. The idea is based on insights from psychological research: Due to cognitive constraints and large amounts of information, people often fail to account for all relevant details when making decisions.[2] Our approach assumes that customers do not take into account all relevant information that determines an expert's payoff.

We employ a simple model to investigate the existence and impact of limited attention in a credence goods market. In the model, customers suffer from either a minor or a major problem. The major service solves both problems but is more costly for a monopolistic expert than the minor service. The minor (and less costly) service can only solve the minor problem. Service costs are common knowledge among experts and customers. By posting an equal-markup price vector, the expert could credibly signal that she has no incentive to over- or undertreat. We assume that customers can verify the treatment applied (that is, we rule out overcharging) but do not fully account for treatment costs. It crucially affects their evaluation of expert profits and, hence, the expert's incentive to defraud them. We predict that customers' limited attention increases the insufficient service provision and raises the markup difference between the major and the minor services. Moreover, customers are more willing to pay for an offer that triggers insufficient service provision if their attention is limited.

We test the predictions in a laboratory experiment. We vary whether a customer observes – in addition to the expert's price vector – the expert's profit vector. A customer then decides whether or not she wants to interact given the posted prices. The expert observes which type of problem her customer has and decides whether to provide either the minor or the major service. The expert charges for the provided service. In the treatment Attention, customers observe the prices, and experts' costs are made salient before deciding on interaction, whereas in the NoAttention treatment, customers only observe prices. Experts and customers are randomly rematched in our lab experiment and hence do not suffer from reputational concerns. We find that experts' price vectors

[2]See Lim and Teoh (2010) for an overview in the context of finance and accounting. Heidhues and Kőszegi (2018) discuss limited attention in the context of applications in industrial organization.

are significantly closer to the equal markup when costs are made salient than when they are not. Customers' interaction probability decreases by around 20 percentage points over time and does not significantly vary across treatments. Controlling for subjects' covariates, experts undertreat customers significantly more often under NoAttention than under Attention.

Attention decreases total welfare calculated as accumulated profits. Due to the rapid decrease in interaction over time, the customer surplus is smaller under Attention than under NoAttention. Experts benefit from limited attention because they can extract the additional surplus generated by more sufficient treatments. When we define welfare as accumulated profit minus the outside option, and random differences in the customers' type of problem (minor or major) are considered, welfare improves under Attention.

In many credence goods markets, there is a call to make experts' financial interests more transparent. Due to limited attention, customers do not fully account for experts' financial incentives. An example of a sector in which more transparency is demanded is the market for health services. In Germany, for instance, many health services are paid for by the patients' insurance companies. The payments are organized bilaterally between the insurance company and the physician without any patient involvement. To increase transparency for such services, patients have had the right to ask for a patient receipt since 2012. This receipt must report the treatments performed and the (expected) costs.[3]

Providers themselves can also advocate for increased transparency. For example, for their car repair services, carmaker Opel introduced a new app-based information service called "MyDigitalService". When car owners have their cars inspected or repaired, they can now more easily follow the different steps in the process and are provided with information regarding additional costs when unanticipated services become necessary.[4]

Our study is directly related to the literature on credence goods. Closest to our paper is the article by Dulleck, Kerschbamer, and Sutter (2011), which employs a large-scale laboratory experiment challenging the seminal model by Dulleck and Kerschbamer

[3]See, for example, `https://www.bundesgesundheitsministerium.de/themen/praevention/patientenrechte/patientenquittung.html`.

[4]See, for example, `https://www.auto-motor-und-sport.de/tech-zukunft/werkstatt/opel-mydigitalservice-transparenz-inspektion-reparatur/`.

(2006). In particular, the authors study the impact of institutions, such as verifiability or liability, on outcomes in credence goods markets and show that liability is an effective tool for improving outcomes in credence goods markets. However, the authors find no evidence that verifiability fosters market results.

There are multiple explanations for why verifiability seems to be less effective for improving market outcomes. Previous work has explained the differences between the prediction of no overtreatment if services are verifiable and the observation in real markets is primarily based on experts' characteristics. Emons (1997, 2001) argues that experts' utilization of capacities drives overtreatment. If demand is low, experts may have the incentive to provide excess services to fill capacities. Gottschalk et al. (2020) provides evidence from a field experiment. Dentists with a low utilization are correlated with a higher probability of receiving an overtreatment recommendation. Hilger (2016) develops a model that accounts for experts' heterogeneity with respect to experts' costs of service provision. If costs are unobservable, experts cannot credibly signal equal markups. Hilger (2016) assumes experts are liable for their services. Hence, experts may have an incentive to overtreat.

To our knowledge, the only paper that is based on customers' characteristics is Kerschbamer, Sutter, et al. (2017). The authors suggest that customers' preferences may drive the deviations observed in Dulleck, Kerschbamer, and Sutter (2011). More precisely, the authors argue that heterogeneity in social preferences may explain the observed behavior. They show theoretically that equal-price equilibria are robust to pro-social but not anti-social preferences. Our study extends this strand of literature by adding the perspective of consumers' limited attention.

Our paper also contributes to the literature on the behavioral industrial organization that investigates market outcomes when consumers have behavioral biases.[5] The closest strand of literature to our setup are studies on add-on pricing, where consumers do not pay attention to the additional price of a two-part tariff (Armstrong and Vickers, 2012; Gabaix and Laibson, 2006; M. D. Grubb, 2015; Heidhues and Koszegi, 2017). Our study

[5]See, for example, M. Grubb (2015) and Heidhues and Kőszegi (2018) for an overview.

contributes by investigating limited attention with regard to a different factor, namely sellers' costs. In particular, customers are fully attentive to the prices of two treatments offered by the sellers, but not to the cost of each treatment. Costs do not directly show up in the customers' payoff function, yet they influence the treatment offered by sellers. The chosen treatment then determines whether consumers receive proper treatment, affecting their payoffs.

The remainder of the paper is as follows. Section 3.2 provides the theoretical framework for the credence goods market. Section 3.3 lays out our experimental design and shows our hypotheses. Section 3.4 displays and discusses our results before we conclude in Section 3.5.

3.2 Theoretical framework

3.2.1 Market

We model a market with verifiability and without liability following Dulleck and Kerschbamer (2006). Consider a market with an expert and a customer. A customer (she) has either a major or a minor problem. The customer knows that she has a problem but does not know whether it is major or minor. However, the customer knows that she has the major problem with an ex-ante probability h and the minor problem with an ex-ante probability $(1-h)$. These probabilities are common knowledge to both the expert and the customer.

The expert (he) can identify the problem at no cost. He can choose to provide either major or minor treatment. The cost of the major treatment is $\bar{c}$ and the cost of the minor treatment is $\underline{c}$, with $\underline{c} < \bar{c}$. The major treatment solves both problems, whereas the minor treatment only solves the minor problem. The expert posts take-it-or-leave-it prices.

The customer has a valuation of $v > 0$ when receiving sufficient treatment. The expert is *not liable* – that is, he can treat a customer who has a major problem with minor treatment. The prices for the major and the minor treatment are denoted as $\bar{p}$ and $\underline{p}$, respectively, with $\underline{p} < \bar{p}$. Due to the *verifiability* of treatment, the expert has to

charge $\bar{p}$ if he provides the major treatment and $\underline{p}$ if he provides the minor treatment (no overcharging). The customer does not know the necessary treatment but knows whether her problem has been solved. We refer to appropriate treatments whenever the customer has a major problem and receives a major treatment or when she has a minor problem and receives a minor treatment. Undertreatment occurs when the customer has a major problem but only receives a minor treatment. Finally, overtreatment means that a customer with a minor problem receives a major treatment.

The game is characterized as follows:

1. The expert posts a price menu $(\bar{p}, \underline{p})$ for the major and minor treatment, respectively.

2. The customer chooses whether to interact with the expert. We refer to this decision as "interaction" or "no interaction", respectively. The presentation of information differs across conditions:

   (a) NoAttention condition: The customer observes the price menu posted by the expert.

   (b) Attention condition: The customer observes the price menu posted by the expert and the expert's (potential) profit for each price.[6]

   If the customer chooses not to interact, the game ends. In that case, the expert and the customer both get the outside option u. If the customer chooses to interact, the game proceeds with stage 3.

3. Nature draws the type of problem that the customer has.[7]

4. The expert observes the problem type of the customer. The expert then provides either major or minor treatment and charges a price according to his treatment recommendation ($\bar{p}$ or $\underline{p}$).

5. The expert observes his payoff, and the customer observes her payoff.

[6]Note that even when a customer cannot directly observe the expert's profit, she can calculate the profit because the costs of both treatments are common knowledge.

[7]As Dulleck and Kerschbamer (2006) point out, it does not make a (game-theoretic) difference whether nature determines the severity of the problem after the customer has consulted an expert (but before the expert has performed the diagnosis) or at the very beginning.

If there is interaction, the expert's payoff (profit) is determined by the price p ($p \in \{\underline{p}, \bar{p}\}$) minus the cost c ($c \in \{\underline{c}, \bar{c}\}$) of the treatment applied, that is, $\pi_e = p - c$. If there is no interaction, the payoff amounts to u.

If the customer chooses to interact and the expert does not undertreat her, she derives her gross valuation of v. If she decides to interact and the expert undertreats her, she derives a valuation of zero. In either case, the customer must pay the price p for the treatment she receives. Hence, for each period, her payoff is either $\pi_c = v - p$ if she is not undertreated or $\pi_c = -p$ if she is undertreated. If the customer decides not to interact, she receives a payoff of u. The game and the payoffs are illustrated in Figure 3.1.

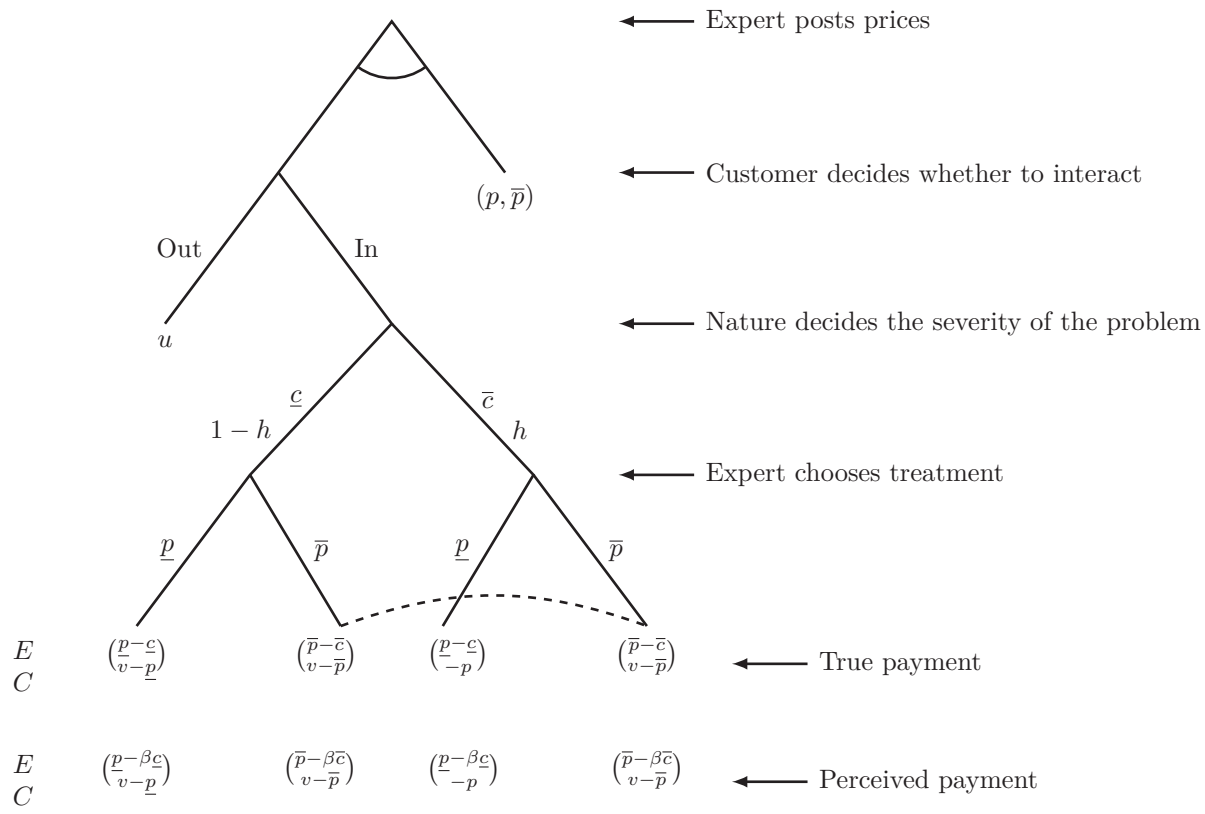


Figure 3.1: Game tree.

3.2.2 Customers with limited attention

We assume that the customers have limited attention. When deciding on interaction, there are three related features of this decision, namely prices $(\underline{p}, \bar{p})$, valuation v, and the likelihood of being undertreated. The likelihood of being undertreated is determined

directly by two factors: the severity of the customer's problem and the action chosen by the expert. With the expert being a profit-maximizing agent, he always chooses the action that gives him the higher profit.

We assume that prices and the valuation are salient features, whereas the probability of being undertreated determined by the expert's profit is a hidden feature. We back up this assumption by three observations. First, several laboratory experiments on credence good markets (see, for example, Dulleck, Kerschbamer, and Sutter, 2011, and screenshots from our treatment NoAttention in Figure 3.A.2) have a design feature that only prices are shown to customers when they decide on interaction. Second, valuation and prices immediately show up in the customer's payoff function. Third, although the expert's profit function and costs are common knowledge, they are communicated to the customer once at the beginning of the experiment. Thus, it is reasonably more difficult for customers to recall this information in every period (see Bordalo et al., 2020). When seeing the information concerning the expert's profit in Decision 3 of the Attention treatment (see Figure 3.A.3), the customer considers the hidden feature when deciding on interaction.

Since the expert's profit equals price minus cost, we consider the cost of each treatment as the direct proxies for the hidden feature of the likelihood to be undertreated. In the experiment, we manipulate the salience of this hidden feature by (not) showing the expert's profits for each treatment at the interaction stage, hence (not) indicating costs. We assume that the expert is aware that the customer has limited attention, but the customer is not aware that the expert knows thereof.

The degree of limited attention is captured by parameter β ($\beta \in (0, 1]$) (see Bordalo et al., 2020). If $\beta = 1$, all features are equally salient. If $\beta \to 0$, the customer takes only salient features into consideration and completely neglects the hidden feature. If the customer decides to interact, the expert's profit is $\pi = p - c$, whereas profit as perceived by a customer with limited attention equals $\pi = p - \beta c$. We differentiate among two cases: $\beta = 1$ and $0 < \beta < 1$.

Case 1: $\beta = 1$

In this case, the customer is equally attentive to all features. As shown by Dulleck and Kerschbamer (2006), in equilibrium:

- The expert posts equal-markup prices:

$$\bar{p} = v + (1-h)(\bar{c} - \underline{c}) - u$$
$$\underline{p} = v - h(\bar{c} - \underline{c}) - u.$$

- The expert provides the appropriate treatment.
- The customer chooses to interact.

Case 2: $0 < \beta < 1$

When customers are inattentive to the hidden feature, equal-markup prices from the customers' point of view take the following form:

$$\bar{p}_c = v + \beta(1-h)(\bar{c} - \underline{c}) - u$$
$$\underline{p}_c = v - \beta h(\bar{c} - \underline{c}) - u.$$

Note, however, that $\forall \beta \in (0,1)$, $\underline{p}_c$ is strictly larger than $\underline{p}$ and $\bar{p}_c$ is strictly smaller than $\bar{p}$. We thus have:

Lemma 1. *When customers have limited attention, the equal-markup tariff $(\bar{p}, \underline{p})$ is perceived by customers as a tariff, such that the markup for the major treatment exceeds that for the minor treatment.*

Now we analyze the optimal price-setting by the expert. To this end, consider the three classes of tariffs as perceived by customers:

(i) The markup for the major treatment exceeds that for the minor treatment ($\bar{p} - \beta\bar{c} > \underline{p} - \beta\underline{c}$),

(ii) the markup of the minor treatment exceeds that for the major treatment ($\bar{p} - \beta\bar{c} < \underline{p} - \beta\underline{c}$), and

(iii) markups are the same for both treatments ($\bar{p} - \beta\bar{c} = \underline{p} - \beta\underline{c}$).

Customers with limited attention expect the following: The expert performs the major treatment if he posts (i), he performs the minor treatment if he posts (ii), and he is indifferent if he posts (iii).[8] The customers observe the price and infer experts' incentives accordingly. Expert's profits in these cases amount to:

(i) $v - u - \beta\bar{c}$

(ii) $(1-h)v - u - \beta\underline{c}$

(iii) $v - u - \beta(h\bar{c} + (1-h)\underline{c})$.

Given that $u > 0, \bar{c} > \underline{c}, v > (\bar{c} - \underline{c}), \beta \in (0,1)$, and $h \in [0,1]$, the equal-markup tariff gives the highest obtainable profit for the expert. We can thus state the following proposition:

Proposition 1. *When the customer has limited attention, conditional on interaction,*

(i) the expert always posts tariffs, such that the markup of the minor treatment exceeds that for the major treatment, and

(ii) the expert always provides the minor treatment.

3.3 Experiment

3.3.1 Experimental design

We build our experimental design on Dulleck, Kerschbamer, and Sutter (2011). Our NoAttention condition replicates the results from the baseline condition with verifiability in Dulleck, Kerschbamer, and Sutter (2011). We introduce salience of the expert's profit in our second condition Attention.

[8] Similarly to Dulleck and Kerschbamer (2006), we assume that the expert is indifferent between two treatments if he posts an equal-markup tariff. Moreover, this is common knowledge.

Subjects are assigned to be either an expert (called Player A in the experiment) or a customer (called Player B in the experiment). Each market consists of eight subjects, with four experts and four customers. In each period, one expert interacts with one customer. The assignment to group and role is random and does not change during the experiment. The stage game is repeated for 16 periods. Subjects are re-matched within their market at the beginning of each period. At the end of each period, subjects are informed about their profit for the current period and their own accumulated profit.

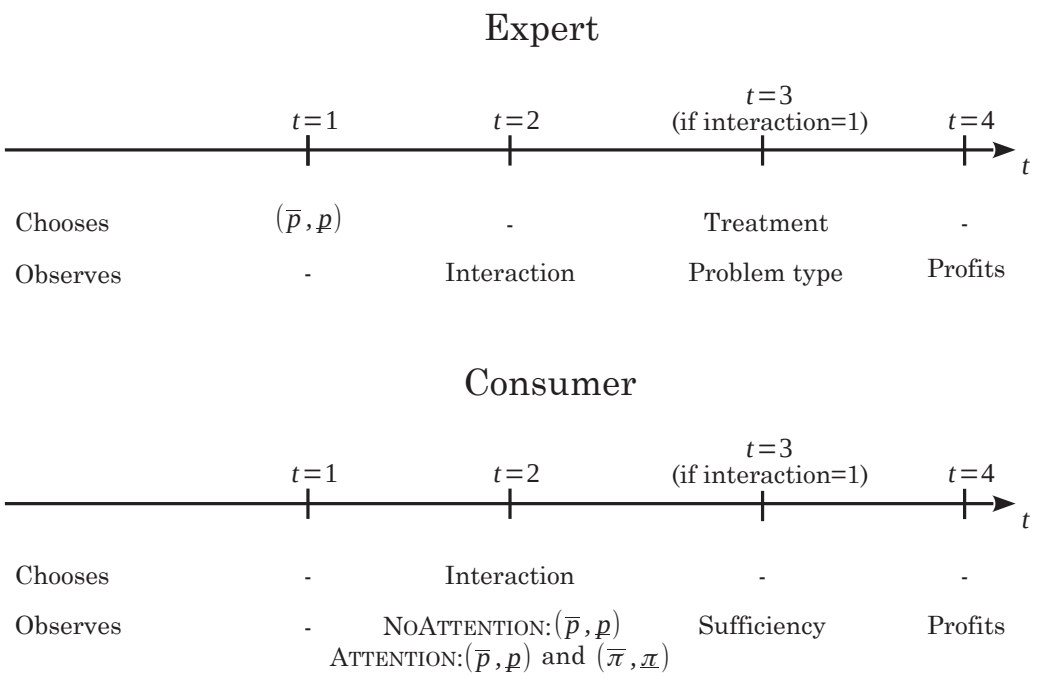


Figure 3.2: Timeline

The timing is displayed in Figure 3.2. In each period, the expert chooses prices $p_i \in [1, 11] \in N$ for each of the two conditions. The customer then chooses whether to interact. If a customer chooses not to interact, the period ends and she and her matched expert both get $u = 1.6$ ECU (outside option). If a customer decides to interact, the expert provides either the minor treatment $\underline{c}$ at costs of 2 ECU (called Action 1 in the experiment) or the major treatment $\bar{c}$ at costs of 6 ECU (called Action 2 in the experiment). The customer derives a utility $v = 10$ ECU if she is sufficiently treated and 0 otherwise. The probability of a customer having a major problem is $h = 0.5$. The expert

and the customer both learn their respective payoffs after every round.

After the experiment, we use two incentivized choices to elicit individuals' risk and loss preferences. We employ the standard choice list by Holt and Laury to measure individuals' level of risk aversion. In a second choice list similar to Karle et al. (2015), we measure individuals' degrees of loss aversion. We ask for individuals' beliefs conditional on the subjects' role of a buyer or a seller.[9] We further complement the incentivized decisions by the validated question on risk aversion by Falk, A. Becker, T. J. Dohmen, et al. (2016). Selected questionnaire items from the preference survey module of Falk, A. Becker, T. J. Dohmen, et al. (2016) serve as a measure of social preferences. We complete the post-experimental part by recording individuals' reasoning for their decision in the experiment and socio-demographics.

Table 3.1 provides an overview on subjects' covariates. The left column shows averages across all participants, the two middle columns show descriptive statistics per condition along with the significance level of the difference in the right column. Our randomly drawn subjects are on average slightly risk-loving (using Holt and Laury (2002) switching point < 5 implies risk-loving). It holds for both conditions. Subjects are loss averse with again virtually no variation across conditions. The Falk, A. Becker, T. Dohmen, et al. (2018) General Preference Survey (GPS) preference measures confirm, consistently with the Holt and Laury (2002), that our subjects are risk-averse.

Table 3.1: Descriptive statistics.

	All	NoAttention	Attention	Difference
Loss aversion (lottery)	4.31	4.33	4.29	$p = 1.000$
Risk aversion (lottery)	4.22	4.23	4.22	$p = 0.953$
Risk aversion (question)	3.78	3.60	3.90	$p = 0.370$
Social preference	4.14	4.08	4.18	$p = 0.679$
Generosity	118.64	140.67	103.96	$p = 0.409$
Belief (buyer)	0.31	0.33	0.29	$p = 0.616$
Belief (seller)	0.42	0.42	0.43	$p = 0.678$
Gender	0.53	0.63	0.47	$p = 0.038$
Age	24.72	25.15	24.43	$p = 0.345$
Number of obs.	120	48	72	15

We classify individuals into two categories based on their social preferences. We define

[9] See the elicitation of beliefs in section 3.B.

a subject as pro-social or selfish based on the median split in the elicited social preference. In both conditions, the means are slightly higher than the median: It is 15.42 in Attention and 15.90 in NoAttention, and, therefore, we assign 15 to the selfish group. Then, our split threshold also conveniently corresponds to giving a stranger the same amount she spent to help. We thus call subjects pro-social if they are willing to give at least the amount the stranger spent to help them (more than 15 ECU), and call them selfish otherwise (15 ECU and less). We then define a market as pro-social based on the number of pro-social experts in this market (from 0 to 4) and treat this measure as a continuous control variable in all following regressions. We only take into account the number of pro-social experts (but not customers), because experts make the two main decisions in the market: pricing and mistreatment. customers, on the other hand, can only accept or reject experts' offers.

3.3.2 Experimental procedure

We conducted our experiment at the DICE Lab of the University of Düsseldorf in June 2019. We programmed the experiment using zTree (Fischbacher, 2007a). Subjects were recruited via ORSEE (Greiner, 2004) and were mostly enrolled as students at the University of Düsseldorf. Upon arriving in the lab, each subject was randomly assigned to a cubicle and provided with instructions. Subjects were given enough time to read the instructions and were allowed to ask experimenters clarifying questions privately. The sessions started after all questions had been addressed.

In total, 120 subjects participated in six sessions of the experiment. Each session lasted for about one and a half hours. On average, subjects earned 18.34 euro. In total, 48 subjects participated in the NoAttention condition, and 72 subjects participated in the Attention condition.

3.3.3 Hypotheses

Based on our theoretical model and the experimental parameterization, we now form our hypotheses for expert and customer behavior. Since customers do not observe the expert's

profit when deciding whether to interact under NoAttention, our model predicts the following expert behavior:

Hypothesis 1. *The expert is more likely to post an undertreatment tariff in* NoAttention *than in* Attention.

Hypothesis 2. *The expert's mark-up difference is larger in* NoAttention *than in* Attention.

Hypothesis 3. *The expert is more likely to undertreat a customer in* NoAttention *than in* Attention.

Customer behavior is predicted as follows:

Hypothesis 4. *A customer is more likely to interact given undertreatment price vectors in* NoAttention *compared to* Attention.

3.4 Results

This section is organized as follows: First, we present how customers' limited attention affects experts' decisions. We analyze the price vectors that experts post and the treatment composition they provide given their posted prices. We then focus on the buyers' side of the market, that is, how limited attention affects their decisions to interact. Finally, we discuss how increased attention influences customers' and experts' welfare in our experiment. Additionally, we look in-depth at how each market outcome varies with the salience of experts' profits for different types of individuals and markets according to the social preferences classification (see Section 3.3).

Table 3.2 provides a first overview of the outcomes on the aggregate level:

In the individual-level data analysis, we control for the price level and previous period market characteristics. We further account for individual experts' characteristics that include age, gender, measures for loss aversion, risk aversion, social preferences, and elicited beliefs.

Table 3.2: Summary statistics.

	Attention		NoAttention	
	Mean	Std. dev.	Mean	Std. dev.
Major price $\bar{p}$	7.93	1.35	7.77	1.57
Minor price $\underline{p}$	5.39	1.73	5.68	1.67
Markup difference Δ	−1.47	1.94	−1.91	1.69
Interaction	0.52	0.50	0.48	0.50
Sufficient	0.38	0.49	0.37	0.48
Number of obs.	1152	1152	768	768

Expert outcomes

Prices and markups

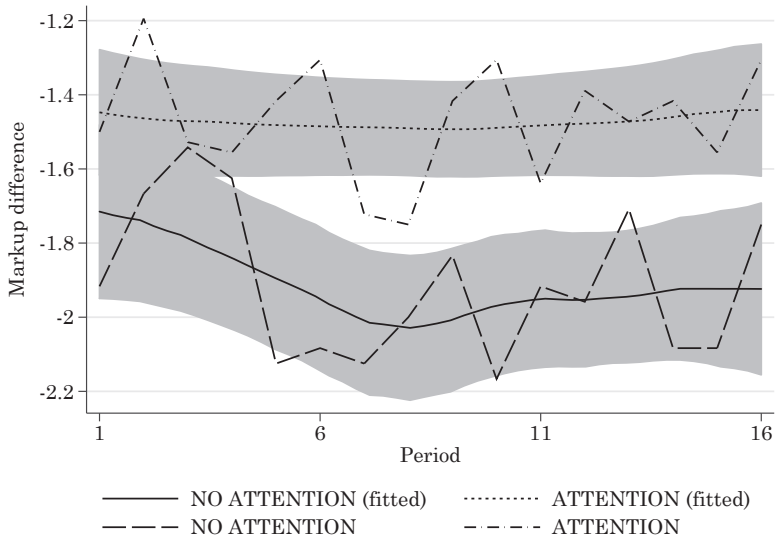
Table 3.3: Probability of posting an undertreatment vector.

Undertreatment vector	All experts	Pro-social experts	Selfish experts
Attention	−0.17***	−0.05	−0.20**
	(0.05)	(0.09)	(0.08)
Pro-social market		0.03	0.04
		(0.04)	(0.05)
Major price	☑	☑	☑
Undertreatment vector $_{t-1}$	☑	☑	☑
Interaction $_{t-1}$	☑	☑	☑
Sufficient $_{t-1}$	☑	☑	☑
Individual controls	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes
Number of obs.	900	360	540

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. All regressions are estimated using probit, average marginal effects are displayed. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and elicited beliefs. We also include time (period) and market fixed effects.

We start by analyzing how the probability of posting undertreatment price vectors varies with the degree of attention. In our experiment, experts do post undertreatment vectors frequently independent of the condition. More precisely, 79.2% of the price vectors in NoAttention and 77.4% of the price vectors in Attention are undertreatment vectors. On the aggregate level, the share of undertreatment vectors is not significantly different ($p = 0.649$, t-test with clustering on subject level)On the individual level, Table 3.3 reveals however that, keeping everything else constant, experts are significantly less

likely to post an undertreatment vector in Attention than inNoAttention. We find that the probability of posting an undertreatment price vector in Attention is 17 percentage points lower compared to NoAttention. Our finding is in line with Hypothesis 1, that is, experts are more likely to post undertreatment price tariffs in NoAttention than in Attention.



Note: Fitted values are estimated using Epanechniov kernel with an optimal bandwidth. Gray areas correspond to 95% confidence intervals.

Figure 3.3: Average markup difference

The impact of salience on the likelihood of posting an undertreatment vector varies for experts with different social preferences. Selfish experts are 20 percentage points more likely to post undertreatment vectors in NoAttention than in Attention. For pro-social experts, the likelihood does not change significantly across conditions. The number of pro-social experts in the market does not seem to matter for price-setting behavior.

Experimental evidence suggests that the markup difference is, on average, negative in both conditions with mean values of -1.91 and -1.47 in NoAttention and Attention, respectively. Figure 3.3 shows that the average markup difference in Attention is less negative than in NoAttention that is, prices set are significantly closer to the equal-markup prices predicted by standard theory. This difference is significant on the aggregate level ($p = 0.027$, t-test with clustering on subject level). Additionally, Table 3.4 shows a substantial effect of Attention on an individual level: Experts whose profits are

displayed to customers post price vectors with a significantly higher markup difference. We account for market characteristics and observe that the markup difference is also heavily affected by the overall price level and inertia. We include them in our regression analysis to control for these possible explanations. However, experts do not account for interaction in the previous period and the sufficiency of the treatment they have previously provided.

Social preference classification provides surprising evidence: salience of the experts' profits has the opposite impact on the average markup difference for pro-social and selfish experts. We find that the increase in the markup difference we have documented on the aggregate and market levels is driven entirely by selfish experts. Customers' attention to their profits has a substantial positive effect on them. Pro-social experts, on the contrary, post price vectors with an even lower markup difference in Attention condition in comparison to NoAttention.

Table 3.4: Markup difference

Markup difference Δ	All experts			Pro-social experts	Selfish experts
	(1)	(2)	(3)		
Attention	2.93***	1.22***	1.22***	−0.78**	1.76***
	(0.34)	(0.26)	(0.26)	(0.31)	(0.50)
Pro-social market				−0.23	−0.26
				(1.37)	(0.29)
Major price		0.55***	0.55***	0.67***	0.43***
		(0.06)	(0.06)	(0.08)	(0.08)
Markup difference $_{t-1}$		0.52***	0.52***	0.19***	0.41***
		(0.04)	(0.04)	(0.07)	(0.06)
Interaction $_{t-1}$			0.01	−0.17	0.12
			(0.13)	(0.13)	(0.19)
Sufficient $_{t-1}$			−0.05	−0.13	−0.11
			(0.14)	(0.16)	(0.20)
Individual controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
Number of obs.	960	900	900	360	540

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences and elicited beliefs. We also include time (period) and market fixed effects.

Next, we want to analyze whether the increase in the markup difference was driven by an increase of $\overline{p}$, a decrease of $\underline{p}$, or both. The price for the minor treatment is on average

5.68 in NoAttention and 5.39 in Attention. The price for the major treatment is on average 7.77 in NoAttention and 7.93 in Attention. Regression analysis in Table 3.5 shows, however, that the higher markup difference in Attention than in NoAttention is driven mainly by the lower price of the minor treatment. Both prices for major and minor treatments are significantly autocorrelated, that is, are highly correlated with the respective prices set in the previous period. Both prices are also positively correlated to the interaction in the previous period.

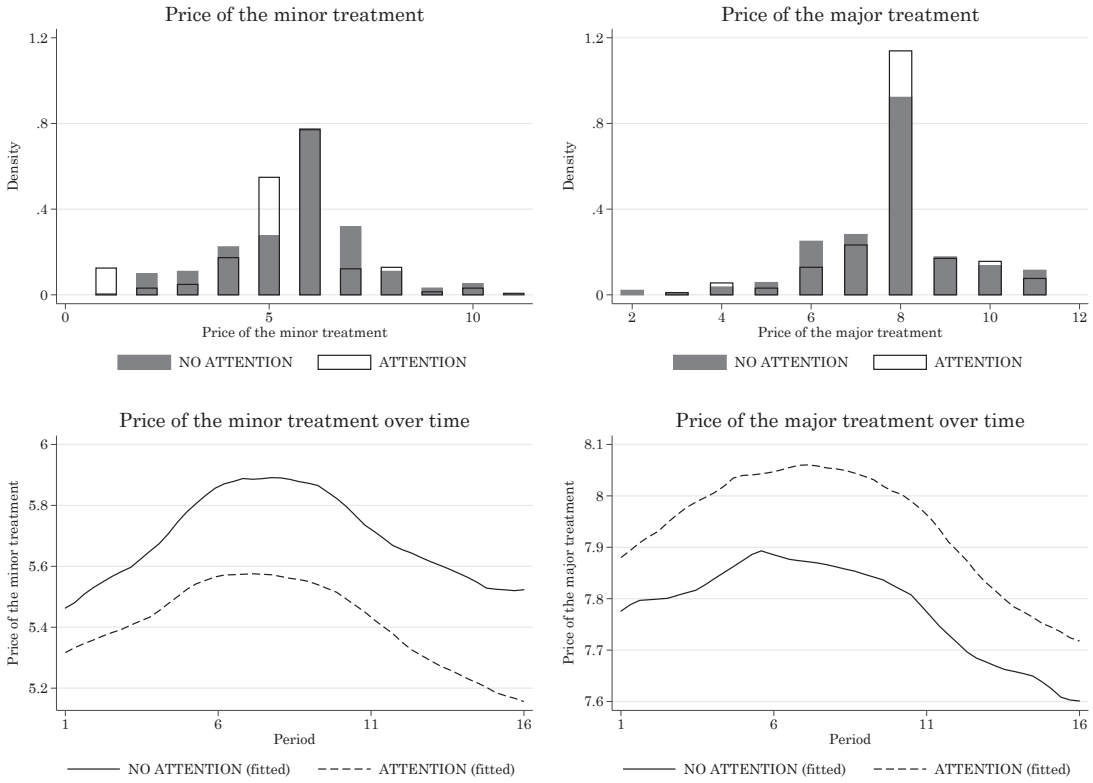
Table 3.5: Prices.

Prices	All experts		Pro-social experts		Selfish experts	
	$\underline{p}$	$\overline{p}$	$\underline{p}$	$\overline{p}$	$\underline{p}$	$\overline{p}$
Attention	−0.93***	0.12	0.81***	0.25	−1.23**	0.23
	(0.26)	(0.22)	(0.28)	(0.25)	(0.48)	(0.38)
Pro-social market			0.44	0.82	0.16	0.23
			(1.61)	(1.78)	(0.26)	(0.22)
Minor price $_{t-1}$	0.54***		0.33***		0.57***	
	(0.05)		(0.07)		(0.07)	
Major price $_{t-1}$		0.47***		0.43***		0.43***
		(0.05)		(0.08)		(0.07)
Markup difference $_{t-1}$	−0.09**	0.03	−0.00	−0.13**	0.05	−0.01
	(0.04)	(0.03)	(0.07)	(0.07)	(0.06)	(0.05)
Interaction $_{t-1}$	0.40***	0.30***	0.35**	0.23	0.39**	0.30**
	(0.11)	(0.11)	(0.14)	(0.15)	(0.16)	(0.14)
Sufficient $_{t-1}$	0.05	0.03	0.19	0.17	0.06	−0.05
	(0.12)	(0.12)	(0.16)	(0.18)	(0.18)	(0.15)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs.	900	900	360	360	540	540

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and elicited beliefs. We also include time (period) and market fixed effects.

Table 3.5 also shows the mechanism of price setting for pro-social and selfish experts. There is no effect of salience on the price of the major treatment: neither for pro-social nor for selfish experts. Instead, the difference in their behavior is captured entirely by $\underline{p}$. Pro-social experts set the price of the minor treatment 0.8 ECU higher in Attention compared to NoAttention, whereas selfish experts, on the contrary, lower it by 1.2 ECU in Attention. Therefore, the price for the minor treatment drives the gap of the salience effect on the markup difference of about 2 ECU.

One could argue that displaying expert's profits to customers decreases the overall



Note: Fitted values are estimated using Epanechniov kernel with optimal bandwidth.

Figure 3.4: Prices

complexity of the experiment. However, as shown in Figure 3.4, time trends in both prices are very similar in Attention and NoAttention, which suggests that price differences can be explained by condition and not by difference in learning. Parallel development of prices over 16 periods helps us to rule out this potential explanation.

Mistreatment

If a customer decided to interact upon posted prices, an expert observes the severity of her problem and chooses which treatment to provide. Given verifiability, there is no scope for overcharging. However, experts may still mistreat customers. Mistreatment can generally occur in two cases: when a customer with a minor problem receives a major treatment (overtreatment), and when a customer with a major problem receives a minor treatment (undertreatment). Under- and overtreatment rates are calculated as a share of all under-/overtreatments given under-/overtreatment was possible (that is, undertreatment rates for only customers with major problems, overtreatment rates for only customers with

minor problems).

Undertreatment rates are 53.66% and 49.69% in NoAttention and Attention, respectively (MWU test, $p = 0.560$). Overtreatment rates are 20.19% and 20.57% in NoAttention and Attention, respectively (MWU test, $p = 0.943$). We estimate the probability of sufficient treatment provision for customers with a major problem and show the results in Table 3.6. We find that overall customers in Attention are more likely to receive a sufficient treatment in comparison to those in NoAttention, which is in line with Hypothesis 3. The impact of salience is insignificant for selfish experts but large and highly significant for pro-social experts: Pro-social experts are almost twice as likely to provide a sufficient treatment in Attention compared to NoAttention. Interestingly, despite posting more undertreatment price vectors, pro-social experts undertreat less. Selfish experts, on the contrary, post fewer undertreatment vectors in Attention; however, their likelihood of sufficient treatment provision does not vary significantly.

Table 3.6: Probability of sufficient treatment provision.

Sufficient	All experts			Pro-social	Selfish
	(1)	(2)	(3)	experts	experts
Attention	0.34***	0.27**	0.23*	0.97***	0.40
	(0.12)	(0.13)	(0.13)	(0.36)	(0.25)
Pro-social market				−0.16	−0.15
				(0.16)	(0.15)
Undertreatment vector		☑	☑	☑	☑
Overtreatment vector			☑	☑	☑
Individual controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
Number of obs.	241	241	241	86	155

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and elicited beliefs. We also include time (period) and market fixed effects.

Generally, experts in Attention make more efficient decisions. The probability of providing a sufficient treatment conditional on a customer having a minor problem is 50.31%, and it differs a lot depending on the price vector, an expert chose in this period: It is less likely that an expert provides sufficient treatment if he posted an undertreatment vector (41.60%) and more likely otherwise (82.35%). In NoAttention this pattern is

much less pronounced. The probability of providing a sufficient treatment conditional on a customer having a minor problem is 46.34%, and it differs rather little depending on the price vector set: When an undertreatment vector has been posted, an expert provides a sufficient treatment with a probability of 45.07%. Otherwise, the probability of sufficient treatment provision is higher (54.55%), but only marginally.

Table 3.7: Overtreatment probability.

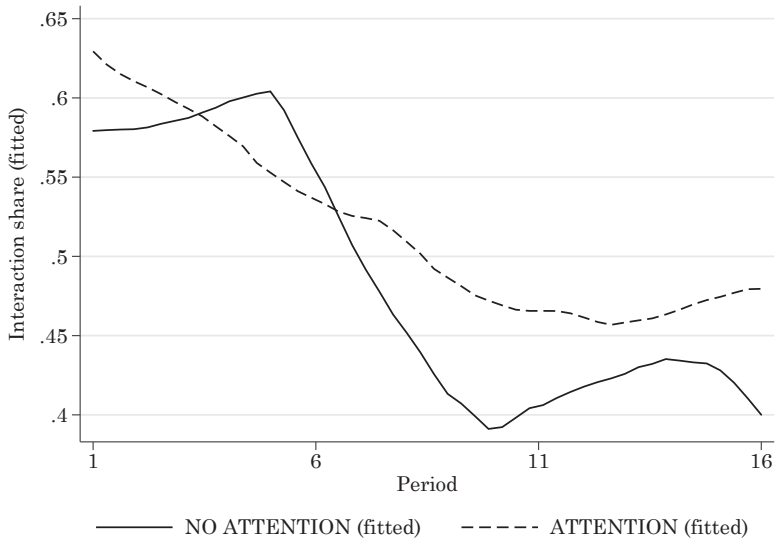
Overtreatment	All experts			Pro-social experts	Selfish experts
	(1)	(2)	(3)		
Attention	0.50***	0.20**	0.14*	0.19	0.38***
	(0.13)	(0.09)	(0.08)	(0.13)	(0.15)
Pro-social market				-0.15**	-0.22**
				(0.07)	(0.10)
Undertreatment vector		☑	☑	☑	☑
Overtreatment vector			☑	☑	☑
Individual controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
Number of obs.	241	241	241	74	158

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences and elicited beliefs. We also include time (period) and market fixed effects.

Average overtreatment probabilities are very similar in Attention (20.57%) and NoAttention (20.19%) (MWU test, $p = 0.943$). However, we find that conditional on other market outcomes, customers in Attention are more likely to be overtreated than in NoAttention (see Table 3.7). More precisely, when an expert in Attention condition posts an overtreatment vector, and a matched customer has a minor problem the expert overtreats with certainty (a probability of 100%), whereas the probability of overtreatment is only 7.44% when another price vector was posted. In NoAttention the pattern is similar but less pronounced. When an overtreatment price vector was posted, customers are 55.56% likely to be overtreated, and 16.84% otherwise. The social expert classification shows that the probability of overtreatment increases with salience but only for selfish experts. Pro-social experts only have an insignificant increase in their likelihood to overtreat, whereas selfish experts are about 38 percentage points more likely to overtreat in Attention.

Customer outcomes

In our experiment, the only decision customers make is whether to interact after observing the prices posted by experts. The trade-off they face is whether to go for a safe outside option of 1.6 ECU or interact and face the risk of being mistreated.



Note: Fitted values are estimated using Epanechnikov kernels with an optimal bandwidth.

Figure 3.5: Interaction probability over time.

In both conditions, customers interact about half of the time: Interaction rates are 48.44% and 52.08% in NoAttention and Attention, respectively. Table 3.8 summarizes interaction probabilities depending on the posted price vector: Customers are most likely to interact when undertreatment price vectors are posted. It is in line with our theoretical predictions: When a customer's attention is limited, she perceives an undertreatment price vector as an equal-markup vector, an equal-markup vector as an overtreatment vector, etc. Therefore, when an expert posts an actual overtreatment price vector, it is perceived as a very unattractive offer by customers with limited attention who thus become less likely to interact.

As shown in Figure 3.5, interaction probability has a rather strong time trend. In the early periods, customers are likely to interact, and this probability decreases over time. For example, in the first period, 62.5% of customers in NoAttention and 72.2% of customers in Attention choose to interact, whereas in the last (16^{th}) period, only

Table 3.8: Interaction probability by price vector posted (%).

	NoAttention	Attention
Undertreatment vector	50.66	52.47
Equal-markup vector	45.45	49.15
Overtreatment vector	33.33	52.11

33.33% of customers in NoAttention and 50% of customers in Attention choose to do so. Our regression analysis in Table 3.9 provides evidence that there is no significant effect of salience on interaction probability.

Table 3.9: Interaction probability.

Interaction	All customers			Pro-social customers	Selfish customers
	(1)	(2)	(3)		
Attention	−0.01	0.01	−0.07	0.10	0.22
	(0.09)	(0.10)	(0.10)	(0.15)	(0.15)
Pro-social market				−0.08	−0.08
				(0.08)	(0.08)
Markup difference	☑	☑	☑	☑	☑
Markup difference $_{t-1}$			☑	☑	☑
Major price	☑	☑	☑	☑	☑
Major price $_{t-1}$			☑	☑	☑
Interaction $_{t-1}$		☑	☑	☑	☑
Sufficient $_{t-1}$		☑	☑	☑	☑
Individual controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
Number of obs.	960	900	900	480	420

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and elicited beliefs. We also include time (period) and market fixed effects.

Welfare

Various key market outcomes, such as markup difference and mistreatment rates, are influenced by attention which can lead to welfare implications. We analyze how welfare differs between conditions, and break it down to customer and expert surplus. In addition to analyzing customers' and experts' profits, we construct a market-level efficiency following Mimra et al. (2016). We calculate efficiency level as cumulative profits in the market less the outside option of all players, and normalize it with respect to the distribution of

Table 3.10: Welfare.

Welfare	Surplus per condition			Surplus per market		
	Consumers	Producer	Total	Consumers	Producer	Total
Attention	−9.38***	4.89***	−4.50***	−1.31***	0.82***	−0.58**
	(0.57)	(0.34)	(0.49)	(0.34)	(0.19)	(0.23)
Major price	0.10	0.27*	0.37*	-0.38***	0.63***	0.30***
	(0.26)	(0.15)	(0.20)	(0.14)	(0.08)	(0.10)
Markup difference	−0.16	−0.22**	−0.38**	−0.05	−0.27***	−0.30***
	(0.17)	(0.10)	(0.15)	(0.12)	(0.05)	(0.08)
Interaction	−2.90***	0.73	−2.17***	−6.66***	2.17***	−4.46***
	(0.76)	(0.47)	(0.72)	(0.48)	(0.25)	(0.33)
Sufficient	4.26***	−0.48	3.78***	7.95***	-0.55**	7.47***
	(0.81)	(0.50)	(0.74)	(0.49)	(0.27)	(0.34)
Individual controls	No	No	No	Yes	Yes	Yes
Fixed effects	No	No	No	Yes	Yes	Yes
Number of obs.	1920	1920	1920	960	960	1920

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and elicited beliefs. We also include time (period) and market fixed effects.

customers in the respective market, which allows accounting for the random differences in total welfare generated by the severity of customers' issues.[10]

We start by analyzing welfare through profits acquired by participants over the course of the experiment and analyze them on the condition and market level. We find that although attention to experts' profits leads to an improvement in a number of market outcomes, cumulative profits go down. Moreover, the loss is driven entirely by the loss in the customer surplus. In contrast, experts benefit greatly from their profits being displayed to customers.

We observe several patterns in total welfare. Total welfare decreases if there is interaction unless it is sufficient, because customers experience a large instant loss from undertreatment. Total welfare remains unchanged if the markup difference increases through $\overline{p}$: In this case, customers lose on average the same amount that experts gain. However, if the markup difference increase comes through the reduction of $\underline{p}$, customer surplus does not change significantly, whereas expert surplus decreases, so total welfare goes down as

[10]Given interaction, every customer is randomly assigned to have a major or a minor problem with a probability of 50%. In case of a minor problem, every customer-expert pair can generate at least $(10-p)+(p-6)=4$ and at most $(10-p)+(p-2)=8$. In case of a major problem, every customer-expert pair can only generate $(0-p)+(p-2)=-2$ in the worst case and $(10-p)+(p-6)=4$ in the best case. We thus account for these differences when calculating the market efficiency measure.

Table 3.11: Efficiency.

Efficiency	All	Pro-social	Selfish
Attention	0.05***	0.02**	0.11***
	(0.01)	(0.01)	(0.01)
Pro-social market	−0.06***	−0.04***	-0.08***
	(0.00)	(0.00)	(0.00)
Major price	☑	☑	☑
Markup difference	☑	☑	☑
Interaction	☑	☑	☑
Sufficient	☑	☑	☑
Individual controls	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes
Number of obs.	1920	896	1024

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Individual controls include age, gender, measures for loss aversion, risk aversion, social preferences, and elicited beliefs. We also include time (period) and market fixed effects.

well.

However, our data also shows that, despite the theoretical probability of customers to have a minor problem is 50%, the minor problem actually arises in 47% and 56% of cases in Attention and NoAttention, respectively. As mentioned above, the severity of the treatment affects crucially the cumulative profits a customer-expert pair can generate, and, thus, it is important to take it into account for estimating efficiency.

We find that efficiency indeed increases if experts' profits are salient to consumers. On average, efficiency increases by 5 percentage points with salience. The effect is significant for sub-samples with different social preferences. However, we find a particularly pronounced efficiency gain from salience for selfish subjects: It accounts for 11 percentage points increase in market efficiency on average.

3.5 Conclusion

There exist contradictions between theoretical predictions and empirical evidence on the role of verifiability in the credence goods market. While theory predicts that under certain conditions, verifiability leads to market efficiency, observations from real markets go against this prediction. We are the first to provide theoretical argument and experimental

evidence that customer's limited attention plays a role in this inconsistency. Our finding goes in line with recent advocacy for more transparency on experts' pay in credence goods markets, such as healthcare or repair services.

Based on the inherent features of lab experiments on the credence goods market, we set up a model of a monopolistic credence goods market in which customers pay limited attention to expert's costs, resulting in a false assessment of the expert's financial incentives. Our model further assumes that the expert knows that customers pay limited attention to their costs, whereas customers are unaware thereof. Our main hypotheses are that an increase in customers' attention with regard to experts' costs results in (i) a decrease in customer interaction given an undertreatment tariff, (ii) a decrease in the number of undertreatment tariffs and insufficient treatments, and (iii) a smaller markup difference between the major treatment and the minor treatment.

We test the hypotheses in a laboratory experiment and find support for the last two hypotheses. We observe less undertreatment, and experts' price vectors were significantly closer to equal mark-up pricing when expert costs are made salient than when they are not. We do not find strong supporting evidence for the first hypothesis. Interestingly, we observe that interaction given an overtreatment tariff under the salience of experts' cost is much higher than under limited attention. We argue that risk aversion and experimental parameterization might account for this effect. In terms of welfare, the salience of experts' costs leads to an increase in accumulated payoffs. Throughout, we observe a heterogeneity of results with regard to social preference.

Overall, our results suggest that customers' limited attention is a possible explanation for the empirical evidence on the inefficiency of verifiability in credence goods market. Furthermore, our study draws a rather nuanced picture when it comes to the merits of introducing more transparency of experts' costs. We observe a positive effect on undertreatment, markups, and welfare, but we do not find an overall increase in interaction compared to the case without transparency. Hence, increasing transparency might serve customers who choose to interact and all experts, but might do more harm than good to customers who interact less, or refrain from interaction altogether. Taken on its own,

our findings explain why providers in healthcare and repair service appear not to object to calls for more transparency. What remains an open question for future research is whether expert providers aim to gain a competitive advantage over their rivals through transparency.

Appendix

3.A Instructions of the Experiment

Thank you for your participation in this experiment. Please do not to talk to any other participants during the experiment. Today's experiment consists of several parts. Your earning is the total income from these parts. In addition, you will receive a show-up fee of 4 Euros for today's participation and for answering the questionnaire.

INSTRUCTIONS

2 Roles and 16 Rounds

This experiment consists of **16 rounds**, each of which consists of the same sequence of decisions. This sequence of decisions is explained in detail below.

There are 2 kinds of roles in this experiment: **player A** and **player B**. At the beginning of the experiment you will be randomly assigned to one of these two roles. On the first screen of the experiment you will see which role you are assigned to. Your role remains the same lthroughout the experiment. In your group there are 4 players A and 4 players B.

One player A always interacts with one player B. However, the pairs **change** after each round. That means you will interact with a **new** player (the other role) every round.

All participants get the same information on the rules of the game, including the costs and payoffs of both players.

Overview of the Sequence of Decisions in a Round

Each round consists of a maximum of 3 decisions which are made consecutively. Decisions 1 abd 3 are made by player A, decision 2 is made by player B.

1. Player A chooses one price for action 1 and one price for action 2.

2. Player B gets to know the prices chosen by player A. Then player B decides whether he/she wants to interact with player A. If not, this round ends for him/her.

If yes...

3. Player A (but **not** player B) is informed whether player B is of type 1 or type 2. Player A chooses thereupon either action 1 or action 2. Player B has to pay the price specified by player A in decision 1 for the action chosen by player A.

Detailed Illustration of the Decisions and Their Consequences Regarding Payoffs

Decision 1

- **Player A** has to choose between 2 actions (action 1 and action 2) at decision 3.
- **Action 1** costs player A **2 points** (= currency of the experiment).
- **Action 2** costs player A **6 points.**
- Player A can charge prices for these actions from player B who decide to interact with him/her. At **decision 1** each Player A has to **set the prices for both actions**. Only (strictly) positive integer numbers are possible, i.e., only 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 are valid prices.
- Note that the price for action 1 must not exceed the price for action 2.

Decision 2

**Instruction of Decision 2 differs between two treatments.*

(NoAttention **treatment)**

- **Player B** gets to know the **prices** of player A for the two actions at decision 1. Then player B decides whether he/she wants to interact with player A or not.
- **If he/she wants to do so**, player A can choose an action at decision 3 and charge a price for that action (see below). **If he/she doesn't want to interact,** this round **ends** for player B and he/she gets a **payoff of 1.6 points** for this round.

(Attention **treatment)**

- **Player B** gets to know the **prices and profits** of player A for the two actions at decision 1. Then player B decides whether he/she wants to interact with player A or not.
- **If he/she wants to do so**, player A can choose an action at decision 3 and charge a price for that action (see below). **If he/she doesn't want to interact,** this round **ends** for player B and he/she gets a **payoff of 1.6 points** for this round.

Decision 3

- Before decision 3 is made (in case player B choses "Yes" at decision 2) a type is randomly assigned to player B. **Player B** can be one of the two types: **type 1** or **type 2**. This type is randomly determined for each player B **in each new round**.
- With a probability of **50% player B is of type 1**, and with a probability of **50% he/she is of type 2**. Imagine that a coin is tossed for each player B in each round. If the result is e.g. "heads", player B is of type 1, if the result is "tails" he/she is of type 2.
- Every **player A** gets to **know the types of player B** who interact with him/her before he makes his decision 3. Then player A chooses an action for each player B, either action 1 or action 2.
- An **action** is **sufficient** in the following cases:

  a) Player B has type 1 and player A chooses either action 1 or action 2.

  b) Player B has type 2 and player A choose action 2.
- An action is **not sufficient**, if player B has type 2 but player A chooses action 1.
- **Player B** receives **10 points**, if the action chosen by player A is **sufficient**. **Player B** receives **0 point**, if the action chosen by player A is **not sufficient**.
- **At no time player B** will be informed whether he/she is of type 1 or a type 2 player in each round, as well as which action player A has chosen.

- **Player A** charges player B the **price** set out in decision 1 for the action chosen in decision 3.

Payoffs

If player B chose not to interact with any of the players A (*decision **"No"** from player B*), both player A and player B get **1.6 points** for this particular round.

Otherwise (*decision "Yes" by player B*) the payoffs are as follows:
Player A receives the according **price** (denoted in points) he/she set at decision 1 **less the costs** for the action chosen at decision 3.

The payoff of **player** B depends on whether the action chosen by player A in decision 3 was sufficient or not:
a) If the action chosen by player A was sufficient, Player B gets 10 points less the price set in decision 1 for the action chosen at decision 3.
b) If the action chosen by player A was not sufficient, Player B has to pay the price set in decision 1 for the action chosen at decision 3.

At the beginning of the experiment you receive an **initial endowment of 6 points**. With this endowment you are able to cover losses that might occur in some rounds. Losses can also be compensated by gains in other rounds. If your total payoff sums up to a loss at the end of the experiment you will have to pay this amount to the supervisor of the experiment. By participating in this experiment you agree to this term. **Please note that there is always a possibility to avoid losses in this experiment.**

To calculate the payoff of this part, the initial endowment and the profits of all rounds are added up. This sum is then converted into cash using the following exchange rate:

1 point = 25 Euro-cents

(i.e. 4 points = 1 Euro)

You will see all further instruction on the computer screen.

3.B Questionnaire

The questionnaire at the end of the experiment contains the following items:

1. **Elicitation of beliefs:**

   (Only for sellers)

   When you set the price, did you expect that Player B will decide to interact? (Yes/No)

   Which action (Action 1 or Action 2) would you choose given the following scenarios?

   Price: 3 for Action 1 and 8 for Action 2

   Price: 4 for Action 1 and 8 for Action 2

   Price: 5 for Action 1 and 8 for Action 2

   Price: 6 for Action 1 and 8 for Action 2

   Price: 7 for Action 1 and 8 for Action 2

   (Only for buyers):

   As you decided to interact, did you expect that Player A will choose a sufficient action?

   Which action (Action 1 or Action 2) do you expect Player A to choose given the following scenarios? Price: 3 for Action 1 and 8 for Action 2

   Price: 4 for Action 1 and 8 for Action 2

   Price: 5 for Action 1 and 8 for Action 2

   Price: 6 for Action 1 and 8 for Action 2

   Price: 7 for Action 1 and 8 for Action 2

2. **Risk preference, general risk question:** same wording as in German Socio-Economic Panel questionnaire (SOEP, see, for example, Wagner et al., 2007)

   How do you evaluate yourself? Are you generally a risk-seeking person or do you try to avoid risks? The leftmost box means "not at all risk-seeking" and the rightmost "very risk-seeking". With the boxes in between, you can graduate your statement.

   not at all risk-seeking ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ *very risk-seeking*

3. **Risk preference, incentivized choice list:** Subjects make eleven, pairwise decisions between a lottery with a fifty-fifty chance of winning either 2 EUR or 7 EUR and a safe payment. The safe payment increases in 0.5 EUR increments, ranging from 2 EUR to 7 EUR.

4. **Loss aversion** similar to Karle et al., 2015.

   You will answer questions related to lotteries. If you accept the lotteries, you can make either a profit or a loss. Below are six different lotteries. For each lottery, you can decide whether to accept or to reject it. If you reject, your payment remains unchanged. If you accept, your earning will make either an additional profit or an additional loss.

   At the end of the experiment, one of the six lotteries will be randomly selected. So you should make every decision as if it were your only decision. The selected lottery is then randomly drawn to determine whether the additional profit or loss will be realized for you.

   (All with the same options: Accept or Reject)

   Lottery 1: With a 50% probability you lose 2 EUR and with a 50% probability you win 6 EUR.

   Lottery 2: With a 50% probability you lose 3 EUR and with a 50% probability you win 6 EUR.

   Lottery 3: With a 50% probability you lose 4 EUR and with a 50% probability you win 6 EUR.

   Lottery 4: With a 50% probability you lose 5 EUR and with a 50% probability you win 6 EUR.

   Lottery 5: With a 50% probability you lose 6 EUR and with a 50% probability you win 6 EUR.

   Lottery 6: With a 50% probability you lose 7 EUR and with a 50% probability you win 6 EUR.

5. **Social preference** (survey question, Falk, A. Becker, T. Dohmen, et al., 2018)

   Question 1: Imagine the following situation: Today you unexpectedly received 1000 EUR. How much of this amount would you donate to a good cause? (Values between 0 and 1000 are allowed).

   Question 2: Please think about what you would do in the following situation. You are in an area you are not familiar with, and you realize that you lost your way. You ask a stranger for directions. The stranger offers to take you to your destination. Helping you costs the stranger about 20 EUR in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs 5 EUR, the most expensive one costs 30 EUR. Do you give one of the presents to the stranger as a "thank you" gift?

   Which present do you give to the stranger?

   1. No, would not give present
   2. The present worth 5 EUR
   3. The present worth 10 EUR
   4. The present worth 15 EUR
   5. The present worth 20 EUR
   6. The present worth 25 EUR
   7. The present worth 30 EUR

6. **Description of reasoning for decisions**

   (Only for sellers)

   Please answer the following questions:

   How did you decide for the prices? Please describe what you thought when you set the prices.

   How did you decide for the actions? Please describe, what you thought when you choose the action.

   Did you change your strategy across periods? When yes, why?

(Only for buyers)

Please describe your thought when you made the decision whether or not to interact.

Did you change your strategy across periods? When yes, why?

7. **Socio-demographics:** age, gender, final grade point average at academic high school, last math grade at academic high school, field of study, monthly disposable amount of money, political orientation, number of experiments already participated in the same lab.

3.C Exemplary screens of stage decisions

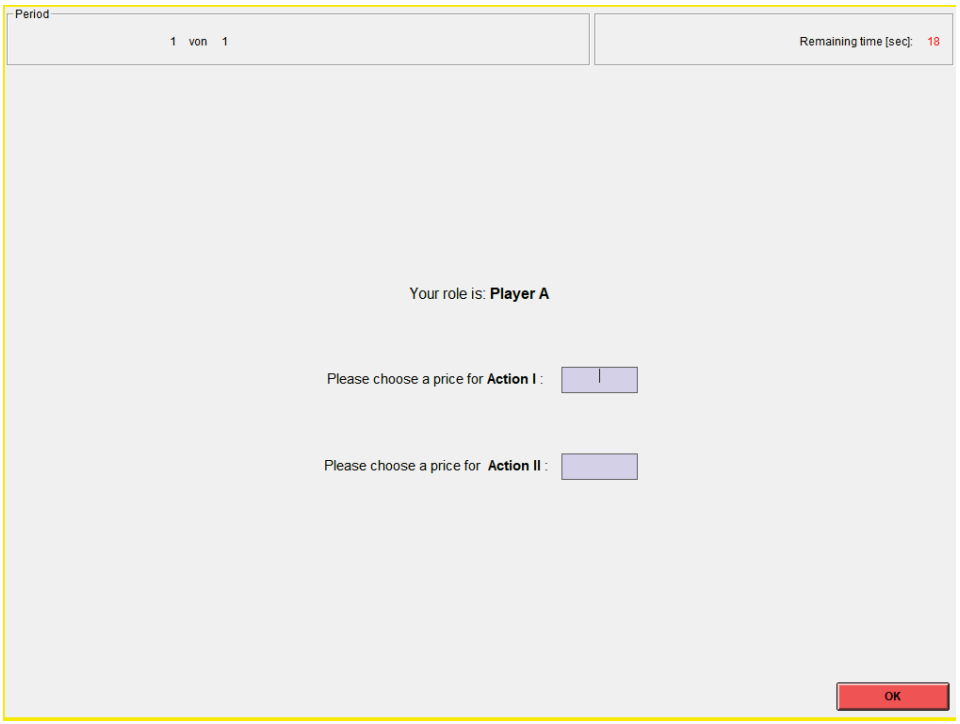


Figure 3.A.1: Exemplary screen (both treatments): experts set prices

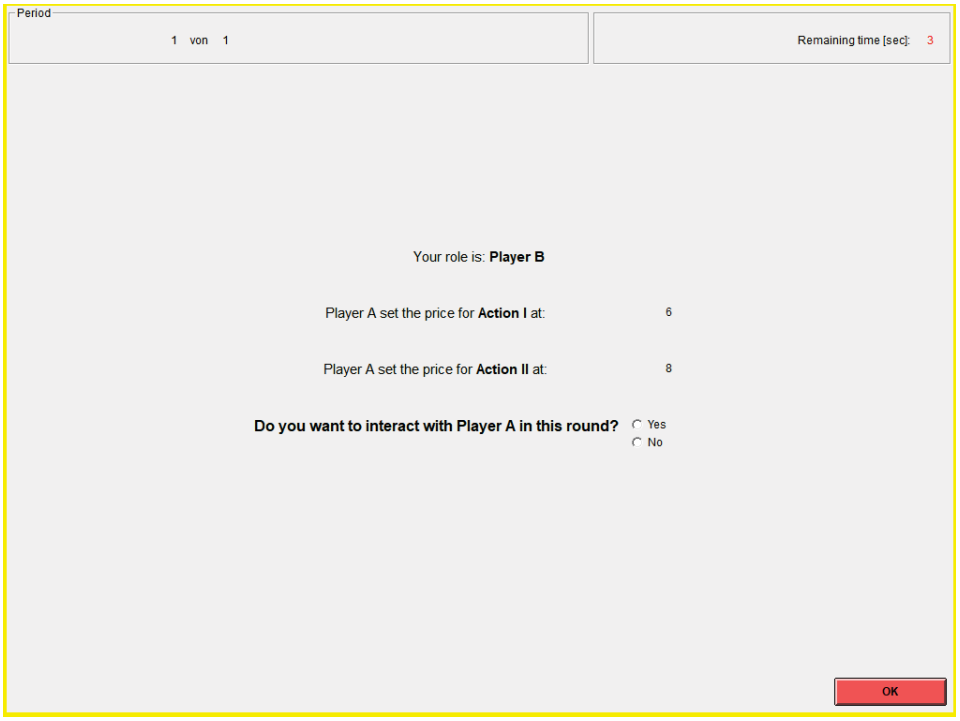


Figure 3.A.2: Exemplary screen (treatment NoAttention): Customers observe prices and decide on interaction

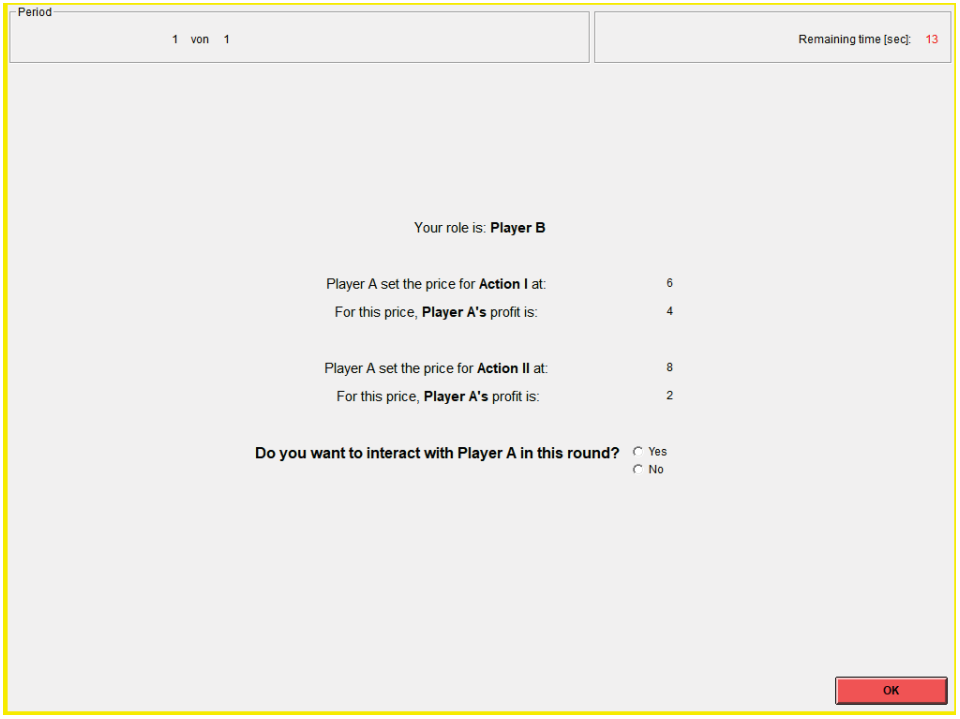


Figure 3.A.3: Exemplary screen (treatment Attention): Customers observe prices and profits, and decide on interaction

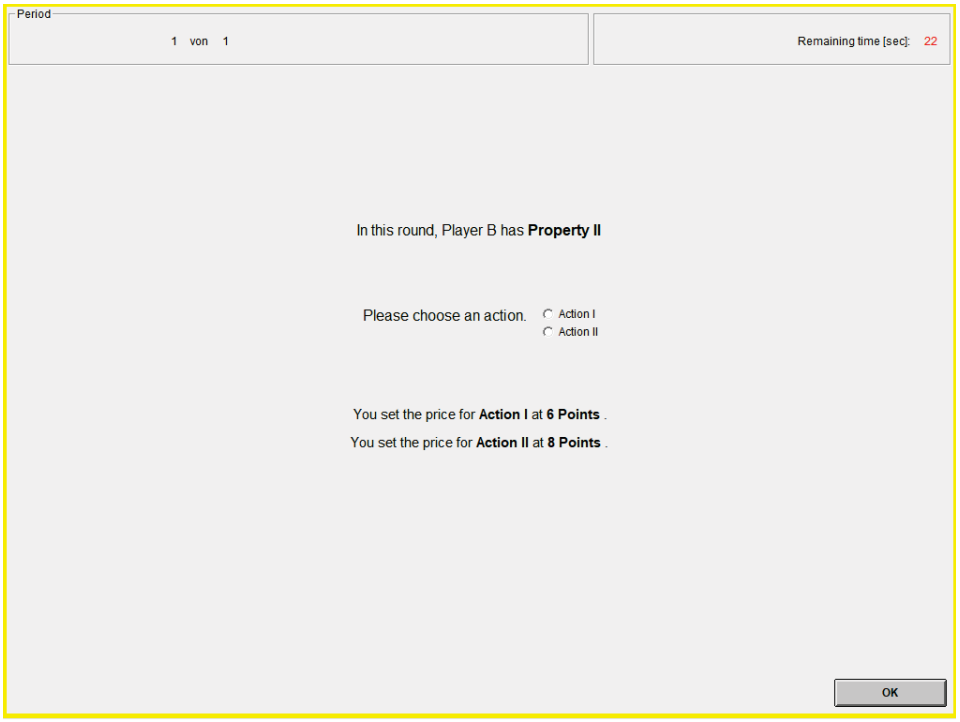


Figure 3.A.4: Exemplary screen (both treatments): Experts choose condition

Chapter 4

Foreclosure and Tunneling with Partial Vertical Ownership

Co-authored with Matthias Hunold

4.1 Introduction

Foreclosure is a major policy concern related to vertical mergers. A vertically integrated entity may not be willing to supply rivals of its downstream unit (input foreclosure), or may not be willing to on-sell the products of a competing upstream firm (customer foreclosure). The Chicago School has argued that an integrated entity that can write efficient contracts does not foreclose other vertically related firms if there are gains from trade. Meanwhile, economists have formally shown that this argument may not apply in certain situations and foreclosure can occur as a result of vertical mergers.

Baumol and Ordover (1994), Spiegel (2013), and Levy et al. (2018) argue that the foreclosure incentives may be even stronger with partial vertical ownership that involves control. For example, if there are voting and non-voting shares of an upstream firm, a downstream firm may own all voting shares and have full control. These articles emphasize that with controlling partial acquisitions, a firm only internalizes parts of another firm's profits and losses, although it can fully distort its strategy to increase its own profit. Consequently, dedicated foreclosure strategies (such as a refusal to supply) can be more attractive when compared to full integration.

In this article, we add to this literature by studying the contracting and corporate governance of partially integrated firms. When a partial owner has control over a target firm, but only obtains part of the dividends, the questions arise whether, how, and to what extent the controlling owner can extract profits from the target firm (*tunneling*). Whereas minority shareholder protection aims at limiting such tunneling, it does take place in practice. Our literature review provides details.

We show that different restrictions on profit shifting lead to distinctively different incentives to foreclose rivals. Certain restrictions indeed cause more incentives to foreclose with partial ownership than in the case of a full vertical merger, in line with the literature mentioned above. However, with other restrictions on tunneling, there are the same or fewer incentives to foreclose in case of partial vertical ownership. For competition policy, it is important to understand under what conditions partial ownership tends to create high foreclosure incentives. We complement the existing literature in this respect.

We focus on studying the restriction on the amount that can be taken out of the target firm (Restriction 1) and the restriction on the amount that must be left in the target firm (Restriction 2). At first sight, it might seem that the restrictions are equivalent. For instance, if the target's profit is 100, one can either specify that at most 20 can be taken out ($t \leq 20$) or that 80 need to be left ($\underline{\pi}^U \geq 80$). However, we will show below that the foreclosure incentives differ substantially. We show that, for different tunneling restrictions, a partial owner's optimal strategy may vary between higher incentives to foreclose than under vertical integration (as discussed in Levy et al. (2018)), the same incentives (because of fully taking into account the target firm's residual profit) and no incentives at all (if the transfer of money into the target firm is sufficiently restricted). We analyze the partial owner's foreclosure incentives for different market environments. In particular, we distinguish between the case where an upstream firm holds shares of a customer (partial forward ownership) and the case where a customer holds shares of its supplier (partial backward ownership).

When a downstream firm partially owns a supplier, we find, in line with Levy et al., that the restriction on the maximal tunneling amount indeed increases partial owner's incentives to foreclose its downstream rivals (input foreclosure) and decreases the incentives to foreclose an upstream target's rivals (customer foreclosure). Interestingly, the alternative restriction on the minimal profit that needs to be left in the target firm yields the same customer and input foreclosure incentives as full integration. Additionally, the restriction on the minimal profit might necessitate shifting profit into the target firm (propping) in order to foreclose. For the case that propping is not feasible at all, or not to a required extent, we find lower incentives for input foreclosure compared to a full integration benchmark.

When an upstream firm partially owns a downstream firm, the restriction on the tunneling amount decreases the incentives of the partial owner to foreclose its target's downstream rivals (input foreclosure) but increases the incentives to force the target to not trade with its own upstream rivals (customer foreclosure). Again, this restriction follows Levy et al. and the results are in line with their findings as well. The minimal profit

restriction, however, yields the same foreclosure incentives as full integration, provided that the partial owner can prop its target firm when the required minimal profit level is relatively high. Additionally, if propping is not feasible at all, or not to a required extent, there are lower customer foreclosure incentives in comparison to a fully integrated firm.

The structure of the remaining text is as follows. Section 4.2 contains the review of the related literature. Section 4.3 studies the input foreclosure incentives under partial backward ownership under different types of restrictions on profit shifting. Section 4.4 contains the analysis for customer foreclosure. We compare the different results in Section 4.5 and also relate them specifically to the article of Levy et al. (2018). Section 4.6 concludes with a discussion of implications for regulation and competition policy.

4.2 Related literature

We relate to and combine mainly two strands of literature, the one on vertical integration and foreclosure, and the other on profit shifting from and to a target firm (tunneling and propping).

Partial vertical ownership. There are crucial differences between a vertical merger and partial controlling backward ownership of the downstream incumbents. Typically, the direction of acquisition does not matter for the competitive effects if the result is a new entity. In particular, a merged entity cannot commit to an internal transfer price above costs (at least the literature on vertical mergers typically assumes this, such as Y. Chen, 2001). This tends to reduce double marginalization within the integrated vertical chain – a pro-competitive effect. The literature has also pointed out the possible anti-competitive effects of vertical mergers. See Rey and J. Tirole (2007) for an overview.

Baumol and Ordover (1994), Spiegel (2013), and Levy et al. (2018) mainly consider the effects of controlling an upstream or downstream firm via partial ownership. They emphasize that, with controlling partial acquisitions, a firm only internalizes parts of another firm's profits and losses, although it can fully distort its strategy to increase its own profit. Consequently, dedicated foreclosure strategies (such as a refusal to supply)

can be more attractive when compared to full integration. A crucial assumption for these results on controlling partial ownership is how the controlling owner can extract profits from the partially owned target firm (tunneling). Our main contribution is to show that the effects of foreclosure depend on the type of tunneling that is feasible in surprising and policy-relevant ways.

Other articles on partial vertical ownership focus more on the case of no or limited control, such that tunneling is less of an issue (Fiocco, 2016; Flath, 1989; Greenlee and Raskovich, 2006; M. Hunold and Stahl, 2016; Matthias Hunold, 2020).

Empirical evidence on tunneling. The second strand of literature deals with tunneling but does not consider partial ownership and foreclosure. Tunneling can take a variety of different forms.[1] The simplest form is shifting profits to the benefit of the controlling shareholder through self-dealing transactions. These may include the sale of over-priced output to the target firm, the purchase of under-priced input from the target firm, excessive salaries, and bonuses for top managers and executives, and even using a corporate jet for private reasons. According to S. Johnson et al. (2000), this form of tunneling is illegal everywhere if it includes theft or fraudulent behavior. However, the controlling shareholders may legally shift profits through asset sales or excessive pricing agreements, or exploit corporate monetary and non-monetary opportunities, or use more complex instruments for profit-shifting.

Especially in countries with weaker investor protection, firms are able to tunnel resources in ways that cannot be prevented by outside investors. A number of studies document empirical evidence for tunneling in various countries like India, China, South Korea, Hong Kong, and Bulgaria. We briefly introduce these studies in turn.

- Bertrand et al. (2002) use the Prowess database to analyze Indian business groups from 1989 to 1999. They compare low-cash-flow to high-cash-flow firms and firms

[1] See Atanasov et al. (2014) for a detailed discussion of three main types of tunneling: cash flow tunneling, asset tunneling, and equity tunneling. Cash flow tunneling is shifting a part of the target firm's current profits (e.g. through transfer pricing, excessive salaries, etc). Asset tunneling is buying the firm's major assets for a price above the market value or selling them for a price below the market value, and thereby influencing the firm's long-term profitability. Equity tunneling is increasing the controller's share at the expense of minority shareholders.

that are a part of a business group to stand-alone firms. They regress a firm's actual reported performance on its predicted performance and the predicted performance of other firms in the same group. They find evidence that tunneling occurs mainly through the firm's non-operating profits and is partly incorporated into the stock market prices.

- Jiang et al. (2010) document the nature and severity of tunneling in China. They analyze 1377 listed companies throughout 1996-2004 and find that controlling shareholders widely use corporate loans to shift profits from listed Chinese companies. They also show that the tunneling problem is the most severe if the control right is significantly larger than the profit right.

- Baek et al. (2006) analyze private placements of firms listed on the Korean Stock Exchange in 1989-2000 and focus on business groups. They compare intragroup deals (deals within one business group) with other deals and provide evidence for tunneling activities within business groups: the firms with favorable past performance sell their securities at a discount to other group members.

- Cheung et al. (2006) analyze transactions between partial owners and target firms of Hong Kong listed companies in 1998-2000. They find that excess returns from those transactions are significantly negative, and negatively related to the percentage ownership of a controlling shareholder. Additionally, they find that the connected party transactions are more likely to be undertaken if the controlling shareholder can be traced to the mainland of China. They explain that those firms find it is easier to expropriate their minority shareholders because rulings by courts in Hong Kong are not enforceable in China and thus Hong Kong investors have little chance to recover shifted assets.

- Atanasov (2005) conducts an econometric analysis of mass privatization in Bulgaria as an extreme case of a lack of mechanisms that can protect minority shareholders.[2] He finds that the absence of regulation allows majority shareholders to extract

[2]He constructs a two-stage estimator which controls for a selection bias. The first stage estimates

up to 85% of the target's firm value to its private benefit. Atanasov provides several examples supporting his evidence: in the year 2000, the national oil refinery Neftochim's stock was only valued at 24% of the price paid by Lukoil for the majority block; Balkanfarma, a holding of three pharmaceutical companies, had a ratio of 21%; and Sodi, the second-largest producer of soda ash in the world, had a ratio of 10.8%. Atanasov argues that controlling shareholders have a strong preference for expropriating minority shareholders rather than adding value through monitoring.

Tunneling also occurs in the context of profit shifting across countries due to tax differences. In their seminal study, Grubert, Mutti, et al. (1991) focus on the ability of firms to shift profits from high-tax to low-tax countries through their foreign affiliates. They use data from 1982 from 33 countries and find that the US-based multinational enterprises shift disproportionally much income to the countries with low statutory tax rates. Moreover, they export more to their foreign affiliates in low-tax countries. More recent examples include Microsoft allegedly shifting profits to its foreign affiliates in Ireland, Puerto Rico, and Singapore to reduce its tax burden in Europe and avoid the US corporate income tax.[3]Another recent example is as well as Apple allegedly using offshore structures to shift billions of dollars out of the United States.[4]

Propping. Opposite to shifting profits from the target firm to the partial owner (tunneling), firms might also shift profits from the owner to the target firm (propping). Partial owners might use it to avoid a potential bankruptcy of the target firm.[5] Friedman et al. (2003) show theoretically that, in case of a moderate negative shock in the market, a partial owner may find it optimal to prop the target firm to prevent its bankruptcy. They also analyze firms hit by the Asian crisis 1997-1998 and provide empirical evidence of propping. Friedman et al. (2003) focus on the Asian crisis 1997-1998, a quasi-natural

whether an investor places a small or a large bid or abstains from bidding at all. The second stage estimates the bid price conditional on bidding.

[3]See United States Congress Senate Committee on Homeland Security and Government Affairs. 2012. Offshore Profit Shifting and the U.S. Tax Code - Part 1 (Microsoft and Hewlett Packard), Hearings, September 20, 2012. 112th Cong. 2nd sess. Washington: GPO.

[4]See United States Congress Senate Committee on Homeland Security and Government Affairs. 2013. Offshore Profit Shifting and the U.S. Tax Code - Part 2 (Apple). Hearings, May 21, 2013. 113th Cong. 1st sess. Washington: GPO.

[5]Similarly, the partial owner might engage in tunneling to protect itself from bankruptcy.

experiment and a shock large and unexpected enough to induce propping. They analyze the effect of debt and corporate governance on firm-level performance by applying difference-in-difference analysis and find evidence for propping, especially pronounced in specific ownership structures, such as pyramids.[6]

Our analysis shows that propping might also facilitate customer foreclosure in the case of partial backward ownership.

4.3 Input foreclosure incentives with partial ownership

4.3.1 Model framework

In this section, we consider a setting with one upstream firm U and two symmetric downstream firms, D_1 and D_2, as shown in Figure 4.1. The upstream firm can sell each downstream firm one unit of the input at prices f_1 and f_2.

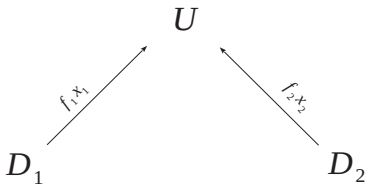


Figure 4.1: Market structure: input foreclosure setup.

The profit of U is

$$\pi^U = f_1 x_1 + f_2 x_2,$$

where $x_i \in \{0, 1\}$ denotes the input sales to firm i. One can interpret an input sale ($x_i = 1$) in several ways. First, one can think of a machine that the downstream firm can use to produce the output. Second, one can think of a per-unit input sold at marginal

[6] In a pyramidal ownership structure, several firms form a business group. This business group is a top-down chain of companies usually controlled by the ultimate shareholder who may only owe a small part of firms located in the lower levels of the pyramidal structure but can control it fully (Riyanto and Toolsema (2008)).

costs and an upfront fee (as may be the case with secret contracting, see Hart and J. Tirole (1990)). We follow Levy et al. (2018) and denote the profit of the downstream firm i as

$$\pi_i = \pi(x_i, x_{-i}) - f_i x_i,$$

where $\pi(x_i, x_{-i})$ is the downstream flow profit before input costs. We allow for the case that a firm cannot make a positive profit without the input. In any case, a firm can produce the output in a more competitive way with the input from U (cheaper or at a higher quality):

Assumption 1. $\pi(1, x_{-i}) > \pi(0, x_{-i})$,

Moreover, a firm's profit decreases if its rival has obtained a unit of input because this intensifies competition:

Assumption 2. $\pi(x_i, 1) \leq \pi(x_i, 0)$, *with the latter holding strictly at least for* $x_i = 1$.

We study the cases of vertical separation, a full merger between U and D_1, and partial vertical ownership where D_1 owns a share $\alpha \in (0; 1)$ of U and can influence the strategy of U to some degree (we explain the restrictions below). For a given ownership structure:

1. Upstream firm U sets input prices f_1 and f_2.
2. Each downstream firm $D_i, i \in \{1; 2\}$, chooses whether to purchase the input ($x_i \in \{0; 1\}$) and then sells its output.

For the following analysis of tunneling, we use for reference the *"market price"* f^*. To allow for different levels of bargaining power, we let the market price have any level in the interval $[\underline{f}, \overline{f}]$. The lower bound $\underline{f}$ is the reservation value of U, which equals its marginal costs of 0, and the upper bound equals the willingness-to-pay of each D_i under vertical separation. It is defined as the maximal price that U can charge each firm, which is equal to the incremental profit from the input, given the other downstream firm also uses the input:

$$\overline{f} = \pi(1, 1) - \pi(0, 1). \quad \textit{(take-it-or-leave-it price)} \tag{4.1}$$

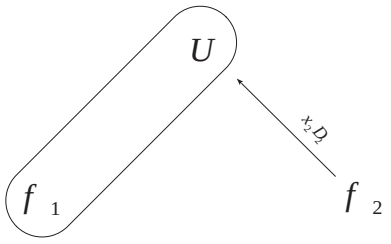


Figure 4.2: Full integration: input foreclosure setup

Definition 1. In the present setting, input foreclosure refers to a situation where U does not sell input to D_2. This implies $x_2 = 0$.

Benchmark: full vertical integration. Full integration between U and D_1 is our benchmark in the subsequent sections where we show that the input foreclosure incentives of partial ownership depend crucially on how we model the restrictions on tunneling and transfer prices (see Figure 4.2). The joint profit of U and D_1 is

$$\pi_{D1}^{U} = \pi(x_1, x_2) + f_2 x_2. \tag{4.2}$$

To start, let us establish

Lemma 1. *It is always optimal for the integrated unit of U and D_1 to supply its downstream business with the input.*

Proof. See Appendix. □

It is optimal for the integrated entity to supply both downstream firms if the joint profit when doing so exceeds the joint profits under foreclosure:

$$\pi(1,1) + f^* \geq \pi(1,0) \tag{4.3}$$

$$\implies f^* \geq \pi(1,0) - \pi(1,1). \tag{4.4}$$

We refer to (4.4) as *"non-foreclosure condition under vertical integration"* in this section.

4.3.2 Partial backward ownership

This section focuses on the case that D_1 has partial ownership of U, as shown in Figure 4.3. This partial ownership entitles D_1 to a share $\alpha \in (0,1)$ of U's profits, which yields for D_1 a total profit of

$$\pi_{D1} = \pi(x_1, x_2) - f_1 x_1 + \alpha(\underbrace{f_1 x_1 + f_2 x_2}_{\pi^U}). \tag{4.5}$$

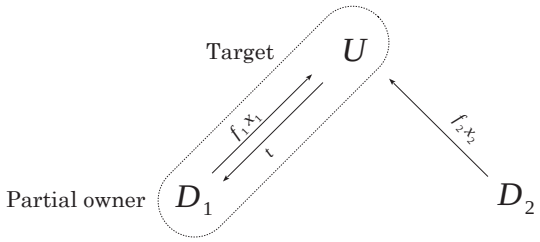


Figure 4.3: Partial backward ownership: D_1 owns stake of U

In line with Levy et al. (2018), we assume that the ownership arrangement allows D_1 to exert control over the strategy of U, subject to different restrictions, which we introduce below. The strategy of U essentially consists of setting the input prices f_1 and f_2 for the two downstream firms.

Firm D_1 can, if the restrictions allow so, use its control to require such a high input price from D_2 that D_2 does not buy the input (input foreclosure). Any price above $\overline{f}$ achieves this, for instance, $f_2 \to \infty$.

As regards the own input price f_1, the partial owner D_1 can generally demand a price that differs from the market price f^*. We speak of *tunneling* in the case of a lower input price ($f_1 < f^*$), whereas we speak of *negative tunneling* or *propping* in the case of a higher input price ($f_1 > f^*$). We denote by t the amount that D_1 tunnels out of U:

$$t = f^* - f_1. \tag{4.6}$$

The profit of supplier U is

$$\pi^U = f_1 x_1 + f_2 x_2 = (f^* - t)\, x_1 + f_2 x_2.$$

In what follows, we focus on the natural case that D_1 never forecloses itself, which means $x_1 = 1$. We can write the profit of D_1 as

$$\pi_{D1} = \pi(1, x_2) - f^* + t + \alpha(\underbrace{f^* - t + f_2 x_2}_{\pi^U}). \tag{4.7}$$

We now present alternative restrictions on tunneling and compare how these restrictions affect the foreclosure incentives. We focus on studying restrictions on the amount to tunnel and on the minimal upstream profit. Both types of restrictions can naturally result from rules that aim at protecting minority shareholders of the upstream firm. This protection might require profits to reach at least the minimum threshold to be satisfied or restrict the amount of money to be transferred downstream. In some cases, however, it might be optimal for the partial owner D_1 to prop U, i.e., to transfer profits upstream. In this case, the minority shareholder protection of the downstream firm can play a role. They can also either restrict the minimal amount of D_1's profits to be left in the firm or the amount of money that can be transferred upstream.

Remark (Potential channels for tunneling in practice.). Although we model tunneling as an adjustment of the input price of D_1, our results also extend to the case that tunneling does not take place through the input price. In general, tunneling could take other forms than through a reduced input price for D_1. For instance, transfer price regulations may put limits on the deviations of the input price for D_1 from the market price that would prevail absent ownership (e.g. "the input price cannot differ more than 5% from f^*"). It might necessitate other forms of tunneling. A very crude way of tunneling would be that the partial owner D_1 physically takes cash out of U.

Tunneling Restriction 1: exogenous limit on the tunneling amount: $t \leq \bar{t}$ (as in Levy et al. (2018)). Following Levy et al. (2018), we assume that tunneling from U to D_1 is limited to an exogenous amount of $\bar{t}$, which yields the restriction $t \leq \bar{t}$. Intuitively,

we expect the limit $\bar{t}$ to be higher if the protection of minority shareholders is weaker: the less the minority shareholders are protected, the easier it should get for the controlling shareholder to shift the profits out of the firm. Similarly, $\bar{t}$ should be higher if the transfer price regulation is weaker.

Lemma 2. *Under the restriction on the absolute tunneling amount, the partial owner D_1 has strictly higher incentives to foreclose its rival than in the case of full integration.*

Proof. The partial owner D_1 is not able to tunnel all profits, neither with nor without foreclosure. This means that D_1 can shift up to $\bar{t}$ out of the upstream firm independent of whether it supplies D_2 or not. Substituting $t = \bar{t}$ in the profit of D_1 yields

$$\pi_{D1}^{F} = \pi(1,0) - f^* + \bar{t} + \alpha\left(f^* - \bar{t}\right) \tag{4.8}$$

in the case of foreclosure, and

$$\pi_{D1}^{S} = \pi(1,1) - f^* + \bar{t} + \alpha\left(2f^* - \bar{t}\right) \tag{4.9}$$

when supplying D_2. Supplying is weakly more profitable than foreclosure if $\pi_{D1}^{S} \geq \pi_{D1}^{F}$, which implies

$$f^* \geq {}^{1}\!/_{\alpha}\left[\pi(1,0) - \pi(1,1)\right]. \tag{4.10}$$

Condition (4.10) implies that foreclosure is more profitable for D_1 than in the case of a vertical merger because $\alpha < 1$. □

For a given tunneling restriction, foreclosure is more profitable when the profit share α from partial ownership is smaller. This condition is similar to the foreclosure incentive condition in Levy et al. (2018) as they assume an exogenous limit on tunneling and restrict the amount of tunneling to be smaller than the downstream gains and upstream losses from not supplying to D_2.[7]

[7] Their assumption A5 reads $t \leq min\{G, L\}$. The assumption implies that the amount to tunnel should not exceed the minimum of downstream gains and upstream losses from foreclosure: authors define the difference between downstream profits with and without foreclosure as G (gains) and the respective difference between upstream profits as L (losses).

Tunneling Restriction 2: minimal upstream profit ($\pi^U \geq \underline{\pi}^U$). Instead of restricting the amount that the downstream firm can tunnel ($t \leq \bar{t}$), one can impose a lower limit $\underline{\pi}^U$ on the profits that need to be left in the upstream firm. Intuitively, the supplier must have at least a certain profit level ($\underline{\pi}^U$), such that the other shareholders (or stakeholders) of the upstream firms do not become suspicious or too unsatisfied. For instance, one can imagine that, in case of a profit level below $\underline{\pi}^U$, these other parties would be able to sue D_1 successfully. So, D_1 needs to leave at least this amount of profit with U. The amount $\underline{\pi}^U$ could be an industry benchmark that provides an indication of what profit to expect under normal circumstances.

We restrict $\underline{\pi}^U$ to the natural upper bound of $2f^*$ because $\underline{\pi}^U > 2f^*$ would mean that $U's$ profits need to be higher than the highest profit achievable at market prices absent vertical ownership.

Assumption 3. $\underline{\pi}^U \leq 2f^*$.

At first sight, it might seem that the restriction on the amount that can be taken out of the target firm (Restriction 1) and the restriction on the amount that must be left in the target firm (Restriction 2) are equivalent. For instance, if the target's profit is 100, one can either specify that at most 20 can be taken out ($t \leq 20$) or that 80 need to be left ($\pi^U \geq 80$). However, we will show below that the foreclosure incentives differ substantially.

In the present case, the tunneling restriction

$$\pi^U \geq \underline{\pi}^U$$

can be written as

$$f^* - t + f_2 x_2 \geq \underline{\pi}^U. \tag{4.11}$$

The restriction implies a maximal tunneling amount of

$$t = f^* + f^* x_2 - \underline{\pi}^U.$$

Assumption 3 implies that the tunneling amount is non-negative if U supplies both downstream firms with input.

Lemma 3. *Under the tunneling restriction of a minimal profit that needs to be left in the upstream firm, the partial owner D_1 has the same incentive to foreclose its downstream rival as under vertical integration.*

Proof. Substituting for t in the profit of D_1 yields

$$\pi_{D1} = \pi(1, x_2) - f^* + \underbrace{\left(f^* + f^* x_2 - \underline{\pi}^U\right)}_{t} + \alpha \underline{\pi}^U, \tag{4.12}$$

and equivalently

$$\pi_{D1} = \pi(1, x_2) + f^* x_2 - (1 - \alpha)\underline{\pi}^U. \tag{4.13}$$

D_1 prefers to supply D_2 if the resulting profits are higher than the profits in the case of foreclosure:

$$\pi(1, 1) + f^* - (1 - \alpha)\underline{\pi}^U \geq \pi(1, 0) - (1 - \alpha)\underline{\pi}^U,$$

Adding $(1 - \alpha)\underline{\pi}^U$ on both sides yields

$$f^* \geq \pi(1, 0) - \pi(1, 1). \tag{4.14}$$

This is the same condition as under full vertical integration (Equation (4.4)). Firm D_1 has the same foreclosure incentives as when U and D_1 are fully integrated. □

The foreclosure condition does not depend on the degree of minority shareholder protection and the share α. This is different from the foreclosure condition (4.10) that we obtained when restricting the amount that D_1 can tunnel with the condition $t \leq \bar{t}$. The latter condition is also the relevant foreclosure condition of Levy et al. (2018) for their partial (backward) ownership case.

Propping and foreclosure. Without profit shifting ($t = 0$), the minimum profit condition (4.11) in the case of foreclosure ($x_2 = 0$) becomes $\underline{\pi}^U > f^*$. To ensure the minimum

profit of U, D_1 would need to engage in negative tunneling ($t < 0$, "*propping*") in the case of foreclosure. Therefore, we specifically analyze the case when $\underline{\pi}^U$ is in the interval $(f^*; 2f^*]$.[8] It is a subset of the cases considered under Lemma 3.

Lemma 4. *If foreclosure is more profitable than supplying D_2 (Condition 4.14 does not hold) and the minimal profit that needs to be left in the upstream firm is relatively large ($\underline{\pi}^U > f^*$), the partial owner D_1 optimally props U to foreclose D_2 by shifting an amount of $\underline{\pi}^U - f^*$ to the target firm.*

Proof. We have shown in the proof of Lemma 3 that foreclosure is profitable in case of the minimal profit restriction under the same condition as under vertical integration (see Equation (4.4)), that is:

$$\pi(1,0) > \pi(1,1) + f^*.$$

Propping is equivalent to $t < 0$ and occurs as part of the foreclosure strategy when the above condition holds and, in addition, $\underline{\pi}^U > f^*$.

To see this, note that in the absence of profit shifting and thus propping ($t = 0$), U supplying both downstream firms at market prices fulfills the restriction $\pi^U \geq \underline{\pi}^U$ as $\underline{\pi}^U \in (f^*; 2f^*]$ and the profit π^U then equals $2f^*$.

Instead, foreclosure of D_2 does not satisfy $\pi^U \geq \underline{\pi}^U$ as the profit π^U then equals f^* and $\underline{\pi}^U > f^*$ by construction of this case. In order so satisfy the minimal profit restriction of U, D_1 must shift profits to U, such that $\pi^U = f^* + t \geq \underline{\pi}^U$. The lowest transfer which satisfies this is given by $\underline{\pi}^U - f^*$, which implies

$$t = f^* - \underline{\pi}^U < 0.$$

which is negative by construction as $\underline{\pi}^U > f^*$.

Therefore, if foreclosure is profitable for D_1, the partial owner will prop U to ensure that its profit level is not below $\underline{\pi}^U$. □

[8] The upper bound of the interval is determined by Assumption 3.

If propping is restricted or not possible, foreclosure may not be feasible with partial ownership, although it would be profitable. For example, suppose that $f^* = 50$, $\underline{\pi}^U = 60$, $\pi(1,1) = 100$, $\pi(1,0) = 200$. Hence, absent foreclosure, $U's$ profit equals

$$2f^* - t = 100 - t \geq \underline{\pi}^U = 60,$$

which implies that D_1 optimally tunnels an amount of $t = 40$ in this case and obtains a profit of

$$\pi(1,1) - f^* + t = 100 - 50 + 40 = 90.$$

With foreclosure, the profit of U becomes

$$f^* - t = 50 - t \geq \underline{\pi}^U = 60,$$

which implies an optimal amount of profit shifting of $t = -10$ and yields a profit for D_1 of

$$\pi(1,0) - f^* + t = 200 - 50 - 10 = 140.$$

Foreclosure is only feasible with propping ($t \leq -10$) and turns out to be profitable for D_1 at $t = -10$ because its foreclosure profit is 140 and thus larger than the profit of 90 absent foreclosure. See Table 4.1 for a summary.

Note that if propping were not possible (which corresponds to $t \geq 0$), then there would not be foreclosure, and D_1 would earn the profit of 90.

Corollary 1. *Foreclosure of the downstream rival does not occur with partial backward ownership in situations where it would occur with a full vertical merger if the target firm's minimum profit level is above the profit obtainable with foreclosure ($\underline{\pi}^U > f^*$) and profit shifting into the target firm (propping) is not feasible at all, or not to the required extent (this corresponds to the restriction $t > \underline{\pi}^U - f^*$).*

This corollary sheds new light on the foreclosure effects of partial vertical ownership: Restrictions on the money a partial owner can prop into the target firm as part of a fore-

	Profit of target firm U	Profit of partial owner D_1
No foreclosure	$\pi^U = 2f^* - t = 100 - t = 60$ $\implies t = 40$	$\pi_{D1} = 100 - f^* + t = 90$
Foreclosure with propping	$\pi^U = f^* - t = 50 - t \geq \underline{\pi}^U = 60$ $\implies t = -10$	$\pi_{D1} = 200 - f^* + t = 140$

Table 4.1: Example with propping in the case of foreclosure where $f^* = 50$, $\underline{\pi}^U = 60$, $\pi(1,1) = 100$, $\pi(1,0) = 200$.

closure strategy may render foreclosure impossible. Even if the vertically related partial owner has full control over the target firm and seemingly more incentives to foreclosure than in the case of a full vertical merger (as argued by Levy et al. (2018)), foreclosure may nevertheless not occur, although it would have occurred with a merger. As propping is a form of expropriation, strong enough minority shareholder protection might assure it is not unlimited. Additionally, transfer price regulations may also limit the scope for propping.

The next proposition summarizes the results on the input foreclosure incentives with partial backward ownership of the Lemmas 2, 3, and 4.

Proposition 1. *Relative to full vertical integration, partial backward ownership (PBO) tends to affect the incentives for input foreclosure in the following ways:*

1. *PBO increases the foreclosure incentives if the absolute amount of tunneling is effectively restricted (Lemma 2);*

2. *PBO has the same effect as full vertical integration if tunneling is restricted by a minimum profit that needs to be left in the target firm, provided that propping is unrestricted (Lemma 3);*

3. *The foreclosure incentives tend to be lower with PBO if tunneling is restricted by a minimum profit that needs to be left in the target firm and if propping is restricted*

as well (Lemma 4).

4.3.3 Partial forward ownership

For the industry structure with one upstream and two downstream firms, we now consider the case where U owns a share $\alpha \in (0,1)$ of $D_1's$ profits. The market structure is shown in Figure 4.4. The partial owner U can exert full control over its target's strategy, subject to tunneling restrictions.

As the derivations are similar to the case of partial backward ownership in the previous section, we present the detailed analysis in the Appendix and only summarize and discuss the result in this section.

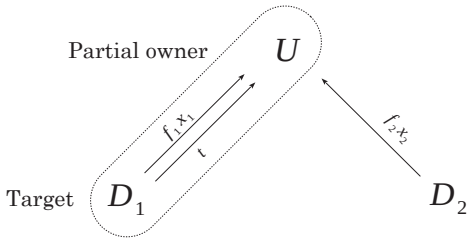


Figure 4.4: Partial forward ownership: U owns stake of D_1

Proposition 2. *Relative to full vertical integration, partial forward ownership (PFO) tends to affect the incentives for input foreclosure in the following ways:*

1. *PFO decreases the foreclosure incentives if the absolute amount of tunneling is effectively restricted (Lemma 8);*

2. *PFO has the same effect as full vertical integration if tunneling is restricted by a minimum profit that needs to be left in the target firm, provided that propping is unrestricted (Lemma 9);*

Proof. See the Appendix for the lemmas and their proofs. □

The intuition for result 1 of the proposition is that the partial owner U internalizes additional upstream profits more than additional downstream profits of D_1. Consequently, it has fewer incentives to foreclose than under full integration where both profits have the

same value. This is in line with Levy et al. (2018). Result 2 is analogous to the result in Proposition 1.

Note that propping is not an issue here as foreclosure requires an upstream action from the partial owner but not from the downstream target and we assume that the owner maximizes its own profit without minority shareholder restrictions within its own entity.

4.4 Customer foreclosure with partial ownership

4.4.1 Model framework

We now study the case of customer foreclosure: An upstream firm being prevented from selling its products. For this, we consider a setting with two symmetric upstream firms, U_1 and U_2, and a downstream monopolist D, as shown in Figure 4.5. We assume that

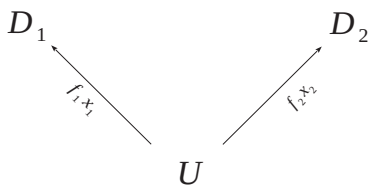


Figure 4.5: Market structure: customer foreclosure setup

the upstream firms produce differentiated input goods. Downstream firm D can use at most two units of input. Those two units can be purchased from a single upstream firm or each input unit from each firm.

Definition 2. In the present setting, customer foreclosure refers to a situation where D buys no input from U_2 and two units of input from U_1.

We further assume that the downstream firm's flow profits before input costs are higher when the input units are differentiated. In particular, we assume

$$\Pi(1,1) > \Pi(2,0) > \Pi(1,0) > \Pi(0,0) = 0, \tag{4.15}$$

where $\Pi(x_1, x_2)$ is the downstream flow profit as a function of the input quantities x_1 and x_2 from U_1 and U_2, respectively.[9] We assume that both upstream firms produce at zero costs.[10] These assumption lead to the natural benchmark where, under vertical separation, D finds it optimal to buy the input from both upstream firms.

Upstream firm $j \in \{1, 2\}$ sells at a unit price of f_j. The profit of upstream firm j when selling one unit is thus

$$\pi^{Uj} = x_j \cdot f_j = 1 \cdot f_j. \tag{4.16}$$

The minimal price at which an upstream firm could sell without making a loss is equal to the cost of producing the input:

$$\underline{f} = 0. \tag{4.17}$$

Such a price might arise if the downstream firm has all the bargaining power.

Lemma 5. *The maximal price at which the downstream firm is best off buying one unit from each upstream firm is*

$$\overline{f} = \min\left[\Pi(1,1) - \Pi(1,0), \Pi(1,1)/2\right]. \tag{4.18}$$

Proof. The downstream firm buys one unit from each upstream firm if the following three requirements hold:

$$\begin{aligned}
&\Pi(1,1) - 2\overline{f} \geq \Pi(2,0) - 2\overline{f} \ \ (i),\\
&\Pi(1,1) - 2\overline{f} \geq \Pi(1,0) - \overline{f} \ \ (ii),\\
&\Pi(1,1) - 2\overline{f} \geq 0 \qquad\quad (iii).
\end{aligned}$$

The first requirement holds by the assumption that $\Pi(1,1) > \Pi(2,0)$.

[9] For homogeneous products (and no non-linear transaction costs, etc.), the first inequality would hold with equality.

[10] We consider zero production costs for the sake of simplicity and comparability to the setup of Section 4.3.1. Our model yields conceptually identical predictions if a firm's production costs are non-decreasing in the number of units produced.

The second requirement implies

$$\begin{aligned}\Pi(1,1) - \overline{f} &\geq \Pi(1,0) \\ \implies \overline{f} &\leq \Pi(1,1) - \Pi(1,0).\end{aligned}$$

Suppose that $\overline{f} = \Pi(1,1) - \Pi(1,0)$. Does this satisfy the third requirement? Substituting in (iii) yields

$$\begin{aligned}\Pi(1,1) - 2\left(\Pi(1,1) - \Pi(1,0)\right) &\geq 0 \\ 2\Pi(1,0) &\geq \Pi(1,1).\end{aligned}$$

The latter inequality should hold for substitutes on the demand side and no costs. It might not hold in the case of economies of scale (e.g. fixed costs that arise once selling products).

In general, the largest price that satisfies all three requirements is

$$\overline{f} = \min\left[\Pi(1,1) - \Pi(1,0), \Pi(1,1)/2\right].$$

□

In the following we use a general *"market price"* f^*, which we restrict to be in the interval $[\underline{f}, \overline{f}]$. For reference, let us describe prices which may arise when the upstream firms non-cooperatively and simultaneously set their prices.

Lemma 6. *When the upstream firms non-cooperatively and simultaneously set their prices, a symmetric price of* $\overline{f}$ *is an equilibrium if product differentiation, measured as the difference* $\Pi(1,1) - \Pi(2,0)$*, is large enough.*

Proof. See Appendix. □

Benchmark: full vertical integration. Full integration between U_1 and D is our benchmark in the subsequent sections where we show that the customer foreclosure incen-

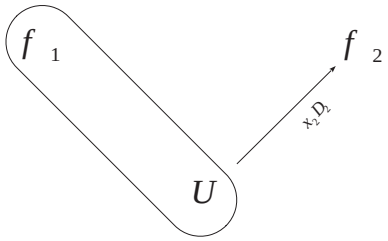


Figure 4.6: Full integration: customer foreclosure setup

tives of partial ownership depend crucially on how we model the restrictions on tunneling and transfer prices (see Figure 4.6).

The joint profit of U and D is

$$\pi_I^S = \Pi(1,1) - f^*$$

when the inputs of both upstream firms are used, and

$$\pi_I^F = \Pi(2,0)$$

in the case where upstream firm 2 is foreclosed. The integrated entity decides to source from U_2 if $\pi_I^S \geq \pi_I^F$,which is equivalent to

$$\Pi(1,1) - f^* \geq \Pi(2,0)$$

$$\implies f^* \leq \Pi(1,1) - \Pi(2,0). \tag{4.19}$$

We refer to equation (4.19) as *the "non-foreclosure condition under vertical integration"*. As $f^* \in [\underline{f}, \overline{f}]$, a necessary condition for foreclosure to arise is that $\overline{f} > \Pi(1,1) - \Pi(2,0)$.

Lemma 7. *The highest feasible input price $\overline{f}$ is larger than the incremental profit of dual sourcing, $\Pi(1,1) - \Pi(2,0)$, if $2 \cdot \Pi(2,0) > \Pi(1,1)$.*

Proof. See Appendix. □

Note that the requirement $2\Pi(2,0) > \Pi(1,1)$ in Lemma 7 is fulfilled in many plausible

cases. In general, it holds if the inputs of the upstream firms are similar enough. Moreover, it may also hold with strong substitutes. An exceptional case, where the condition might not hold, would be when it is not profitable to sell both units of the same kind, such that essentially $\Pi(2,0) = \Pi(1,0)$ and if there are fixed costs of selling products, such that $2 \cdot \Pi(1,0)$ would be smaller than $\Pi(1,1)$.

Corollary 2. *Together, lemmas 5, 6 and 7 imply that the competitive input price may well be at the level $\overline{f}$ where foreclosure of U_2 is jointly profitable for U_1 and D when they are vertically integrated.*

4.4.2 Partial forward ownership

As regards customer foreclosure, the partial forward ownership is the more interesting case. Suppose that U_1 owns a share $\alpha \in (0,1)$ of D's profits. The partial owner U_1 can exert full control over its target's strategy, subject to tunneling restrictions. See Figure 4.7 for on illustration. Our results under these assumptions are summarized in Proposition

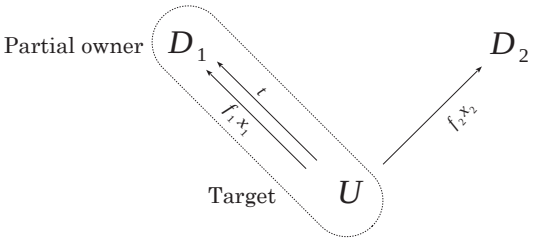


Figure 4.7: Partial forward ownership: U_1 owns a stake of D

3.

Proposition 3. *Relative to full vertical integration, partial forward ownership (PFO) tends to affect the incentives for customer foreclosure in the following ways:*

1. *PFO increases the foreclosure incentives if the absolute amount of tunneling is effectively restricted (Lemma 10);*

2. *PFO has the same effect as full vertical integration if tunneling is restricted by a minimum profit that needs to be left in the target firm ($\pi_D \geq \underline{\pi}_D$), provided that propping is unrestricted (Lemma 11);*

3. *The foreclosure incentives tend to be lower with PFO if tunneling is restricted by a minimum profit that needs to be left in the target firm and if propping is restricted as well (Lemma 12).*

Proof. See the Appendix for the lemmas and their proofs. □

The mechanism for result 1 of the proposition is analog to the case of input foreclosure and PBO in Proposition 1. When the partial ownership values own profits more than the target's profits, then commanding a foreclosure action that hurts the target is more profitable than under full integration where both profits have the same value.

With the minimal profit restriction, the partial owner becomes the claimant of the full incremental profits of the target and thus has the same foreclosure incentives as under full integration (result 2). However, when the partial owner has to ensure a higher profit of the target D than would arise under foreclosure ($\underline{\pi}_D > \Pi(2,0) - 2f^*$) but propping is not possible, foreclosure is harder than under full integration (result 3). This result is relevant as the competitive input price may well be at the level $\overline{f}$ where foreclosure of U_2 is jointly profitable for U_1 and D (Corollary 2).

4.4.3 Partial backward ownership

Downstream firm D owns a share $\alpha \in (0,1)$ of $U_1's$ profits. The partial owner D can exert full control over its target's strategy, subject to tunneling restrictions (see details on the market structure in Figure 4.8).

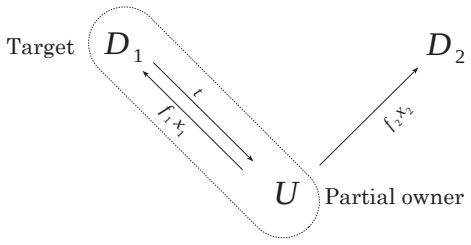


Figure 4.8: Partial backward ownership: D owns stake of U_1

Absent foreclosure and absent tunneling ($t = 0$), the profit of each upstream firm equals f^*. With customer foreclosure of U_2 and absent tunneling ($t = 0$), the profit of U_1

equals $2f^*$ whereas the profit of U_2 equals 0.

Analog to Assumption 3, we assume that the minimal profit $\underline{\pi}_{U1}$ is not larger than the equilibrium profit of the upstream firm under vertical separation (see Equation (4.16)).

We summarize D's incentives to foreclose U_2 subject to different tunneling restrictions in Proposition 4.

Proposition 4. *Relative to full vertical integration, partial backward ownership (PBO) tends to affect the incentives for customer foreclosure in the following ways:*

1. *PBO decreases the foreclosure incentives if the absolute amount of tunneling is effectively restricted (Lemma 13);*

2. *PBO has the same effect as full vertical integration if tunneling is restricted by a minimum profit that needs to be left in the target firm, provided that propping is unrestricted (Lemma 14).*

Proof. See the Appendix for the lemmas and their proofs. □

The intuition for result 1 of the proposition is that when the partial owner D internalizes additional downstream profits more than additional upstream profits of U_1, there is less incentive than under full integration to sacrifice downstream profits to the benefit of upstream profits.

Note that propping is not an issue here as foreclosure requires a downstream action from the partial owner but not from the upstream target and we assume that the owner maximizes its own profit without minority shareholder restrictions within its own entity.

4.5 Discussion

4.5.1 Overview of results

For Restriction 1 on the amount that a partial owner can tunnel, our results are in line with the existing literature (Baumol and Ordover, 1994; Spiegel, 2013; Levy et al., 2018). Compared to full integration, partial backward ownership leads to higher input foreclosure

incentives than full integration but lower customer foreclosure incentives. Partial forward ownership has the opposite effects. See Table 4.2 for an overview of our main results.

We add to this the insight that the restriction on the minimal profit leads to the same foreclosure incentives as full integration. The reason is that the partial owner becomes a residual claimant of the joint profits – which implies the same incentives as full integration.

When the minimal profit that needs to be left in the target firm is higher than the profit obtainable in the case a foreclosure strategy is in place, the latter equivalence result relies on the assumption that propping is feasible. Propping means that the partial owner can shift funds into the target firm. The partial owner may need to prop to induce the target firm to foreclosure a rival of the owner. A foreclosure action, which may be profitable for the partial owner, can reduce the target's profit below the critical level, such that propping may be necessary for foreclosure to be feasible. When propping is not feasible, the foreclosure incentives are eliminated under the minimal profit restriction and, thus, can be lower than with full integration.

A key distinction between Restriction 1 on the tunneling amount and Restriction 2 on the minimal profit of the target firm is whether or not propping might occur. Intuitively, Restriction 2 sets a target profit level that the partial owner has to assure, which means that if this target profit level is high enough, the partial owner cannot satisfy the restriction without additional transfers to the target firm. Under Restriction 1, the mechanism is different: The non-controlling shareholders of the target firm can only impose restrictions on how much value is tunneled out of the firm. Profit shifting into the target firm is thus not an issue when there is solely a restriction on the amount that can be tunneled out of the target firm. Of course, in a real-world case, several restrictions on tunneling can be in place simultaneously, including the restrictions 1 and 2 that we study. Indeed, a restriction on propping is essentially a restriction on negative tunneling. Table 4.2 summarizes our results.

Table 4.2: Overview of results

Input foreclosure (not serving the downstream rival)

Benchmark – non-foreclosure condition with full integration: $f^* \geq \pi(1,0) - \pi(1,1)$

	Partial backward ownership	Partial forward ownership
Restriction 1: tunneling amount	$f^* \geq 1/\alpha\,[\pi(1,0) - \pi(1,1)]$ Higher incentives to foreclose than with full integration;	$f^* \geq \alpha\,[\pi(1,0) - \pi(1,1)]$ Lower incentives to foreclose than with full integration
Restriction 2: minimal profit	$f^* \geq \pi(1,0) - \pi(1,1)$ Same incentives to foreclose as with full integration; Propping needed if $\underline{\pi}^U > f^*$.	$f^* \geq \pi(1,0) - \pi(1,1)$ Same incentives to foreclose as with full integration; No propping needed.$^+$

Customer foreclosure (not buying rival's input)

Benchmark – foreclosure condition with full integration: $f^* \leq [\Pi(1,1) - \Pi(2,0)]$

	Partial backward ownership	Partial forward ownership
Restriction 1: tunneling amount	$f^* \leq 1/\alpha\,[\Pi(1,1) - \Pi(2,0)]$ Less incentives to foreclose than with full integration;	$f^* \leq \alpha\,[\Pi(1,1) - \Pi(2,0)]$ More incentives to foreclose than with full integration;
Restriction 2: minimal profit	$f^* \leq [\Pi(1,1) - \Pi(2,0)]$ Same incentives to foreclose as with full integration; No propping needed.$^+$	$f^* \leq [\Pi(1,1) - \Pi(2,0)]$ Same incentives to foreclose as with full integration; Propping needed if $\underline{\pi}_D > \Pi(2,0) - 2f^*$.

$^+$No propping is needed in the sense that foreclosure requires an action from the partial owner and we assume that the owner maximizes its own profit without minority shareholder restrictions within its own entity.

4.5.2 A review of the results in Levy et al. (2018)

Levy et al. (2018) base their analysis on comparing the downstream gains (G in their notation) and upstream losses (L) of foreclosing D_2. Our model is sufficient to replicate their findings and can naturally extend to their setting with N upstream suppliers. We can rearrange Condition (4.3) to show that the fully integrated entity chooses to supply D_2 if the downstream gains of foreclosure (G) do not exceed the foregone upstream profits

from supplying an additional retailer (L):

$$\underbrace{\pi(1,0)-\pi(1,1)}_{G} \leq \underbrace{\pi(1,1)-\pi(0,1)}_{L}.$$

What we call exogenous restriction on the tunneling amount, $t \leq \bar{t} < f^*$, corresponds to the case considered in Levy et al. Their Assumption 5 requires that the effect of tunneling on D_1's and U's payoffs is smaller than the effect of foreclosure, i.e., $t \leq \min\{G, L\}$. The partial owner has stronger incentives to foreclose its rival in comparison to the full integration case, namely, D_1 chooses to let U supply D_2 with an input if

$$\underbrace{\pi(1,0)-\pi(1,1)}_{G} \leq \underbrace{\alpha\left[\pi(1,1)-\pi(0,1)\right]}_{\alpha L}.$$

We argue that the way one specifies the restriction on tunneling plays a crucial role in shaping the incentives of the partial owner to foreclose its rival. By restricting the minimal profit which has to stay in the upstream firm (what we call Restriction 2) instead of imposing an exogenous limit on tunneling (what we call Restriction 1), the foreclosure condition becomes

$$\underbrace{\pi(1,0)-\pi(1,1)}_{G} \leq \underbrace{\pi(1,1)-\pi(0,1)}_{L}.$$

This condition is the same as it would have been for the full merger with U and is strictly lower than under an exogenous tunneling restriction.

Levy et al. (2018) implicitly assume that the tunneling amount t is non-negative.[11] We show in Corollary 1 that propping restrictions may eliminate the incentives to foreclose D_2 completely. If the minimal profit which has to stay in the upstream firm is large enough, i.e. $\underline{\pi}^U$ is in the interval $(\pi(1,1)-\pi(0,1); 2(\pi(1,1)-\pi(0,1))]$, and tunneling is restricted to be non-negative, it becomes impossible for the partial owner to foreclose its rival. Foreclosure is not feasible, although it could be profitable for the partial owner.

Therefore, the ability and incentives to foreclose depend crucially on the assumptions on the minority shareholder protection structure and the types of tunneling restrictions

[11] Levy et al. (2018) write on page 14: "D_1 pays for [U's] input the same amount it pays under non-integration, but minus a discount t if D_1 controls [U]".

minority shareholders may impose. As Levy et al. (2018) show, restrictions on the tunneling amount in partial backward ownership may increase the input foreclosure incentives compared to the full integration case. In this article, we show that other tunneling restrictions may leave the foreclosure incentives of partial vertical owners unchanged or even eliminate them.

4.6 Conclusion

We review the incentives of a firm that holds partial vertical ownership to foreclose rivals. The partial owner only obtains the part of its target's profits but it may substantially change its strategy and foreclosure incentives. We focus on the phenomena of tunneling and propping, that is shifting profits out of and into the target firm, and demonstrate how the different restrictions imposed on these activities alter the downstream firm's incentives to foreclose a rival. This phenomenon has, to our knowledge, so far received only limited and, arguably, insufficient attention in theoretical competition policy analyses.

We show that, depending on the type of tunneling, a partial owner's optimal strategy may vary between higher incentives to foreclose than under vertical integration (as discussed in Levy et al. (2018)), the same incentives (because of fully taking into account the target firm's residual profit) and no incentives at all (if propping is sufficiently restricted). We analyze the partial owner's foreclosure incentives for a variety of market environments.

For partial backward ownership, we find that the restriction on the maximal tunneling amount indeed increases the partial owner's incentives to foreclose its downstream rivals (input foreclosure) and decreases the incentives to foreclosure the rivals of the upstream target (customer foreclosure). This is in line with Levy et al. who exclusively use this kind of tunneling restriction. Interestingly, the alternative restriction on the minimal profit that needs to be left in the target firm yields the same customer and input foreclosure incentives as full integration. Additionally, the restriction on the minimal profit might necessitate propping money into the target firm in order to foreclose. If propping is not feasible at all, or not to a required extent, the partial backward owner faces lower incentives for input foreclosure compared to a full integration benchmark.

For partial forward ownership, the restriction on the tunneling amount decreases the incentives of the partial owner to foreclose its target's downstream rivals (input foreclosure) but increases the incentives to foreclose its own upstream rivals (customer foreclosure). This restriction follows the setup of Levy et al. and our results are in line with their findings as well. The minimal profit restriction, however, yields the same foreclosure incentives as full integration, provided that the partial owner can prop its target firm if the minimal profit level is relatively high. Additionally, if propping is not feasible at all, or not to a required extent, the partial forward owner has lower customer foreclosure incentives in comparison to a fully integrated firm.

In summary, the way tunneling is modeled can substantially affect the results of a foreclosure analysis in the case of partial vertical ownership. A precise understanding of the tunneling restrictions is thus crucial for a correct assessment of possible foreclosure incentives. Albeit, as our literature review reveals, tunneling is a common phenomenon, it so far appears to be less clear how one should precisely think of the restrictions on tunneling in a vertical relations framework. We have shed light from a theory perspective. It would be fruitful for future research to look more closely at different institutional contexts to provide guidance about what kind of tunneling restrictions are most relevant in practice.

Appendix

4.A Additional lemmas and proofs

4.A.1 Input foreclosure

Proof of Lemma 1. Suppose that the integrated entity can commit to not supplying itself (for instance, by setting a fee of $f_1 = \infty$ if that is public). The integrated entity's profit when not supplying itself becomes

$$\pi^U_{D1}(x_1 = 0, x_2 = 1) = \pi(0, 1) + f^*.$$

If the entity does not supply D_2, but only D_1, its joint profits are

$$\pi^U_{D1}(x_1 = 1, x_2 = 0) = \pi(1, 0).$$

It is weakly more profitable for the integrated unit to supply itself than only D_2 because

$$\begin{aligned}
&\pi^U_{D1}(x_1 = 0, x_2 = 1) \leq \pi^U_{D1}(x_1 = 1, x_2 = 0) \\
\iff &\pi(1, 0) \geq \pi(0, 1) + f^* \\
\iff &f^* \leq \pi(1, 0) - \pi(0, 1).
\end{aligned}$$

The latter condition holds due to Assumption (2) and Condition 4.1.

Moreover, if f_1 and f_2 are set secretly (downstream firm 1 does not see f_2 when accepting the contract and vice versa), the integrated unit simply cannot commit to not supplying itself. Thus, it cannot charge D_2 a transfer price above f^* in equilibrium as it would do better with charging a price at which the downstream firm buys the input. □

Forward ownership: lemmas for Proposition (2) and their proofs

Lemma 8. *Under the restriction on the absolute tunneling amount, the partial owner U has strictly lower incentives to foreclose its target's rival D_2 than in the case of a full integration.*

Proof. The upstream profits without and with foreclosure are

$$\pi_U^S = 2f^* + \bar{t} + \alpha\left(\pi(1,1) - f^* - \bar{t}\right),$$

$$\pi_U^F = f^* + \bar{t} + \alpha\left(\pi(1,0) - f^* - \bar{t}\right).$$

The upstream owner is better off when supplying D_2 if

$$\pi_U^S \geq \pi_U^F$$

$$\implies f^* \geq \alpha\left[\pi(1,0) - \pi(1,1)\right].$$

The foreclosure incentives for the upstream firm are lower than in the case of full integration (condition (4.3)). □

Lemma 9. *Under the tunneling restriction of a minimal profit that needs to be left in the upstream firm, the partial owner U has the same incentive to foreclose its target's downstream rival D_2 as under vertical integration.*

Proof. If both tunneling and propping are feasible, the downstream firm D_1 ends up with the profit of $\underline{\pi}_{D1}$ in any case, but the amount of tunneling , t^S and t^F, differ in general. The upstream profits are

$$\pi_U^S = 2f^* + \underbrace{\left(\pi(1,1) - f^* - \underline{\pi}_{D1}\right)}_{\mathrm{t}^S} + \alpha\underline{\pi}_{D1},$$

$$\pi_U^F = f^* + \underbrace{\left(\pi(1,0) - f^* - \underline{\pi}_{D1}\right)}_{\mathrm{t}^F} + \alpha\underline{\pi}_{D1}.$$

The upstream owner is better off when supplying D_2 if

$$\pi_U^S \geq \pi_U^F$$

$$\implies f^* \geq \pi(1,0) - \pi(1,1). \tag{4.20}$$

The foreclosure incentives are the same as in the full integration case. □

4.A.2 Customer foreclosure

Proof of Lemma 6. By construction, it is optimal at the price $\overline{f}$ for the downstream firm to source one unit from each downstream firm. Can an upstream firm deviate profitably? It could benefit from selling two units by lowering the price. Would is the largest deviation price p which leads to this outcome?

The price p needs to satisfy the following:

$$\Pi(2,0) - 2p \geq \Pi(1,1) - f^* - p \quad (i)$$
$$\Pi(2,0) - 2p \geq \Pi(1,0) - p \qquad (ii)$$
$$\Pi(2,0) - 2p \geq 0 \qquad\qquad (iii).$$

Case 1: Suppose that

$f^* = \min\left[\Pi(1,1) - \Pi(1,0), \Pi(1,1)/2\right] = \Pi(1,1) - \Pi(1,0)$

This corresponds to no economies of scale – substitute in isolation is better than selling them together:

$$\Pi(1,1) - \Pi(1,0) < \Pi(1,1)/2$$
$$\implies \Pi(1,1) < 2\Pi(1,0).$$

The first condition (i) from above becomes

$$\Pi(2,0) - 2p \geq \Pi(1,1) - \Pi(1,1) + \Pi(1,0) - p$$
$$\implies \Pi(2,0) - \Pi(1,0) \geq p.$$

This is equivalent to the second condition.

At $p = \Pi(2,0) - \Pi(1,0)$, the third condition holds as

$\Pi(2,0) - 2\Pi(2,0) + 2\Pi(1,0) = 2\Pi(1,0) - \Pi(2,0) > 2\Pi(1,0) - \Pi(1,1) > 0.$

Is such a price cut profitable? It is not if

$$\Pi(1,1) - \Pi(1,0) > 2[\Pi(2,0) - \Pi(1,0)]$$
$$\Longrightarrow \Pi(1,1) - \Pi(2,0) > \Pi(2,0) - \Pi(1,0)],$$

that is if the differentiation effect is larger than the quantity expansion effect.

Case 2: Suppose that

$f^* = \min[\Pi(1,1) - \Pi(1,0), \Pi(1,1)/2] = \Pi(1,1)/2.$

This corresponds to economies of scale: Selling more units but substitutes is better than selling each substitute in isolation:

$$\Pi(1,1) - \Pi(1,0) > \Pi(1,1)/2$$
$$\Longrightarrow \Pi(1,1) > 2\Pi(1,0). \tag{4.21}$$

The first condition (i) from above becomes

$$\Pi(2,0) - 2p \geq \Pi(1,1) - \Pi(1,1)/2 - p$$
$$\Longrightarrow \Pi(2,0) - \Pi(1,1)/2 \geq p.$$

Together with the second condition (ii) from above, the highest possible deviation price is

$$p = min\left[\Pi(2,0) - \Pi(1,1)/2, \Pi(2,0) - \Pi(1,0)\right].$$

The first argument of the minimum function is smaller as:

$$\Pi(2,0) - \Pi(1,1)/2 < \Pi(2,0) - \Pi(1,0)$$
$$\Longrightarrow \Pi(1,1) > 2\Pi(1,0),$$

which corresponds to condition (4.21) which constitutes this case. Hence the price has to satisfy $p \leq \Pi(2,0) - \Pi(1,1)/2$.

At the price $p = \Pi(2,0) - \Pi(1,1)/2$, the third condition (iii) holds:

$$\begin{aligned}\Pi(2,0) - 2p &= \Pi(2,0) - 2[\Pi(2,0) - \Pi(1,1)/2] \\ &= \Pi(1,1) - \Pi(2,0)] > 0.\end{aligned}$$

Is such a price cut profitable? It is NOT if

$$\begin{aligned}\Pi(1,1)/2 &> 2[\Pi(2,0) - \Pi(1,1)/2] \\ \implies \Pi(1,1) * 3/4 &> \Pi(2,0),\end{aligned}$$

that is if the differentiation effect is large enough. □

Proof of Lemma 7. Case 1: $\overline{f} = \min[\Pi(1,1) - \Pi(1,0), \Pi(1,1)/2] = \Pi(1,1) - \Pi(1,0)$.

$$\begin{aligned}\overline{f} &= \Pi(1,1) - \Pi(1,0) < \Pi(1,1) - \Pi(2,0) \\ \implies \Pi(2,0) &< \Pi(1,0).\end{aligned}$$

The latter condition contradicts the assumption in condition (4.15) whereby selling two units is more profitable than selling one.

Case 2: $\overline{f} = \min[\Pi(1,1) - \Pi(1,0), \Pi(1,1)/2] = \Pi(1,1)/2$

$$\begin{aligned}\overline{f} &= \Pi(1,1)/2 < \Pi(1,1) - \Pi(2,0) \\ \implies \Pi(2,0) &< \Pi(1,1)/2.\end{aligned}$$

The latter condition implies $\Pi(1,1) > 2\Pi(2,0) > 2\Pi(1,0)$, where the latter inequality follows from the assumption in condition (4.15) again. Case 2 arises under condition $\Pi(1,1) > 2\Pi(1,0)$ from Equation (4.21), which is implied by the previous condition

already. □

Forward ownership: lemmas for Proposition 3 and their proofs.

Lemma 10. *Under the restriction on the absolute tunneling amount ($t \leq \bar{t}$), the partial owner U_1 has strictly higher incentives to foreclose its rival than in the case of full integration.*

Proof. Partial owner U_1which owns a share α of its target's profits, may want D to source from both upstream competitors and get:

$$\pi_{U1}^{S} = f^* + \bar{t} + \alpha \left(\Pi(1,1) - 2f^* - \bar{t}\right),$$

or, alternatively, supply input to its downstream firm only by itself and obtain:

$$\pi_{U1}^{F} = 2f^* + \bar{t} + \alpha \left(\Pi(2,0) - 2f^* - \bar{t}\right).$$

D gets input from both downstream firms if

$$\pi_{U1}^{S} \geq \pi_{U1}^{F}$$

$$\implies f^* \leq \alpha \left[\Pi(1,1) - \Pi(2,0)\right].$$

Foreclosure is more profitable than under full integration because the partial owner U_1 puts relatively less weight on the downstream losses from foreclosure. □

Lemma 11. *Under the tunneling restriction of a minimal profit that needs to be left in the downstream firm ($\pi_D \geq \underline{\pi}_D$), the partial owner U_1 has the same incentive to foreclose its rival as under vertical integration.*

Proof. When minimal profit which has to be left in the downstream firms is restricted, U_1 gets the following profits if D sources from both upstream firms:

$$\pi_{U1}^{S} = f^* + \alpha\underline{\pi}_D + \underbrace{(\Pi(1,1) - 2f^* - \underline{\pi}_D)}_{t_{U1}^{S}},$$

or only from its partial owner:

$$\pi_{U1}^{F} = 2f^{*} + \alpha\underline{\pi}_{D} + \underbrace{\left(\Pi(2,0) - 2f^{*} - \underline{\pi}_{D}\right)}_{t_{U1}^{F}}.$$

D gets input from both downstream firms if

$$\pi_{U1}^{S} \geq \pi_{U1}^{F}$$

$$\implies f^{*} \leq \left[\Pi(1,1) - \Pi(2,0)\right]. \tag{4.22}$$

The condition is the same as in the full integration case. □

Lemma 12. *If sourcing from U_2 is less profitable than foreclosing it (condition 4.22 does not hold) and the minimal profit that needs to be left in the downstream firm is relatively large ($\underline{\pi}_D > \Pi(2,0) - 2f^*$), the partial owner U_1 optimally props D in order to foreclose U_2. If propping is not feasible, no foreclosure takes place in this case.*

Proof. Propping is needed if for foreclosure if the target firm's minimal profit restriction can only be met if input comes from both suppliers, i.e.,

$$\Pi(1,1) - 2f^{*} > \underline{\pi}_{D} > \Pi(2,0) - 2f^{*}.$$

As $\Pi(1,1) > \Pi(2,0)$, the above condition can be reduced to

$$\underline{\pi}_{D} > \Pi(2,0) - 2f^{*}.$$

Foreclosure of U_2 is profitable for the partial owner U_1 if

$$\Pi(2,0) > \Pi(1,1) - f^{*}.$$

Conversely, if propping is limited or impossible, the partial owner U_1 would want to foreclose U_2 but D has to source from it if $\underline{\pi}_D > \Pi(2,0) - 2f^*$. □

Backward ownership: lemmas for Proposition 4 and their proofs.

Lemma 13. *Under the restriction on the absolute tunneling amount, the partial owner D has strictly lower incentives to foreclose its target's rival than in the case of full integration.*

Proof. The partial owner D can choose to source from both upstream firms and obtain the following profits:

$$\pi_D^S = \Pi(1,1) - 2f^* + \bar{t} + \alpha\left(f^* - \bar{t}\right).$$

Alternatively, D may only obtain input from its target firm and get:

$$\pi_D^F = \Pi(2,0) - 2f^* + \bar{t} + \alpha\left(2f^* - \bar{t}\right).$$

The partial owner D sources from both upstream firms if

$$\pi_D^S \geq \pi_D^F$$

$$\implies f^* \leq {}^1\!/\!_\alpha\left[\Pi(1,1) - \Pi(2,0)\right]$$

The foreclosure condition is stricter than under full integration: The partial owner D is more affected from a downstream loss of customer foreclosure relative to the upstream gains and thus has fewer incentives to foreclose U_2 than under full integration. □

Lemma 14. *Under the tunneling restriction of a minimal profit that needs to be left in the upstream firm, the partial owner D has the same incentive to foreclose its target's downstream rival as under vertical integration.*

Proof. The downstream firm's profits when sourcing from either both or only one upstream firm are given by

$$\pi_D^S = \Pi(1,1) - 2f^* + \underbrace{\left(f^* - \underline{\pi}_{U1}\right)}_{t^S} + \alpha\underline{\pi}_{U1},$$

$$\pi_D^F = \Pi(2,0) - 2f^* + \underbrace{(2f^* - \underline{\pi}_{U1})}_{t^F} + \alpha\underline{\pi}_{U1}.$$

Partial owner D sources from both upstream firms if

$$\pi_D^S \geq \pi_D^F$$

$$\implies f^* \leq [\Pi(1,1) - \Pi(2,0)]. \tag{4.23}$$

The foreclosure incentives are the same as in the full integration case. □

Bibliography

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2019). "Preferences for Truth-Telling". *Econometrica* 87.4, pp. 1115–1153.

Alsop, Ron (2008). *The Trophy Kids Grow Up: How the Millennial Generation is Shaking Up the Workplace*. John Wiley & Sons.

Andreoni, James and B Douglas Bernheim (2009). "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects". *Econometrica* 77.5, pp. 1607–1636.

Aquino, Karl and Americus Reed II (2002). "The self-importance of moral identity." *Journal of Personality and Social Psychology* 83.6, p. 1423.

Araujo, Felipe A, Erin Carbone, Lynn Conell-Price, Marli W Dunietz, Ania Jaroszewicz, Rachel Landsman, Diego Lamé, Lise Vesterlund, Stephanie W Wang, and Alistair J Wilson (2016). "The slider task: an example of restricted inference on incentive effects". *Journal of the Economic Science Association* 2.1, pp. 1–12.

Ariely, Dan, Anat Bracha, and Stephan Meier (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially". *American Economic Review* 99.1, pp. 544–55.

Armstrong, Mark and John Vickers (2012). "Consumer Protection and Contingent Charges". *Journal of Economic Literature* 50.2, pp. 477–93.

Atanasov, Vladimir (2005). "How much value can blockholders tunnel? Evidence from the Bulgarian mass privatization auctions". *Journal of Financial Economics* 76.1, pp. 191–234.

Atanasov, Vladimir, Bernard Black, and Conrad S Ciccotello (2014). "Unbundling and measuring tunneling". *University of Illinois Law Review*, p. 1697.

Aurand, Timothy, Wayne Finley, Vijaykumar Krishnan, Ursula Sullivan, Jackson Abresch, Jordyn Bowen, Michael Rackauskas, Rage Thomas, and Jakob Willkomm (2018). "The VW Diesel Scandal: A Case of Corporate Commissioned Greenwashing". *Journal of Organizational Psychology* 18.1, pp. 23–32.

Austen-Smith, David and Roland G Fryer Jr (2005). "An economic analysis of "acting white"". *The Quarterly Journal of Economics* 120.2, pp. 551–583.

Baek, Jae-Seung, Jun-Koo Kang, and Inmoo Lee (2006). "Business groups and tunneling: Evidence from private securities offerings by Korean chaebols". *The Journal of Finance* 61.5, pp. 2415–2449.

Balafoutas, Loukas, Adrian Beck, Rudolf Kerschbamer, and Matthias Sutter (2013). "What drives taxi drivers? A field experiment on fraud in a market for credence goods". *The Review of Economic Studies* 80.3, pp. 876–891.

Balafoutas, Loukas, Rudolf Kerschbamer, and Matthias Sutter (2015). "Second-Degree Moral Hazard in a Real-World Credence Goods Market". *The Economic Journal* 127.599, pp. 1–18.

Barberis, Nicholas C (2013). "Thirty years of prospect theory in economics: A review and assessment". *Journal of Economic Perspectives* 27.1, pp. 173–96.

Baumol, W.J. and J.A. Ordover (1994). "On the Perils of Vertical Control by a Partial Owner of a Downstream Enterprise". *Revue d'Économie Industrielle* 69.1, pp. 7–20.

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak (1964). "Measuring utility by a single-response sequential method". *Behavioral Science* 9.3, pp. 226–232.

Bénabou, Roland and Jean Tirole (2002). "Self-confidence and personal motivation". *The Quarterly Journal of Economics* 117.3, pp. 871–915.

— (2006). "Incentives and prosocial behavior". *American Economic Review* 96.5, pp. 1652–1678.

Bertrand, Marianne, Paras Mehta, and Sendhil Mullainathan (2002). "Ferreting out tunneling: An application to Indian business groups". *The Quarterly Journal of Economics* 117.1, pp. 121–148.

Bleichrodt, Han, Jose Luis Pinto, and Peter P Wakker (2001). "Making descriptive use of prospect theory to improve the prescriptive use of expected utility". *Management Science* 47.11, pp. 1498–1514.

Booij, Adam S and Gijs Van de Kuilen (2009). "A parameter-free analysis of the utility of money for the general population under prospect theory". *Journal of Economic Psychology* 30.4, pp. 651–666.

Bordalo, Pedro, Nicola Gennaioli, Andrei Shleifer, et al. (2020). "Memory, Attention, and Choice". *The Quarterly Journal of Economics* 135.3, pp. 1399–1442.

Burks, Stephen V, Jeffrey P Carpenter, Lorenz Goette, and Aldo Rustichini (2013). "Overconfidence and social signalling". *The Review of Economic Studies* 80.3, pp. 949–983.

Bursztyn, Leonardo, Georgy Egorov, and Robert Jensen (2019). "Cool to be smart or smart to be cool? Understanding peer pressure in education". *The Review of Economic Studies* 86.4, pp. 1487–1526.

Bursztyn, Leonardo and Robert Jensen (2017). "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure". *Annual Review of Economics* 9, pp. 131–153.

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky (2022). "Measuring the welfare effects of shame and pride". *American Economic Review* 112.1, pp. 122–68.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik (2014). "Robust nonparametric confidence intervals for regression-discontinuity designs". *Econometrica* 82.6, pp. 2295–2326.

Camerer, Colin F (1998). *Prospect theory in the wild: Evidence from the field*. Tech. rep. California Institute of Technology.

Carpenter, Jeffrey and Caitlin Knowles Myers (2010). "Why volunteer? Evidence on the role of altruism, image, and incentives". *Journal of Public Economics* 94.11-12, pp. 911–920.

Castagnetti, Alessandro and Renke Schmacker (2022). "Protecting the ego: Motivated information selection and updating". *European Economic Review*, p. 104007.

Chen, Daniel L, Martin Schonger, and Chris Wickens (2016). "oTree—An open-source platform for laboratory, online, and field experiments". *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.

Chen, Si and Hannah Schildberg-Hörisch (2019). "Looking at the bright side: The motivational value of confidence". *European Economic Review* 120, p. 103302.

Chen, Y. (2001). "On Vertical Mergers and their Competitive Effects". *The RAND Journal of Economics* 32.4, pp. 667–685.

Cheung, Yan-Leung, P Raghavendra Rau, and Aris Stouraitis (2006). "Tunneling, propping, and expropriation: evidence from connected party transactions in Hong Kong". *Journal of Financial Economics* 82.2, pp. 343–386.

Corgnet, Brice, Roberto Hernán-González, and Eric Schniter (2015). "Why real leisure really matters: Incentive effects on real effort in the laboratory". *Experimental Economics* 18.2, pp. 284–301.

Coutts, Alexander (2019). "Good news and bad news are still news: Experimental evidence on belief updating". *Experimental Economics* 22.2, pp. 369–395.

Dai, Zhixin, Fabio Galeotti, and Marie Claire Villeval (2018). "Cheating in the lab predicts fraud in the field: An experiment in public transportation". *Management Science* 64.3, pp. 1081–1100.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang (2007). "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness". *Economic Theory* 33.1, pp. 67–80.

Danz, David, Lise Vesterlund, and Alistair J Wilson (2020). *Belief elicitation: Limiting truth telling with information on incentives*. Tech. rep. National Bureau of Economic Research.

Darby, Michael R. and Edi Karni (1973). "Free Competition and the Optimal Amount of Fraud". *Journal of Law and Economics* 16.1, pp. 67–88.

Deshpande, Rohit and Ajay K Kohli (1989). "Knowledge disavowal: structural determinants of information-processing breakdown in organizations". *Knowledge* 11.2, pp. 155–169.

Dulleck, Uwe and Rudolf Kerschbamer (2006). "On doctors, mechanics, and computer specialists: The economics of credence goods". *Journal of Economic Literature* 44.1, pp. 5–42.

— (2009). "Experts vs. Discounters: Consumer Free-Riding and Experts Withholding Advice in Markets for Credence Goods". *International Journal of Industrial Organization* 27.1, pp. 15–23.

Dulleck, Uwe, Rudolf Kerschbamer, and Matthias Sutter (2011). "The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition". *The American Economic Review* 101.2, pp. 526–555.

Eckartz, Katharina, Oliver Kirchkamp, and Daniel Schunk (2012). "How do incentives affect creativity?" *Jena Economic Research Papers, 6(68).*

Eil, David and Justin M Rao (2011). "The good news-bad news effect: asymmetric processing of objective information about yourself". *American Economic Journal: Microeconomics* 3.2, pp. 114–38.

Emons, Winand (1997). "Credence Goods and Fraudulent Experts". *RAND Journal of Economics* 28.1, pp. 107–119.

— (2001). "Credence Goods Monopolists". *International Journal of Industrial Organization* 19.3-4, pp. 375–389.

Ewers, Mara and Florian Zimmermann (2015). "Image and misreporting". *Journal of the European Economic Association* 13.2, pp. 363–380.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2018). "Global Evidence on Economic Preferences". *The Quarterly Journal of Economics* 133.4, pp. 1645–1692.

Falk, Armin, Anke Becker, Thomas J Dohmen, David Huffman, and Uwe Sunde (2016). "The preference survey module: A validated instrument for measuring risk, time, and social preferences". *Netspar discussion paper*.

Falk, Armin and Nora Szech (2020). *Competing Image Concerns: Pleasures of Skills and Moral Values*. Working paper.

Federal Bureau of Investigation (2011). *Financial Crimes Report to the Public: Fiscal Year 2010-2011*. United States Department of Justice.

Fehr, Ernst and Lorenz Goette (2007). "Do workers work more if wages are high? Evidence from a randomized field experiment". *American Economic Review* 97.1, pp. 298–317.

Fein, Steven and Steven J Spencer (1997). "Prejudice as self-image maintenance: Affirming the self through derogating others." *Journal of Personality and Social Psychology* 73.1, p. 31.

Fiocco, R. (2016). "The Strategic Value of Partial Vertical Integration". *European Economic Review* 89, pp. 284–302.

Fischbacher, Urs (2007a). "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments". *Experimental Economics* 10.2, pp. 171–178.

— (2007b). "z-Tree: Zurich toolbox for ready-made economic experiments". *Experimental Economics* 10.2, pp. 171–178.

Fischbacher, Urs and Franziska Föllmi-Heusi (2013). "Lies in disguise—an experimental study on cheating". *Journal of the European Economic Association* 11.3, pp. 525–547.

Flath, D. (1989). "Vertical Integration by means of Shareholding Interlocks". *International Journal of Industrial Organization* 7.3, pp. 369–380.

Fong, Yuk-fai, Ting Liu, and Donald J. Wright (2014). "On the role of verifiability and commitment in credence goods markets". *International Journal of Industrial Organization* 37, pp. 118–129.

Frankel, Alexander and Michael Schwarz (2014). "Experts and Their Records". *Economic Inquiry* 52.1, pp. 56–71.

Friedman, Eric, Simon Johnson, and Todd Mitton (2003). "Propping and tunneling". *Journal of Comparative Economics* 31.4, pp. 732–750.

Friedrichsen, Jana and Dirk Engelmann (2018). "Who cares about social image?" *European Economic Review* 110, pp. 61–77.

Fudenberg, Drew and David K Levine (2006). "A dual-self model of impulse control". *American Economic Review* 96.5, pp. 1449–1476.

Gabaix, Xavier and David Laibson (2006). "Shrouded attributes, consumer myopia, and information suppression in competitive markets". *The Quarterly Journal of Economics* 121.2, pp. 505–540.

Gächter, Simon, Lingbo Huang, and Martin Sefton (2016). "Combining "real effort" with induced effort costs: the ball-catching task". *Experimental Economics* 19.4, pp. 687–712.

Gächter, Simon, Eric J Johnson, and Andreas Herrmann (2021). "Individual-level loss aversion in riskless and risky choices". *Theory and Decision*, pp. 1–26.

Garbarino, Ellen, Robert Slonim, and Marie Claire Villeval (2019). "Loss aversion and lying behavior". *Journal of Economic Behavior and Organization* 158, pp. 379–393.

Gibson, Rajna, Carmen Tanner, and Alexander F Wagner (2013). "Preferences for truthfulness: Heterogeneity among and within individuals". *American Economic Review* 103.1, pp. 532–48.

Gneezy, Uri, Agne Kajackaite, and Joel Sobel (2018). "Lying Aversion and the Size of the Lie". *American Economic Review* 108.2, pp. 419–53.

Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel van Veldhuizen (2020). "Bribing the self". *Games and Economic Behavior* 120, pp. 311–324.

Goerg, Sebastian J, Sebastian Kube, and Jonas Radbruch (2019). "The effectiveness of incentive schemes in the presence of implicit effort costs". *Management Science* 65.9, pp. 4063–4078.

Gollier, Christian and Alexander Muermann (2010). "Optimal choice and beliefs with ex ante savoring and ex post disappointment". *Management Science* 56.8, pp. 1272–1284.

Golman, Russell, David Hagmann, and George Loewenstein (2017). "Information avoidance". *Journal of Economic Literature* 55.1, pp. 96–135.

Gottschalk, Felix, Wanda Mimra, and Christian Waibel (2020). "Health Services as Credence Goods: a Field Experiment". *The Economic Journal* 130.629, pp. 1346–1383.

Greenlee, P. and A. Raskovich (2006). "Partial Vertical Ownership". *European Economic Review* 50.4, pp. 1017–1041.

Greiff, Matthias (2019). "Team production and esteem: A dual selves model with belief-dependent preferences". *Games* 10.3, p. 33.

Greiner, Ben (2004). *The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics*. Report. University of Cologne.

— (2015). "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE". *Journal of the Economic Science Association* 1.1, pp. 114–125.

Grolleau, Gilles, Martin G Kocher, and Angela Sutan (2016). "Cheating and loss aversion: do people cheat more to avoid a loss?" *Management Science* 62.12, pp. 3428–3438.

Grossman, Zachary and Joel J Van Der Weele (2017). "Self-image and willful ignorance in social decisions". *Journal of the European Economic Association* 15.1, pp. 173–217.

Grubb, Michael (2015). "Failing to Choose the Best Price: Theory, Evidence, and Policy". *Review of Industrial Organization* 47.3, pp. 303–340.

Grubb, Michael D. (2015). "Consumer Inattention and Bill-Shock Regulation". *Review of Economic Studies* 82.1, pp. 219–257.

Grubert, Harry, John Mutti, et al. (1991). "Taxes, tariffs and transfer pricing in multinational corporate decision making". *The Review of Economics and Statistics* 73.2, pp. 285–293.

Grubiak, Kevin (2019). *Exploring Image Motivation in Promise Keeping: An Experimental Investigation*. Tech. rep. School of Economics, University of East Anglia, Norwich, UK.

Gul, Faruk (1991). "A theory of disappointment aversion". *Econometrica*, pp. 667–686.

Halleck, Rebecca (2019). "Who's Been Charged in the College Admissions Cheating Scandal? Here's the Full List". *New York Times, 2019, March 12*.

Hanna, Rema and Shing-Yi Wang (2017). "Dishonesty and Selection into Public Service: Evidence from India". *American Economic Journal: Economic Policy* 9.3, pp. 262–290.

Hart, O. and J. Tirole (1990). "Vertical Integration and Market Foreclosure". *Brookings Papers on Economic Activity. Microeconomics*, pp. 205–286.

Heck, Patrick R, Daniel J Simons, and Christopher F Chabris (2018). "65% of Americans believe they are above average in intelligence: Results of two nationally representative surveys". *PloS one* 13.7, e0200103.

Heidhues, Paul and Botond Koszegi (2017). "Naïveté-Based Discrimination". *The Quarterly Journal of Economics* 132.2, pp. 1019–1054.

Heidhues, Paul and Botond Kőszegi (2018). "Behavioral industrial organization". In: *Handbook of Behavioral Economics - Foundations and Applications 1*. Ed. by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson. Vol. 1. Handbook of Behavioral Economics: Applications and Foundations 1. North-Holland. Chap. 6, pp. 517–612.

Hilger, Nathaniel G. (2016). "Why Don't People Trust Experts?" *The Journal of Law and Economics* 59.2, pp. 293–311.

Holt, Charles A. and Susan K. Laury (2002). "Risk Aversion and Incentive Effects". *American Economic Review* 92.5, pp. 1644–1655.

Hossain, Tanjim and Ryo Okui (2013). "The binarized scoring rule". *Review of Economic Studies* 80.3, pp. 984–1001.

Hunold, M. and K. Stahl (2016). "Passive Vertical Integration and Strategic Delegation". *The RAND Journal of Economics* 47.4, pp. 891–913.

Hunold, Matthias (2020). "Non-Discriminatory Pricing, Partial Backward Ownership, and Entry Deterrence". *International Journal of Industrial Organization* 70, p. 102615.

Jiang, Guohua, Charles MC Lee, and Heng Yue (2010). "Tunneling through intercorporate loans: The China experience". *Journal of Financial Economics* 98.1, pp. 1–20.

Johnson, Simon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2000). "Tunneling". *American Economic Review* 90.2, pp. 22–27.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler (1990). "Experimental tests of the endowment effect and the Coase theorem". *Journal of Political Economy* 98.6, pp. 1325–1348.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler (1991). "Anomalies: The endowment effect, loss aversion, and status quo bias". *Journal of Economic Perspectives* 5.1, pp. 193–206.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica* 47.2, pp. 263–292.

Karle, Heiko, Georg Kirchsteiger, and Martin Peitz (2015). "Loss aversion and consumption choice: Theory and experimental evidence". *American Economic Journal: Microeconomics* 7.2, pp. 101–20.

Karlsson, Niklas, George Loewenstein, and Duane Seppi (2009). "The ostrich effect: Selective attention to information". *Journal of Risk and Uncertainty* 38.2, pp. 95–115.

Kerschbamer, Rudolf, Daniel Neururer, and Matthias Sutter (2016). "Insurance coverage of customers induces dishonesty of sellers in markets for credence goods". *Proceedings of the National Academy of Sciences* 113.27, pp. 7454–7458.

Kerschbamer, Rudolf, Matthias Sutter, and Uwe Dulleck (2017). "How social preferences shape incentives in (experimental) markets for credence goods". *The Economic Journal* 127.600, pp. 393–416.

Khalmetski, Kiryl and Dirk Sliwka (2019). "Disguising lies—Image concerns and partial lying in cheating games". *American Economic Journal: Microeconomics* 11.4, pp. 79–110.

Kőszegi, Botond (2006). "Ego utility, overconfidence, and task choice". *Journal of the European Economic Association* 4.4, pp. 673–707.

Kőszegi, Botond and Matthew Rabin (2007). "Reference-dependent risk attitudes". *American Economic Review* 97.4, pp. 1047–1073.

Leibenstein, Harvey (1950). "Bandwagon, snob, and Veblen effects in the theory of consumers' demand". *The Quarterly Journal of Economics* 64.2, pp. 183–207.

Levy, Nadav, Yossi Spiegel, and David Gilo (2018). "Partial Vertical Integration, Ownership Structure, and Foreclosure". *American Economic Journal: Microeconomics* 10.1, pp. 132–80.

Li, Jiawei, Stephen Leider, Damian Beil, and Izak Duenyas (2021). "Running online experiments using web-conferencing software". *Journal of the Economic Science Association* 7.2, pp. 167–183.

Lim, Sonya S. and Siew Hong Teoh (2010). "Limited attention". In: *Behavioral Finance: Investors, Corporations, and Markets*. Ed. by H. Kent Baker and John R. Nofsinger. The Robert W. Kolb Series in Finance. Wiley. Chap. 16, pp. 295–312.

Liu, Ting (2011). "Credence Goods Markets with Conscientious and Selfish Experts". *International Economic Review* 52.1, pp. 227–244.

Lovett, Benjamin J (2020). "Disability Identification and Educational Accommodations: Lessons From the 2019 Admissions Scandal". *Educational Researcher* 49.2, pp. 125–129.

McGuire, Thomas G. (2000). "Physician agency". In: *Handbook of Health Economics*. Ed. by J. Culyer Anthony and P. Newhouse Joseph. Vol. 1, Part A. Elsevier. Chap. 9, pp. 461–536.

Mimra, W., A. Rasch, and C. Waibel (2016). "Price competition and reputation in credence goods markets: Experimental evidence". *Games and Economic Behavior* 100, pp. 337–352.

Niehaus, Paul (2014). "A theory of good intentions". *San Diego, CA: University of California and Cambridge, MA: NBER* 111.

Olafsson, Arna, Michaela Pagel, Olivier Coibion, and Yuriy Gorodnichenko (2018). "The ostrich in us: Selective attention to personal finances". *NBER working paper* 23945.

Oster, Emily, Ira Shoulson, and E Dorsey (2013). "Optimal expectations and limited medical testing: evidence from Huntington disease". *American Economic Review* 103.2, pp. 804–30.

Pennings, Joost ME and Ale Smidts (2003). "The shape of utility functions and organizational behavior". *Management Science* 49.9, pp. 1251–1263.

Petrishcheva, Vasilisa, Gerhard Riener, and Hannah Schildberg-Hörisch (2020). "Loss aversion in social image concerns". *IZA Discussion Paper*.

Pettit, Nathan C., Sarah P. Doyle, Robert B. Lount, and Christopher To (2016). "Cheating to get ahead or to avoid falling behind? The effect of potential negative versus positive status change on unethical behavior". *Organ. Behav. Hum. Decis. Process.* 137, pp. 172–183.

Rasch, Alexander and Christian Waibel (2018). "What Drives Fraud in a Credence Goods Market? Evidence From a Field Study". *Oxford Bulletin of Economics and Statistics* 80, pp. 605–624.

Raven, John C (1983). "Manual for Raven's progressive matrices and vocabulary scales". *Standard Progressive Matrices*.

Rey, P. and J. Tirole (2007). "A Primer on Foreclosure". *Handbook of Industrial Organization* 3, pp. 2145–2220.

Riyanto, Yohanes E and Linda A Toolsema (2008). "Tunneling and propping: A justification for pyramidal ownership". *Journal of Banking & Finance* 32.10, pp. 2178–2187.

Ropka, Mary E, Jennifer Wenzel, Elayne K Phillips, Mir Siadaty, and John T Philbrick (2006). "Uptake rates for breast cancer genetic testing: a systematic review". *Cancer Epidemiology and Prevention Biomarkers* 15.5, pp. 840–855.

Santos-Pinto, Luis and Joel Sobel (2005). "A model of positive self-image in subjective assessments". *American Economic Review* 95.5, pp. 1386–1402.

Schindler, Simon and Stefan Pfattheicher (2017). "The frame of the game: Loss-framing increases dishonest behavior". *Journal of Experimental Social Psychology* 69, pp. 172–177.

Schneider, Henry S. (2012). "Agency Problems and Reputation in Expert Services: Evidence From Auto Repair". *The Journal of Industrial Economics* 60.3, pp. 406–433.

Schulz-Hardt, Stefan, Dieter Frey, Carsten Lüthgens, and Serge Moscovici (2000). "Biased information search in group decision making." *Journal of Personality and Social Psychology* 78.4, p. 655.

Serra-Garcia, Marta and Nora Szech (2021). "The (in) elasticity of moral ignorance". *Management Science*.

Sexton, Steven E and Alison L Sexton (2014). "Conspicuous conservation: The Prius halo and willingness to pay for environmental bona fides". *Journal of Environmental Economics and Management* 67.3, pp. 303–317.

Soetevent, Adriaan R (2011). "Payment choice, image motivation and contributions to charity: Evidence from a field experiment". *American Economic Journal: Economic Policy* 3.1, pp. 180–205.

Spiegel, Y. (2013). "Backward Integration, Forward Integration, and Vertical Foreclosure".

Sweeny, Kate, Darya Melnyk, Wendi Miller, and James A Shepperd (2010). "Information avoidance: Who, what, when, and why". *Review of General Psychology* 14.4, pp. 340–353.

Taylor, Curtis R. (1995). "The Economics of Breakdowns, Checkups, and Cures". *Journal of Political Economy* 103.1, pp. 53–74.

Thornton, Rebecca L (2008). "The demand for, and impact of, learning HIV status". *American Economic Review* 98.5, pp. 1829–63.

Wagner, Gert G, Joachim R Frick, and Jürgen Schupp (2007). "The German Socio-Economic Panel Study (SOEP) - evolution, scope and enhancements". *SOEP papers on Multidisciplinary Panel Data Research, No. 1.*

Wakker, Peter P (2010). *Prospect theory: For risk and ambiguity.* Cambridge University Press.

Zaltman, Gerald (1983). "Construing knowledge use". In: *Realizing Social Science Knowledge.* Springer, pp. 236–251.

Zimmermann, Florian (2020). "The dynamics of motivated beliefs". *American Economic Review* 110.2, pp. 337–61.