Essays in Behavioral Economics

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

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

vorgelegt von

Si Chen

aus Longyan, VR China

Bonn, 2020

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Thomas Dohmen
Zweitreferent:	Prof. Dr. Sebastian Kube
Tag der mündlichen Prüfung:	21. August 2020

Acknowledgements

This dissertation would not have been possible without the support of Bonn Graduate School of Economics, Collaborative Research Center Transregio 224, and the Institute for Applied Microeconomics at Bonn University. I am also greatly grateful to those who have been there for me throughout the journey.

I am blessed to have great advisors and mentors at the BGSE: Thomas Dohmen, Sebastian Kube, Florian Zimmermann, Lorenz Götte, and Hannah Schildberg-Hörisch. I thank Thomas Dohmen and Sebastian Kube for unreservedly and constantly showing their faith in me. People say ' To build is more difficult than to tear apart'. Their faith in me has built towards my confidence and perseverance as an early-stage economist. Thomas has been an incredible advisor and mentor to me. He has been never stingy with his feedback on my work and coaching for my career. Meanwhile, his wisdom and devotion to work inspire me to try harder and do my best. I also thank Thomas for making me part of the Collaborative Research Center Transregio 224, from which I benefit greatly through the financial support and the conference participations. I thank Sebastian Kube for always keeping his door open for me, which is a privilege as well as an encouragement. I also thank Sebastian for being as light-hearted as he is, a character that took much steam off me in some overwhelming moments. I am grateful to Florian Zimmermann and Lorenz Götte for the time that they devoted to advising me on my research and my career. I thank Florian for being such an inspiring young economist who is a walking proof of the value of hard-work, intelligence, and modesty. I thank Lorenz for all the fun and intellectually nourishing time we spent together. I am also thankful to my co-author Hannah Schildberg-Hörisch for working with me all the way through my very first research project, a project that kick-started my academic endeavour.

I am indebted to numerous administrative staff at the Institute for Applied Microeconomis and the BGSE for their support. A special shout-out goes to Simone Jost and Andrea Reykers, who meticulously and patiently supported me on the tedious preparation of my experiments and my job search.

I am grateful for the friendship that supported me through the dark moments and made the nice moments even nicer. I thank Birgit Mauersberger, Rebecca Hader and Gasper Ploj for being the most accepting friends from whom I feel cared about. I thank Axel Wogrolly and Shanzhi Ye for the often intellectually challenging conversations which I always found refreshing. I thank Lukas Kiessling for his kind and insightful comments on my research and his help in compiling this dissertation. I am thankful to Deniz Kattwinkel for spending long afternoon hours having bubble teas with me when I needed the company the most. I thank Xueying Liu, Jingyu Song, Lingwei Wu, Donghai Zhang and Jingjing Li for letting me unload the stress and miraculously still sticking around after.

I am deeply thankful to my late mother Wei Chen and my father Yulai Chen. Not for a single day have they doubted about me. Words cannot express my gratitudes to them. I am indebted to my father and his wife Anna Li for their understanding during all these years I spent far away from them. I also thank Barbara Heese and Hans-Christian Heese for becoming my family in Germany.

Finally, I thank my husband Carl Heese for being the wonderful person he is. His diligence and kindness inspire me everyday. Thank you for sharing this journey with me. Thank you for being part of my life.

Contents

List of Figures

List of Tables

Introduction

This dissertation presents three chapters revolving around the common theme '*motivated beliefs*'. Motivated reasoning occurs when individuals trade-off accurate beliefs and desirable beliefs (Bénabou and Tirole, 2016).

The research on motivated beliefs has been expanding rapidly in the last two decades, when the field of behavioral economics starts to seek the rationale behind the seemingly irrational biases. Overconfidence, for example, is one of the most widely observed biases in individuals. Research shows that overconfidence prevails despite feedback (Moore and Small, 2008). The motivated reasoning literature put forward the explanation that the prevalence of overconfidence might be by choice and a result of the trade-off between the accuracy and desirability of individuals' self-views. Another context, in which motivated reasoning can occur, is in decisions where benefiting oneself might harm others. In this context, holding the belief that benefiting themselves does not harm others, individuals can feel moral while behaving selfishly (for a review, see Gino, Norton, and Weber, 2016).

Across various contexts, two research questions are at the core of the research on motivated beliefs: first, why would people desire certain beliefs? Second, by what means do they gain and maintain beliefs that they desire? Understanding these questions would help us to gain a better idea about not only what drives beliefs, but also how beliefs affect decisions.

In this dissertation, I present my experimental and theoretical research on motivated beliefs. I use experimental tools for the empirical research on motivated beliefs for two reasons. First, beliefs are usually not observed nor recorded in observational data. Second, beliefs are endogenous, which makes its effect on the outcomes of interest hard to be disentangled from the effect of other personal characteristics like gender, age, personalities, etc. On top of the experiments, theoretical tools come in handy for a deeper understanding of the empirical evidence, by offering it a psychological and micro foundation. I also take advantage of the theoretical analyses to generate testable predictions that are of empirical interests.

In Chapter 1, I present an experimental examination on the effect of self-confidence on the effort that an individual exerts into a task. As posited in many theoretical models, one reason why people find high self-confidence desirable might be that it can motivate them to work harder, i.e. the motivational value of confi-

dence (e.g. Bénabou and Tirole, 2002; Gervais and Goldstein, 2007; Krähmer, 2007; Santos-Pinto, 2008; Santos-Pinto, 2010; Ludwig, Wichardt, and Wickhorst, 2011a; Ludwig, Wichardt, and Wickhorst, 2011b). Despite the prevalent use in theory, the supporting empirical evidence of the motivational value of confidence is scarce. Does confidence motivate effort? In this chapter, I present a laboratory experiment that shows that higher confidence does lead to higher effort provision. In the experiment, we use feedback to exogenously manipulate the subjects' confidence in their ability in a modified slider task, in which the subjects' objective is to posit a 0-100 slider to its middle point. Then in a further stage, we measure the subjects' effort provision in the same task by the length of time that they spend working on the task. We find that higher ability beliefs lead to more effort. Consequently, for overconfident individuals, de-biasing information hurts their effort provision. This finding offers empirical evidence for the models involving the motivational value of confidence. It also points out boosting self-confidence as a way to motivate effort at work place.

In Chapter 2, I present a theoretical analysis of the information acquisition strategy of an agent who values not only her material well-beings but also her belief in the innocuousness of her decision. In many decisions, a selfish choice can harm others. For example, doctors prescribing drugs for which they receive commissions might harm the patients for whom those drugs are inappropriate. Empirical observations like charitable giving show that people do not always maximize their material benefits. Recent empirical findings also show that when people can behave selfishly while feeling moral, they jump on the opportunity (Dana, Weber, and Kuang, 2007; Gino, Norton, and Weber, 2016, etc). It implies that people do not genuinely care about externalities of their decisions. So, if people do not genuinely care about the externalities of their decisions, what drives the deviation from maximizing their own material interests? In this chapter, we propose a model in which the agent values her *beliefs* that her decisions are innocuous, on top of her material benefits. The model provides a unifying theoretical framework for analyzing information preferences in social decisions. This framework does not only explain the previous empirical findings, but also generates novel testable predictions for future empirical research. The main result of the model shows that the agent's optimal information signal cannot be positively skewed, i.e. it cannot be more likely to show evidence against the innocuousness of the self-benefiting action. The result also extends to a dynamic setting where the agent has access to continuous information flow and can decide when to stop acquiring information. In this setting, we compare between two scenarios: one in which the agent's material interest is not affected by her decision; and the other in which there is a self-benefiting action that the agent can take to increase her material payoff. The model shows that when the material interest of the agent is involved, having received mostly information supporting the innocuousness of the selfish action, more agent types would stop acquiring information, while having received mostly information against the innocuousness of the selfish action, more agent types would continue acquiring information. Besides, the model also shows

that the information acquisition strategy motivated by the selfish desire for material benefits can reduce the negative externalities imposed by the agent's decision. This counter-intuitive result stems from a principal agent problem: if the agent has no material interest in the decision, she might slack when acquiring information. In the presence of a self-benefiting action, the agent might become better informed, in the hope to persuade herself to behave selfishly, and hence does less harm.

In Chapter 3, I present an experimental investigation on the dynamics of information acquisition in social decisions. When gathering information for a decision, people can often acquire more than one piece of information and it is at their discretion to decide when to stop. For example, people can decide when to stop reading news articles, gathering medical evidences, or interviewing candidates for a job opening. The information decision can affect beliefs based on which people make their decisions. Joining the literature on motivated beliefs, we experimentally investigate how people decide when to stop acquiring information, if they want to feel moral while behaving selfishly. In this laboratory experiment, a dictator can decide between two actions, one of which harms a receiver. While the dictators know how the actions affect their own payoffs, they are uncertain about the action that is harmful to the receiver. Before the decision, the dictators can sequentially acquire information about the harmful action, and freely decide when to stop. We compare between two treatments: in the control, the actions do not affect the dictators' payoffs; in the treatment, one of the actions generates additional payments for the dictators themselves. We find evidence supporting the model prediction in Chapter 2: compared to the dictators in the control, more dictators in the treatment stop acquiring information, having received mostly information suggesting that the self-benefiting action is harmless; and more dictators in the treatment continue acquiring information, having received mostly information suggesting that the self-benefiting action is harmful to the receivers. We also show in our data that this information acquisition strategy improves the welfare of the receiver, as predicted possible by the model in Chapter 2.

In summary, this dissertation contributes evidence that biased beliefs can sometimes be beneficial. For example, a rosy self-view can boost motivation. One way through which people can cultivate desirable beliefs is by acquiring information strategically. These beliefs in turn affect their decisions on, for example, how much effort to exert in a task or whether to undertake a self-benefiting action. Joining the research on motivated beliefs, this dissertation is dedicated to better understanding the strategic formation of beliefs, the psychological reasons behind people's desire for certain beliefs and how they affect peoples' decisions.

References

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics* 117 (3): 871–915. [2]

Bénabou, Roland, and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30 (3): 141–64. [1]

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33 (1): 67–80. [2]

Gervais, Simon, and Itay Goldstein. 2007. "The Positive Effects of Biased Self-Perceptions in Firms." *Review of Finance* 11 (3): 453–496. [2]

Gino, Francesca, Michael I Norton, and Roberto A Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives* 30 (3): 189–212. [1, 2]

Krähmer, Daniel. 2007. "Equilibrium Learning in Simple Contests." *Games and Economic Behavior* 59 (1): 105–131. [2]

Ludwig, Sandra, Philipp C. Wichardt, and Hanke Wickhorst. 2011a. "On the Positive Effects of Overconfident Self-Perception in Teams." *Available at SSRN 1854465*, [2]

Ludwig, Sandra, Philipp C. Wichardt, and Hanke Wickhorst. 2011b. "Overconfidence Can Improve and Agent's Relative and Absolute Performance in Contests." *Economics Letters* 110: 193–196. [2]

Moore, DA., and DA. Small. 2008. "When It Is Rational for the Majority to Believe that They Are Better Than Average." *Rationality and Social Responsibility: Essays in Honor of Robyn M. Dawes. Mahwah, NJ: Erlbaum*, [1]

Santos-Pinto, Luís. 2008. "Positive Self-Image and Incentives in Organisations." *Economic Journal* 118 (531): 1315–1332. [2]

Santos-Pinto, Luís. 2010. "Positive Self-Image in Tournaments." *International Economic Review* 51 (2): 475–496. [2]

Chapter 1

Confidence and Effort

Joint with Hannah Schildberg-Hörisch

1.1 Introduction

Overconfidence is a widespread phenomenon (Plous, 1993; Moore and Healy, 2008) and has been found to affect manifold decisions of economic relevance. On the one hand, overconfidence may distort decision making: for example, overconfidence can induce excessive competitiveness (Camerer and Lovallo, 1999; Bartling, Fehr, Maréchal, and Schunk, 2009; Dohmen and Falk, 2011; Danz, 2014), suboptimal financial and health-related decisions of individuals (Benartzi, 2001; Sandroni and Squintani, 2007), and poor judgment in firms regarding investments, market entry, and mergers (Camerer and Lovallo, 1999; Malmendier and Tate, 2005; Malmendier and Tate, 2008). On the other hand, recent research also documents possible merits of overconfidence: it may promote social status (Kennedy, Anderson, and Moore, 2013), convincingness (Schwardmann and Weele, 2018), and innovativeness (Galasso and Simcoe, 2011; Hirshleifer, Low, and Teoh, 2012).

This paper focuses on a different effect of higher confidence in general and overconfidence in particular that has attracted special attention by a growing number of microeconomic models with overconfident agents (e.g. Bénabou and Tirole, 2002; Gervais and Goldstein, 2007; Krähmer, 2007; Santos-Pinto, 2008; Santos-Pinto, 2010; Ludwig, Wichardt, and Wickhorst, 2011a; Ludwig, Wichardt, and Wickhorst, 2011b). These models argue that individuals with higher beliefs on their own ability, even the overconfident ones, exert higher effort. Intuitively, individuals with higher beliefs on their own ability anticipate greater return to effort and hence work harder. Whether this implication holds true empirically is however unobvious: the less confident might work harder in the hope of compensating their perceived lack of ability, while the overconfident may become smug and slack.

The prominence of the motivational value of confidence as a basic ingredient of microeconomic models with overconfident agents contrasts a lack of empirical evidence in favor of it. As a first contribution, our paper aims at closing that gap

by testing the motivational value of absolute confidence empirically, i.e. the hypothesis that individuals with a higher belief on their own ability exert higher effort (Hypothesis 1). This relation is supposed to hold for overconfident individuals who have exaggerated ability beliefs as well. Moreover, we test an important implication of the motivational value of confidence: informing overconfident individuals about their own ability will reduce their effort provision (Hypothesis 2). We hypothesize that such negative de-biasing information induces overconfident individuals to adjust ability beliefs downwards, which in turn, decreases their effort provision. Finally, we move beyond existing research on overconfidence by offering an empirical strategy to identify significant absolute overconfidence at the individual level.[1] It is based on a definition of absolute overconfidence that takes into account that beliefs on one's own ability correspond to a belief distribution and acknowledges that observational measures of ability are noisy albeit informative about the actual underlying ability. Combining these two insights, we define an individual as overconfident if the median of her belief distribution exceeds the upper limit of the 95% confidence interval around her actual underlying ability, which we construct based on noisily measured ability. This definition implies that overconfident individuals assign more than 50% probability mass of the belief distribution to ability levels higher than their actual ability. In other words, they believe that it is more likely that they have a higher ability than their true ability than vice versa.

We test our two main hypotheses in a laboratory experiment with 5 stages. Stage 1 of the experiment measures individual ability in a modified version of the Gill and Prowse (2019) slider task. The modified slider task does not allow subjects to perfectly monitor their own ability, which offers scope for over- or underconfidence. In Stage 2, subjects' belief distributions on their own ability in the modified slider task as measured in Stage 1 are elicited using a visualized "ball allocation task". In the ball allocation task, subjects are asked to allocate 100 balls that each represent one percentage point probability into 11 bins that illustrate intervals of increasing abilities. Combining the data on observed ability from Stage 1 and median ability beliefs from Stage 2, we can identify overconfident subjects. In Stage 3, subjects are randomly assigned to a treatment with information (INFO) or a treatment with no information (NOINFO) about their own ability in the modified slider task measured in Stage 1. In Stage 4, subjects work on the same modified slider task as in Stage 1. They can, however, choose individually how much effort to exert by stopping working on the task. Finally, in Stage 5 we again use the ball allocation task to elicit

1. Moore and Healy (2008) categorize three kinds of overconfidence: overestimation, overplacement, and overprecision. This paper focuses on overestimation of one's own ability compared to an objective measure of it (absolute overconfidence). In contrast, overplacement refers to an overestimate of oneself relative to others (relative overconfidence); overprecision refers to excessive certainty regarding the accuracy of one's belief.

subjects' belief distributions on their ability in the modified slider task as measured in Stage 4.

In line with the motivational value of confidence, we find that subjects with a higher belief on their own ability exert higher effort in Stage 4. This relation also holds and is particularly strong for overconfident subjects. The exogenous variation in information provision across treatments INFO and NOINFO induces exogenous changes in ability beliefs, which in turn provide causal evidence on the motivation value of confidence. Moreover, we show that informing overconfident subjects about their own ability results in lower effort provision.

Our results provide an empirical backing for a number of economic models that rely on the motivational value of confidence. In some of these models, higher confidence leads to higher effort because of complementarity between ability and effort (Bénabou and Tirole, 2002; Gervais and Goldstein, 2007; Krähmer, 2007; Santos-Pinto, 2008, 2010; Gervais, Heaton, and Odean, 2011; Ludwig, Wichardt, and Wickhorst, 2011a; Rosa, 2011). In Ludwig, Wichardt, and Wickhorst (2011b), the motivational value of overconfidence originates from an underestimation of effort costs. Here we review the related theory models separately in four contexts: principal-agent models, tournaments, teamwork, and contests. Regarding principal-agent models, Santos-Pinto (2008) shows that by motivating effort, the agent's overconfidence can sometimes benefit the principal even when the agent's effort is unobservable. Rosa (2011) studies equilibrium incentive contracts for overconfident agents and finds that overconfidence increases effort implemented by the equilibrium contracts. In Gervais, Heaton, and Odean (2011), overconfident managers can be more easily motivated to exert costly effort to learn about risky projects as they overestimate the quality of their private information, and are therefore sometimes preferable to their rational counterparts for firms. Bénabou and Tirole (2002) show that overconfidence can help a time-inconsistent agent to overcome her self-control problem by motivating effort. Studying tournaments as a form of providing incentives in firms, Santos-Pinto (2010) shows that if higher confidence increases effort firms can benefit from workers' overconfidence by structuring prizes in tournaments. Studying overconfidence in teamwork, Gervais and Goldstein (2007) and Ludwig, Wichardt, and Wickhorst (2011a) argue that, when team members' efforts are complements, the presence of an overconfident agent who exerts excessive effort can benefit all team members. Finally, in Krähmer (2007) and Ludwig, Wichardt, and Wickhorst (2011b), a worse but overconfident contestant can prevail in contests by exerting high effort.

While we provide empirical evidence for the effect of confidence on real effort provision, Sautmann (2013) and Fischer and Sliwka (2018) study the effect of confidence on monetary investments as a proxy of effort. Their experiments focus on the contexts of principal-agent problems and human capital acquisition, respectively. In the laboratory, Sautmann (2013) investigates principals' tendency to exploit agents' measured absolute over- or underconfidence when offering a contract and the result-

ing monetary investments that the agents make to improve the contracted outcome. She finds no effect of agents' over- or underconfidence on their investment decisions, and attributes it to the fact that most principals do not adjust their offers according to agents' over- or underconfidence. Fischer and Sliwka (2018) study the effect of relative confidence in one's knowledge and one's learning ability on costly investment in learning materials. They show that students with higher, exogenous beliefs in their learning ability invest more in learning material when preparing for a test. In contrast, higher beliefs in knowledge only makes those with lower-than-median prior knowledge increase their investment in costly learning material, while those with higher-than-median prior knowledge decrease their investment. In comparison to these two papers, we focus on the effect of confidence on real effort provision and therefore use a real effort task that has no scope for learning over time. Also using a real effort task, Barron and Gravert (2018) study in the laboratory how beliefs on relative cognitive ability affect workers' selection into jobs with different incentive schemes, and their subsequent effort choice. They do not find a significant effect of confidence on effort since effort is high regardless of the beliefs on relative cognitive ability. Barron and Gravert (2018) measure effort at the intensive margin in a laboratory real effort task, which has been commonly observed to be subject to the ceiling effect, i.e. subjects tend to exert close to maximum effort for a fixed and relatively short working time (e.g. Eckartz, 2014; Corgnet, Hernán-González, and Schniter, 2015; Araujo, Carbone, Conell-Price, Dunietz, Jaroszewicz, et al., 2016; Gächter, Huang, and Sefton, 2016; Goerg, Kube, and Radbruch, forthcoming). Aiming to moderate the ceiling effect, we follow Goerg, Kube, and Radbruch (forthcoming) in allowing subjects to leave the laboratory early if they decide to work less, and thus measure effort on the extensive margin.

Our findings also contribute to the empirical literature on factors that motivate effort (DellaVigna and Pope, 2017). In particular, our results suggest boosting confidence as an effective and potentially cost-efficient way to enhance effort provision as argued in the models by Santos-Pinto (2008, 2010), Gervais, Heaton, and Odean (2011), Rosa (2011). The negative impact of de-biasing information on the effort provision of overconfident individuals is of obvious relevance in diverse principal-agent contexts such as interactions between employers and employees or teachers and students. For example, employers could restrain from providing accurate feedback to an overconfident employee in order to motivate high effort. Our findings also offer an explanation for why teachers are often reluctant or, for younger students, sometimes even prohibited to provide clear-cut, but possibly worse than expected feedback on students' skills, namely to avoid demotivating their students in future learning efforts.

In terms of research methods, we offer a clean conceptual definition of absolute overconfidence and an empirical strategy to identify significant absolute overconfidence at the individual level. Previous papers on absolute overconfidence have directly compared point beliefs on absolute performance to measured performance

to identify overconfidence (e.g. Blavastkyy, 2009; Clark and Friesen, 2009; Urbig, Stauf, and Weitzel, 2009; Ludwig and Nafziger, 2011; Sautmann, 2013; Hollard, Massoni, and Vergnau, 2016). Only Ludwig and Nafziger (2011) discuss the risk of misclassifying individuals due to measurement error in observed performance and compare average point beliefs on absolute performance to average performance, but they identify overconfidence only at the group level. The other papers use that approach to identify absolute overconfidence at the individual level, ignoring possible misclassification due to measurement error. Moreover, Malmendier and Tate (2005, 2008), Galasso and Simcoe (2011), Malmendier, Tate, and Yan (2011), Hirshleifer, Low, and Teoh (2012) use indirect approaches to categorize CEOs as overconfident based on their options exercise behavior or their portrayal in the press.[2] our paper proposes a definition of absolute overconfidence that regards observed ability as a noisy measure of actual ability and accounts for the fact that individuals hold a belief distribution on their ability. Based on that definition, we offer a strategy for identifying significant overconfidence at the individual level that can be applied more broadly in future work.

The remainder of the paper is structured as follows. Section 1.2 outlines our two main hypotheses. The experimental design is described in section 1.2. Section 1.4 provides a definition of absolute overconfidence and shows how one can build on that definition in order to empirically identify absolute overconfidence at the individual level. Section 1.5 presents results and several robustness checks. We discuss our findings and conclude in section 1.6.

1.2 Hypotheses

When faced with an effort-intensive task, individuals have to decide how much effort to exert. Without knowledge about their true ability in the task, they need to rely on their belief on their ability to make this decision. Whether diligence is induced by higher or lower confidence in one's own ability constitutes the first research question that we aim to answer.

Hypothesis 1 (motivational value of confidence) A higher belief on own ability leads to higher effort provision.

Utility maximizers choose their effort levels by balancing expected marginal benefits and marginal costs of effort provision. In a task in which ability and effort are

2. In particular, Malmendier and Tate (2005) and Galasso and Simcoe (2011) classify CEOs as overconfident if they, e.g., hold nontradeable in-the-money executive stock options until expiration rather than exercising them after the vesting period or if they exercise options of their own company later than suggested by a rational benchmark, since such behaviors suggest overconfidence in the own ability to keep the company's stock price rising. Malmendier and Tate (2008), Malmendier, Tate, and Yan (2011), and Hirshleifer, Low, and Teoh (2012) additionally rely on a CEO's characterization as "confident" or "optimistic" in the press.

complements, those with higher ability beliefs will exert higher effort since they expect higher marginal benefits from effort provision.[3] This motivational value of confidence is particularly relevant for modelling individuals who overestimate their own ability, for its potential to offset suboptimal decision making through exaggerated effort (e.g. Bénabou and Tirole, 2002; Compte and Postlewaite, 2004). In contrast, when ability and effort are substitutes, higher ability implies lower marginal returns to effort, and therefore higher ability beliefs will lead to lower effort provision.

Hypothesis 2 Informing overconfident individuals on their own ability reduces their effort provision.

When overconfident individuals receive feedback on their actual ability, they adjust their ability beliefs downwards. If Hypothesis 1 holds, this downward adjustment of ability beliefs due to de-biasing information is predicted to reduce effort provision.

Understanding the potential trade-off between accurate feedback and reduced effort provision is an important step towards an effective handling of feedback, especially when high effort is desirable due to positive externalities. For example, if Hypothesis 2 holds, in teamwork where diligence is key to success and efforts of the team members are complementary, avoiding negative feedback to one another might benefit the group performance more than providing fully honest feedback (Gervais and Goldstein, 2007; Ludwig, Wichardt, and Wickhorst, 2011a). As another example, when teachers are convinced that high effort in learning promotes future school performance and valuable traits like conscientiousness, Hypothesis 2 suggests that they might be reluctant to give perfectly accurate feedback to their students.

3. A formal model that derives Hypotheses 1 and 2 is presented in the Online Appendix.

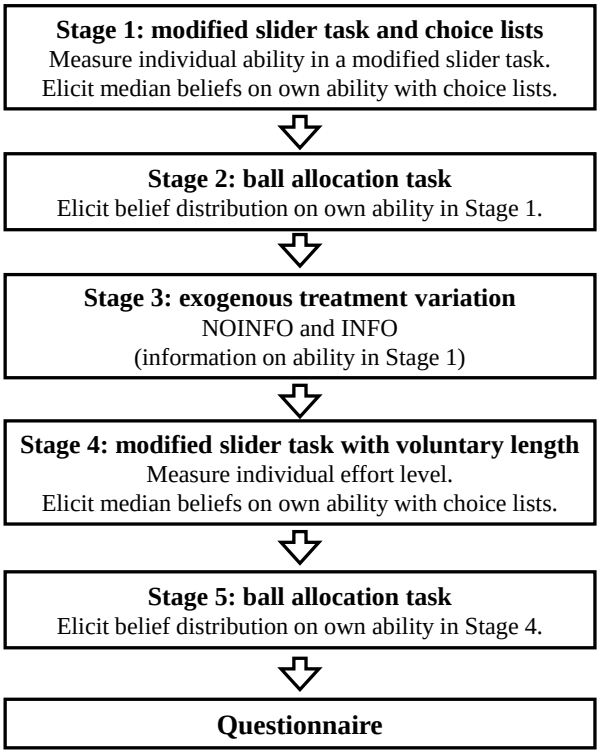


Figure 1.1. Overview of Experimental Design

1.3 Experimental Design and Implementation

We conducted a laboratory experiment to test Hypothesis 1 and 2. The experiment has five stages (Figure 1.1).

Stage 1: In Stage 1, we measured individual ability using an adapted version of the well-established slider task (Gill and Prowse, 2019) and elicited subjects' median

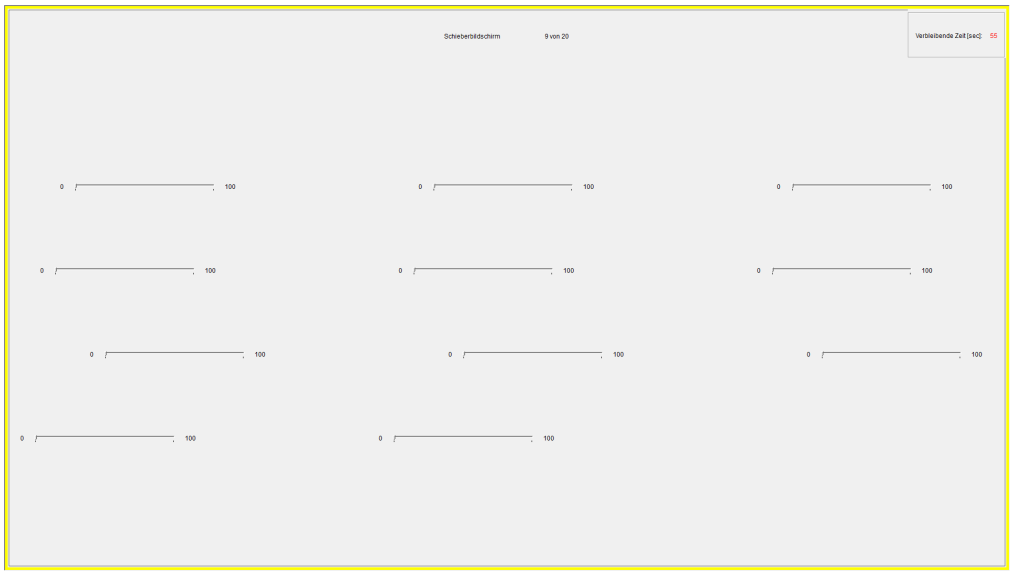


Figure 1.2. A Slider Screen

beliefs on their own ability in an incentive compatible way using "choice list screens". Slider screens and choice list screens were shown alternately, each 20 times.

Each slider screen displayed 11 sliders (see Figure 1.2), each with a scale from 0 to 100. We used 11 sliders per screen to eliminate obvious focal points in the later belief elicitation. The subjects' task was to position each slider into the interval $[49.5, 50.5]$ at the middle of the scale.[4] For each subject, the proportion of correctly positioned sliders serves as the measure of individual ability in the modified slider task. Subjects earned a piece rate of 1 point for each successfully positioned slider. At the end of the experiment, each point was exchanged into 0.05 Euro.

In contrast to the original version of the slider task by Gill and Prowse (2019), the numerical position of each slider was not displayed on screen and subjects could only guess the slider's position by eye-balling. Thus, subjects could not perfectly monitor their ability, which offers scope for over- or underconfidence. Further advantages of the slider task are that it does not require prior knowledge and does not exhibit a time trend in performance, in line with the assumption that underlying true ability is constant over time.[5]

Subjects had 55 seconds to work on each slider screen. Fixing an upper time limit ensures that our measure of individual ability, the share of correctly positioned

4. A slider's position on the scale was measured in 0.25 increments.

5. In a Tobit panel regression of the number of correctly positioned sliders per screen on a screen sequence number and an additional dummy for the last screen, the coefficient of the screen sequence number is not significant (-0.006, $p = 0.13$) and the dummy for the last screen is marginally significant only (-0.19, $p = 0.07$).

sliders, is comparable across subjects. Only after 55 seconds, subjects could proceed to the next choice list screen that appeared automatically. Our data suggest that 55 seconds were sufficient for subjects to work on all 11 sliders: on average, subjects left only 5 out of 220 sliders untouched in Stage 1.

The choice list screens were designed to elicit the median of a subject's belief distribution on the number of correctly positioned sliders on the preceding slider screen. In each choice list (see Table 1.1), subjects faced two payment alternatives. Alternative A was a two-outcome lottery with possible payments of 0 or 3 points, each with 50% probability. It remained constant in all rows of a choice list table. Choosing alternative B, a subject earned 3 points if she had positioned at least a given number of sliders correctly on the previous slider screen and 0 points otherwise. Starting from 1, the required number of correctly positioned sliders increased by 1 in each row of the choice list when moving from top to bottom.

Table 1.1. A Choice List

	Alternative A	Alternative B
Decision 1	3 points with 50% probability 0 points with 50% probability	3 points if you positioned at least 1 slider correctly 0 points if you positioned less than 1 slider correctly
Decision 2	3 points with 50% probability 0 points with 50% probability	3 points if you positioned at least 2 sliders correctly 0 points if you positioned less than 2 sliders correctly
⋮	⋮	⋮
Decision 11	3 points with 50% probability 0 points with 50% probability	3 points if you positioned 11 sliders correctly 0 points if you positioned less than 11 sliders correctly

In any given row h, alternative B yielded a higher expected payoff than alternative A, if and only if a subject's performance on the previous slider screen exceeded h with more than 50% probability. In line with this reasoning, a subject should choose alternative B if and only if she believed that with more than 50% probability she had positioned h sliders correctly. Choosing alternative B in row h, a subject should have chosen alternative B in all the rows above row h (single switching point).[6] Since subjects chose between two risky alternatives, they should always prefer a higher winning probability. As a consequence, the switching point is independent of subjects' risk attitudes. In contrast to non-incentivized measures often found in the psychology literature, economists typically use incentives to measure beliefs. To incentivize truthful revelation, one row of each choice list was randomly selected for payment at the end of the experiment. Subjects had to click a "finish" button to exit a choice list screen. After leaving a choice list screen, the next slider screen was

6. In order to reduce the number of clicks subjects had to make, after a subject had selected alternative A in one row, alternative A would automatically be selected in all rows below. Subjects still had the opportunity to revise the resulting switching point.

shown automatically. After 20 pairs of slider and choice list screens, the experiment moved on to Stage 2.

Stage 2: In Stage 2, we elicited each subject's belief distribution on the overall number of correctly set sliders in Stage 1 using a ball allocation task Delavande and Rohwedder (adapted from 2008, see Figure 3).[7] In the ball allocation task, each subject had 100 balls, each representing one percentage point of belief. Subjects had to allocate them among 11 bins representing the intervals [0, 20], [21, 40],..., [201, 220]. The number of balls a subject allocated into each bin indicates the probability, with which the subject believed that the actual number of correct sliders fell in the bin's interval. The allocation of balls therefore approximates a subject's belief distribution.

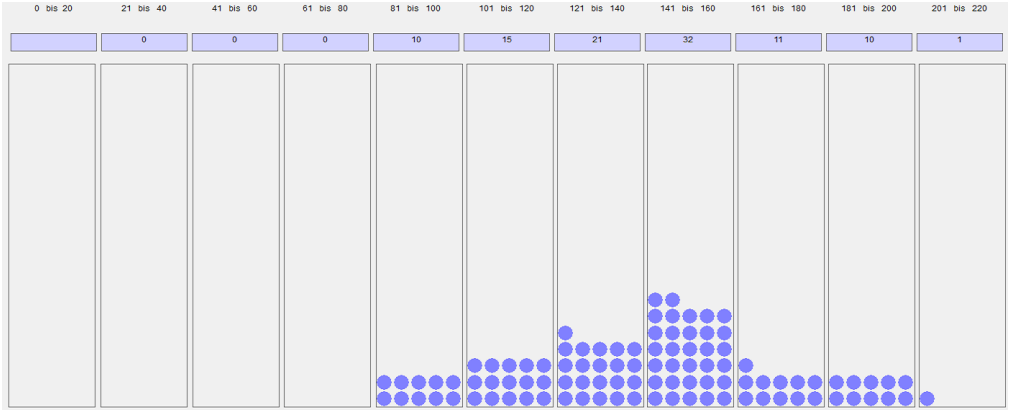


Figure 1.3. Ball Allocation Task

The ball allocation task was incentivized using the randomized Quadratic Scoring Rule (rQSR) adapted from Drerup, Enke, and Von Gaudecker (2017) and Schlag and Weele (2013). For each subject i, we first computed a number Y_i following the formula below:

$$Y_i = \sum_{j=1}^{11} (b_i^j - 100 \times 1_j)^2,$$

where $j \in \{1, 2, 3, ..., 11\}$ denotes the respective bin, b_i^j denotes the number of balls subject i assigned to bin j, indicator 1_j equals 1 for the bin that contains the actual number of correctly positioned sliders and 0 otherwise. Y_i is increasing in the number of balls a subject allocated into the *wrong* bins. Y_i has a minimum of 0 and a maximum of 20,000. Subject i obtained 30 points if and only if $Y_i < X_i$, where X_i is

7. We implemented both methods of belief elicitation (choice lists and ball allocation task) since we were not sure which method would be easier for the subjects to understand. In Section 1.5, we report the main results using beliefs elicited by the ball allocation task, and use the beliefs elicited by the choice lists as a robustness check. Please see subsection 1.5.4 for the reasons for doing so.

a random number drawn from the uniform distribution $U[0, 20000]$. The probability of winning the lottery increases in the number of balls allocated to the correct bin, while the magnitude of the reward remains fixed. This payment procedure incentivizes subjects to reveal their beliefs truthfully without imposing assumptions on their risk preferences (Schlag and Weele, 2013), since all subjects alike will strive for maximizing the winning probability of the lottery by allocating the balls in accordance with their true belief distribution.

Stage 3: Subjects were randomly assigned to one of two treatments (INFO or NOINFO), in which subjects either received feedback on their own ability or not. In the INFO treatment, the computer screen of each subject displayed her own actual number of correctly set sliders on each slider screen in Stage 1 and their aggregate, along with the corresponding beliefs. Subjects could read this private information for up to 2 minutes and could proceed to Stage 4 by clicking a "continue" button. After 2 minutes, they received a reminder urging them to click "continue". In the NOINFO treatment, subjects did not receive any information. To keep the treatments similar, they were given a break of up to 2 minutes, which was announced on the screen.

Stage 4: The set-up of Stage 4 was similar to the one in Stage 1: At most twenty slider and choice list screens were shown alternately and subjects earned a piece rate of 1 point for each successfully positioned slider.[8]

In contrast to Stage 1, after completing each pair of a slider and choice list screen in Stage 4 subjects could choose between continuing to the next slider screen and terminating the slider task. Subjects knew beforehand that they could work on up to 20 slider screens. Subjects could also skip the slider task in Stage 4 altogether and enter Stage 5 directly by clicking a "terminate" button on the instruction screen at the beginning of the stage. The number of slider screens a subject worked on serves as the measure of her *effort* level, with a minimum of 0 and a maximum of 20.

Subjects could leave the laboratory once they had individually gone through all stages of the experiment. Therefore, exerting less effort made the experiment shorter. In order to avoid potential spillovers of one departure on the leave or stay decision of the remaining subjects, we invited the subjects to come to the laboratory any time within a 3-hour time range (either from 9 am to 12 am or from 2 pm to 5 pm). Observing a departure of a fellow subject did not provide information on how long she had worked. By offering subjects flexibility on when to start and end working in an experiment, we aim at mitigating ceiling effects, i.e. participants' tendency to exert maximum effort in experimental real effort tasks independent of

8. One might be concerned that subjects might have idled on the sliders screens in order to estimate the number of correct sliders accurately in the choice lists. Our data mitigate that concern: in both Stage 1 and Stage 4, subjects placed on average 96% of the sliders in the interval [40,60]. Only one subject in Stage 1 and 4 subjects in Stage 4 (out of 176 subjects) stated a belief of 0 correctly positioned sliders in all choice lists and positioned 0 or 1 slider correctly.

the incentives (e.g. Eckartz, 2014; Corgnet, Hernán-González, and Schniter, 2015; Araujo et al., 2016; Gächter, Huang, and Sefton, 2016; Goerg, Kube, and Radbruch, forthcoming).

Stage 5: We used the same ball allocation task as in Stage 2 to elicit each subject's belief distribution on the total number of correctly positioned sliders in Stage 4. Again, each subject had 100 balls, symbolizing 100 percentage points. Unlike in Stage 2, the length of intervals represented by each bin was determined by the number of screens a subject had worked on in Stage 4. For example, after working on 4 screens, a subject saw bins representing the intervals $[0, 4], [5, 8], ..., [41, 44]$. The individual-specific upper bound was the total number of sliders a subject had worked on. Subjects were incentivized in the same way as in Stage 2 such that the allocation of balls approximated the belief distribution on the number of correctly positioned sliders in Stage 4. Subjects who had skipped Stage 4 skipped also Stage 5.[9]

Final questionnaire: After Stage 5, subjects answered a questionnaire on, among other things, socio-demographics, risk and ambiguity preferences, personality traits and survey measures of absolute overconfidence, relative overconfidence, and over-precision.

Payments: Subject were informed about their level of earnings and paid in cash right after they had finished the experiment. Total earnings were the sum of the following components: the amount earned in the slider task and choice lists in Stages 1 and 4, and in the ball allocation task in Stages 2 and 5; a random payoff of either 0, 1, or 2.5 Euro for revealing risk preferences in a Holt and Laury table (Holt and Laury, 2002) and a random payoff of either 0 or 2 Euro for revealing ambiguity aversion in the questionnaire; a 1 Euro reward for answering the questionnaire and a 2 Euro show-up fee. On average, subjects earned 11.6 Euro.

Instructions and control questions: Detailed paper instructions were handed out before Stage 1 and Stage 2. Subjects read the instructions privately, they kept and could refer to them until the end of the experiment. In addition, subjects answered two control questions designed to test and improve their understanding of the corresponding tasks before each of Stage 1 and Stage 2. The correct answer to each control question consisted of more than one element. Only when all correct elements were ticked, the answer was considered correct. When a correct answer was submitted, the experiment proceeded. If an answer was wrong on the first try, a subject learned that the answer was wrong and was encouraged to try again. If subjects failed again on the second try, the correct answer was shown along with

9. Only two subjects possibly anticipated that we would ask for beliefs in Stage 5 again, sat strategically idly in Stage 4 and allocated all 100 balls to bin 1 in Stage 5 in order to earn the reward in Stage 5 with certainty. However, this strategy did not pay off: These two subjects earned 9.05 Euro and 10.60 Euro, respectively, which is less than the average payment of 11.60 Euro.

an explanation. The recorded answers to the control questions show that the vast majority subjects understood the tasks well before carrying them out.

Implementation: We run six sessions in the BonnEconLab in Bonn, Germany in October and November 2016. 180 participants aged 17 to 61 took part in the experiment (average age of 23, with 19 and 28 being the 10% and 90% quantiles, respectively). 73 of them were male and 107 were female. The subject pool consisted mainly of students from various majors in University of Bonn (89%). 89 subjects were randomly assigned to the INFO treatment and 91 to the NOINFO treatment. Treatments were randomized within sessions to balance the data with respect to time of the day, weekday, and weather etc. The experiment lasted about one hour on average. We used z-tree (Fischbacher, 2007) to implement the experiment and hroot (Bock, Baetge, and Nicklisch, 2014) to invite subjects and to record their participation. Instructions and interfaces on the client computers were written in German, as subjects were either German natives or German speaking. The Online Appendix contains an English translation of the instructions.

1.4 Definition and Identification of Absolute Overconfidence

Identifying absolute overconfidence requires two elements: an individual's belief on her own ability and information on her actual underlying ability.[10]

To identify overconfident individuals, we propose a definition of absolute overconfidence that takes into account individual *belief distributions on ability* and regards *observed ability as a noisy measure of actual ability*. Based on that definition, we offer a strategy for identifying significant overconfidence at the individual level.

Assuming ability belief distributions as opposed to point beliefs is more plausible and robust for several reasons. First, point beliefs imply that individuals are certain in their beliefs, which is often not the case. Second, when individuals are asked to reveal a point belief, it is often unclear what is elicited: mean, median, or mode of their belief distribution. Even when the moment to be measured is explicitly specified, a moment like a median or mode can be complicated to understand and measurement may eventually fail to elicit the moment accurately. Finally, a framework built on belief distributions is more general and contains point beliefs as a special case. For those reasons, belief distributions have become widely used in surveys since the early 1990s (Manski and Neri, 2013). Experimental research using belief distributions is also growing (e.g. Eil and Rao, 2011; Manski and Neri, 2013; Neri, 2015;

10. Overconfidence can result from overestimating own ability for a realistic assessment of task difficulty and from underestimating exogenous task difficulty (compare Heidhues, Kőszegi, and Strack (2018) who portray an agent who simultaneously holds beliefs on her ability and an external fundamental, which, together with effort, determine her performance). Our use of the term overconfidence covers both possible sources of absolute overconfidence.

Bruhin, Santos-Pinto, and Staubli, 2018; Gee and Schreck, 2018). Experimental economists often favor belief distribution elicitation for its superior predictive power of choice behavior, e.g. Nyarko and Schotter (2002). Eil and Rao (2011) and Bruhin, Santos-Pinto, and Staubli (2018) have used belief distributions to identify relative confidence. We build on their work by using beliefs distributions to identify absolute overconfidence at the individual level.

We define an individual as overconfident if the median of her belief distribution exceeds her actual underlying ability. The intuition behind this definition is that an individual is overconfident if she assigns more than 50% probability mass of the belief distribution to ability levels higher than her true ability, i.e. if she believes that it is more likely that her ability exceeds her true ability than vice versa. We will infer median beliefs from the ball allocation task that elicits the complete belief distribution.

We now turn to the difference between observed and actual underlying ability. Any measurement of ability is subject to noise, and therefore only partially represents the actual ability underlying the measurement. Due to temporary variation in unobserved factors such as luck or distraction, measuring ability repeatedly may reveal different observed values for the same individual given the same actual underlying ability. This measurement error could result in misclassifying individuals in terms of overconfidence. Observed ability alone is not a reliable benchmark to compare the ability beliefs with.

In order to avoid misclassification due to measurement error, we consider observed ability, i.e. the observed percentage of correct sliders, as a random draw from a distribution that is shaped by the actual underlying ability. Assume that for an individual i, the outcome of a task is binary: either a success with probability a_i or a failure with probability $1 - a_i$. a_i is the actual underlying ability of individual i. The number of realized successes then obeys the binomial distribution $B(a_i, n)$, where n corresponds to the number of observed task outcomes. By the Central Limit Theorem, the observed success rate (observed ability) is asymptotically normal,

$$q_i \sim N(a_i, \frac{a_i(1-a_i)}{n}),$$

where q_i is the observed value. With 95% probability, the actual ability falls into the confidence interval around the observed value $[q_i - 1.96\sigma_i, q_i + 1.96\sigma_i]$, where $\sigma_i = \sqrt{\frac{q_i(1-q_i)}{n}}$. The boundaries of the confidence interval can be computed using the observed ability q_i and the number of observed outcomes n. We identify an individual as significantly overconfident if the median of her belief distribution m_i exceeds the upper limit of this confidence interval, i.e. $m_i > q_i + 1.96\sigma_i$. Analogously, an individual is classified as underconfident if $m_i < q_i - 1.96\sigma_i$. Individuals with $q_i - 1.96\sigma_i \leq m_i \leq q_i + 1.96\sigma_i$ are well-calibrated since their ability beliefs are either accurate or very close to their actual ability. We summarize our definition

of significant absolute overconfidence at the individual level in definition 1.

Definition 1 In tasks with a binary outcome, an individual i is significantly overconfident if $m_i > q_i + 1.96\sigma_i$, where m_i is the median of her belief distribution, q_i is observed ability, $\sigma_i = \sqrt{\frac{q_i(1-q_i)}{n}}$, and n is the number of observed task outcomes.

1.5 Results

In this section, we first summarize key features of our data. We then identify overconfident, underconfident, and well-calibrated subjects by applying our definition of significant over- and underconfidence at the individual level and compare our classification to the one resulting from the common approach that does not address measurement error in observed ability. Finally, we provide empirical evidence on Hypotheses 1 and 2 and present several robustness checks.

The analysis relies on observations from 176 subjects, 88 in the INFO and 88 in the NOINFO treatment. We exclude one subject who stated in the final questionnaire that she exited stage 4 accidentally by pressing the wrong button and three subjects who gave wrong answers to all four control questions.

In our main analysis, we focus on median beliefs elicited by the ball allocation task instead of the choice lists. We provide the reasons for this decision and discuss the results based on median beliefs elicited in the choice lists in the section on robustness checks.

As a result of random assignment, observed abilities and median ability beliefs in Stage 1 do not differ significantly across treatments (Mann-Whitney-U test, $p = 0.85$ and $p = 0.48$, respectively).[11] Out of a total of 220 sliders, the mean number of correctly set sliders in Stage 1 is 40 in INFO and 39 in NOINFO. The average median ability belief is 108 in INFO and 107 in NOINFO. Table 1.2 below provides further summary statistics.

11. Throughout the paper, we report p-values for two-sided tests. The next paragraph describes in detail how we infer median ability beliefs from the ball allocation task.

Table 1.2. Summary Statistics (Means and Standard Deviations)

	Overall	Treatment INFO	Treatment NOINFO
Stage 1: # correctly positioned sliders	39.41	40.02	38.80
	(15.81)	(17.38)	(14.14)
Stage 1: Earnings sliders	1.97	2.00	1.94
	(0.79)	(0.87)	(0.71)
Stage 1: Median belief, choice list	98.22	98.99	97.45
	(34.08)	(33.87)	(34.47)
Stage 1: Earnings choice lists	1.34	1.32	1.37
	(0.41)	(0.42)	(0.41)
Stage 2: Median belief, ball allocation task	107.35	107.90	106.79
	(33.09)	(32.65)	(33.71)
Stage 2: Earnings ball allocation task	0.66	0.61	0.70
	(0.75)	(0.74)	(0.75)
Stage 4: Number of screens worked on	14.36	13.10	15.63
	(7.53)	(8.03)	(6.80)
Stage 4: # correctly positioned sliders	27.11	25.58	28.65
	(20.61)	(22.14)	(18.97)
Stage 4: Earnings sliders	1.36	1.28	1.43
	(1.03)	(1.11)	(0.95)
Stage 4: Median belief, choice list	53.44	32.84	74.05
	(43.81)	(29.68)	(46.05)
Stage 4: Earnings choice lists	1.02	1.04	0.99
	(0.65)	(0.71)	(0.60)
Stage 5: Median belief, ball allocation task	63.17	36.39	89.03
	(49.17)	(31.12)	(49.68)
Stage 5: Earnings ball allocation task	0.57	0.60	0.55
	(0.73)	(0.74)	(0.73)

Numbers refer to means and standard deviations are reported in brackets below the means. Median beliefs refer to correctly positioned sliders.

1.5.1 Identification of Overconfident Subjects

We first compute each subject's median belief on her own ability in the Stage 1 modified slider task, using the corresponding histogram of the belief distribution from the allocation of balls in Stage 2. The bins of the histograms are the same as the bins in the ball allocation task. We compute the median of this belief distribution m_i as

$$m_i = \beta_i - d_i \times \frac{\sum_{j=1}^{k_i} b_i^j - 50}{b_i^{k_i}},$$

where i indicates the subject, j denotes the serial number of the bins. b_i^j is the number of balls subject i allocates to bin j. d_i represents the length of the intervals. k_i

denotes the serial number of the bin that contains the median and thus satisfies $\sum_{j=1}^{k_i-1} b_i^j < 50 \leq \sum_{j=1}^{k_i} b_i^j$. $b_i^{k_i}$ is the number of balls that subject i allocates to the bin that contains the median. β_i is the upper bound of subject i's k_ith interval.

Following the identification strategy outlined in Section 1.4, we classify 166 subjects as overconfident (83 in INFO and 83 in NOINFO), 5 subjects as underconfident (3 in INFO and 2 in NOINFO) and 5 as well-calibrated (2 in INFO and 3 in NOINFO).[12] Remember that we identify an individual as significantly overconfident if the median of her belief distribution m_i exceeds the upper limit of the 95% confidence interval around her observed ability, i.e. if $m_i > q_i + 1.96\sigma_i$, as significantly underconfident if $m_i < q_i - 1.96\sigma_i$, and as well-calibrated if $q_i - 1.96\sigma_i \leq m_i \leq q_i + 1.96\sigma_i$. In our Stage 1 data, the average σ is 0.025 (standard deviation 0.005), which corresponds to 14% of the size of the average q (0.179, standard deviation 0.071). On average, this results in a substantial confidence interval around the median that classifies more subjects as well-calibrated than the standard approach. If we compare the median beliefs directly to the observed ability, 170 subjects would be classified as overconfident, 0 as well-calibrated, and 6 as underconfident. Despite the substantial confidence interval, the difference in classification using the two methods is rather small due to the strongly exaggerated beliefs elicited in Stage 2 of our experiment.

However, in principle addressing measurement error in observed ability can make a marked difference in subjects' classification as over-, underconfident, or well-calibrated. In the INFO treatment, information on their ability measured in Stage 1 reduced the upward bias in subjects' beliefs on their Stage 4 ability (elicited in Stage 5, see Table 1 in the Online Appendix). Using our overconfidence definition, which addresses measurement error in observed ability, on Stage 4 and 5 data, we would classify 40 subjects as significantly overconfident, 36 as well-calibrated, and 7 as significantly underconfident. In contrast, neglecting measurement error and comparing the median belief elicited in Stage 5 directly to observed ability in Stage 4 results in 51 overconfident, 0 well-calibrated, and 32 underconfident subjects.[13]

12. Even if the vast majority of subjects in our sample is overconfident, our analysis still provides a valid empirical test of the theoretical argument that higher beliefs on own ability induce higher effort provision. Importantly, this argument does not concern how precise the beliefs are, but their consequences for effort choice. Inspired by the comments of a referee, we run further treatments ("Mirror Image" (N=40) and "Prior Information" (N=73), each with an INFO and NOINFO version) to investigate whether the exaggerated belief distributions in our data could be driven by (i) the exact nature of the illustrative picture of the ball allocation task in the instructions or (ii) a lack of prior experience with the difficulty of the slider task. We report details on the design of these treatments and results in section 4 of the Online Appendix. In brief, we find that the share of overconfident subjects remains very high and our main results are robust to pooling all data.

13. 7 subjects (5 in treatment INFO and 2 in treatment NOINFO) skipped Stage 4 without working on a single slider. They therefore also did not participate in Stage 5. We omit these observations in the analysis of Stage 4 and 5 data, which is more conservative than setting their ability beliefs equal to zero. There is no indication of selection on observables: these 7 subjects do not differ significantly from

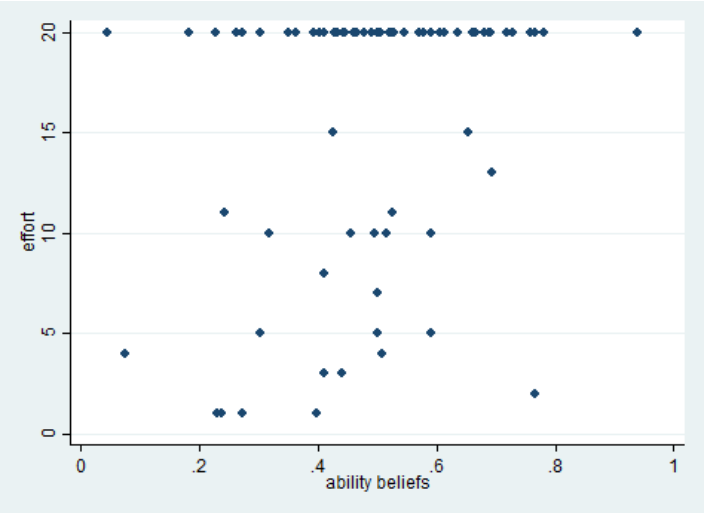
1.5.2 Result 1: The Motivational Value of Confidence

According to Hypothesis 1, higher beliefs in one's own ability lead to higher effort provision (motivational value of confidence). Figure 1.4 illustrates the correlation between individual effort choices and ability beliefs in treatment NOINFO. As hypothesized, ability beliefs and effort provision are significantly positively correlated (Pearson correlation, $r = 0.25$, $p = 0.02$). Considering overconfident subjects only, the Pearson correlation between ability beliefs and effort choices increases to $r = 0.32$, $p < 0.01$.

Although we offered subjects flexibility on when to end working to mitigate ceiling effects in effort choices, 60% of subjects exerted maximum effort by working on the maximum number of 20 slider screens (67% in NOINFO, 53% in INFO). In line with the motivation value of confidence, subjects who exerted maximum effort had significantly higher beliefs in their abilities than those who worked less: the average median beliefs on the proportion of correctly positioned sliders in Stage 4 were 39% and 31%, respectively (Mann-Whitney-U test, $p = 0.02$). This suggests that the censored nature of our data likely biases our results *against* finding a positive relation between ability beliefs and effort provision. We use Tobit regressions to address censoring in our data. Table 2 in the Online Appendix displays a Tobit regression of effort choices on ability beliefs and ability in treatment NOINFO. The results show that higher ability beliefs (but not higher ability) predict higher effort provision.

---

the remaining ones in terms of age and gender (Mann-Whitney-U tests yield $p = 0.21$ and $p = 0.15$, respectively).



Notes: The vertical axis displays chosen effort in the slider task in Stage 4, measured by the number of slider screens worked on. The horizontal axis represents subjects' median ability belief on the share of correctly positioned sliders that is elicited by the ball allocation task in Stage 5, i.e. m_i/(number of screens worked on in Stage 4 * 11 sliders). Only observations from the NOINFO treatment are displayed.

Figure 1.4. The relation between beliefs on ability and effort provision

On top of correlational evidence, our experimental design allows for providing causal evidence on the motivational value of confidence using an instrumental variable (IV) approach. An IV approach is preferable for two reasons: First, it avoids omitted variable bias due to unobserved factors that possibly affect both ability beliefs and effort provision such as optimistic predisposition.[14] Second, it eliminates potential simultaneity bias: on top of ability beliefs affecting effort provision, effort provision could possibly also affect ability beliefs, since ability beliefs were measured in Stage 5 after effort provision in Stage 4.

Our instrument is a dummy variable indicating whether a subject is assigned to treatment INFO or NOINFO. Random assignment to treatment ensures instrument exogeneity. The first stage F-statistic in Table 1.3 confirms the relevance of our instrument. The first stage results imply that informing subjects about their actual ability in treatment INFO reduces their ability beliefs by 26 percentage points on average compared to no feedback in treatment NOINFO. Thus, the random assign-

14. Personality traits like locus of control and conscientiousness could also affect both ability beliefs and effort provision but were measured explicitly. Our measure of locus of control comprises ten items adapted from Rotter (1966) that are used in the 2005 wave of the German Socio-Economic Panel. To measure conscientiousness, we use the two items proposed by Rammstedt and John (2007). Neither locus of control nor conscientiousness are significantly correlated with a subject's belief on own ability (Spearman correlations are 0.001 and 0.021, respectively, both $p > 0.78$).

ment to treatment effectively introduces substantial, exogenous variation in ability beliefs in Stage 5. The exclusion restriction that needs to be met for our instrument to be valid is that the exogenous variation in information provision in Stage 3 affects effort choice in Stage 4 only through beliefs on own ability measured in Stage 5, which we deem highly plausible: Since our instrument relies on random assignment of subjects to treatments, it is orthogonal to all kinds of subjects' characteristics such as personality traits that could affect effort choice.

The IV estimates in Table 1.3 confirm that individuals with higher ability beliefs exert more effort: on average, subjects with a 10 percentage points higher ability belief work on 0.8 additional screens according to the IV OLS estimates in column (1) of Table 1.3, i.e. they increase their effort choice by 4 percentage points. According to the IV Tobit regression in column (2) that addresses censoring, 10 percentage points higher ability beliefs increase the number of slider screens worked on by 2 (10 percentage points, $p = 0.055$).

To sum up, we find that higher ability beliefs lead to higher effort. Thus, our results are in line with the motivational value of confidence.

Table 1.3. IV regression: Causal evidence on the motivational value of confidence

	IV OLS regression (1)	IV Tobit regression (2)
Second stage: Effort level in Stage 4		
Ability belief in Stage 5	8.03**	20.19*
	(4.05)	(10.53)
Constant	12.06***	16.51***
	(0.17)	(4.14)
R^2	0.055	—
First stage: Ability belief in Stage 5		
Information	−0.26***	−0.26***
	(0.02)	(0.02)
Constant	0.49***	0.49***
	(0.02)	(0.02)
Adjusted R^2	0.404	—
N	169	169
First stage F-statistic	115.70	—

The variable *effort level* is measured by the number of screens worked on in Stage 4. The variable *information* takes a value of 1 in the INFO treatment and 0 in the NOINFO treatment, *ability belief* refers to the median belief in percentages, elicited by the ball allocation task in Stage 5. Ability beliefs are missing for 7 subjects who decided to work on zero screens in Stage 4 (5 in INFO, 2 in NOINFO).

1.5.3 Result 2: Information Reduces Overconfident Subjects' Effort Provision

In line with Hypothesis 2, providing overconfident individuals with accurate feedback on their ability led to substantially lower effort provision. On average, informed overconfident subjects worked on 13 instead of 16 screens in Stage 4 – a difference of 19% (see Figure 1.5). A Mann-Whitney-U test comparing distributions of effort provision across treatments yields $p = 0.04$.[15]

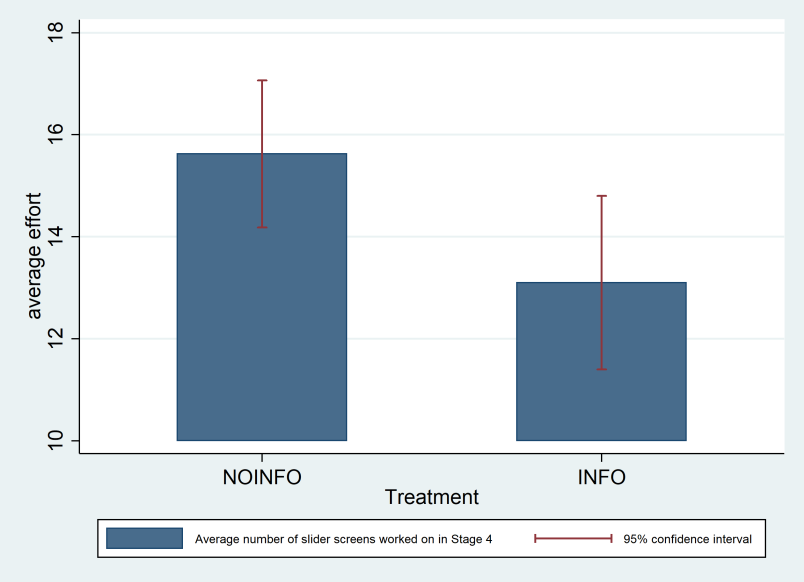


Figure 1.5. Effort provision in Stage 4

1.5.4 Robustness Checks

Before discussing the implications of our findings, we exclude several alternative explanations of our data and provide additional robustness checks.

Wealth effects: A first concern might be that receiving information about own Stage 1 ability enables subjects to infer their payment from Stage 1, which, despite the small stakes, might raise the salience of wealth effects. 83 out of 88 of subjects in treatment INFO received information that induces a downward adjustment of expected earnings. Thus, in the presence of wealth effects information should induce

15. Table 1 in the Online Appendix documents that information provision also caused a decrease in overconfident subjects' exaggerated ability beliefs. In treatment INFO, the Pearson correlation of their ability beliefs in Stage 2 and Stage 5 decreases to 0.22, $p = 0.05$, compared to 0.81, $p < 0.001$ in treatment NOINFO.

higher levels of effort provision in Stage 4 if the marginal utility of money is decreasing. Such an increased effort provision would counteract the hypothesized negative effect of information on effort provision in Stage 4. However, we find a strongly negative effect of information on effort provision. Moreover, the correlation between overall performance in Stage 1 and effort provision in Stage 4 is low and not significant in treatment INFO (Pearson correlation, $r = 0.03$, $p = 0.72$). Wealth effects had no major impact on effort provision in Stage 4.

Emotion effects of information provision: A second concern could be that information affects beliefs and consequentially effort provision not only through its content, but also through emotions like disappointment. While we consider emotions as a possible consequence of information provision, our paper focuses on the instrumental value of information content and its effect on confidence. Table 1.4 displays results of a Tobit regression of Stage 4 effort provision on the corresponding ability beliefs and a dummy variable *information* that takes the value 1 for subjects in the INFO treatment and 0 otherwise. Ability beliefs are a highly significant predictor of effort provision, while the information dummy is not (column (1)). This result remains qualitatively the same when restricting the sample to overconfident subjects only, for whom information conveys bad news (column (2)). These results suggest that information in the form of truthful feedback on ability affected effort provision through ability beliefs, while emotions played at most a subordinate role.

Table 1.4. Tobit regression

Dependent variable: Effort level in Stage 4		
	All	OC only
	(1)	(2)
Ability belief in Stage 5	19.55**	22.89**
	(9.01)	(9.34)
Information	-0.17	1.71
	(3.54)	(3.67)
Constant	16.83***	14.33***
	(4.70)	(4.90)
N	169	160
Pseudo R^2	0.012	0.013

The variable *effort level* is measured by the number of screens worked on in Stage 4. The variable *information* takes a value of 1 in the INFO treatment and 0 in the NOINFO treatment, *ability belief* refers to the median belief in percentages, elicited by the ball allocation task in Stage 5. The ability beliefs are missing for 7 subjects who worked on zero screens in Stage 4.

Median beliefs based on choice list screens: In our main analysis, we identify overconfident subjects based on median beliefs elicited by the ball allocation task instead of the choice lists for two reasons: First, median beliefs elicited by the

choice lists cannot be used to identify overconfidence as outlined in Section 1.4, which requires the median belief on a large number of sliders in order to apply the Central Limit Theorem. While the ball allocation task elicited the belief distribution regarding all 220 sliders, each choice list elicited the median of subjects' belief distribution regarding only 11 sliders. Due to the discontinuity of the support space, beliefs elicited by the choice lists are not addable – the sum of each subject's 20 medians elicited by the choice lists in Stage 1 does not necessarily equal the median belief on all 220 sliders. Second, measuring beliefs via the choice lists seems to be less intuitive for the subjects than using the ball allocation task. In the final questionnaire, 56 subjects indicated that the choice lists were more intuitive, while 124 subjects indicated the ball allocation task. Besides, subjects' answers to the control questions also suggest that the ball allocation task was easier to understand.[16]

It is reassuring that the correlation of within-subject median beliefs across the two elicitation tools is high (Pearson correlation $r = 0.70$, $p < 0.001$ in Stage 1 and $r = 0.54$, $p < 0.001$ in Stage 4). Average choice list elicited median beliefs in Stages 1 and 4 are 4.91 (std. dev. 1.70) and 3.92 (std. dev. 2.70), respectively, compared to 5.37 (std. dev. 1.65) and 3.97 (std. dev. 2.25) in the ball allocation task. The latter numbers are obtained by dividing the overall median belief by the number of screens each subject worked on in the respective stage.

We report the IV regressions with the median beliefs elicited by the choice lists in Table 1.5. In line with the IV regressions with median beliefs elicited by the ball allocation task in Table 1.3, they show a positive effect of ability beliefs on effort provision.

16. Regarding the choice lists, 48 out of 180 participants gave a wrong answer to control question 1, 15 to control question 2. When it comes to the ball allocation task, 7 subjects gave wrong answers to control question 3, and 9 subjects failed on control question 4. 14 subjects failed both control questions for the choice lists, while only 4 gave wrong answers to both control questions concerning the ball allocation task.

Table 1.5. IV regression with median beliefs elicited by the choice lists

	IV OLS regression	IV Tobit regression
Second stage: Effort level in Stage 4		
Ability belief in Stage 4	10.18*	24.85*
	(5.86)	(14.66)
Constant	11.33***	14.78***
	(2.24)	(5.46)
First stage: Ability belief in Stage 4		
Information	−0.21***	−0.21***
	(0.03)	(0.03)
Constant	0.46***	0.46***
	(0.02)	(0.02)
Adjusted R^2	0.18	—
N	169	169
First stage F-statistic	35.98	—

The variable *effort level* is measured by the number of screens worked on in Stage 4. The variable *information* takes a value of 1 in the INFO treatment and 0 in the NOINFO treatment, *ability belief* refers to the median belief in percentages, average across choice lists in Stage 4. The ability beliefs are missing for 7 subjects who worked on zero screens in Stage 4.

Using mean instead of median beliefs: A further advantage of the ball allocation task is that we can easily provide robustness checks based on moments of the belief distribution other than the median. As a final robustness check, we show that all our results remain robust when we use the mean instead of the median of the belief distribution to measure confidence. Comparing the means of the individual belief distributions elicited in Stage 2 to the 95% confidence intervals around the observed abilities in Stage 1, 167 subjects are classified as overconfident, 4 as underconfident, and 5 as well-calibrated.[17] Using information provision as an instrument for mean beliefs, an IV regression replicates the result that a 10 percentage points higher mean belief increases effort provision by 0.8 screens ($p = 0.05$, first stage F-statistic = 118). According to the corresponding IV Tobit regression that addresses censoring, 10 percentage points higher mean ability beliefs increase the number of slider screens worked on by 2.1 (10.5 percentage points, $p = 0.055$). Our data also support Hypothesis 2: overconfident subjects identified based on mean beliefs exert significantly lower effort when they are informed about their own ability. On average, they work on 13 screens in the INFO treatment as opposed to 16 screens in the NOINFO treatment (Mann-Whitney-U test, $p = 0.02$).

17. That means, one subject that was classified as well-calibrated (underconfident) using the median-based definition is re-classified as overconfident (well-calibrated) using the mean-based definition.

1.6 Conclusion

Our results provide first empirical evidence for a motivational value of absolute confidence that numerous microeconomic models build on: higher beliefs on own ability lead to higher effort provision (motivational value of confidence) and this relationship also holds for overconfident individuals. In simple decision-theoretic models (like the one in the Online Appendix), the over-provision of effort by overconfident individuals is to their detriment, since marginal costs of effort exceed its marginal benefits. In richer settings, however, overconfident individuals (Bénabou and Tirole, 2002; Compte and Postlewaite, 2004) or other parties such as their employers, team members or partners (Gervais and Goldstein, 2007; Ludwig, Wichardt, and Wickhorst, 2011a) may benefit from the exaggerated effort provision. In contrast to the general notion that accurate self-assessment enhances decisions, our results support the possibility that overconfidence can be beneficial.

We also show that de-biasing overconfident individuals by informing them about their true ability hurts their effort provision. This result sheds light on designing incentives in contexts in which diligence is appreciated. For example, to keep an overconfident employee motivated, the employer could restrain from providing accurate performance feedback.[18] Similarly, teachers may avoid providing clear-cut feedback to overconfident students, in order not to dampen their learning efforts that may still pay off in the long run.

Our findings also contribute to the empirical literature on the consequences of overconfidence and add insights on factors that motivate effort (DellaVigna and Pope, 2017). In particular, our results suggest boosting confidence as an effective and potentially cost-efficient way to enhance effort provision.

In terms of research methods, we contribute by proposing a definition of overconfidence which takes both measurement error in ability into account and the fact that individuals typically hold belief distributions on own ability. Based on this definition, we develop a method for empirically identifying significant absolute overconfidence at the individual level that can be applied more broadly in future work.

18. This implication is in line with plenty empirical evidence that subjective performance evaluations in firms often tend to be too lenient (Prendergast, 1999).

1.A Additional Tables and Results

Table 1.6. Ability and beliefs on ability of overconfident subjects

Treatment	(1) Ability Stage 1	(2) Belief on ability Stage 2	(3) Ability Stage 4	(4) Belief on ability Stage 5
NOINFO	17%	50%	16%	50%
INFO	18%	51%	16%	23%

Ability refers to the average % of correctly positioned sliders in the respective stage and beliefs reflect the corresponding average median beliefs. Beliefs on ability in Stage 1 and 4 were measured in Stages 2 and 5, respectively. We exclude 7 subjects who did not participate in Stage 4. In Stage 5, the ability beliefs of overconfident subjects in the INFO treatment were significantly lower than in treatment NOINFO (Mann-Whitney-U test, $p < 0.01$). Within the INFO treatment, overconfident subjects significantly reduced their ability beliefs after receiving information on their own ability (Wilcoxon signed-ranks test, $p < 0.01$). In contrast, in the NOINFO treatment overconfident subjects' beliefs remained stable over time (Wilcoxon signed-ranks test, $p = 0.95$). Both in treatment INFO and NOINFO, the Pearson correlations between ability measured in Stage 1 and the corresponding ability beliefs measured in Stage 2 are small and not significant (both $p > 0.18$). In treatment INFO, the Pearson correlation between ability measured in Stage 4 and the corresponding ability beliefs measured in Stage 5 is 0.31 ($p < 0.01$), while it remains insignificant ($p = 0.68$) in treatment NOINFO.

Table 1.7. Tobit regression of effort choice on ability beliefs and ability

Dependent variable: Effort level in Stage 4		
	(1)	(2)
Ability belief in Stage 5	26.64**	26.49**
	(12.92)	(12.91)
Ability measured in Stage 4		14.24
		(25.08)
Constant	13.80**	11.55
	(6.25)	(7.37)
N	86	86
Pseudo R^2	0.015	0.017

The variable *effort level* is measured by the number of screens worked on in Stage 4. The variable *ability belief* refers to the median belief in percentages elicited by the ball allocation task in Stage 5. The variable *ability* refers to the average % of correctly positioned sliders per screen in Stage 4. The regression uses observations from treatment NOINFO only ($N = 86$). 59 observations are censored from above at 20. Results in column (1) confirm that higher ability beliefs predict higher effort provision. On average, subjects with a 10 percentage points higher ability belief work on 2.6 additional screens, i.e. increase their effort choice by 13 percentage points (2.6/20, i.e. the maximum number of screens). In contrast, actual ability is not a significant predictor of effort choice (see column (2)).

1.B Model

An agent decides on an effort level $e \in [0, \bar{e}]$ to exert in a task with production function $Q(e, a)$, where a denotes her *a priori* unknown ability in this task, $a \in (0, 1]$.[19] For each unit produced, the agent gains a utility increment of $r > 0$, e.g. a piece rate payment. Effort provision induces a cost represented by the loss function $L(e)$. Suppose $Q(e, a)$ and $L(e)$ are continuous and twice differentiable. The agent's utility function is[20]

$$U(e, a) = rQ(e, a) - L(e). \tag{1.1}$$

We introduce the following assumptions:
Assumptions (i) $Q_e > 0, Q_{ee} \leq 0, \forall e, \forall a$; (ii) $Q_a > 0, \forall e > 0, \forall a$; (iii) $L_e > 0, L_{ee} \geq 0, \forall e$; (iv) $Q_{ea} > 0, \forall e, \forall a$.
Part (i) implies that given any positive ability the marginal return to effort is positive and weakly monotonically decreasing. Part (ii) assumes that production is strictly monotonically increasing in ability for any given positive effort level. Part (iii) guar-

19. For ease of exposition, we focus on positive abilities. In our experiment, zero ability leads to zero production. Proposition 1 still holds.

20. For simplicity, the utility function refers to a risk-neutral individual. Results remain qualitatively the same if we introduce risk aversion or risk proclivity.

antees that the marginal effort cost is positive and weakly monotonically increasing in effort. Part (iv) formalizes complementarity between effort and ability.

In our experiment, effort e is measured by the number of slider screens a subject works on in Stage 4 and a is operationalized as the percentage of correctly positioned sliders. Ability a is the result of various individual skills that affect performance in the modified slider task, mainly motor and visual skills, but also other personal skills like the ability to concentrate. In our data, the percentage of correct sliders per screen does not exhibit a time trend (see footnote 7), which is in line with the notion that ability should be constant over time.

In our view, the assumptions listed above are likely to be met for the following reasons. Part (i): With effort e being the number of slider screens a subject worked on, Q_e is the number of correct sliders on an additional screen. $Q_{ee} \leq 0$ captures a weakly downward trend of this number across screens. In our data, the average number of correct sliders for an additional screen worked on is 1.95, i.e. $Q_e > 0$, and the lack of time trend in the number of correct sliders per screen suggests that $Q_{ee} = 0$ (see footnote 8). Part (ii): For any positive effort level e (number of slider screens worked on), the higher the ability a (measured by the percentage of correct sliders), the higher was the total number of correct sliders $Q(e,a)$. That is $Q_a > 0$. Part (iii): In the slider task, effort cost includes time costs and the cognitive cost of concentrating. We took care that $L_e > 0$ holds in Stage 4 by letting subjects finish and leave the lab early if they decided to exert lower e by working on fewer slider screens. Given the lack of trend in the marginal return to effort ($Q_{ee} = 0$), the fact that many subjects started Stage 4 but chose to terminate it early suggests that marginal effort costs increased over time, $L_{ee} > 0$. Part (iv): In our experiment, $Q_{ea} > 0$ implies that the higher the percentage of correctly set sliders, the higher is the benefit of working on one more slider screen, a highly plausible assumption.

Under these assumptions, the following proposition holds:

Proposition 1 *A higher belief on own ability leads to weakly higher effort provision.*

Proof: Let a_l and a_h denote two ability beliefs with $a_l < a_h$, e_l and e_h the respective utility maximizing effort levels.

In a corner solution with $e_h = 0$, suppose $e_l > 0$, then the following must hold:

$$rQ_e(0)_{|a_h} < L_e(0) \tag{1.2}$$

$$rQ_e(e_l)_{|a_l} = L_e(e_l)$$

Since $Q_{ea} > 0$, $Q_{ee} \leq 0$ and $L_{ee} \geq 0$,

$$L_e(0) \leq L_e(e_l) = rQ_e(e_l)_{|a_l} < rQ_e(e_l)_{|a_h} \leq rQ_e(0)_{|a_h},$$

which contradicts 1.2. Therefore, $e_l = 0$. That is, when e_h is 0, e_l must be 0.

In an interior solution with $e_l > 0$, the following conditions must hold:

$$rQ_e(e_l)_{|a_l} = L_e(e_l). \tag{1.3}$$

As proved before, if $e_l > 0$, $e_h > 0$. Therefore

$$rQ_e(e_h)_{|a_h} = L_e(e_h)$$

Suppose $e_l \geq e_h$. Since $Q_{ea} > 0$, $Q_{ee} \leq 0$ and $L_{ee} \geq 0$,

$$rQ_e(e_l)_{|a_l} \leq rQ_e(e_h)_{|a_l} < rQ_e(e_h)_{|a_h} = L_e(e_h) \leq L_e(e_l),$$

which contradicts 1.3. Therefore, $e_l < e_h$. That is, when e_l is positive, e_h must be greater than e_l. *qed*

To introduce belief updating, we consider the following three-period model. At the center of our interest is an overconfident agent, whose prior belief on her own ability is unrealistically high. In period 1, let her ability belief be $\widehat{a}^0$, $\widehat{a}^0 > a$, where a is her actual ability.[21] The agent exerts her utility maximizing effort e^0, which results in output $q = Q(e^0, a)$, while she anticipates to produce $\widehat{q}^0 = Q(e^0, a^0)$. Since Q is monotonically increasing in a, $q < \widehat{q}^0$. In period 2, the agent is informed about the real output q and, as a response, adjusts her ability belief to $\widehat{a}^1$, which satisfies $Q(e^0, \widehat{a}^1) = q$. Given $Q_a > 0$, it must hold that $\widehat{a}^1 < \widehat{a}^0$. That is, faced with adverse feedback, the agent adjusts her ability belief downwards. In period 3, the agent exerts her utility maximizing effort e^1 regarding $\widehat{a}^1$. As $\widehat{a}^1 < \widehat{a}^0$, it follows directly from Proposition 1 that $e^1 < e^0$ if $e^0 > 0$ and $e^1 = 0$ if $e^0 = 0$.

Corollary 1 *For a positive initial effort level, an unexpectedly low ability feedback causes a decrease in effort provision.*

Median Utility Maximization

When belief distributions are taken into account, $\widehat{a}^0$ and $\widehat{a}^1$ refer to the medians of individual belief distributions m^0 and m^1. Given the monotonicity of the production function in argument a, inserting a median belief into the utility function gives the median of the agent's ex-ante belief distribution on her ex-post utility, implying that the agent exerts the effort level that maximizes the median of her utility distribution. That is the agent chooses the effort level that satisfies the following first order conditions:

$$rQ_e(e^0)_{|m^0} = L_e(e^0) \quad and \quad rQ_e(e^1)_{|m^1} = L_e(e^1).$$

The study of quantile maximization dates back to Manski (1988) who pointed out that "if actions are characterized by probability measures of outcomes, then we should consider rational any pattern of behavior consistent with the existence of a preference ordering on the space of these probability measures." More recently, quantile maximization was axiomatized by Rostek (2010).

21. For expositional clarity, we set aside belief distributions for now, allowing us to proceed without imposing any assumption on the belief probability distribution. We will later introduce belief distribution by using the median of the belief distribution as a in the model.

In our set-up, the intuition for assuming median utility maximization is the following: Let the optimal effort corresponding to median belief on ability be e^*. An agent who holds an ability belief distribution believes that it is unlikely (less probable than 50%) that her ability is lower than her median belief. Due to the positive monotonicity of optimal effort in ability belief (Proposition 1), she would not exert lower effort than e^*. At the same time, she believes that it is also unlikely (less likely than 50%) that her ability is higher than the median belief. Hence, the effort she exerts will not exceed e^*. Combining the arguments above, her optimal effort level is e^*, which means the agent chooses her effort provision to maximize median utility.

1.C Experimental Instructions

Instructions that were printed on paper

This section contains an English translation of the experimental instructions that were originally written in German. Subjects received a printed version of the instructions for part 1 once they sat down in the laboratory.

The experiment
General explanations

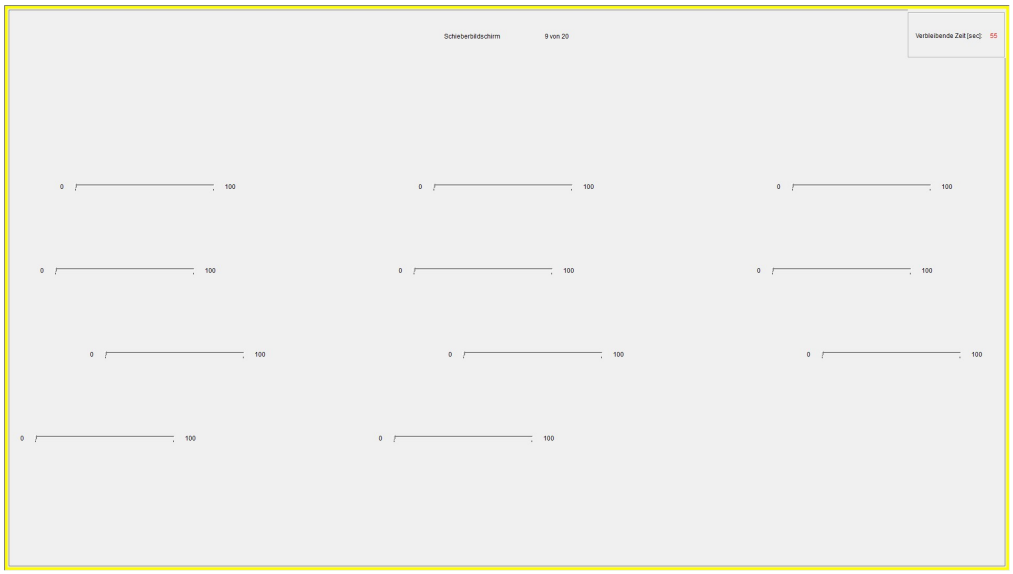
Welcome to this economic experiment.
In the course of this experiment you can earn a nonnegligible amount of money. The exact amount strongly depends on your decisions. So please read the following instructions carefully! If you have any questions, please raise your hand and we will come to your seat.
During the whole experiment, it is not allowed to talk to the other participants, to use cell phones, or to launch any programs on the computer. Disregarding any of these rules will lead to your exclusion from the experiment and from all payments.
The earnings resulting from your decisions will be paid out to you in cash at the end of the experiment. During the experiment we do not talk about Euro but points. Consequently, our total payment will be calculated in points first. At the end of the experiment, your total points will be converted into Euro, using the following rule:

1 point = 5 cents

Additionally, you will receive 40 points for showing up on time that you will be paid at the end of the experiment independent from your other decisions in the experiment. On the following pages, we will describe the exact experimental procedure. The experiment consists of 5 consecutive parts.

Your decision in part 1

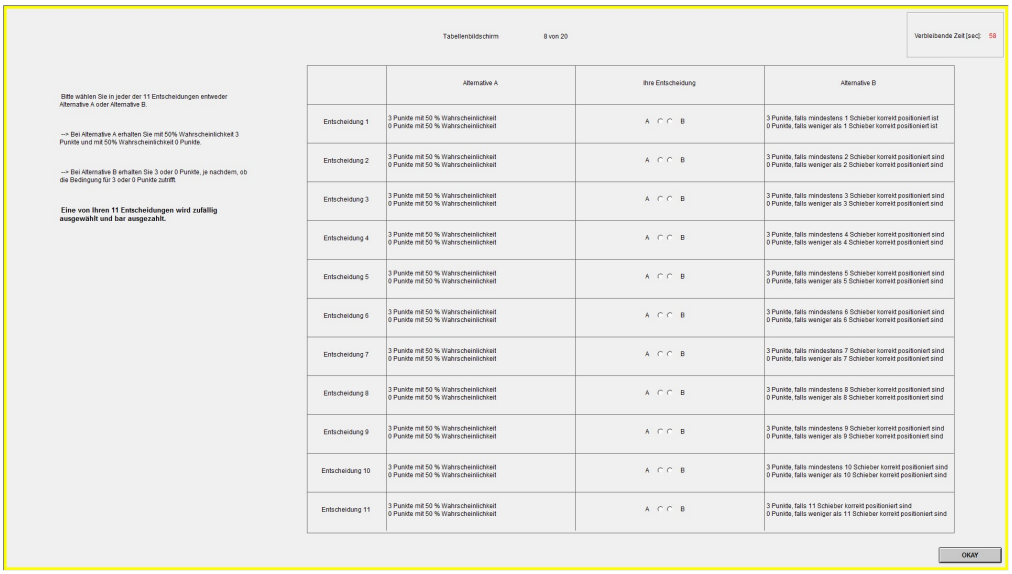
In **part 1** you will see a screen with 11 sliders ("slider screen") and a "table screen" alternately, each screen 20 times.
You have 55 seconds of time to work on each slider screen. The remaining time will be shown in the upper right corner of the screen. When the time is over, the next table screen will be shown automatically.
Your task on each **slider screen** is to **position as many of the 11 sliders in the centre of the respective scale as possible**.



Initially, each slider is at the very left end of the scale (position 0) and can be moved on the scale with the help of the mouse, maximally to the very right end of the scale (position 100). To move the slider you have to press the left mouse button. However, the current position of the slider is not shown, so that you have to estimate on your own where the middle of the scale is. The arrows on the keyboard and the mouse wheel are deactivated. When the working time for a slider screen is over the number of sliders you positioned correctly will be counted and saved by the computer automatically. At this point in time you do not yet get feedback on how many sliders you positioned correctly.

A slider will count as correctly positioned at the middle position if you have positioned it between positions 49.5 and 50.5 (so either exactly at the middle position 50 or very nearby position 50). **For each correctly positioned slider you will get a point.**

Every slider screen is followed by a **table screen** which looks like this:



For each of the 11 lines (decision 1 to decision 11), please click on either alternative A or B with your mouse to decide which alternative you prefer. Both alternatives are lotteries. Lottery A is the same for every decision: With a probability of 50%, you will earn 3 points and with a probability of 50%, you will win 0 points.
For alternative B, you will win either 3 points or 0 points as well. **For alternative B, the number of sliders that were correctly positioned at the middle position on the previous slider screen determines whether you will get 3 points or 0 points.** The conditions that need to be met to earn 3 points become more and more demanding with each decision: If you choose alternative B in the first decision, you will earn 3 points if you positioned at least one slider correctly on the previous slider screen; in case that you did not position any sliders correctly, you will earn 0 points. If you choose alternative B in the second (third etc.) decision, you will earn 3 points if you positioned at least two (three etc.) sliders correctly; in case you positioned less than two (three etc.) sliders correctly, you will earn 0 points. If you choose alternative B in the last decision, you will earn 3 points if you positioned all sliders correctly; in case you positioned less than all 11 sliders correctly, you will earn 0 points. **Thus, it depends on your personal estimate how many sliders you have positioned correctly whether alternative A or B is more attractive for you.**
An example: You think that the probability that you positioned 6 sliders correctly is higher than 50% and the probability that you positioned 7 sliders correctly is lower than 50%. In that case, your earnings will be highest if you choose alternative B in decision 1 to 6 and alternative A in decision 7 to 11. In decision 1 to 6, you will then earn 3 points with a probability higher than 50%. If you had chosen alternative A in decisions 1 to 6, you would have earned 3 points only with a probability of 50%. If you choose alternative A in decisions 7 to 11, you will earn 3 points with a probability of 50%. If you had chosen alternative B in decisions 7 to 11, you would

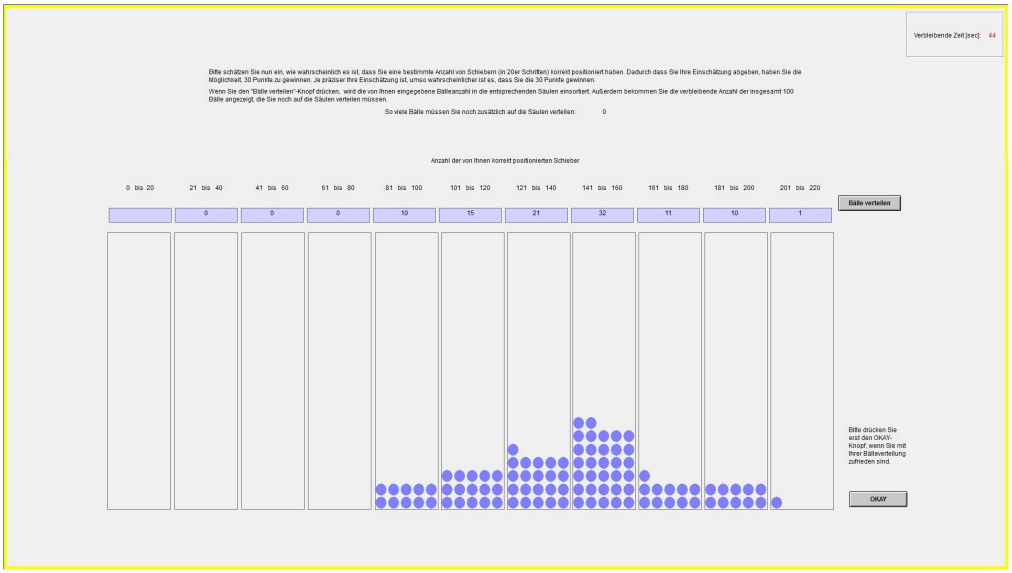
have earned 3 points with less than 50%.
As soon as you have changed from alternative B to A, the lower rows of a table will fill automatically, because alternative B's attractiveness decreases top-down. As long as you do not click the "OKAY" button you can still change your decision. As soon as you have made all eleven decisions of a given table to your entire satisfaction, please click the "OKAY" button downright. Then, the next slider screen will be displayed.
Your payoff from part 1 of the experiment is composed of the following parts:
Slider screens: For each correctly positioned slider (out of altogether 20*11=220 sliders) you will receive one point - thus, a maximum of 220 points!
Table screens: For each table screen, only one decision will be paid. So you will receive a maximum of 3 points for answering a table screen. First, one of the 11 decisions is drawn randomly for each table screen. All decisions are selected with the same probability (i.e. 1/11). Only the selected decision determines your payment. **This implies that you should make your decision in every line of each table as if this was your only decision.** In the next step, it is checked whether you have chosen alternative A or B in the selected decision. If you have chosen alternative A, the 50-50 lottery will be played and will determine your payoff. If you have chosen alternative B, you will either get 3 points or 0 points, depending on the number of sliders you positioned correctly on the respective slider screen.
An example: Suppose that in the 1st step the 4th decision of a table screen is drawn randomly. In decision 4, alternative A was chosen. In the 2nd step, it will be determined randomly whether you earn 3 points or 0 points. Both payoffs are equally likely (both have a probability of 50%). As soon as you have worked on all 20 slider and table screens, the next parts of the experiments will follow. Details on **parts 2 to 5** will be provided in the course of the experiment.

Training tasks and control questions

Before part 1 of the experiment begins, we would like to kindly ask you to answer some questions concerning your understanding of the tasks and decisions. Answering those questions will help to get acquainted with the situation, so that you can make good decisions later on.
At the end of today's experiment - right after part 5 - some screens with questions and the like will follow before you will receive your earnings.
If you have any questions right now or during the time you work on the control questions, or if you would now like to start with the control questions and the experiment, please raise your hand. We will then come to your seat to answer your questions and to start the experiment. Please do not pose your questions loudly!
Please do not press the START button on your own!

Your decisions in part 2

In part 2 of the experiment, we would like you to provide an estimate concerning the probability that you positioned a certain number of sliders correctly (in steps of 20). By providing that estimate, you have the possibility to earn 30 points. **The more precise your estimate is, the more likely it is that you will earn the 30 points.**
The screen which we use for asking for your estimate looks like this:



There are **11 pillars which each represent a certain quantity (in steps of 20) of correctly positioned sliders in part 1.** Reminder: In part 1 of the experiment you worked on 220 sliders in total. Thus, the first pillar stands for 0-20, the second pillar for 21-40 and the last pillar for 201-220 correctly positioned sliders. **Your task is to distribute 100 balls across these pillars. Each ball represents one percentage point.** If you place, e.g., 50 balls in the second pillar, that implies that you assume that with 50% probability you positioned 21-40 sliders of all 220 sliders correctly. If you place, e.g., 23 balls in the ninth pillar, that implies that with 23% probability you assume that you positioned between 161 and 180 sliders of all 220 sliders correctly. **The more likely you deem a certain pillar to contain the number of sliders you positioned correctly, the more balls should be put into this pillar.** The task will only be finished when you have distributed exactly 100 balls into the 11 pillars and you feel confident about the resulting probability distribution because it is a good fit to your estimate concerning the correctly positioned sliders. In that case, please press the “OKAY” button downright to continue with part 3 of the experiment.
To put balls into a pillar please fill in the respective number into the input field above the pillar. When you press the “Distribute balls” button, the number of balls you filled in will be put into the respective pillars. Furthermore, the remaining number of the

100 balls which you still have to distribute among the pillars will be shown. You can change the number of balls in a pillar until you press the "OKAY" button.
To sum up: The more precise your estimate - i.e. the more balls you placed in the correct pillar and the less balls you placed in the wrong pillars - the more likely it is that you will earn 30 points.
(Only) for those who are interested in the exact payoff scheme: After you have distributed all 100 balls into the 11 pillars a number A is calculated in the following way:

$$A = \sum_{i=1}^{11} (balls\ in\ pillar\ i - 100 * I_i)^2 \tag{1.4}$$

where i=1, ..., 11 refers to the different pillars and I_i equals 1 for the pillar which contains the number of the correctly positioned sliders and 0 otherwise. The more your estimate deviates from the actual number of correctly positioned sliders, the higher is A. Then, a number X is drawn randomly from the interval [0, 20000]. If $A < X$, you will win the additional 30 points. If $A > X$, you will not win further points.

Instructions provided on screen

Stage 2

Please now estimate how likely it is that you have positioned a certain number of sliders (in steps of 20) correctly. By giving your estimate, you have the possibility to win 30 points. The more precise your estimate is, the more likely it is that you will win the 30 points. When you click on the "Distribute balls" button, the balls will be assigned to the respective pillars according to the number of balls you have allocated to each pillar. Moreover, the remaining number of the overall 100 balls that you still have to assign to a pillar will be displayed on the screen.

[visual display of the ball allocation task, see above]

Stage 3

This is now Part 3 of the experiment. The table below offers you feedback how many of the 11 sliders you have positioned correctly at the middle of the scale on each of the 20 screens and how many sliders you have positioned correctly in total. You now have a maximum of 120 seconds to read the feedback. Please click on the "OKAY" button, when you would like to proceed to Part 4 of the experiment.

Screen	1	2	3	...	20	Total
Number of correctly set sliders	6	5	4	...	4	46
Self-assessment*	5	6	4	...	4	50

* According to your own answers on the choice list screens you believe that you have correctly positioned the stated or a lower number of sliders with at least 50% probability.
Comment: The numbers in the second and third row of the table are just examples.

Stage 4

This is now Part 4 of the experiment. Exactly like in Part 1 of the experiment, in Part 4 you will see one slider screen and one choice list screen alternately. The screens' layout and your task on the respective screens are exactly the same as in Part 1.

The only difference is that in Part 4 you will be shown an additional screen with two buttons before you will see each slider screen and the corresponding choice list screen. If you click on the button "Additional slider screen", one more slider screen will be displayed and you can keep on working. If you click on the button "Finish slider task", no more slider screen will be displayed; one last choice list screen will be displayed, followed by Part 5 of the experiment. **You can stop the slider task at any time. The earlier you finish working on the slider task, the earlier the experiment will end for you, and the earlier you can receive your final payment. You will earn all the points that you have collected until the end of the slider task. The more points you collect, the higher the payment, which you will receive from us in cash at the end of the experiment.** If you did not finish the slider task earlier, Part 4 of the experiment will end automatically after 20 slider and choice list screens.

As in Part 1, in Part 4 one row of each choice list will be randomly selected for payment. This means that you should make your decision in each row of each choice list as if it was your only decision. For the randomly selected row, the payment is determined by your choice between alternatives A and B. You will again receive a maximum of 3 points for answering one choice list screen. All payments in Part 4 will be added to the payment in Part 1 and 2.

Stage 5

This is now Part 5 of the experiment.
In total you have worked on 44 (example number) sliders.
Please now estimate how likely it is that you have positioned a certain number of sliders correctly. By giving your estimate, you have the possibility to win 30 points. The more precise your estimate is, the more likely it is that you will win the 30 points.
When you click on the "Distribute balls" button, the balls will be assigned to the respective pillars according to the number of balls you have allocated to each pillar. Moreover, the remaining number of the overall 100 balls that you still have to assign to a pillar will be displayed on the screen.

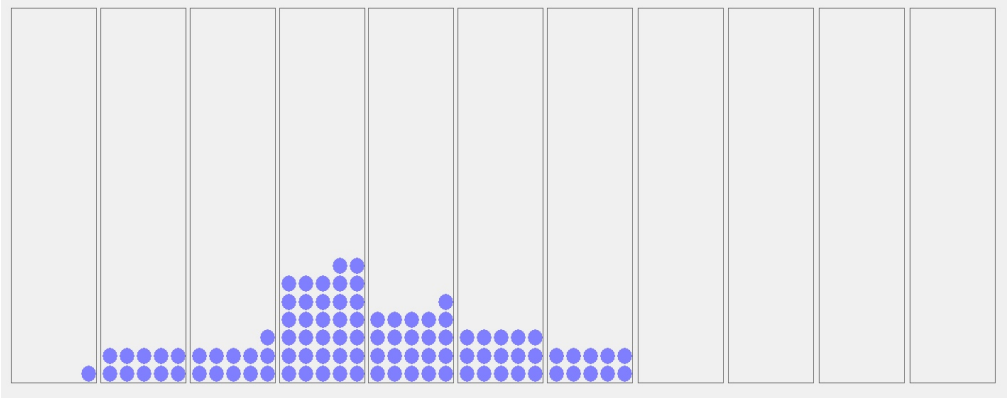
[visual display of the ball allocation task, pillar intervals adjusted to total number of sliders worked on]

1.D Robustness check treatments

Inspired by the comments of a referee, we ran further treatments at the BonnEcon-Lab in December 2018 to investigate whether the exaggerated belief distributions in our data could be driven by (i) the exact nature of the illustrative picture of the ball allocation task in the instructions or (ii) a lack of prior experience with the difficulty of the slider task.

Design

Treatments "Mirror Image" (N=21 for Mirror Image NOINFO, N=19 for Mirror Image INFO): Design and instructions were exactly as above, except for the illustrative picture of the ball allocation task in the instructions being replaced by its mirror image:



Treatments "Prior Information" (N=33 for Prior Information NOINFO, N=40 for Prior Information INFO): Design and instructions were exactly as above, except for amending the first paragraph of section "Training tasks and control questions" in the instructions in the following way: "Before Part 1 of the experiment begins, you will have the opportunity to get acquainted with the slider task by working on a slider screen with 9 sliders. You have 45 seconds of time to do so, i.e. 5 seconds per slider (exactly as in Part 1). Subsequently, you will get a brief feedback how many of the 9 sliders you positioned correctly. Moreover, we would like to kindly ask you to answer some questions concerning your understanding of the tasks and decisions. Answering those questions will help to get acquainted with the situation, so that you can make good decisions later on."

Results

Compared to the average median ability belief elicited in Stage 2 in the main treatments (49%), the median ability belief elicited in Stage 2 is 5 percentage points lower in treatments Mirror Image (44%) and 15 percentage points lower in treatments Prior Information (34%, Mann-Whitney-U test $p = 0.06$ and $p < 0.01$). These

results suggest that the exact nature of the picture of the ball allocation task has a moderate effect on beliefs, and a lack of prior experience and feedback regarding the slider task affect ability beliefs to a larger extent. Similarly, the share of overconfident, well-calibrated and underconfident individuals does not differ significantly in the main treatments and treatments Mirror Image (Chi^2 test, $p = 0.55$), while fewer subjects are overconfident and more well-calibrated or underconfident in treatment Prior Information than in our main treatments (Chi^2 test, $p < 0.01$). Table 1.8 displays the exact shares:

Table 1.8. Subject classification

	Overconfident	Well-calibrated	Underconfident
Main treatments	166 (94%)	5 (3%)	5 (3%)
Treatments Mirror Image	39 (97.5%)	1 (2.5%)	0 (0%)
Treatments Prior Information	54 (74%)	10 (14%)	9 (12%)

The table displays shares of subjects who fall in each of the respective categories. As in the main text, we classify subjects according to their performance in Stage 1 and beliefs in Stage 2.

Reassuringly, our two main results remain robust when pooling data from all treatments. In a treatment-wise analysis, results are quantitatively similar in the main, Mirror Image and Prior Information treatments, however not significant in treatments Mirror Image and Prior Information due to the low number of observations.

Pooling all data from all six treatments, we can replicate the result that individual effort provision is increasing in beliefs on one's own ability. In an IV Tobit regression of effort on ability beliefs using the information dummy as instrument (as in Table 1.3), the pooled data yield a coefficient of 21.1, $p = 0.02$ similar to 20.2, $p = 0.06$ in the main treatments (Mirror Image treatments only: 23.7, $p = 0.30$; Prior Information treatments only, in which our instrument is only weak: 23.6, $p = 0.51$).

The pooled data also reflect that negative debiasing information on individual ability diminishes effort provision. We report in Result 2 that, on average, overconfident subjects in treatment INFO worked on 13 screens in Stage 4, compared to 16 screens the NOINFO subjects worked on (Mann-Whitney-U test, $p = 0.04$). The corresponding numbers are 13 compared to 15 screens in the pooled data (Mann-Whitney-U test, $p < 0.01$), 13 compared to 15 screens in treatments Mirror Image (Mann-Whitney-U test, $p = 0.35$) and 11 compared to 13 screens in treatments Prior Information (Mann-Whitney-U test, $p = 0.19$).

References

Araujo, Felipe A., Erin Carbone, Lynn Conell-Price, Marli W. Dunietz, Ania Jaroszewicz, Rachel Landsman, Diego Lame, 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: 1–12. [8, 16]

Barron, Kai, and Christina Gravert. 2018. "Confidence and Career Choices: An Experiment." *WZB Discussion Paper, SP II 2018-301r*, [8]

Bartling, Björn, Ernst Fehr, Michel André Maréchal, and Daniel Schunk. 2009. "Egalitarianism and Competitiveness." *American Economic Review* 99 (2): 93–98. [5]

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics* 117 (3): 871–915. [5, 7, 10, 29]

Benartzi, Shlomo. 2001. "Excessive Extrapolation and the Allocation of 401 (k) Accounts to Company Stock." *Journal of Finance* 56 (5): 1747–1764. [5]

Blavastkyy, Pavlo R. 2009. "Betting on Own Knowledge: Experimental Test of Overconfidence." *Journal of Risk and Uncertainty* 38: 39–49. [9]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "Hroot-Hamburg Registration and Organization Online Tool." *European Economic Review* 71: 117–120. [17]

Bruhin, Adrian, Luís Santos-Pinto, and David Staubli. 2018. "How Do Beliefs About Skill Affect Risky Decisions?" *Journal of Economic Behavior & Organization* 150: 350–371. [17, 18]

Camerer, Colin, and Dan Lovallo. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review* 89 (1): 306–318. [5]

Clark, Jeremy, and Lana Friesen. 2009. "Overconfidence in Forecasts of Own Performance: An Experimental Study." *Economic Journal* 119 (1): 229–251. [9]

Compte, Olivier, and Andrew Postlewaite. 2004. "Confidence-Enhanced Performance." *American Economic Review* 94 (5): 1536–1557. [10, 29]

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): 284–301. [8, 16]

Danz, David. 2014. "The Curse of Knowledge Increases Self-Selection into Competition: Experimental Evidence." *WZB Discussion Paper SP II 2014-207*, [5]

Delavande, Adeline, and Susann Rohwedder. 2008. "Eliciting Subjective Probabilities in Internet Surveys." *Public Opinion Quarterly* 72 (5): 866–891. [14]

DellaVigna, Stefano, and Devin Pope. 2017. "What Motivates Effort? Evidence and Expert Forecasts." *Review of Economic Studies* 85 (2): 1029–1069. [8, 29]

Dohmen, Thomas, and Armin Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101 (2): 556–590. [5]

Drerup, Tilman, Benjamin Enke, and Hans-Martin Von Gaudecker. 2017. "The Precision of Subjective Data and the Explanatory Power of Economic Models." *Journal of Econometrics* 200 (2): 378–389. [14]

Eckartz, Katharina. 2014. "Task Enjoyment and Opportunity Costs in the Lab: The Effect of Financial Incentives on Performance in Real Effort Tasks." *Jena Economic Research Papers 2014-005*, [8, 16]

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): 114–38. [17, 18]

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171–178. [17]

Fischer, Mira, and Dirk Sliwka. 2018. "Confidence in Knowledge or Confidence in the Ability to Learn: An Experiment on the Causal Effects of Beliefs on Motivation." *Games and Economic Behavior* 111: 122–142. [7, 8]

Gächter, Simon, Lingbo Huang, and Martin Sefton. 2016. "Combining Real Effort with Induced Effort Costs: the Ball-Catching Task." *Experimental Economics* 19 (4): 687–712. [8, 16]

Galasso, Alberto, and Timothy S. Simcoe. 2011. "CEO Overconfidence and Innovation." *Management Science* 57 (8): 1469–1484. [5, 9]

Gee, Laura K., and Michael J. Schreck. 2018. "Do Beliefs About Peers Matter for Donation Matching? Experiments in the Field and Laboratory." *Games and Economic Behavior* 107: 282–297. [18]

Gervais, Simon, and Itay Goldstein. 2007. "The Positive Effects of Biased Self-Perceptions in Firms." *Review of Finance* 11 (3): 453–496. [5, 7, 10, 29]

Gervais, Simon, James B Heaton, and Terrance Odean. 2011. "Overconfidence, Compensation Contracts, and Capital Budgeting." *Journal of Finance* 66 (5): 1735–1777. [7, 8]

Gill, David, and Victoria Prowse. 2019. "Measuring Costly Effort Using the Slider Task." *Journal of Behavioral and Experimental Finance* 21: 1–9. [6, 11, 12]

Goerg, Sebastian J., Sebastian Kube, and Jonas Radbruch. Forthcoming. "The Effectiveness of Incentive Schemes in the Presence of Implicit Effort Costs." *Management Science*, [8, 16]

Heidhues, Paul, Botond Kőszegi, and Philipp Strack. 2018. "Unrealistic Expectations and Misguided Learning." *Econometrica* 86 (4): 1159–1214. [17]

Hirshleifer, David, Angie Low, and Siew Hong Teoh. 2012. "Are Overconfident CEOs Better Innovators?" *Journal of Finance* 67 (4): 1457–1498. [5, 9]

Hollard, Guillaume, Sebastien Massoni, and Jean-Christophe Vergnau. 2016. "In Search of Good Probability Assessors: An Experimental Comparison of Elicitation Rules for Confidence Judgments." *Theory and Decision* 80: 363–387. [9]

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92 (5): 1644–1655. [16]

Kennedy, Jessica A., Cameron Anderson, and Don A. Moore. 2013. "When Overconfidence is Revealed to Others: Testing the Status-Enhancement Theory of Overconfidence." *Organizational Behavior and Human Decision Processes* 122 (2): 266–279. [5]

Krähmer, Daniel. 2007. "Equilibrium Learning in Simple Contests." *Games and Economic Behavior* 59 (1): 105–131. [5, 7]

Ludwig, Sandra, and Julia Nafziger. 2011. "Beliefs About Overconfidence." *Theory and Decision* 70 (1): 475–500. [9]

Ludwig, Sandra, Philipp C. Wichardt, and Hanke Wickhorst. 2011a. "On the Positive Effects of Overconfident Self-Perception in Teams." *Available at SSRN 1854465*, [5, 7, 10, 29]

Ludwig, Sandra, Philipp C. Wichardt, and Hanke Wickhorst. 2011b. "Overconfidence Can Improve and Agent's Relative and Absolute Performance in Contests." *Economics Letters* 110: 193–196. [5, 7]

Malmendier, Ulrike, and Geoffrey Tate. 2005. "CEO Overconfidence and Corporate Investment." *Journal of Finance* 60 (6): 2661–2700. [5, 9]

Malmendier, Ulrike, and Geoffrey Tate. 2008. "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction." *Journal of Financial Economics* 89 (1): 20–43. [5, 9]

Malmendier, Ulrike, Geoffrey Tate, and Jon Yan. 2011. "Overconfidence and Early-Life Experiences: the Effect of Managerial Traits on Corporate Financial Policies." *Journal of Finance* 66 (5): 1687–1733. [9]

Manski, Charles F. 1988. "Ordinal Utility Models of Decision Making under Uncertainty." *Theory and Decision* 25 (1): 79–104. [33]

Manski, Charles F., and Claudia Neri. 2013. "First- and Second-Order Subjective Expectations in Strategic Decision-Making: Experimental Evidence." *Games and Economic Behavior* 81: 232–254. [17]

Moore, Don A., and Paul J. Healy. 2008. "The Trouble with Overconfidence." *Psychological Review* 115 (2): 502. [5, 6]

Neri, Claudia. 2015. "Eliciting Beliefs in Continuous-Choice Games: a Double Auction Experiment." *Experimental Economics* 18 (4): 569–608. [17]

Nyarko, Yaw, and Andrew Schotter. 2002. "An Experimental Study of Belief Learning Using Elicited Beliefs." *Econometrica* 70 (3): 971–1005. [18]

Plous, Scott. 1993. *The Psychology of Judgment and Decision Making*. McGraw-Hill Book Company. [5]

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1): 7–63. [29]

Rammstedt, Beatrice, and Oliver P. John. 2007. "Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German." *Journal of Research in Personality* 41: 203–212. [23]

Rosa, Leonidas Enrique de la. 2011. "Overconfidence and moral hazard." *Games and Economic Behavior* 73 (2): 429–451. [7, 8]

Rostek, Marzena. 2010. "Quantile Maximization in Decision Theory." *Review of Economic Studies* 77 (1): 339–371. [33]

Rotter, Julian B. 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs: General and Applied* 80 (1): 1–28. [23]

Sandroni, Alvaro, and Francesco Squintani. 2007. "Overconfidence, Insurance, and Paternalism." *American Economic Review* 97 (5): 1994–2004. [5]

Santos-Pinto, Luís. 2008. "Positive Self-Image and Incentives in Organisations." *Economic Journal* 118 (531): 1315–1332. [5, 7, 8]

Santos-Pinto, Luís. 2010. "Positive Self-Image in Tournaments." *International Economic Review* 51 (2): 475–496. [5, 7, 8]

Sautmann, Anja. 2013. "Contracts for Agents with Biased Beliefs: Some Theory and an Experiment." *American Economic Journal: Microeconomics* 5 (3): 124–156. [7, 9]

Schlag, Karl H., and Joël van der Weele. 2013. "Eliciting Probabilities, Means, Medians, Variances and Covariances Without Assuming Risk Neutrality." *Theoretical Economics Letters* 3 (1): 38–42. [14, 15]

Schwardmann, Peter, and Joël van der Weele. 2018. "Deception and Self-Deception." *Working Paper*, [5]

Urbig, Diemo, Julia Stauf, and Utz Weitzel. 2009. "What is Your Level of Overconfidence? A Strictly Incentive Compatible Measurement of Absolute and Relative Overconfidence." *Utrecht School of Economics Discussion Paper Series 09-20*, [9]

Chapter 2

Preferences for Information in Social Decisions

Joint with Carl Heese

2.1 Introduction

People often can acquire information that is instrumental for their decision making. For example, doctors can conduct medical examinations for information before determining what drug to prescribe to patients; voters can turn to news outlets for information before voting on ethically controversial policies; and employers can screen candidates' resumes and interview them for information before a hiring decision. In these decisions, the information people acquire plays an important role in both the decision-making and the resulting welfare outcomes.

This paper theoretically analyzes how people acquire information, in decisions where pursuing one's own material benefits *might* harm others. Our theoretical analyses speak to many decisions across a wide span of contexts: how do doctors examine the patients, if for one of the drugs they would receive a kickback? How do voters select news to read, if one of the policies entails them paying more taxes? How do discriminatory employers screen job candidates, if they personally prefer candidates of a certain gender or race?

Much empirical evidence has shown that many people depart from maximizing their self-interest, if doing so benefits others.[1] This means that these individuals' decisions are not sorely governed by their material desires. The recent research on *motivated reasoning* shows that many people deviate from complete egoism in order

1. For example, people donate to charity (e.g. DellaVigna, List, and Malmendier, 2012), pay postage to return misdirected letters (e.g. Franzen and Pointner, 2013), and share wealth with strangers in laboratory dictator games (e.g. Forsythe, Horowitz, Savin, and Sefton, 1994).

to 'feel moral' (for a review, see Gino, Norton, and Weber, 2016).[2] It argues that, in social decisions, individuals can behave selfishly without a guilty conscience if they can make themselves *believe* that the selfish decision harms no others (for a review, see Gino, Norton, and Weber, 2016).

In this paper, we propose a model of an agent who gains utility from not only her material benefits, but also her *beliefs* in the innocuousness of her decision. The model offers a useful tool for empirical scientists from different literatures, particularly the literatures on social preferences, information preferences, and motivated reasoning. It provides a positive theory that does not only organize many empirical findings from the existing literatures, but also generates new insights. The model is stylized: there is an uncertain, binary state of the world. The agent chooses between two actions. In the first state, the first action is harmless to others, and in the second state the second action. The agent has two motives: first, she gains material utility $r \geq 0$ from the first action. Second, she receives utility from believing that her action does not harm somebody else. The more likely she thinks her action is harmful, the higher her belief utility. The agent can acquire *costless* information about the state. Building on the Bayesian persuasion literature (Kamenica and Gentzkow, 2011), we render the information acquisition process a process of self-persuasion: the agent's sender-self attempts to persuade her receiver-self to behave selfishly by strategically acquiring a signal about the state. After the signal realizes, the agent's receiver-self chooses the action that maximizes her expected utility given the realized (Bayesian) posterior belief. The geometric approach of the Bayesian persuasion literature yields a versatile tool for studying *information preferences* that arise from self-persuasive motives in social decisions where benefiting oneself can harm others.

We provide four sets of results. First, we solve for the optimal signal globally, and then solve the optimal signal in a given subset of signals of interest. The material motive for one action offers a natural interpretation for information that increases the belief in the innocuousness of this action as "good news". We show that, when the agent has access to all possible signals, the material motive $r \geq 0$ causes her to never choose a positively skewed signal, i.e. a signal where good news are more informative about the state than bad news. We provide testable predictions on *preferences for skewness*, which suggest that offering negatively skewed information instead of positively skewed information can encourage more individuals to acquire information. In the laboratory experiment presented in Chapter 3, we find empirical evidence in line with this prediction.

2. Other explanations of the deviation include increasing utility in others' payoffs (Andreoni, 1990; Andreoni and Miller, 2002), preference for equity (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Andreoni and Miller, 2002), social image concerns (Andreoni and Bernheim, 2009), and social efficiency (Charness and Rabin, 2002a). These explanations focus on preferences over final distribution of payment and assume expected utility maximizer. They cannot explain the context-dependent information acquisition observed in for example Dana, Weber, and Kuang (2007), Feiler (2014), and Grossman (2014).

Second, we provide results regarding the externalities that might not be obvious at first sight. Although one might think that strategic information acquisition motivated by selfish interests must lead to more negative externalities, our model shows that also the reverse can happen: for some agent types, motivated information acquisition *improves* the welfare of the others affected by the decision. This counter-intuitive result rests on the observation that an "unmotivated" agent faces a moral hazard problem: when unmotivated, some agent types acquire only a small amount of information due to, for example, the satisficing behavior (Simon, 1955). The agent's selfish preference for one option over the other can mitigate this moral hazard problem by causing her to acquire more information in order to choose her least-preferred option only when she is certain that it is harmless to others. This result implies that delegating information acquisition to a neutral investigator might lower the welfare of the others affected by the decision.

Third, we provide a dynamic interpretation for the persuasion model and results about the dynamics of information acquisition in social decisions. The static persuasion model can be understood as a dynamic model where an agent observes a continuous flow of information and can commit to a stopping strategy, which, in expectation, generates a distribution of stopped posterior beliefs. Importantly, we argue that the typical commitment assumption is particularly weak in models of self-persuasion: we show that it is equivalent to assuming that the sender-self releases the received information flow unmodified to the receiver-self. However, such modifications we deem to be empirically unrealistic since they require a considerable degree of sophistication. The dynamic model generates the testable prediction that more individuals would continue acquiring information having received mostly information supporting the innocuousness of the lucrative action, and less individuals would continue acquiring information having received mostly information against the innocuousness of the lucrative action. In the laboratory experiment presented in Chapter 3, we find empirical evidence in line with this prediction.

Finally, we propose a parametric family of other-regarding preference types and further results that are useful for structural analysis.

The paper relates to three streams of the empirical literature. First, the paper relates to the recent research on information preferences. Joining the discussion on information preferences (e.g. Masatlioglu, Orhun, and Raymond, 2017), our results show that in social decisions where benefiting oneself can harm others, the agent can never favor a positively skewed information. The resulting policy suggestion is that offering negatively skewed information can reduce socially harmful information avoidance. Second, it relates to the literature on motivated reasoning in social decisions. We show that the model predicts the avoidance of perfectly revealing information, as observed by Dana, Weber, and Kuang (2007) and Feiler (2014). Third, the paper provides a model of social preferences under uncertainty, and thereby relates to the literature on social preferences. We discuss the related literature in more detail in Section 2.7 and Section 2.8.

2.2 Model Setup

An agent (she) has to make a decision between two options x and y. There is an unknown binary state $\omega \in \{X, Y\} = \Omega$ and the prior belief is that the probability of X is $p_0 \in (0,1)$. A passive agent, whom we hereafter refer to as *the other* (he), can be affected by the agent's decision between x and y – when the agent chooses an action that does not match the state, i.e. x in Y or y in X, the action has a negative externality of -1 on the other (he) and otherwise not.[3] The agent dislikes the belief that her decision harms the other. When the agent believes that state X holds with probability p and chooses $a \in \{x,y\}$, her utility is given by

$$U(a,p;r) \quad = \quad \begin{cases} u(p)+r & \text{if} \quad a=x \\ u(1-p) & \text{if} \quad a=y. \end{cases} \tag{2.1}$$

If choosing x, she receives a state-independent *remuneration* $r \geq 0$[4] and belief utility $u(p)$ for believing that her choice x is harmless for the other agent with probability p. The belief utility u is weakly increasing, and continuously differentiable; we normalize $u(1) = 0$. That is, the agent feels no disutility if she is certain that the action of her choice does not harm the receiver.[5] If choosing y, she only receives belief utility $u(1-p)$ for believing that her choice y is harmless for the other with probability $1-p$. We call u the *(other-regarding) preference type* of the agent.

Before deciding between x and y, the agent has unrestricted access to information about the state at no cost.[6] Formally, she can choose any *signal structure*, i.e., a joint distribution of a set of signals $s \in S$ and the state. For any signal structure, the distribution of her posterior beliefs $\Pr(X|s)$ conditional on the realized signal s must be Bayes-plausible.[7] In the following, we model her choice of a signal structure as the choice of a posterior belief distribution $\tau \in \Delta(\Delta(\Omega))$ from the set of Bayes-plausible distributions.[8] After choosing τ, a posterior $p \in \text{supp}(\tau)$ is drawn

3. The negativity of the externality is only a matter of normalization. Our model applies to all situations where one of the options is better for the other agent, and one worse.

4. A remuneration $r > 0$ might arise in situations where she receives a choice-contingent monetary payment, e.g. a commission, a prize or it might arise from choice-contingent non-monetary rewards, e.g. an increase in the reputation within a group or the feeling of satisfaction from a particular choice.

5. This normalization is without lost of generality. Our results hold as long as u is weakly increasing, and continuously differentiable.

6. Later, in the Online Supplement, we describe how the model naturally generalizes to the situation when information is costly.

7. A distribution of posteriors is called Bayes-plausible if the expected posterior equals the prior, that is $\sum_{p \in supp(\tau)} p \Pr_\tau(p) = p_0$.

8. It has been shown in the literature on Bayesian persuasion (Kamenica and Gentzkow, 2011) that the model where the agent can choose any signal structure and the model where she can choose any Bayes-plausible distribution of posteriors are equivalent. It is because, for any Bayes-plausible distribution of beliefs τ, there is a signal structure such that the distribution of posterior belief $\Pr(X|s)$ is τ.

by nature and privately observed by the agent. Then, she decides on $a \in \{x, y\}$ to maximize her utility given the realized posterior belief p.

2.3 Optimal Information Acquisition with Belief Utility

2.3.1 The Optimal Information Acquisition Strategy

In this subsection, we analyse the optimal information acquisition strategy of the agent. We discuss the case where $r = 0$ and the case where $r > 0$ respectively.

Preliminaries. A posterior belief determines the utility in two steps. First, it determines the agent's choice of action between x and y – the agent chooses the action that maximizes $U(a, p; r)$ for any given belief. Then, together with the chosen option, it determines the utility. Hence, we can get rid of the argument a in the utility function and directly express utility as a function of posterior belief p:

$$V(p; r) = \max_{a \in \{x,y\}} U(a, p; r). \tag{2.2}$$

$V(p; r)$ is the continuation value for any posterior realization p given remuneration r. The optimal posterior belief distribution τ is the one maximizing her expected continuation value, i.e.

$$E_\tau V(p; r) = \sum_{p \in \text{supp}(\tau)} \Pr_\tau(p) V(p; r). \tag{2.3}$$

Before analyzing the optimal posterior belief distribution, we first narrow down the space of the optimal τ in Lemma 1. It shows that for any other-regarding preference type, there exists an optimal posterior distribution τ^* that is supported on *two* (potentially identical) beliefs $\underline{p} \leq \overline{p} \in [0, 1]$. We show in the proof of Lemma 1 that the agent chooses x when her posterior belief realizes as $\overline{p}$ and chooses y when her posterior realizes as $\underline{p}$, whenever $\underline{p} < \overline{p}$. The proof is in the Appendix.

Lemma 1. *For any $r \geq 0$ and any u, there is an optimal posterior distribution τ^* with binary support $\text{supp}(\tau^*) = \{\underline{p}, \overline{p}\}$ and $\underline{p} \leq p_0 \leq \overline{p} \in [0, 1]$, where p_0 is the agent's prior belief.*

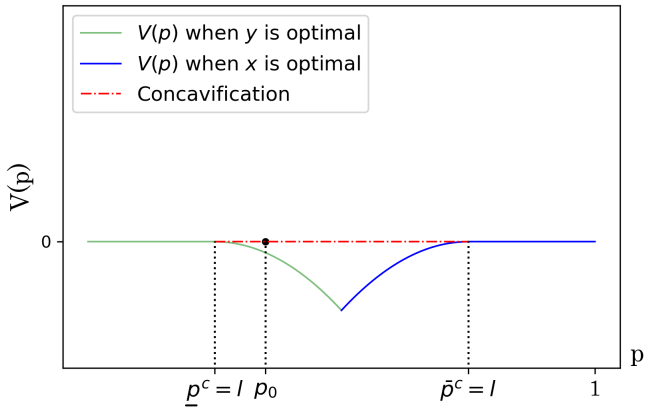
The Optimal Information Structure Given $r = 0$. First, we introduce an important threshold of belief

$$l = \min \{q : u(q) = 0\}.$$

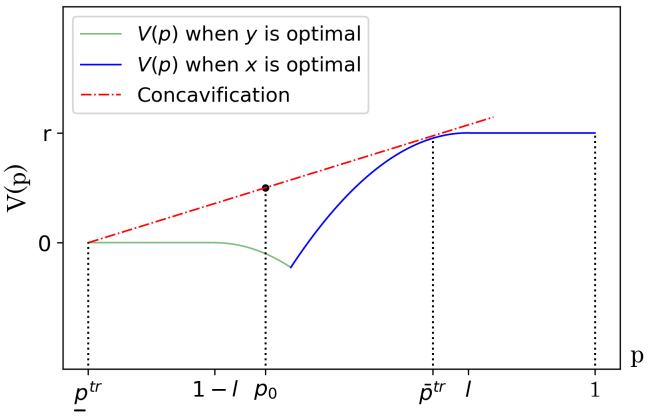
$l \leq 1$ is the threshold above which any further certainty that her chosen option is harmless does not increase her belief utility any more;[9] whereas when her belief that

9. See Simon (1955) for seminal literature on satisficing. One feature of the satisficing behavior in our setting is that the agent exhibits satisficing behavior for beliefs instead of outcomes.

her chosen option is harmless is lower than l, her utility increases when she gains additional certainty that the option is harmless. We term l the agent's *moral standard*. A moral standard $l < 1$ captures the idea of satisficing – the agent is "satisficed" if she is certain with l probability that her chosen option does not harm others, and any further certainty no longer brings her additional utility.



(a) $l < 0$ and $r = 0$



(b) $l < 0$ and $r > 0$

Figure 2.1. Illustration of Optimal Cutoffs

Figure 2.2a illustrates $V(p)$ in the scenario without remuneration, i.e. $r = 0$. The agent's only concern is her belief utility u. Whenever she is more certain than her moral standard l that her decision does not harm the other ($\underline{p}^c < 1 - l$ or $\bar{p}^c > l$), her belief utility is at its highest value 0. Therefore, any information acquisition strategy

that always makes her more certain in the state than her moral standard is optimal for her. Formally:

Theorem 2. *When $r = 0$, any cutoff pair $(\underline{p}^c, \overline{p}^c)$ with $\underline{p}^c \in [0, 1-l]$ and $\overline{p}^c \in [l, 1]$ is optimal.*

The Optimal Information Structure Given $r > 0$. Next, we turn to the optimal cutoffs when x is remunerative, i.e., $r > 0$.

In the presence of remuneration, the agent values her belief utility that she does no harm on the other, as well as the remuneration utility r given by choosing x. Regarding $\underline{p}^{tr}$, we first observe that the agent chooses the non-remunerative option y only if she is certain that y is the option harmless to the other, i.e. $\underline{p}^{tr} = 0$. This is because if choosing y for any $\underline{p}^{tr} > 0$, she can always improve her belief utility by choosing y at $\underline{p}^{tr} = 0$. Meanwhile, $\underline{p}^{tr} = 0$ minimizes the probability that $\underline{p}^{tr}$ is realized, *ceteris paribus*, so that she can choose the remunerative option x with the highest probability.

Regarding $\overline{p}^{tr}$, when she considers choosing the remunerative option x, she faces a trade-off: on the upside, she appreciates higher $\overline{p}^{tr}$, as it increases her belief utility from believing that x is harmless with higher certainty; on the downside, the higher $\overline{p}^{tr}$ is, the lower is the probability that it is realized, and hence the lower is the probability that she can choose x. This tradeoff determines the optimal cutoff $\overline{p}^{tr}$.

Among those who acquire information, there are two classes of agent types. The types in the first class acquire complete information given $r > 0$.[10] These types must have moral standard $l = 1$, i.e., they are not satisfied by any belief lesser than certainty. Besides, they value additional certainty in their beliefs so much that their marginal belief utility still exceeds the remuneration r even when their belief is already very close to certainty. Since their moral standard is equal to 1, these types also acquire complete information when $r = 0$.

The second class of types does not acquire complete information. We first formally express the tradeoff between the belief utility and the risk of the undesirable realization of $\underline{p}^{tr}$. Recall the agent's maximization problem (2.3); given that $\underline{p}^{tr} = 0$[11] and $V(0) = 0$, her maximization problem is

$$\overline{p}^{tr} = argmax_{p \in [p_0, 1]} \Pr(p) V(p; r), \tag{2.4}$$

subject to the Bayes-plausibility constraint. Bayes-plausibility, together with $\underline{p}^{tr} = 0$, implies that $\Pr(\overline{p}^{tr}) = \frac{p_0}{\overline{p}^{tr}} \in [0, 1]$. The $\overline{p}^{tr}$ therefore satisfies the following first order condition:

10. Formally they are those who satisfy the condition $u'(1) \geq r$. $u'(1) \geq r$ implies that $l = 1$. Since u is continuously differentiable, if $l < 1$, then $r > u'(1) = 0$.

11. Proof see Appendix.

$$Pr(\bar{p}^{tr})u'(\bar{p}^{tr}) + \frac{\partial Pr(\bar{p}^{tr})}{\partial p}V(\bar{p}^{tr};r) = 0 \tag{2.5}$$
$$\Leftrightarrow \quad \frac{p_0}{\bar{p}^{tr}}u'(\bar{p}^{tr}) - \frac{p_0}{(\bar{p}^{tr})^2}V(\bar{p}^{tr};r) = 0$$

. The intuition of (2.5) is that its first term describes the marginal increase in belief utility u for being more certain that the chosen option x is harmless; and its second term captures the marginal undesirable risk that the information can make the remunerative option unacceptable.

Figure 2.2b illustrates the problem geometrically. It is easy to see in (2.5) that its solution $\bar{p}^{tr}$ is where the linear line connecting point $(0,0)$ and point $(\bar{p}^{tr}, V(\bar{p}^{tr}))$ is exactly tangential to $V(p)$ at the latter. This linear line is the smallest concave function lying weakly above $V(p)$, which we refer to as the concavification of $V(p)$. The expected utility given belief cutoffs $(0,\bar{p}^{tr})$ is given by the intersection of the concavification and the vertical line above p_0.

Theorem 3 shows that when the interior solution $\tilde{p}$ exists, it must be the optimal upper cutoff $\bar{p}^{tr}$ and it must be smaller than l, i.e. $\bar{p}^{tr} = \tilde{p} < l$.[12]

Theorem 3. *When $r > 0$, for any type u with $u'(1) < r$, let $\tilde{p}$ be the interior solution of (2.5). When $p_0 \geq \tilde{p}$, the agent acquires no information; When $p_0 < \tilde{p}$, $\underline{p}^{tr} = 0$ and $\bar{p}^{tr} = \tilde{p} < l$.*

The proof of Theorem 3 is in the Appendix.

The Effect of Remuneration $r > 0$. Now we compare the optiomal information structure between the scenario where the agent has no selfish motive in the decision $(r = 0)$, and the one where she has $(r > 0)$. Corollary 4 shows that, for *all* preference types that acquire information in both scenarios, both $\underline{p}$ and $\bar{p}$ are weakly smaller when there is a remuneration for option x, i.e. $r > 0$.

Corollary 4. *Take any u and any optimal posterior belief distribution supported on $(\underline{p}^c, \bar{p}^c)$ given $r = 0$. If it is not optimal to acquire no information given $\bar{r} > 0$, then for any optimal posterior belief distribution supported on $(\underline{p}^{tr}, \bar{p}^{tr})$ given $\bar{r} > 0$,*

$$\underline{p}^{tr} \leq \underline{p}^c, \tag{2.6}$$
$$\bar{p}^{tr} \leq \bar{p}^c. \tag{2.7}$$

When the interior solution $\tilde{p}$ exists, the corollary directly follows from Theorem 2 and 3. Recall that regarding $\bar{p}$, $\bar{p}^c \geq l$ and $\bar{p}^{tr} < l$; regarding $\underline{p}$, $\underline{p}^{tr} = 0$ and $\underline{p}^c \leq 1 - l$. When the interior solution $\tilde{p}$ does not exist, the corollary is trivially true: no

12. In the Appendix we show that an interior solution of (2.5) exists when $u'(1) < r$, i.e. whenever the agent does not acquire full information.

matter $r = 0$ or $r > 0$, she always strictly prefers accurate beliefs, i.e. $\underline{p}^{tr} = \underline{p}^{c} = 0$ and $\bar{p}^{tr} = \bar{p}^{c} = 1$. [13]

Intuitively, the theorem predicts that facing a remuneration $r > 0$, individuals favor information that lead to posterior beliefs that is weakly more certain in the state Y and weakly less certain in the state X. This is because only in scenario $r > 0$ the agent wants to persuade herself to choose x, which is harmless only in state X. Lower $\bar{p}$ or lower $\underline{p}$ makes $\bar{p}$ more likely to be realized, and only after $\bar{p}$ is realized the agent choose the remunerative option x.

2.3.2 Preference for Negatively Skewed Information

In this subsection, we discuss the agent's preference for information skewness. We first define the skewness and informativeness of information. Then we derive 4 results in regard of the agent's preference for negatively skewed information.

Information Skewness. We follow Masatlioglu, Orhun, and Raymond (2017) and define the *skewness* of an information structure π as the standardized third moment of its posterior belief distribution, [14]

$$skew(\pi) = \frac{\mu_3}{\sigma^3}, \tag{2.8}$$

where $\mu_3 = \mathrm{E}((\Pr(X|s) - \Pr(X))^3|\pi)$ and $\sigma^3 = \mathrm{E}\{[\Pr(X|s) - \Pr(X)]^2|\pi\}^{\frac{3}{2}}$. Intuitively, a binary information structure is positively (negatively) skewed if it resolves more positive (negative) uncertainty. One can show that a binary signal with posteriors $\underline{p}$ and $\bar{p}$ is negatively skewed if and only if $1 - \bar{p} > \underline{p}$; and positively skewed if and only if $1 - \bar{p} < \underline{p}$.

Equally Informative Signals. Information structures do not only vary in their skewness, but also in their informativeness. The informativeness of an information signal can be measured by the variance of its posterior belief distribution.

Preference for Negatively Skewed Information. Regarding the preference for skewed information structures, our model shows the following results, each of which offers insight on encouraging information acquisition through offering (more) negatively skewed information. Formal proofs of these results are provided in the Appendix.

13. In the companion experiment (Chapter 3), we find that most individuals stop acquiring information when their beliefs are far from certainty. It suggests that the belief utility is likely concave. The concavity of the belief utility captures the following psychological mechanism: it is increasingly more uncomfortable for the individual to choose an option, as she becomes more certain that her chosen option is the one worse for the other.

14. Given a symmetric prior and for the signal structures used in our experimental design, this notion of skewness coincides with other notions from the theoretical literature, including the central third moment, third-order stochastic dominance, third-degree risk-order and the Dillenberger and Segal (2017) notion of skewness.

First, the model suggests that reversing a positively skewed information signal whose posterior belief distribution is supported on $(\underline{p}, \bar{p})$ to the analogue negatively skewed information $(1-\bar{p}, 1-\underline{p})$ would make more agent types acquire information. This reversal is particularly interesting since it preserves the informativeness of the information. The reversal increases the agent's utility of acquiring the information, because both signals generates the same belief utility while the negatively skewed information signal entails a higher chance to realize in $\bar{p}$, after which the agent chooses the remunerative option x and gains additional r utils. Therefore, if the agent prefers no information over negatively skewed information $(1-\bar{p}, 1-\underline{p})$, she also prefers no information over positively skewed information $(\underline{p}, \bar{p})$ of the same informativeness. Conversely, the agent can prefer negatively skewed information $(1-\bar{p}, 1-\underline{p})$ over no information, but does not prefer positively skewed information $(\underline{p}, \bar{p})$ over no information. Formally:

Proposition 5. *Between no information, a negatively skewed information* $(1-\bar{p}, 1-\underline{p})$ *and the positively skewed information* $(\underline{p}, \bar{p})$*, the agent acquires either no information or the negatively skewed information* $(1-\underline{p}, 1-\bar{p})$.

Proposition 5 suggests that principals can convince more people to listen to some equally informative source of information by reversing positively skewed to positively skewed information.

Next, the model suggests that the agent prefers more informative bad news, i.e., smaller $\underline{p}$, when holding the informativeness of good news, i.e. $\bar{p}$ fixed. This is because the more accurate the bad news is, the higher is the belief utility when the agent chooses the non-self-rewarding option y after receiving a bad news. Besides, the more accurate the bad news is, the more likely it is for the agent to receive good news so that she can choose the self-rewarding option x. Formally,

Proposition 6. *Consider information structures* π *and* π' *that lead to the binary posterior belief distributions supported on* $(\underline{p}, \bar{p})$ *and* $(\underline{p}', \bar{p})$ *respectively. If* $\underline{p} < \underline{p}'$*, the agent prefers information structure* π.

Proposition 6 suggests that if the agent avoids information π, she also avoids information π'. By offering more accurate bad news, an information source can broader its audience. In contrast, offering more accurate good news does not necessarily lead to more information acquisition. This is because more accurate positive information on the one hand increases the agent's belief utility when $\bar{p}$ is realized and x is chosen, but on the other hand reduces the probability that $\bar{p}$ is realized and hence the chance that the agent chooses the lucrative option x. Therefore, whether an increase in the accuracy of the good news attracts or deters an agent depends on this tradeoff, which in turn depends on the type of the agent.

Third, regarding the agent's optimal choice of information structure, the following corollary directly follows from Theorem 2 and 3.

Corollary 7. *The agent either chooses to receive no information, full information, or a piece of negatively skewed information.*

The Effect of Increasing Remuneration r. Finally, we analyze the effect of an increase in the remuneration r from choosing x, in particular generalizing Corollary 4.

Proposition 8. *When the reward r is higher,*

1. *individuals are more likely to avoid information;*
2. *when they do not avoid information, the information is more negatively skewed on average.*

2.4 The Externalities

How does a remuneration $\bar{r} > 0$ affect the welfare of the other? An option being remunerative does not only directly affect the agent's decision between the options (*the decision effect*), but also indirectly affects how she acquires information (*the information effect*). In what follows, we theoretically show that, while the decision effect of the remuneration is always negative on the welfare of the other, the information effect is positive for some agent types. This information effect can sometimes offset the decision effect and lead to an overall neutral or even positive effect of the remuneration on the welfare of the other.

This result arises from a moral hazard problem: when impartial between options, the agent might acquire little information. Therefore she sometimes mistakenly chooses the harmful option because she is ill-informed about the state. We show that one option being remunerative can mitigate this moral hazard problem. Although she now more often falsely chooses x, the agent *less* often falsely chooses y because she now requires higher certainty in the innocuousness of y before choosing it.

First, let us formally express the expected utility of the other – the passive person affected by the agent's decision between x and y. Let $v(a,\omega)$ be the utility of the other when the agent chooses $a \in \{x,y\}$ in $\omega \in \{X,Y\}$. Recall that the other has negative utility of -1 if the chosen option does not match the state and has utility 0 otherwise, i.e. $v(x,Y) = v(y,X) = -1$ and $v(x,X) = v(y,Y) = 0$.

For any given belief, the agent chooses the option $a \in \{x,y\}$ that maximizes her own utility $U(a,p;r)$ (see (2.1)). Hence for given r, we write the chosen option as a function of her belief p, i.e. $a_r(p) = \max_{a\in[x,y]} U(a,p;r)$. We call a_r the *decision rule* given r. τ pins down the joint distribution of the posterior belief realization and the state ω, given the prior belief p_0 and the Bayes-plausibility constrain. We hence can write the expected utility of the other given posterior belief distribution τ as

$$E_\tau v \equiv Ev(a_r(p),\omega)|\tau. \tag{2.9}$$

Notice in (2.9) that the agent determines the expected utility of the other by making two decisions: first, she chooses the decision rule $a_r(p)$; second, she chooses the information acquisition strategy and hence the posterior belief distribution τ. A remuneration for choosing x, i.e. $\bar{r} > 0$, affects both decisions. We call the effect of $\bar{r}$ on $E_\tau v$ through changing the decision rule $a_r(p)$ the *decision effect*; and we call the one through changing the posterior belief distribution τ the *information effect*.

We write the overall effect of the remuneration $\bar{r}$ on the expected utility of the other as

$$Ev(a_{tr}(p),\omega)|\tau^{tr} - Ev(a_{co}(p),\omega)|\tau^{co}, \tag{2.10}$$

where τ^{tr} is the optimal information acquisition strategy given $\bar{r} > 0$ and τ^{co} the optimal strategy given $r = 0$; a^{tr} is the decision rule when $\bar{r} > 0$ and a^{co} the decision rule given $r = 0$.

Next, we discuss the decision effect and the information effect of $\bar{r}$ on the expected utility $Ev(a_r(p),\omega)|\tau$ respectively.

The Decision Effect. We first discuss the decision effect, i.e. the effect of $\bar{r} > 0$ on the expected utility of the other through affecting the agent's choice between x and y given posterior beliefs. This effect can be expressed by the difference of expected utility of the other when keeping the posterior belief distribution fixed at τ^{co} and changing the decision rule:

$$\text{DE} \equiv Ev(\boldsymbol{a_{tr}}(p),\omega)|\tau^{co} - Ev(\boldsymbol{a_c}(p),\omega)|\tau^{co} \tag{2.11}$$

For any belief p, the agent chooses x over y iff

$$u(p) + r > u(1-p). \tag{2.12}$$

When there is no remuneration, i.e. $r = 0$, for any belief the agent always chooses the option that is less likely to be harmless for the other. She is indifferent between the two options at belief $p = 0.5$. However, when x is remunerative, i.e., $\bar{r} > 0$, the agent's indifferent point becomes lower – she chooses x for less certainty that it is harmless. Theorem 9 shows that this change of decision rule makes the other weakly worse off. The proof is in the Appendix.

Theorem 9. *For any $\bar{r} > 0$, any agent type $u(.)$,*

$$Ev(\boldsymbol{a_{tr}}(p),\omega)|\tau^{co} \leq Ev(\boldsymbol{a_{co}}(p),\omega)|\tau^{co},$$

i.e. the decision effect is weakly negative.

The Information Effect. Next we discuss the effect of remuneration $\bar{r} > 0$ on the expected utility of the other that is due to change of the agent's optimal information acquisition strategy τ, i.e. the information effect. This effect can be expressed by

keeping fixed the decision rule that is optimal given $\bar{r} > 0$ and changing the information acquisition strategy from τ^{co} to τ^{tr}:

$$\text{IE} \equiv Ev(a_{tr}(p), \omega)|\boldsymbol{\tau^{tr}} - Ev(a_{tr}(p), \omega)|\boldsymbol{\tau^{co}}. \tag{2.13}$$

Recall that Theorem 2 shows that when $r = 0$, an agent with moral standard $l < 1$ has optimal information acquisition strategy that does not yield perfect beliefs, i.e., cutoffs other than 0 and 1 can be optimal for the agent. It implies that when $r = 0$, if the agent is satisfied before her belief reaches certainty, a *true* moral hazard problem arises: the agent only acquires partial information about the state. Consequently, she sometimes mistakenly chooses x when the state is Y, and she also sometimes mistakenly chooses y when the state is X.

Theorem 10 shows that a self-reward of an option can serve as a motivation device and mitigate the moral hazard problem. The intuition is that when x is remunerative, the agent makes no mistakes when she chooses option y – she only chooses y when she is certain that it is the option harmless to the other. The information effect, therefore, can be positive. We also show that the positive information effect can dominate the decision effect and result in an overall positive effect of $\bar{r} > 0$ on the expected utility of the other.

Theorem 10. *There are agent types u such that the presence of a remuneration $\bar{r} > 0$ has a positive information effect on the expected utility of the other and the overall effect of $\bar{r} > 0$ on the expected utility of the other is positive.*

The proof is in the Appendix. Note that the overall effect is

$$\text{DE} + \text{IE} = Ev(a_{tr}(p), \omega)|\boldsymbol{\tau^{tr}} - Ev(a_{co}(p), \omega)|\boldsymbol{\tau^{co}},$$

namely, the difference of the other's expected utility between $\bar{r} > 0$ and $r = 0$.

2.5 The Dynamics of Motivated Information Acquisition

The model of Section 2.2 is static. Here we explain that it also has a dynamic interpretation. Consider the alternative model where the agent can dynamically acquire information from a fixed information source and freely decide when to stop, and the flow of information follows a continuous martingale.[15] Choosing the optimal posterior belief distribution supported on $(\underline{p}, \bar{p})$ translates into the following dynamic behaviour: an agent chooses *belief cutoffs* $\underline{p}$ and $\bar{p}$ and acquires information until her belief reaches either $\bar{p}$ or $\underline{p}$. She then chooses x if $\bar{p}$ is reached, and y if $\underline{p}$ is reached.

15. As an example of such an environment, consider the continuous version of Wald's sequential sampling model, where information is generated by a diffusion process, but costless (see e.g. Morris and Strack, 2017).

Notably, this information acquisition strategy is dynamically incentive compatible, as long as the agent cannot modify or hide the received information from her (receiver) self: to see why, suppose that the agent cannot commit to stopping at the beliefs $\underline{p}$ and $\overline{p}$, but sincerely releases the received information to her (receiver) self. We argue that using $(\underline{p}, \overline{p})$ is a Markov perfect equilibrium (MPE) of the dynamic game. At each instance, the agent decide whethers to continue or stop information acquisition by comparing the continuation values. For any given current belief p, the continuation value from continuing is given by $\tilde{V}(p)$ where $\tilde{V}$ is the concavification of V. The continuation value from stopping is $V(p)$, her utility of choosing the action that maximizes her utility at the current belief. As can be seen in Figure 2.1, the former is larger than the latter if and only if the current belief lies between $\underline{p}$ and $\overline{p}$. That means, it is better for the agent to continue acquiring information until her current belief reaches either$\underline{p}$ or $\overline{p}$.

Corollary 4 predicts that when most of the information received so far indicates that the remunerative option x is *harmless* to others ($p > p_0$), weakly more individuals facing $r > 0$ stop acquiring information; when most of the information received so far indicates that the remunerative option x is *harmful* to others ($p < p_0$), weakly more individuals facing $r > 0$ continue acquiring information. In the companion laboratory experiment presented in Chapter 3, we find evidence in line with this prediction.

2.6 Structural Analysis

2.6.1 A Parametric Family of Other-Regarding Preferences

In this section, we propose the following two-dimensional family of other-regarding preference types:

$$u(p) = -\alpha(1-p)^n$$

for some $n \geq 1$. The parameter α moderates how an individual with perfect knowledge would value the interest of others relative to her own. The parameter n captures how the agent values the interest of others when his belief changes; formally, n is the elasticity of the belief utility function as a function of $q = 1 - p$, that is as a function of the belief that the option is harmful to the other.[16] In Section 2.C of the Appendix, we show how the optimal information strategy varies with the parameters.

2.6.2 An Order of Other-Regarding Preferences

The next result shows that there is a simple ordering other-regarding preference types that translates into an ordering of the predicted degree of self-deceptive behaviour: the lower the curvature of the belief utility, the more self-deceptive the

16. The elasticity of a differentiable function $f(k)$ at k is defined as $\frac{\partial f}{\partial k}\frac{k}{f(k)}$.

agent behaves across *all* possible situations. We say that a preference type with belief utility u is *more self-deceptive* than a type with belief utility v if $u' < v'$ and write $u \succ_{dec} v$. For any type u, let

$$\underline{\delta}(u;\bar{r}) = \max\left[\underline{p}^c(u) - \underline{p}^{tr}(u)\right], \tag{2.14}$$

$$\overline{\delta}(u;\bar{r}) = \max\left[\overline{p}^c(u) - \overline{p}^{tr}(u)\right]. \tag{2.15}$$

where we take the maximum over all pairs of optimal belief cutoffs $(\underline{p}^c(u), \overline{p}^c(u))$ given $r = 0$ and all pairs pf optimal belief cutoffs $(\underline{p}^{tr}(u), \overline{p}^{tr}(u))$ given $\bar{r}$.

Theorem 11. *Let $u \succ_{dec} v$. Then for all $\bar{r} > 0$, the following holds.*

1. *If it is optimal for the v-type to avoid information completely given $\bar{r}$, then, this is also true for the u-type. The converse is not true.*
2. *If it is not optimal for the v-type to avoid information completely given $\bar{r}$, then either it is optimal for the u-type to avoid information completely given $\bar{r}$ or*

$$\overline{\delta}(u;\bar{r}) > \overline{\delta}(v;\bar{r}),$$
$$\underline{\delta}(u;\bar{r}) \geq \underline{\delta}(v;\bar{r}).$$

Note that, given the normalization $u(1) = v(1) = 0$, the relation $u \succ_{dec} v$ implies that $v(0) < u(0)$. It follows from (2.1) that under certainty about the state $\omega = B$, type v chooses the other-regarding action y whenever u does. We see that the ordering $\succ_{dec}$ is an extension of the natural ordering of other-regarding preference types under certainty. The proof is in Section 2.B.8 of the Appendix.

2.7 Related Literature

Literature on Belief-Dependent Utility. Featuring an agent who cares about her own belief that her decision harms no others, our model relates to the literature on belief-dependent utility. Deviating from the outcome-based utility, the economics research has put forward concepts of utility directly derived from beliefs, including the utility derived from memories (remembered utility, Kahneman, Wakker, and Sarin, 1997; Kahneman, 2003, etc), the anticipation of future events (anticipatory utility, Loewenstein, 1987; Brunnermeier and Parker, 2005; Brunnermeier, Gollier, and Parker, 2007; Schweizer and Szech, 2018, etc), ego-relevant beliefs (ego utility, Köszegi, 2006, etc), and belief-dependent emotions (Geanakoplos, Pearce, and Stacchetti, 1989, etc). We suggest that individuals receive utility from believing that their decisions impose no harm on others. This approach is most similar to the idea of belief utility from a moral self-identity proposed by Bénabou and Tirole (2011) in self-signalling games.

Self-Signalling Models. A strand of the literature proposes self-signalling as the main concern in social decisions (Akerlof and Kranton, 2000; Bodner and Prelec,

2003; Bénabou and Tirole, 2006; Bénabou and Tirole, 2011; Grossman and Weele, 2017). Assuming a high level of individual rationality, a self-signaling model features intrapersonal signaling games in which one self of the agent knows her prosocial type and makes decisions, including the decision on what information to collect, and the other self observes the decisions to infer her prosocial type. Addressing *whether* people acquire perfect information, Grossman and Weele (2017) endogenize the decision to avoid perfectly revealing information and show that the avoidance of *perfect* information can be an equilibrium outcome in a self-signalling model. In contrast, we model the process of acquiring information as the process of a person persuading herself to behave selfishly. When doing so, we only assume that the agent cannot modify or hide from herself any information that she has already acquired. Leveraging insights from the Bayesian persuasion, our model is tractable. It goes beyond the binary decision of acquiring or avoiding a certain type of information and generates a complete decscription of the person's *information preferences* over a large range of information environments. Further, it predicts the avoidance of perfect or noisy information (Theorem 12 and Theorem 13), but points out that information avoidance is just one side of self-persuasive behaviour, whereas the other side is that people seek information when current beliefs are not "desirable" (Section 2.5).

Other-Regarding Preferences Literature. By modelling social decisions as driven by utility based on beliefs in one's righteousness, we add to the discussion of an important yet less-understood aspect of social preference, namely social preference under uncertainty. In social decisions with uncertainty, an expected-utility-maximizing agent with intrinsic valuation for the welfare *outcome* of others always prefers complete knowledge in social decisions (for example, the agents in Andreoni, 1990; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002b). It contradicts our empirical finding of strategic information acquisition in Chapter 3 and the avoidance of perfect information observed by, for example, Dana, Weber, and Kuang (2007). Reassessing individuals' motives in social decisions, some models deviate from outcome-based social preferences. Andreoni and Bernheim (2009) propose that individuals act fairly to signal to others that they are fair, which has a similar spirit as the self-signaling models discussed above. Niehaus (2014) proposes a model with an agent who receives a warm glow from her *perceived* social outcomes of her decision. Rabin (1994), Konow (2000), and Spiekermann and Weiss (2016) suggest cognitive dissonance to be a factor for prosocial decisions. In these models, the conflicting desires for selfish interests and fairness create an unpleasant tension, which the agents can reduce by deceiving themselves that a selfish option is fair. A model proposed by Rabin (1995) views moral dispositions as "internal constraints on the agent's true goal of pursuing her self-interest." It shows that for an agent who only engages in a self-benefiting action if she is certain enough that this action harms no one else, partial information or information avoidance can be optimal.

In comparison to these studies, connecting the literature of belief-based utility and Bayesian persuasion (Kamenica and Gentzkow, 2011), we model an agent who gains utility directly from her beliefs and attempts to persuade herself to behave selfishly. Mathematically, our model includes the agent in Rabin (1995) as a special case. Our model allows for convenient analyses of the agent's prosocial behaviour when both information and prosocial actions are choice variables, and lends itself to further structural empirical analysis: We propose a two-dimensional parametric family of belief utility function, where the first parameter captures how much the agent values the interest of others relative to his own interest when certain that an action harms others, and the second parameter captures the elasticity of this valuation as the agent's belief changes. Further, we show that we can order general other-regarding preference types along the model predicted prosocial behaviour simply by ordering the curvature of the belief utility function.

2.8 Concluding Remarks

The model in this paper offers a unifying framework for analyzing individual information preference in social decisions. It can be used to analyze information preferences over information structures of all degrees of informativeness and skewness. Basing the general analyses on the agent's preferences over the posterior belief distributions, the model sheds light on the information acquisition strategies in dynamic and static information environments alike. The framework closely connects several empirical literatures: the literatures on information preferences, on motivated reasoning in social decisions, and on other-regarding preferences.

Building on techniques from the Bayesian persuasion literature (Kamenica and Gentzkow, 2011), the theoretical analyses are geometrically intuitive and hence easy to be adapted for empirical needs. For example, we can use the model to explain the noisy information acquisition strategy found in our experiment, presented in Chapter 3, and the avoidance of perfect information observed by Dana, Weber, and Kuang (2007) and Feiler (2014) (see Appendix 2.A). We hope that it is a step towards a more comprehensive understanding of other-regarding preferences, and might facilitate the empirical research in related settings in the future.

Based on the insight on individuals' preferences of skewed information, we point out several possible ways to encourage information acquisition in social decisions. First, reversing a positively skewed information to an analogue negatively skewed information can make more people acquire information, while preserving the informativeness of the information. Besides, increasing the informativeness in the signal against a self-benefiting action can also attract more people to acquire information – if people are going to give up their own benefits for the sake of others, they would rather be certain that doing so actually benefits others.

Our result that motivated information acquisition can improve the welfare of the other affected by the decision is particularly relevant for policy-makers. Under the opposite intuition that strategic information acquisition motivated by selfish incentives must increase negative externalities, it might seem to be a good idea to de-bias the information acquisition behavior by involving an independent investigator whose compensation is not related to the decision. However, our model suggest that sometimes such strategic information acquisition motivated by selfish incentives can make the other party affected by the decision *better-off*. We present empirical evidence in line with this theoretical result in Chapter 3. This finding offers the novel insight that assigning the job of collecting information to an independent investigator, who is disinterested in the decision, can sometimes lead to worse decision making and more negative externalities.

2.A Additional Theoretical Results

In this section, we discuss two additional results of our model and the respective empirical evidence: the avoidance of noisy and perfect information, which is information that reveals the truth in one piece.

Our model predicts that with or without a remunerative option, there are agents who acquire no noisy information at all (Section 2.A.1). The experiment presented in Chapter 3 show evidence in line with this prediction.

Regarding perfect information, our model predicts that there are agents who avoid *perfectly revealing* information. This result is consistent with the empirical finding of Dana, Weber, and Kuang (2007). Besides, we theoretically show that the higher is the prior belief that the self-rewarding option is harmless to others, the more agent types would avoid perfect information. This prediction is in line with the experimental finding of Feiler (2014).

2.A.1 Avoidance of Noisy Information

Our model predicts that both in decisions with or without a remunerative option, some agent types move on to the decision without acquiring any noisy information.

Theorem 12. *1. When $r = 0$, for any prior $p_0 \in (0,1)$, there is a set $S^{co}(p_0)$ of preference types u that avoid information completely, i.e. the belief cutoffs $\underline{p}^{co} = \overline{p}^{co} = p_0$ are optimal.*

2. When $r > 0$, for any prior $p_0 \in (0,1)$, there is a set $S^{tr}(p_0)$ of preference types u that avoid information completely, i.e. the belief cutoffs $\underline{p}^{tr} = \overline{p}^{tr} = p_0$ are optimal.

The types of the agent who acquire no information, when no option is remunerative, are those with moral standard $l \leq p_0$ or $l \leq 1 - p_0$, i.e., those for whom there is already no gain in belief utility for more certain beliefs at the prior belief. In the decision with remuneration, the agent decides not to acquire information only if she would choose x at the prior belief. The further information then poses an undesirable risk that it might reverse her decision from x to y. She avoids noisy information only when this risk outweighs her utility gain from more certain beliefs that she does not harm the other.

In the experiment presented in Chapter 3, we find that in the presence of a remunerative option, 15% individuals do not acquire any information and in absence of the remunerative option 7% individuals do not acquire information (Chi-2 $p = 0.00$). When there is a remunerative option, among those who avoid noisy information completely 96% choose the remunerative action x (25/26). In contrast, when there is no remunerative option, only 17% of those who avoid noisy information choose x (2/12).

2.A.2 Avoidance of Perfect Information

The model also makes predictions about how people acquire information that reveals the truth at once – *perfectly revealing information*.

Recall that in the theoretical model, the agent can choose any signal structure (Section 2.2). Perfectly revealing information is a special case of the signal structures that the model encompasses. Let $p_0 \in (0,1)$ be any uncertain prior belief. The decision whether or not to acquire a piece of perfectly revealing information is formally the preference between the posterior belief distribution τ^{p_0} that has mass 1 on the prior belief p_0 and the posterior distribution τ^{ce} with $supp(\tau^{ce}) = \{0,1\}$. Theorem 13 shows that in the presence of remuneration, for any uncertain prior belief, there are some agent types who would avoid perfectly revealing information. The higher is the prior belief in the alignment between the agent's and the other's material interests, the more agent types would avoid perfect information.

Theorem 13. *1. When $r > 0$, for any prior $p_0 \in (0,1)$, there is a set $S(p_0)$ of preference types u that avoid perfectly revealing information, i.e. $\tau^{p_0} \succ \tau^{ce}$.*

2. For any prior beliefs $p_0^l < p_0^h \in (0,1)$, it holds that $S(p_0^l) \subset S(p_0^h)$.

A piece of perfectly revealing information either makes the agent certain that the remunerative option is harmless, or makes her certain that it is harmful. For an agent who would choose the remunerative option at the prior belief, if the realized signal is that the remunerative option is harmless, the agent gains in belief utility, as she becomes more certain that she is not harming the other. But on the other hand, she faces the risk that the realized signal would make her certain that the remunerative option is harmful so that she would have to forgo the remuneration and choose the other option instead.

The first item of Theorem 13 shows that for any uncertain prior belief, there are some types of agents for whom the risk of having to forgo the remuneration outweighs the potential gain in belief utility so that they would rather avoid the perfect information and make a decision based on their prior beliefs. Trivially, agent types with weakly convex preference type u will always acquire perfect information. These agents who avoid perfect information must have strictly concave belief utility u.

The second item of Theorem 13 predicts that when the prior belief is higher, it is optimal for more agent types to avoid the perfect information and choose the remunerative option directly. When the prior belief increases, on the one hand, the additional belief utility from being certain that the preferred option is indeed harmless decreases, so the perfect information becomes less attractive; but on the other hand, the probability that the remunerative option is harmless increases, so the perfect information becomes more attractive. Since these agent types who avoid perfect information have strictly concave belief utilities u, the magnitude of the first negative effect becomes larger with increasing prior, while the magnitude of the second

positive effect is linear in the prior belief. Therefore, as the prior increases, the perfect information becomes overall less attractive and more agent types would avoid perfect information.

These predictions are consistent with previous empirical findings. In a dictator environment, Dana, Weber, and Kuang (2007) find that a significant fraction of dictators avoids information that reveals the *ex-ante* unknown state all at once. Feiler (2014) further documents that the fraction of dictators who avoid such perfectly revealing information increases with the dictators' prior belief that a self-benefiting option has no negative externality.

2.B Proofs

2.B.1 Proof of Lemma 1 and Theorem 2

Proof of Lemma 1. The statement holds trivially when u is strictly convex since then the agent strictly prefers Black-well more informative information and the unique optional posterior distribution has support on $\underline{p} = 0$ and $\overline{p} = 1$. It remains to prove the lemma when u is weakly concave. Consider any optimal posterior distribution τ. Suppose that there are two beliefs $p_1, p_2 \geq p_0$ with $\Pr_\tau(p_1) > 0, \Pr_\tau(p_2) > 0$. Let $\hat{p} = \Pr_\tau(p_1) + \Pr_\tau(p_2)p_2$. Then $\hat{p} \geq p_0$. Also,

$$\begin{aligned} & V(\hat{p}) - (\Pr_\tau(p_1)V(p_1) + \Pr_\tau(p_2)V(p_2)) \\ = \; & u(\hat{p}) - (\Pr_\tau(p_1)u(p_1) + \Pr_\tau(p_2)u(p_2)) \\ \geq \; & 0, \end{aligned}$$

since u is weakly concave. So, we see that she is weakly better off with the posterior distribution that arises from τ when shifting the mass from p_1 and p_2 to $\hat{p}$. Suppose that there are two beliefs $p_1, p_2 \leq p_0$ with $\Pr_\tau(p_1) > 0, \Pr_\tau(p_2) > 0$. The analogous argument shows that shifting mass from p_1 and p_2 to $\hat{p} = \Pr_\tau(p_1) + \Pr_\tau(p_2)p_2$ makes her weakly better off. This finishes the proof of the Lemma.

Proof of Theorem 2. When $r = 0$, any pair of beliefs $(\underline{p}^c, \overline{p}^c)$ with $\underline{p}^c \in [1-l, l]^c$ and $\overline{p}^c \in [1-l, l]^c$ implies an expected continuation value $\mathrm{E}_{(\underline{p}^c, \overline{p}^c)}V(p)$ of 0. Since, given $r = 0$, the expected continuation value for any posterior distribution τ is weakly negative, any such pair of belief cutoffs is optimal. This finishes the proof of Theorem 2.

2.B.2 Proof of Theorem 12, Corollary 4, and Theorem 3

Let $r > 0$. Any optimal pair of belief cutoffs $\underline{p} \leq \overline{p}$ satisfies Bayes-plausibility,

$$\overline{p}\Pr_\tau(\overline{p}) + \underline{p}(1 - \Pr_\tau(\overline{p})) = p_0, \tag{2.16}$$

which pins down how likely it is that she stops at the upper cutoff $\overline{p}$ and how likely it is that she stops at the lower cutoff $\underline{p}$, given the prior belief. The likelihood of the upper belief cutoff is negatively proportional to its relative distance to the prior,

$$\Pr_\tau(\overline{p}) = \frac{p_0 - \underline{p}}{\overline{p} - \underline{p}}. \tag{2.17}$$

The expected continuation value, given belief cutoffs $(\underline{p}, \overline{p})$ is therefore

$$\mathrm{E}_{(\underline{p},\overline{p})} V(p) = \frac{p_0 - \underline{p}}{\overline{p} - \underline{p}} V(\overline{p}) + \frac{\overline{p} - p_0}{\overline{p} - \underline{p}} V(\underline{p}), \tag{2.18}$$

which is simply the value of the affine function connecting $V(\overline{p})$ and $V(\underline{p})$ through the prior. Since $r > 0$, there is a unique pair of beliefs $(\underset{\sim}{p}, \tilde{p})$ that support the concave envelope.[17] Note that

$$\underset{\sim}{p} = 0. \tag{2.19}$$

The following lemma shows that the pair of belief cutoffs $(\underset{\sim}{p}, \tilde{p})$ is the unique optimal strategy whenever it is not optimal to acquire no information.

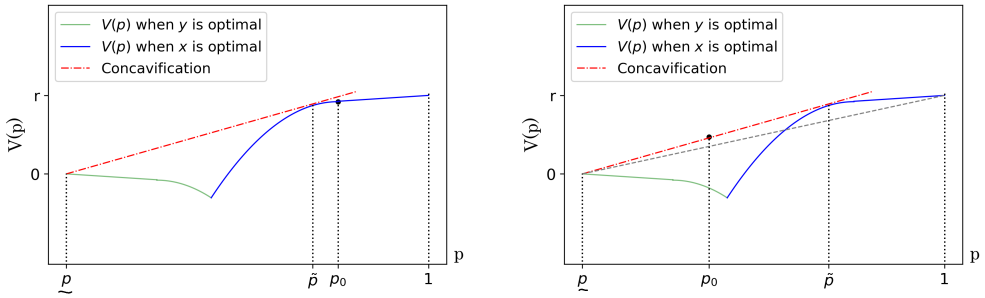
Lemma 14. *Let $r > 0$.*

1. *When $p_0 \in [\underset{\sim}{p}, \tilde{p})$, then there is a unique pair of optimal belief cutoffs, given by $(\underline{p}^{tr}, \overline{p}^{tr}) = (\underset{\sim}{p}, \tilde{p})$.*
2. *When $p_0 \notin [\underset{\sim}{p}, \tilde{p})$, then acquiring no information is optimal, i.e. the belief cutoffs $\underline{p}^{tr} = \overline{p}^{tr} = p_0$ are optimal.*

Proof. Consider any two belief cutoffs $\underline{p} \leq p_0 \leq \overline{p}$ and the value of the connecting function at the prior. The optimal belief cutoffs maximize (2.18).
The claim can be seen geometrically: when $p_0 \in [\underset{\sim}{p}, \tilde{p}]$, the optimal belief cutoffs are given by the unique pairs of beliefs $\underline{p}$ and $\tilde{p}$ that support the concave envelope of V (see Figure 2.3a). Whenever $p_0 \notin [\underset{\sim}{p}, \tilde{p})$, the maximum of (2.18) is achieved through no information acquisition (see Figure 2.3b). □

Proof of Theorem 12. The first item of Theorem 12 follows from Theorem 2 since, for any prior $p_0 \in (0,1)$, there is an open set of preference types u for which $p_0 \in [l, 1-l]^c$.
Lemma 14 together with (2.19) implies that for $r > 0$, the optimal lower belief cutoff is $\underset{\sim}{p} = 0$. Therefore, it follows from Lemma 14 that, for any prior $p_0 \in (0,1)$, the set $S(p_0)$ of types u for which no information acquisition is optimal is given by the types for which $p_0 \geq \tilde{p}$. This shows the second item of Theorem 12. Also note that this set is strictly smaller when the prior is larger.

17. The smallest concave function that lies weakly above V is called the concave envelope of V; compare to Figure 2.2.



(a) Complete Information Avoidance **(b)** Optimal Belief Cutoffs Other Than the Prior

Note: The green and the blue line show the continuation value function V which is defined componentwise. The beliefs $\underset{\sim}{p}$ and $\tilde{p}$ are the unique beliefs supporting the concave envelope of V. The agent cannot improve on the belief cutoffs $(\underset{\sim}{p}, \tilde{p})$ when $p_0 \in [\underset{\sim}{p}, \tilde{p}]$: This can be seen geometrically: given (2.18), the optimal cutoffs maximize the value of the connecting function at the prior. Any other choice implies that the connecting line (grey) takes a value at the prior lower than $V(p_0)$ (left) or lower than the line connecting $\underset{\sim}{p}$ and $\tilde{p}$ (right). When $p_0 \notin [\underset{\sim}{p}, \tilde{p}]$, it is optimal for the agent to acquire no information.

Figure 2.2. Optimal Belief Cutoffs

Proof of Corollary 4. It remains to show for the case when u is weakly concave; see the discussion after Theorem 2. Take any weakly concave u, any optimal strategy $(\underset{\sim}{p}^c, \overline{p}^c)$ given $r = 0$. Suppose that it is not optimal to acquire no information given $\overline{r} > 0$.

Given Lemma 14, there are unique optimal belief cutoffs $\underset{\sim}{p}^{tr} < p_0 < \overline{p}^{tr}$, and, given (2.19), it holds $\underset{\sim}{p}^{tr} = 0$. Given (2.18), the upper belief cutoff $\overline{p}^{tr}$ maximizes

$$\max_{p \in [p_0, 1]} \Pr(p) V(p; \overline{r}) \tag{2.20}$$

subject to the Bayes-plausibility constraint that $\Pr(p)p = p_0$. Plugging in the Bayes-plausibility constraint gives the objective function

$$\max_{p \in [p_0, 1]} \frac{p}{p_0} V(p; \overline{r}), \tag{2.21}$$

and taking derivatives gives the first-order condition

$$\begin{aligned} & \frac{p_0}{p} u'(p) - \frac{p_0}{p^2} V(p; \overline{r}) = 0 \\ \Leftrightarrow \quad & p u'(p) - V(p; \overline{r}) = 0. \end{aligned} \tag{2.22}$$

The maximization problem (2.20) has a solution since continuous functions take maxima on compact sets. Note that the second derivative of the objective function (2.21) is weakly negative,

$$\frac{\partial}{\partial p} (p u'(p) - u(p) - \overline{r}) = p u''(p) \leq 0, \tag{2.23}$$

where we used that u is weakly concave.

Case 1. $u'(1) \geq \bar{r} > 0$

The condition $u'(1) \geq \bar{r}$ implies that $l = 1$. Therefore, without remuneration, $r = 0$, the optimal belief cutoffs are $(\underline{p}^c, \overline{p}^c) = (0, 1)$. Since $\underline{p}^{tr} = 0$, the inequalities (2.6) and (2.7) follow. This finishes the proof of the theorem for Case 1.

Case 2. $u'(1) < \bar{r}$

The condition $u'(1) < \bar{r}$ is equivalent to

$$1u'(1) - V(1) < 0. \tag{2.24}$$

If $l < 1$, then $u'(p) = 0$ for $p \geq l$ since u is continuously differentiable. In any case,

$$lu'(l) - V(l) < 0. \tag{2.25}$$

Suppose that the derivative of the objective function is weakly negative for all $p \in [p_0, 1]$; this is equivalent to

$$p_0 u'(p_0) - V(p_0; \bar{r}) \leq 0, \tag{2.26}$$

given (2.23). Then, the objective function is maximized at the boundary $\overline{p} = p_0$. Bayes-plausibility implies that $\Pr(\overline{p}) = 1$. We conclude, that no information acquisition is optimal. However, we excluded this case by assumption. Therefore,

$$p_0 u'(p_0) - V(p_0; \bar{r}) > 0, \tag{2.27}$$

and it follows from the intermediate value theorem, (2.25), and (2.27) that the first-order condition (2.22) is satisfied by some $\tilde{p}$ with $p_0 < \tilde{p} < l$. It follows from (2.23) that the derivative of the objective function is weakly positive for $p < \tilde{p}$ and weakly negative for $p > \tilde{p}$ such that $\tilde{p}$ maximizes the objective function. We conclude that the belief cutoffs

$$(\underline{p}^{tr}, \overline{p}^{tr}) = (0, \tilde{p}) \tag{2.28}$$

are optimal. Moreover, given (2.25), any optimal upper belief cutoff satisfies $\overline{p}^{tr} < l$ and the first-order condition.

Now, we finish the proof of the theorem for Case 2. The inequality (2.7) follows directly from $\underline{p} = 0$. Theorem 2 states that, without remuneration, the optimal belief cutoffs are the pairs of beliefs $(\underline{p}^c, \overline{p}^c)$ that satisfy $\overline{p}^c \geq l$ and $\underline{p} \leq 1 - l$. Since $\tilde{p} < l$, the inequality (2.6) holds strictly. When $l < 1$, then, without remuneration, there are optimal belief cutoffs $(\underline{p}^c, \overline{p}^c)$ with $\underline{p}^c > 0 = \underline{p}^{tr}$. Hence, the inequality (2.6) holds strictly. This finishes the proof of the theorem.

Proof of Theorem 3. See the proof of Corollary 4 above.

2.B.3 Proof of Theorem 9 and Theorem 10

Proof of Theorem 9. Let p^* solve

$$u(p) + \bar{r} = u(1-p). \tag{2.29}$$

It is easy to see that $p^* < 0.5$. Recall that $\bar{p}^c \geq l$ and $\underline{p}^c \leq 1-l < 0.5$. If $1-l < p^*$, then at the posterior belief cutoffs $\bar{p}^c$ and $\underline{p}^c$, the agent chooses x and y according to decision rule a_{tr} just like according to a_c. There is no decision effect and $Ev(\boldsymbol{a_{tr}}(p),\omega)|\tau^c = Ev(\boldsymbol{a_c}(p),\omega)|\tau^c$.
If $1-l \geq p^*$, then for any lower belief cutoff $\underline{p}^c \in [0,p^*]$, there is also no the decision effect because the agent chooses x at $\bar{p}^c$ and y at $\underline{p}^c$.
However, for any lower belief cutoff $\underline{p}^c \in [p^*, 1-l]$, the agent chooses x at $\underline{p}^c$ instead if they decide according to a_{tr}. Therefore with a_{tr}, the expected utility of the other if the lower cutoff is realized is

$$-Pr(Y|\underline{p}^c) = -(1-\underline{p}^c). \tag{2.30}$$

Whereas with a_c, the expected utility of the other if the lower cutoff is realized is

$$-Pr(X|\underline{p}^c) = -\underline{p}^c. \tag{2.31}$$

Recall $\underline{p}^c < 0.5$, hence $-(1-\underline{p}^c) < -\underline{p}^c$, i,e, the expected utility of the other if the lower cutoff is realized is lower with a_{tr} than with a_c. Since the probability that the lower cutoff is realized is pinned down by τ^c, and the expected utility of the other if the upper cutoff is realized is the same between the scenario with a_c and the scenario with a_{tr}, the expected utility of the other is strictly *lower* keeping τ^c fixed and changing a_c to a_{tr}. The decision effect is strictly negative, i.e. $Ev(\boldsymbol{a_{tr}}(p),\omega)|\tau^c < Ev(\boldsymbol{a_c}(p),\omega)|\tau^c$.

Proof of Theorem 10. We prove the theorem with an example. Consider u such that $p_0 < l < 1$ and $u''(x) \to \infty$ for $p \in [1-l, l]$. When $r = 0$, it follows from Theorem 2 that a pair of optimal cutoffs are $\underline{p}^c = 1-l$ and $\bar{p}^c = l$. When $r > 0$, $\underline{p}^{tr} = 0$ and $\bar{p}^{tr} \to l$, since these two points support the concave envelope.

First, we prove that for this agent, the information effect is strictly positive.

What does the agent choose at each cutoff? It follows from (2.12) that, for any $r > 0$, the belief p^* where she is indifferent between x and y, i.e. $u(p^*) + r = u(1-p^*)$, converges to $\frac{1}{2}$. So the agent chooses y at the two lower cutoffs $\underline{p}^{tr}$ and $\underline{p}^c$, and x at the two upper cutoffs $\bar{p}^{tr}$ and $\bar{p}^c$. That is, the decision effect (2.11) converges to zero.

Let us now derive the expected utility of the other with τ^c and τ^{tr}, when fixing $a_{tr}(p)$.

First, with τ^c,

$$\begin{aligned} Ev(a_{tr}(p),\omega)|\tau^c &= -1 \cdot Pr(\underline{p}^c)Pr(X|\underline{p}^c) + (-1) \cdot Pr(\overline{p}^c)Pr(Y|\overline{p}^c) && (2.32)\\ &= -(Pr(\underline{p}^c)(1-l) + Pr(\overline{p}^c)(1-l)) && (2.33)\\ &= -(Pr(\underline{p}^c) + Pr(\overline{p}^c))(1-l) && (2.34)\\ &= -(1-l). && (2.35) \end{aligned}$$

Then with τ^{tr},

$$\begin{aligned} Ev(a_{tr}(p),\omega)|\tau^{tr} &\to -1 \cdot Pr(\underline{p}^{tr}Pr(X|\underline{p}^{tr}) + (-1) \cdot Pr(\overline{p}^{tr})Pr(Y|\overline{p}^{tr}) && (2.36)\\ &= -(Pr(\underline{p}^{tr}) \cdot 0 + \frac{p_0}{l}(1-l)) && (2.37)\\ &= -\frac{p_0}{l}(1-l). && (2.38) \end{aligned}$$

Since $p_0 < l$,

$$Ev(a_{tr}(p),\omega)|\tau^{tr} > Ev(a_{tr}(p),\omega)|\tau^c, \tag{2.39}$$

In other words, we have proven that for this agent if $p_0 < l$, the information effect (2.13) is strictly positive.

We have already discussed above that the decision effect converges to zero for this agent. We hence can conclude that the overall effect is strictly positive. It finishes the proof.

2.B.4 Proof of Theorem 13

Proof of Theorem 13. Item 1 of Theorem 13 is a corollary of the second item of Theorem 12: if a preference type prefers no information over all possible information structures, clearly, she prefers no information over the fully revealing signals.
We prove item 2 of Theorem 13 by contradiction.
For any prior belief $p_0^l < p_0^h \in (0,1)$, if an agent type prefers the prior to the fully revealing signals at this prior then

$$\begin{aligned} \tau^{p_0^l} &\succ \tau^{ce} && (2.40)\\ \Leftrightarrow r + u(p_0^l) &> rp_0^l. && (2.41) \end{aligned}$$

Suppose this agent prefers the fully revealing signals to prior p_0^h, then

$$\begin{aligned} \tau^{p_0^h} &\prec \tau^{ce} && (2.42)\\ \Leftrightarrow r + u(p_0^h) &< rp_0^h. && (2.43) \end{aligned}$$

Subtract 2.41 from 2.43 and rearrange,we get:

$$\frac{u(p_0^h) - u(p_0^l)}{p_0^h - p_0^l} < r. \tag{2.44}$$

Since $u(\cdot)$ is concave and $p_0^l < p_0^h < 1$,

$$\frac{u(1) - u(p_0^h)}{1 - p_0^h} < r. \tag{2.45}$$

Since $u(1) = 0$, we get

$$\begin{aligned} \frac{-u(p_0^h)}{1 - p_0^h} &< r, && (2.46)\\ \Rightarrow r + u(p_0^h) &> r p_0^h, && (2.47)\\ \Rightarrow \tau^{p_0^h} &\succ \tau^{ce}. && (2.48) \end{aligned}$$

Contradiction. Hence $\tau^{p_0^h} \succeq \tau^{ce}$. In other words, $S(p_0^l) \subset S(p_0^h)$.

2.B.5 Proof of Proposition 5

For any binary signal, denote by g the signal realization leading to the larger posterior $\bar{p}$ ("good news") and by b the signal realization leading to the smaller posterior ("bad news").

Proof. The result of the proposition is driven by the presence of the selfish reward r from taking the action x. First, note that a signal is negatively skewed if good news are more likely; hence,

$$\Pr(s = g|(1 - \bar{p}, 1 - \underline{p})) > \Pr(s = b|(1 - \bar{p}, 1 - \underline{p}))$$

and a signal is positively skewed if good news are less likely; hence,

$$\Pr(s = g|(\underline{p}, \bar{p})) < \Pr(s = b|(\underline{p}, \bar{p}))$$

Second, note that posteriors are ordered as follows,

$$1 - \bar{p} < \underline{p} < \frac{1}{2} < 1 - \underline{p} < \bar{p},$$

which reflects that $(\underline{p}, \bar{p})$ is positively skewed and $(1 - \bar{p}, 1 - \underline{p})$ is positively skewed. Third, note that after a g-signal, the agent believes that she is less likely to harm the other when choosing g, and additionally receives a reward from choosing x, thus chooses x for both information structures.

There are two cases.

Case 1. *The positively skewed information $(\underline{p}, \bar{p})$ is such that the agent y after a b-signal.*

Then, when receiving negatively skewed information, after a b-signal the agent is even more convinced of not harming the other, and therefore also chooses y. As a

consequence, given the symmetry of the posterior beliefs, the expected belief utility from negatively and positively skewed information is the same. However, when receiving negatively skewed information, the likelihood to receive a g-signal is strictly higher, $\Pr(s = a|(1-\bar{p}, 1-\underline{p}) > \Pr(s = a|(\underline{p}, \bar{p})$, hence, the likelihood of receiving remuneration r. The agent chooses the negatively skewed information to maximize the likelihood of receiving the reward.

Case 2. *The positively skewed information* $(\underline{p}, \bar{p})$ *is such that the agent chooses* x *after a b-signal.*

Then, when the agent chooses to receive no information and the action x, she receives the remuneration with the maximal probability 1, and a higher expected belief utility since the belief utility u is concave. Hence, the agent would (weakly) prefer to receive no information over the positively skewed information.

2.B.6 Proof of Proposition 6

Proof. The result of the proposition is driven by the presence of the selfish reward r from taking the action x. Recall that posteriors are ordered as follows,

$$\underline{p} < \underline{p}' < \frac{1}{2} < \bar{p} \tag{2.49}$$

which reflects that $(\underline{p}, \bar{p})$ is more negatively skewed and $(\underline{p}', \bar{p})$ less negatively skewed. After a g-signal, the agent believes that she is less likely to harm the other when choosing g, and additionally receives a reward from choosing x, thus chooses x for both information structures. There are two cases.

Case 1. *The less negatively skewed information* $(\underline{p}', \bar{p})$ *is such that the agent chooses* y *after a b-signal.*

Then, when receiving the more negatively skewed information $(\underline{p}, \bar{p})$, after a b-signal the agent is even more convinced of not harming the other, and therefore also chooses y. As a consequence, the expected belief utility from $(\underline{p}, \bar{p})$ is higher than from $(\underline{p}', \bar{p})$. Additionally, when receiving the more negatively skewed information, the likelihood to receive a g-signal is strictly higher, hence, the likelihood of receiving remuneration r. Together, this shows that the agent prefers $(\underline{p}, \bar{p})$ over $(\underline{p}', \bar{p})$.

Case 2. *The less negatively skewed information* $(\underline{p}', \bar{p})$ *is such that the agent chooses* x *after a b-signal.*

Then, when the agent chooses to receive no information and takes the action x, she receives the remuneration with the maximal probability 1, and a higher expected belief utility since the belief utility u is concave. Hence, the agent would (weakly) prefer to receive no information over the less negatively skewed information.

2.B.7 Proof of 8

First, note that it is optimal for an agent to acquire no information if and only if

$$p_0 \leq \bar{p}^{\text{conc}}(r) \tag{2.50}$$

where $(\underline{p}^{\text{conc}}(r), \bar{p}^{\text{conc}})(r)$ is the belief pair supporting the concave envelope of V given a reward $r \geq 0$. We show

Claim 1. *$\bar{p}^{conc}$ is weakly decreasing in r.*

This proves the first item of Proposition 8. For the second item, recall that whenever the individual does not prefer to avoid information, then the optimal information structure is given by $(\underline{p}^{\text{conc}}(r), \bar{p}^{\text{conc}})(r)$, and that $\underline{p}^{\text{conc}}(r) = 0$. Then, we note that an information structure $(\underline{p}, \bar{p})$ with $\underline{p} = 0$ is the more negatively skewed, the smaller $\bar{p}$. As a consequence, also the second item follows from Claim 1.

Proof of Claim 1. Let $r' > r$. We consider two cases.

Case 1. $\bar{p}^{conc}(r) = 1$.

Note that full information acquisition is optimal if and only if $l = 0$ and $u'(1) \geq r''$ for any given reward r''. Hence, the assumption of the case implies $u'(1) \geq r$. Either $u'(1) \geq r'$ and full information acquisition is optimal given r'. Then, $1 = \bar{p}^{\text{conc}}(r') = \bar{p}^{\text{conc}}(r)$. or $u'(1) < r'$. Then, the first-order condition (2.5) has an interior solution, and $\bar{p}^{\text{conc}}(r') < 1$, hence $\bar{p}^{\text{conc}}(r') < \bar{p}^{\text{conc}}(r)$.

Case 2. $\bar{p}^{conc}(r) < 1$.

Hence, the first-order condition (2.5) has an interior solution $\bar{p}^{\text{conc}}(r) < 1$. Denote $\bar{p}(r'') = \bar{p}^{\text{conc}}(r'')$ for any given r''. Implicit differentiation shows

$$\begin{aligned} \frac{\partial \bar{p}(r'')}{\partial r''} &= 1 \cdot \frac{1}{u'(\bar{p}(r'') + \bar{p}(r'')u''(\bar{p}(r'')) - u'(\bar{p}(r''))} \\ &< 0, \end{aligned} \tag{2.51}$$

since $u'' < 0$. Note that (2.51) implies $\bar{p}^{\text{conc}}(r') < \bar{p}^{\text{conc}}(r)$.

2.B.8 Proof of Theorem 11

Proof. Let $u \succ_{dec} v$ and consider the situation with remuneration $\bar{r} > 0$. Let $\underset{\sim}{p}(u)$ and $\tilde{p}(u)$ be the unique pair of beliefs supporting the concave envelope of the continuation value function V of the u-type and $\underset{\sim}{p}(v)$ and $\tilde{p}(v)$ be the unique pair of beliefs supporting the concave envelope of the continuation value function V of the v-type. Recall that $\underset{\sim}{p}(u) = \underset{\sim}{p}(v) = 0$, given Lemma 14 and (2.19).
Consider the first item of the theorem. Lemma 14 says that it is optimal for the v-type to avoid information completely, given $\bar{r}$, if and only if $p_0 \notin [\underset{\sim}{p}(v), \tilde{p}(v)]$. Similarly, it is optimal for the u-type to avoid information completely, given $\bar{r}$, if and only

if $p_0 \notin [\underset{\sim}{p}(u), \tilde{p}(u)]$. Since $\underset{\sim}{p}(u) = \underset{\sim}{p}(v) = 0$, to prove the first item of the theorem it suffices to show that

$$\tilde{p}(u) < \tilde{p}(u). \tag{2.52}$$

Note that the beliefs $\underset{\sim}{p}(u)$ and $\tilde{p}(u)$ supporting the concave envelope of V satisfy

$$V(\tilde{p}(u); \bar{r}) - V(\underset{\sim}{p}(u); \bar{r}) = u'(\tilde{p})(\tilde{p}(u) - \underset{\sim}{p}(u). \tag{2.53}$$

Since $\underset{\sim}{p}(u) = \underset{\sim}{p}(v) = 0$ and $V(0, \bar{r}) = 0$, this implies that $\tilde{p}(u)$ satisfies the condition

$$pu'(p) - V(p; \bar{r}) = 0; \tag{2.54}$$

compare to the first-order condition (2.22). Similarly, $\tilde{p}(v)$ satisfies

$$pv'(p) - V(p; \bar{r}) = 0; \tag{2.55}$$

Therefore, if

$$\begin{aligned} \forall p : pu'(p) - V(p; \bar{r}) &> pv'(p) - V(p; \bar{r}), \\ \Leftrightarrow \forall p : pu'(p) - u(p) &> pv'(p) - v(p), \end{aligned} \tag{2.56}$$

then (2.52) holds. We rewrite (2.56) using $u(1) = v(1) = 0$,

$$\forall p : pu'(p) + \int_p^1 u'(p)dp > pv'(p) + \int_p^1 v'(p)dp. \tag{2.57}$$

Clearly, $u' > v'$ implies (2.57). This finishes the proof of the first item.

Consider the second item of the theorem. Given Lemma 14, if it is not optimal for the type to avoid information completely, then, $p_0 < \tilde{p}(v)$ and the optimal belief cutoffs of the v-type given $\bar{r}$ are $\underset{\sim}{p}(v) = 0$ and $\tilde{p}(v)$. Given (2.52), we have to distinguish two cases.

Case 1. $\tilde{p}(u) \leq p_0$

Then, it follows from given Lemma 14 that it is optimal for the u-type not to acquire any information. This finishes the proof of the second item in this case.

Case 2. $p_0 < \tilde{p}(u) < \tilde{p}(v)$

Then, it follows from Lemma 14 that the optimal belief cutoffs of the u-type given $\bar{r}$ are $\underset{\sim}{p}(u) = 0$ and $\tilde{p}(u)$. Consider $l(v) = \min\{p \in [0,1] : v(p) = 0\}$ and $l(u) = \min\{p \in [0,1] : u(p) = 0\}$. Note that $u \succ_{dec} v$ implies $l(u) > l(v)$. The claim of the second item of the theorem follows from the characterization of the optimal belief cutoffs without remuneration in Corollary 4 and from (2.52). □

2.C Parametric Examples

Isoelastic Belief Utility. Let $u(p) = -\alpha(1-p)^n$ for some $n > 1$. Given Corollary 4, the agent's optimal belief cutoffs are weakly smaller with remuneration $\bar{r} > 0$ than without if it is not optimal to avoid information completely given $\bar{r}$. Note that $u'(1) = 0$ for all $n > 1$ such that it follows from the proof of Corollary 4 (see Case 2, in particular (2.27)) that it is not optimal to avoid information completely if

$$\begin{aligned} 0 &< p_0 u'(p_0) - u(p_0) - r \\ \Leftrightarrow r &< \alpha(1-p_0)^{n-1}((1-p_0)-n) \\ \Leftrightarrow \alpha &> \frac{r}{(1-p_0)^{n-1}((1-p_0)-n)}. \end{aligned} \tag{2.58}$$

Since $n > 1$, the right hand side is negative and the condition (2.58) is generally fulfilled. So, the condition of Corollary 4 is fulfilled. Since $u'(1) = 0$ for all $n > 1$, it follows from Corollary 4 that the upper belief cutoff is strictly smaller with remuneration $\bar{r} > 0$ than without.

Linear Belief Utility. [18] If $u(p) = \alpha(p-1)$, then, given (2.1), she chooses x at belief $p = \Pr(\alpha)$ if and only if

$$\begin{aligned} \alpha(p-1) + r &\geq -\alpha p \\ \Leftrightarrow 2\alpha p &\geq \alpha - r \\ \Leftrightarrow p &\geq \frac{1}{2} - \frac{r}{2\alpha}. \end{aligned} \tag{2.59}$$

She always prefers x regardless of her belief if

$$\begin{aligned} \frac{1}{2} - \frac{r}{2\alpha} &\leq 0 \\ \Leftrightarrow \alpha &\leq r. \end{aligned} \tag{2.60}$$

If (2.60) holds, it is optimal not to acquire any information and choose x. Conversely, when she is sufficiently altruistic, i.e. when $\alpha > r$, it is optimal to get fully informed about the state and choose the action that is harmless to the other in both states, i.e. x in X and y in Y.

18. The assumption of linear belief utility means that the agent is an expected utility maximizer.

References

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics* 115 (3): 715–753. [63]

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal* 100 (401): 464–477. [50, 64]

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): 1607–1636. [50, 64]

Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70 (2): 737–753. [50]

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–1678. [64]

Bénabou, Roland, and Jean Tirole. 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2): 805–855. [63, 64]

Bodner, Ronit, and Drazen Prelec. 2003. "Self-Signaling and Diagnostic Utility in Everyday Decision Making." *Psychology of Economic Decisions* 1: 105–26. [63]

Bolton, Gary E, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90 (1): 166–193. [50, 64]

Brunnermeier, Markus K, Christian Gollier, and Jonathan A Parker. 2007. "Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns." *American Economic Review* 97 (2): 159–165. [63]

Brunnermeier, Markus K, and Jonathan A Parker. 2005. "Optimal Expectations." *American Economic Review* 95 (4): 1092–1118. [63]

Charness, Gary, and Matthew Rabin. 2002a. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117 (3): 817–869. [50]

Charness, Gary, and Matthew Rabin. 2002b. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117 (3): 817–869. [64]

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33 (1): 67–80. [50, 51, 64, 65, 67, 69]

DellaVigna, Stefano, John A List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127 (1): 1–56. [49]

Dillenberger, David, and Uzi Segal. 2017. "Skewed noise." *Journal of Economic Theory* 169: 344–364. [57]

Fehr, Ernst, and Klaus M Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114 (3): 817–868. [50, 64]

Feiler, Lauren. 2014. "Testing Models of Information Avoidance with Binary Choice Dictator Games." *Journal of Economic Psychology* 45: 253–267. [50, 51, 65, 67, 69]

Forsythe, Robert, Joel L Horowitz, Nathan E Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic behavior* 6 (3): 347–369. [49]

Franzen, Axel, and Sonja Pointner. 2013. "The External Validity of Giving in the Dictator Game." *Experimental Economics* 16 (2): 155–169. [49]

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. "Psychological Games and Sequential Rationality." *Games and Economic Behavior* 1 (1): 60–79. [63]

Gino, Francesca, Michael I Norton, and Roberto A Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives* 30 (3): 189–212. [50]

Grossman, Zachary. 2014. "Strategic ignorance and the robustness of social preferences." *Management Science* 60 (11): 2659–2665. [50]

Grossman, Zachary, and Joël van der Weele. 2017. "Self-Image and Willful Ignorance in Social Decisions." *Journal of the European Economic Association* 15 (1): 173–217. [64]

Kahneman, Daniel. 2003. "Experienced Utility and Objective Happiness: A Moment-Based Approach." In *The Psychology of Economic Decisions*. Vol. 1, Oxford: Oxford University Press, 187–208. [63]

Kahneman, Daniel, Peter P Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics* 112 (2): 375–406. [63]

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (6): 2590–2615. [50, 52, 65]

Konow, James. 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American economic review* 90 (4): 1072–1091. [64]

Köszegi, Botond. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4 (4): 673–707. [63]

Loewenstein, George. 1987. "Anticipation and the Valuation of Delayed Consumption." *Economic Journal* 97 (387): 666–684. [63]

Masatlioglu, Yusufcan, A Yesim Orhun, and Collin Raymond. 2017. "Preferences for Non-Instrumental Information and Skewness." *Working Paper*, [51, 57]

Morris, Stephen, and Philipp Strack. 2017. "The Wald problem and the equivalence of sequential sampling and static information costs." *Unpublished manuscript, June*, [61]

Niehaus, Paul. 2014. "A Theory of Good Intentions." *San Diego, CA: University of California and Cambridge, MA: NBER*, [64]

Rabin, Matthew. 1994. "Cognitive dissonance and social change." *Journal of Economic Behavior & Organization* 23 (2): 177–194. [64]

Rabin, Matthew. 1995. "Moral Preferences, Moral Constraints, and Self-Serving Biases." [64, 65]

Schweizer, Nikolaus, and Nora Szech. 2018. "Optimal Revelation of Life-Changing Information." *Management Science* 64 (11): 5250–5262. [63]

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99–118. [51, 53]

Spiekermann, Kai, and Arne Weiss. 2016. "Objective and Subjective Compliance: A Norm-Based Explanation of 'Moral Wiggle Room'." *Games and Economic Behavior* 96: 170–183. [64]

Chapter 3

Dynamic Information Acquisition in Social Decisions: An Experiment

Joint with Carl Heese

3.1 Introduction

The motivated reasoning literature demonstrates that people often trade off the accuracy against the desirability of their beliefs (for a review, see Bénabou and Tirole, 2016). The desirability of beliefs can arise in decisions where benefiting oneself *might* harm others. In these situations, individuals can behave selfishly without a guilty conscience if they believe that the selfish decision harms no others (for a review, see Gino, Norton, and Weber, 2016). In this paper, we analyze how individuals acquire information about the externalities of the decisions that they are about to make.

To shed light on the dynamics of the information acquisition process, we focus on information that unveils the unknown externalities gradually (i.e., *noisy* information). Whereas a piece of perfect one-shot information uncovers the truth immediately, noisy information increases one's belief accuracy bit by bit. Individuals can not only decide whether to start acquiring noisy information but also *when* to stop the inquiry. Compared to perfect information, situations with noisy information offer individuals a higher chance to end up with beliefs more desirable than their initial beliefs, by allowing them to choose when to stop their inquiries strategically.

In many economic decisions with potential externalities, individuals can acquire noisy information to guide their decisions. Examples include medical examinations that help a doctor to decide between treatments with different profits, media consumption before voting on ethically controversial but personally costly policies, or candidate screening and interviewing by discriminatory employers on the labor market. In these decisions, when individuals decide to stop acquiring noisy information

plays an important role in both the decision-making and the resulting welfare outcomes.

In this paper, we *experimentally* show how individuals strategically decide when to stop acquiring noisy information about their options' externalities when an option benefits themselves. We also show in our experimental data that strategic information acquisition motivated by selfish interests can *reduce* the negative externalities resulting from the decision.

Using a laboratory experiment, we address three challenges that render an investigation of noisy information acquisition in the field, using observational data, difficult. First, individuals' often unknown and heterogeneous prior beliefs can act as a confounding factor; in our laboratory experiment, we fix the prior beliefs of all subjects such that they begin with the same known prior belief. Second, the information history of each individual is usually hard to monitor; our experiment allows us to monitor the entire information history of each subject. Third, the access to information and interpretation of it are often heterogeneous; the information in our experiment has a clear Bayesian interpretation and is costless for all subjects. Besides, we provide the subjects with the Bayesian posterior beliefs after each piece of information to address heterogeneous ability to interpret information rationally.

More specifically, our subjects take part in a modified binary dictator game, in which each dictator has to decide between two options. The dictators know each option's outcome for themselves. In our baseline, the two options pay the dictators themselves equally. In the treatment, in contrast, one option pays the dictators more than the other option. For each dictator, contingent on an unknown binary state, one of the options reduces the payoff of the receiver, while the other does not. Before making the decision, each dictator can acquire as much noisy information as they want about which option harms the receiver. The information is costless. If one option generates a higher payoff for the dictators, they can opt for the extra payoff without a guilty conscience, as long as they believe that this option does not harm others. Whereas when the options pay themselves equally, the dictators do not have this incentive to prefer certain beliefs about the harmful option. Hence, the dictators in the latter case serve as the baseline.

In the laboratory experiment, we find that compared to the baseline, dictators facing a self-benefiting option *exploit* information: when most of the information received up to that point suggests that the self-benefiting option *harms* the receivers, a higher proportion of them *continue* acquiring information; when most of the information received up to that point suggests that the selfish option causes *no harm* to the receivers, a higher proportion of them *stop* acquiring information. How does this information acquisition strategy arise? Intuitively, having received dominant information suggesting that the self-rewarding option harms the receivers, the dictators become more inclined to forsake the additional payment. In this case, the further information might present supporting evidence for a selfish decision favorably and make them choose the self-benefiting option instead. In contrast, having received

dominant information supporting the innocuousness of the self-rewarding option, individuals face the undesirable risk that further information might challenge the previous evidence. This intuition is formalized in our theoretical model.

We empirically show results regarding receiver welfare that might not be obvious at first sight. Although one might think that strategic information acquisition motivated by selfish interests must lead to more negative externalities, our model in Chapter 2 shows that also the reverse can happen: for some agent types, motivated information acquisition *improves* the welfare of the others affected by the decision. Our experimental data provide evidence consistent with this prediction. This counter-intuitive result arises from a moral hazard problem: when disinterested, some agent types acquire only a small amount of information due to, for example, the satisficing behavior (Simon, 1955). The agent's selfish preference for one option over the other can mitigate this moral hazard problem by causing her to choose her least-preferred option only when she is certain that it is harmless to others. This result implies that delegating information acquisition to a neutral investigator might lower the welfare of the others affected by the decision.

This paper contributes insights into how people engage in motivated reasoning. To the best of our knowledge, we are the first to show that individuals strategically decide when to stop acquiring noisy information, even when they interpret information rationally. The existing literature on motivated beliefs has largely focused on biases in processing *exogenous* information and find that people react to exogenous information in a self-serving manner (Eil and Rao, 2011; Mobius, Niederle, Niehaus, and Rosenblat, 2011; Falk and Szech, 2016; Gneezy, Saccardo, Serra-Garcia, and Veldhuizen, 2016; Exley and Kessler, 2018; Zimmermann, forthcoming). In the literature on excusing selfish behavior without involving information, individuals have been found to manipulate their beliefs and avoid being asked for good deeds (Haisley and Weber, 2010; DellaVigna, List, and Malmendier, 2012; Di Tella, Perez-Truglia, Babino, and Sigman, 2015; Andreoni, Rao, and Trachtman, 2017). An early psychology paper of Ditto and Lopez (1992) documents that individuals require less supportive information to reach their preferred conclusion, possibly due to the bias of overreacting to their preferred information. In comparison, the psychology behind our finding is the tradeoff between a more informed vs. a more desirable decision, rather than the fact that information deemed more valid leads to a conclusion faster. Our experiment shows evidence that individuals use strategic information acquisition itself as an instrument for motivated reasoning.

Our empirical investigation of endogenous information choice relates to the empirical studies on the avoidance of perfectly revealing information in social decisions (Dana, Weber, and Kuang, 2007; Feiler, 2014; Grossman, 2014; Golman, Hagmann, and Loewenstein, 2017; Serra-Garcia and Szech, 2019). In contrast to information avoidance, we find that when it comes to noisy information, individuals *seek* further information if the previously received information is predominantly against the innocuousness of their selfish interests. The avoidance of perfect information docu-

mented in the previous studies importantly reveals that individuals have information preferences in social decisions. Delving into *how* people acquire information, our investigation sheds light on *what* the individuals' information preferences are in social decisions. Our model provides a unified framework for analyzing the acquisition of information, with the avoidance of perfect information as a special case.

Another related strand of the empirical literature is the one focusing on rational inattention, showing that individuals who allocate *costly* attention rationally might make decisions based on incomplete information (e.g. Bartoš, Bauer, Chytilová, and Matějka, 2016; Ambuehl, 2017; Masatlioglu, Orhun, and Raymond, 2017). As pointed out by Bénabou and Tirole (2016), when the nature of the decision so determines that some beliefs are more desirable than others, the decision-makers might engage in motivated reasoning and lean towards these beliefs. This is a different psychology than the undirected inattention. For inattention to be rational, information must be costly. In contrast, in our experiment, information entails no monetary cost and a highly limited time cost. We also limit the cognitive cost to interpret the information by providing Bayesian posterior beliefs to subjects after each piece of information.

We organize the rest of the paper as follows: In Section 3.2, we first detail the experimental design and then empirically analyze the dictators' information acquisition strategy in our experiment. In Section 3.3, we show that in our experiment strategic information acquisition motivated by the dictator's selfish interests improves the receiver welfare. In Section 3.4, we conclude and propose some ideas for future research.

3.2 Motivated Information Acquisition

In this section, we focus on how individuals acquire information about their options' externalities in a decision. In Section 3.2.1, we provide details of the experimental design that features a dictator game and costly information. In Section 3.2.2, we empirically analyze the dictators' information acquisition strategies.

3.2.1 A Laboratory Experiment With Modified Dictator Games

We conduct a laboratory experiment with modified binary dictator games. Contingent on an unknown state, one of the two options in the dictator game reduces the receivers' payoffs, and the other does not. Before deciding, the dictators can acquire information about the harmful option at no cost.

3.2.1.1 The Treatment Variations

Our experiment has a 2×2 design and 4 treatments, as illustrated in Table 3.1. The treatments vary on two dimensions: (i) whether one of the dictator game options

increases the dictators' payoffs; (ii) whether the dictators can proceed to the dictator game without acquiring any information on the externalities of their options.

The key treatment variation in our experiment is whether the dictators' selfish interests are concerned in the dictator game. We present the dictator games in Section 3.2.1.2. In the "*Tradeoff*" treatments, one option increases the dictators' payoffs, while the other does not. In the "*Control*" treatments, neither option affects the dictators' payoffs. The comparison between the *Tradeoff* and *Control* pins down the causal effect of having a self-benefiting option on the dictators' information acquisition behavior. We describe the details of this treatment variation below when we present the dictator game.

The second treatment variation concerns the dictators' freedom to acquire no information. It serves two purposes: (i) In the "*NoForce*" treatments, dictators are *not forced* to acquire any information. These treatments allow us to examine the proportion of dictators who do not acquire any information, but they also leave room for self-selection into the information processes. (ii) In the "*Force*" treatments, the dictators are *forced* to acquire at least one piece of information before making their decisions in the dictator game. This modification eliminates the potential self-selection into the information process.[1]

Table 3.1. Treatments

	With Selfish Interests	No Selfish Interests	Shorthand
No Forced Draw	*Tradeoff–NoForce*	*Control–NoForce*	***NoForce***
A Forced Draw	*Tradeoff–Force*	*Control–Force*	***Force***
Shorthand	***Tradeoff***	***Control***	-

This table presents our four treatments with a two by two design. *Tradeoff* vs. *Control* is our key treatment variation. Dictators in *Tradeoff* can gain additional payment by choosing a particular option in the modified dictator game, while those in *Control* cannot. *Force* vs. *NoForce* differ in that in the former the dictators have to acquire at least one piece of information, while in the latter they can choose to acquire no information.

3.2.1.2 The Dictator Game

Table 3.2 presents the payment scheme of the dictator game in the *Tradeoff* and the *Control* treatments respectively. In all treatments, the dictators choose between two options, x and y. There are two states of the world, *"x harmless"* or *"y harmless"*. Depending on the state, either option x or y reduces the receivers' payments by 80 points, while the other one does not affect the receivers' payment. Note that each option harms the receiver in one of the states. This design makes sure that the dictators cannot avoid the risk of harming the receiver without learning the state.

1. We explain in details this selection effect when we analyze the data in Section 3.2.2.

Table 3.2. Dictator Decision Payment Schemes

(a) *Control* Treatments

	Good state (x harmless)	Bad state (y harmless)
x	$(\mathbf{0}, 0)$	$(\mathbf{0}, -80)$
y	$(0, -80)$	$(0, 0)$

(b) *Tradeoff* Treatments

	Good state (x harmless)	Bad state (y harmless)
x	$(+\mathbf{25}, 0)$	$(+\mathbf{25}, -80)$
y	$(0, -80)$	$(0, 0)$

These tables present the dictator games in *Control* and *Tradeoff* treatments. The number pairs in the table present (dictator's payment, receiver's payment).

In *Control*, the dictators receive no additional points regardless of their choices and the state. In *Tradeoff*, x is self-benefiting for the dictators: they receive 25 additional points when choosing x, but no additional points when choosing y.

***Good* State vs *Bad* State.** For the ease of exposition, we hereafter refer to the state *"x harmless"* as the *"Good state"*, and the state *"y harmless"* as the *"Bad state"*. It is because in state *x harmless*, the dictator's and the receiver's interests are aligned in *Tradeoff*: option x is better for both of them. The dictator can claim the additional payment of 25 points without harming the receiver. Reversely, in state *y harmless*, if the dictator decides to choose x to gain the additional payment, she makes the receiver worse-off. The dictator is in a dilemma between less payment for herself or hurting the receiver. Although this contrast between states does not apply to the *Control* treatments, we will refer to *"x harmless"* as the *Good state* and *"y harmless"* as the *Bad state* for consistency.

Note that in treatments *Tradeoff*, dictators would prefer to believe that they are in the *Good state*, such that they can choose option x and gain the additional payment without having a bad conscious; whereas in the *Control* treatments, dictators are indifferent about which state they are in, since their payments are not affected by their decisions in either state.

The dictators start the experiment without knowing the state that they are in individually. They only know that in every twenty dictators, seven are in the *Good state*, and thirteen are in the *Bad state*. That is, the dictators start the experiment with a prior belief of 35% on that they are in the *Good state* and 65% in the *Bad state*. Before making the decision, they can update their beliefs by drawing information described in the next subsection.

3.2.1.3 The Noisy Information

We design a noisy information generator for each state, which generates information that is easily interpretable according to the Bayes' rule. Specifically, each piece of information is a draw from a computerized box containing 100 balls. In *Good state*, 60 of the balls are white and 40 are black; in *Bad state*, 40 balls are white and 60 are

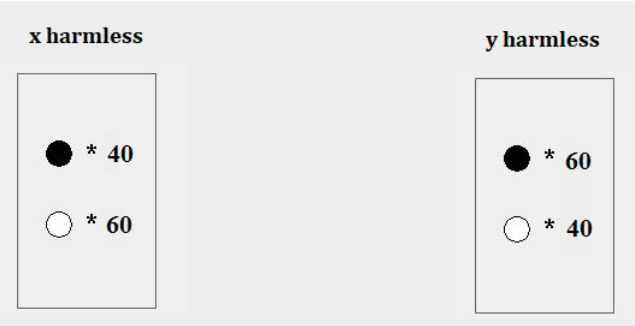


Figure 3.1. The Noisy Information Generators

black (Figure 3.1). The draws are with replacement from the box that matches to each dictator's actual state. After each draw, we display the Bayesian posterior belief on the individual computer screen, to reduce the cognitive cost of interpreting the information and reduce non-Bayesian updating.

Good News vs. Bad News. For the ease of exposition, we refer to a white ball as a piece of "good news" and a black ball as a piece of "bad news". It is because, in the *Good state*, dictators draw a white ball with a higher probability. A white ball hence supports the dictators to believe in the *Good* state, in which the dictators in treatments *Tradeoff* can choose x and gain the additional payment without reducing the payment of the receiver. Reversely, in the *Bad state*, dictators would draw a black ball with higher probability. A black ball is an evidence for the *Bad* state, in which option x rewards the dictators in *Tradeoff* at the cost of the receivers. Although dictators in *Control* do not have a preference over the two states, and hence unlikely to have a preference for black or white balls, we will still refer to a white ball as good news and a black ball as bad news for consistency.

3.2.1.4 The Experimental Procedure

The experiment consists of three parts: the preparation stage, the main stage, and the supplementary stage.

The Preparation Stage. (i) The dictators read paper-based instructions on the dictator decision, and the noisy information. (ii) We also describe in written the Bayes rule and tell the dictators that later in the experiment, we are going to help them to interpret the information by showing them the Bayesian posterior beliefs after each ball that they draw. (iii) Besides, the instructions specify that each experiment participant starts the experiment with 100 points of an endowment. (iv) We also inform them that option x is harmless for 7 out of 20 of the dictators and y for 13 out of 20. That is, the dictators' prior beliefs on the states are 35% and 65% on the *Good* state and the *Bad* state.

After reading the instructions, the dictators answer five control questions designed to check their understanding of the instructions. They keep the paper instructions for reference throughout the experiment.

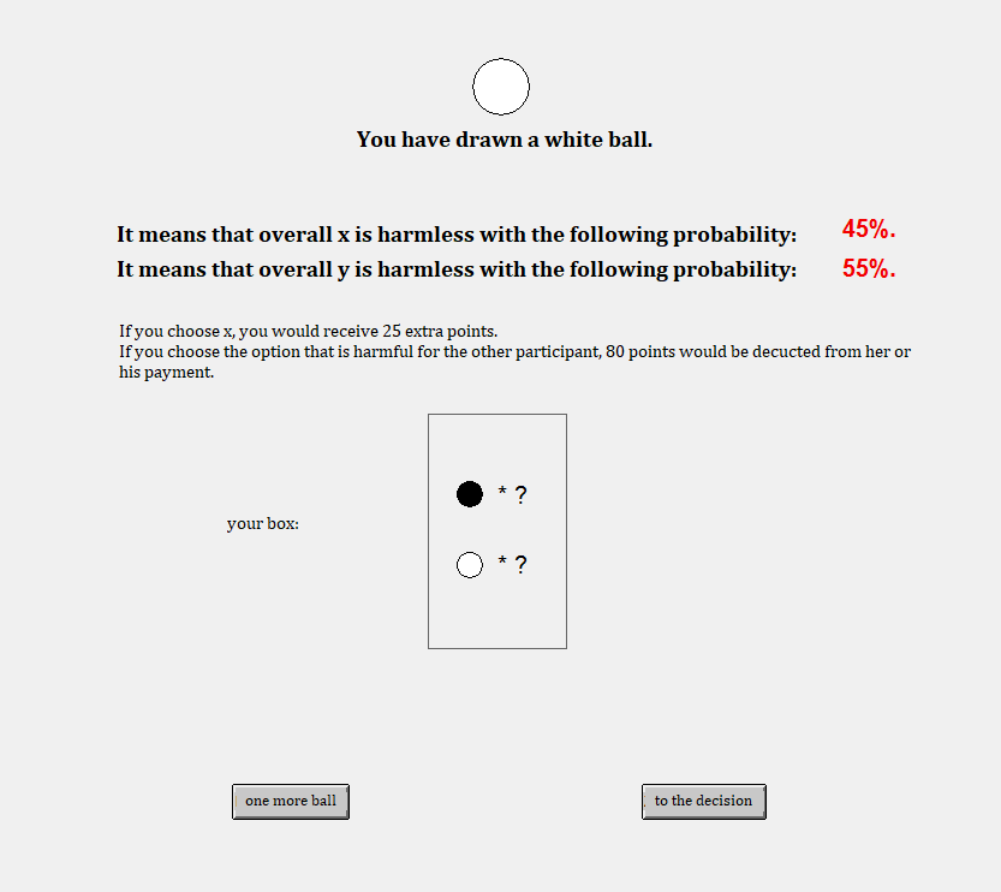


Figure 3.2. Screenshot of the Information Stage

The Main Stage. In the main stage, (i) dictators can acquire information about the state that they are individually in; (ii) they choose between x and y in the dictator game.

Specifically, the dictators can acquire a piece of information by clicking a button that makes the computer draw a ball randomly from the box matched to their actual individual state (see Figure 3.1). The draws are with replacement. After each draw, the screen displays the latest ball drawn and the Bayesian posterior beliefs on the *Good state* and the *Bad state* given all the balls drawn so far (rounded to the second decimal, see Figure 3.2). There are two buttons on the screen: one to draw an additional ball, and the other to stop drawing and proceed to the dictator game. Either to draw a ball or to stop drawing, a dictator must click on one of the buttons.

The draws do not impose any monetary cost on the dictators. The time cost of acquiring information is limited: between draws, there is a mere 0.3 second time

lag to allow the ball and the Bayesian posterior belief to appear on the computer screen. It means that a dictator can acquire 100 balls within 30 seconds, which would almost surely yield certainty.

In the *NoForce* treatments, the dictators can draw from zero to infinitely many balls. That is, they can proceed directly to the dictator game without drawing any ball, and if they decide to acquire information, the information acquisition can only be ended by them. In the *Force* treatments, the dictators must draw at least one ball, and after the first draw, they have full autonomy regarding when to stop drawing just like in *NoForce*. Besides drawing balls, the dictators have no other way to learn about the true state that they are in throughout the experiment. It is common knowledge that the receivers do not learn the information acquired by the dictators throughout the experiment.

Having ended information acquisition, dictators choose between x and y in the dictator game in Table 3.1a (in the *Control* treatments) or Table 3.1b (in the *Tradeoff* treatments). Next in the implementation state, the dictator' choices are implemented and the payments are calculated.

The Supplementary Stage. (i) We elicit the dictators' posterior beliefs on the state after the dictator game. The belief elicitation is incentivized by using the randomized Quadratic Scoring Rule. We compare the elicited and the Bayesian posterior beliefs in Appendix 3.F and find that for the majority of dictators, their elicited posterior beliefs and their Bayesian posterior beliefs coincide. (ii) The subjects take part in the Social Value Orientation (SVO) slider measure, which measures "the magnitude of concern people have for others' and categorizes subjects into altruists, prosocials, individualists, and competitive type (Murphy, Ackermann, and Handgraaf, 2011). (iii) The subjects answer a questionnaire consisting of socio-demographics, preferences, a selection of HEXACO personality inventory (Lee and Ashton, 2018), and a 5-item Raven's progressive matrices test (Raven et al., 1998). We report the details of the questionnaire in Appendix 3.F.

Implementation. We randomize within each laboratory session: (i) the *Tradeoff* and *Control* treatments, (ii) the states: we randomly assign 35% of the laboratory terminals to the *Good state*, and 65% to the *Bad state*. The subjects are then randomly seated and randomly matched in a ring for the dictator game. The subjects are told that their decisions would affect the payment of a random participant in the same experimental session other than themselves. After all the subjects have decided in the dictator game, the experiment moves on to the implementation stage, where we inform the subjects that the dictator game decisions are being implemented and their payments are affected according to another participant's dictator game decision. Each subject plays the dictator game only once.

We conducted the experiment in October and December 2018 at the BonnEconLab (*NoForce* and *Force* treatments respectively). 496 subjects took part (168 in *Tradeoff–NoForce*, 167 in *Control–NoForce*, 82 in *Tradeoff-Force* and 79 in *Control-*

Force). Among the subjects, 60% are women, and 93% are students. They are, on average, 24 years old, the youngest being 16 and the oldest being 69. The subjects are balanced between treatments, concerning gender, student status, and age (see Appendix 3.F). We used z-tree (Fischbacher, 2007) to implement the experiment and hroot (Bock, Baetge, and Nicklisch, 2014) to invite subjects and to record their participation. Instructions and interfaces on the client computers were written in German, as all subjects were native German speakers.

Payments. In the experiment, payments are denoted in points. One point equals 0.05 EUR. At the end of the experiment, the details of the points and the equivalent payments earned in the experiment are displayed on the individual computer screens. The subjects received payments in cash before leaving the laboratory. The total earnings of a subject were the sum of the following components: an endowment of 5 EUR, an additional 1.25 EUR if the subject was in treatments *Tradeoff* and chose x, a 4 EUR reduction if the subject's randomly assigned dictator made a decision that reduces her payments, a random payment of either 1.5 EUR or 0 for revealing their posterior beliefs, a payment ranging from 1 to 2 EUR depending on the subject's decisions in the SVO slider measure, a payment ranging from 0.3 to 2 EUR depending on the decisions in the SVO slider measure of another random subject in the same laboratory session, and a fixed payment of 3 EUR for answering the questionnaire. A laboratory session lasted, on average, 45 minutes, with an average payment of 11.14 EUR.

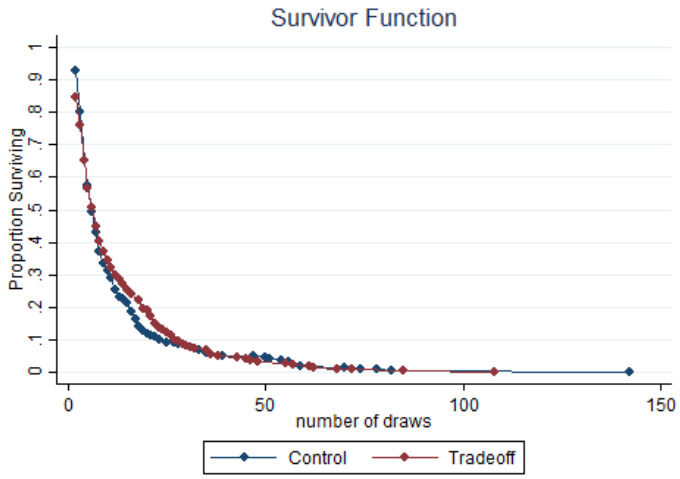
3.2.2 Empirical Analyses of Motivated Information Acquisition

In this section, we analyze the data from our experiment to investigate the effect of having a selfishly preferred option on how individuals acquire information about their options' externalities. The median number of balls drawn by the dictators is 6 (*Tradeoff*: 6, *Control*: 5; Mann-Whitney-U $p = 0.98$). We summarize our data in Appendix 3.A and proceed below with the analyses of the dictators' information acquisition behavior.

Do dictators acquire information?

Finding 1. *The proportion of dictators who do not acquire any information is 15% in Tradeoff–NoForce and 7% in Control–NoForce.*

In the *NoForce* treatments, where the dictators are allowed to draw no information before the dictator game, 38 out of 335 proceed to the dictator game without drawing information (*Tradeoff–NoForce:* 15%; *Control–NoForce:* 7%). Among them, in *Tradeoff–NoForce*, 25 out of 26 choose x, the option with additional payments for themselves; in *Control– NoForce*, where neither option produces additional payments for the dictators themselves, only 2 out of 12 choose x.



This figure plots the fraction of dictators remaining in the information acquisition process over the number of draws.

Figure 3.3. Life Table Survival Function

Table 3.3. Proportion of Dictators Drawing No Ball

	No Info%	Their Choices	
Tradeoff–NoForce	15%	x: 96%	y: 4%
Control–NoForce	7%	x: 17%	y: 83%
Chi-2 p-value	0.02	0.00	

This table displays in each treatment (i) the proportion of dictators who do not draw any ball before making their decisions between x and y; (ii) the proportion *among them* who choose option x. Note that in treatment *Tradeoff–NoForce*, dictators who choose option x receive additional payment, while those in treatment *Control–NoForce* do not.

Do dictators stop earlier in *Tradeoff* than in *Control*?

Finding 2. *Overall, the proportions of dictators who continue acquiring information after each draw do not differ between treatments.*

Figure 3.3 presents in *Tradeoff* and *Control* the proportions of dictators surviving over time, i.e. the proportion of dictators who are still acquiring information over time. The survival function does not differ between *Tradeoff* and *Control* (log-rank test for equality of survivor functions, $p = .63$).

Finding 2 speaks against an overall lower propensity to acquire noisy information when individuals' selfish interests are involved in the decision. It contrasts the avoidance of perfect information found by the previous literature (e.g. Dana, Weber, and Kuang, 2007).

When do dictators stop acquiring information? We now turn to the 458 dictators who did acquire information and focus on the role of the information history in their decisions to continue acquiring information after each draw of ball.

Specifically, we predict:
Having an option that generates additional payoffs for the dictators themselves (i) increases their tendency to *continue* acquiring information, when a dominant amount of information received up to that point is *bad* news against the innocuousness of this option; (ii) but increases their tendency to *stop*, when a dominant amount of information received is *good* news supporting the innocuousness of this option.

The intuition of the prediction is that when the dictators are inclined to forgo their selfish interests upon receiving dominant bad news, continuing the inquiry might reverse the previous bad news favorably and make them choose the self-rewarding x instead. This possibility might encourage dictators to continue drawing balls. However, when the dictators have received dominant desirable good news and are inclined to behave selfishly, the further information might be bad news that deteriorates their current desirable beliefs. This risk might discourage the dictators from drawing further information. This intuition is formalized in the theoretical model presented in Chapter 2.

In what follows, we first compare the decisions to stop acquiring the information directly after the first draw between *Tradeoff* and *Control*. Then, we analyze the entire information histories, levering insights from the research of survival analysis.

3.2.2.1 The First Draw of Ball

For dictators, whose first ball is good news and those whose is bad news, we respectively compare between *Tradeoff* and *Control* their decisions to continue acquiring information right after the first draw. The good and bad nature of the first draw is exogenous in our experiment since the composition of the 100 balls in the boxes depends solely on the exogenous state, and the draws are random.

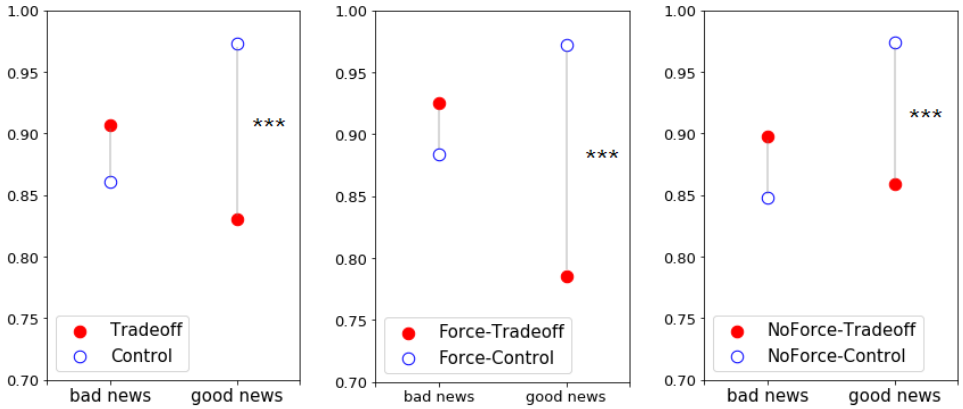
Finding 3. *(i) When the first draw is bad news, the proportion of dictators who continue drawing balls right after it is similar across treatments. (ii) In case of good news, the proportion is smaller in Tradeoff treatments than in Control treatments.*

Finding 3 shows evidence that having a self-rewarding option causes individuals to be more likely to stop acquiring further information when the previous information supports the innocuousness of this option. On the opposite, when the information received up to that point suggests that the selfish decision harms others, individuals continue acquiring information similarly with or without the self-rewarding option. Table 3.4 and Figure 3.4 present the exact proportions of dictators who continue acquiring information right after the first draw.

Table 3.4. Proportion of Dictators Continuing After the First Ball

	First News **Good**			First News **Bad**		
Treatment	Pooled	*Force*	*NoForce*	Pooled	*Force*	*NoForce*
Tradeoff	83%	79%	86%	91%	93%	90%
Control	97%	97%	97%	86%	88%	85%
Chi-2 p-value	.00	.01	.00	.26	.52	.22

This table displays the proportions of dictators who continue acquiring information after the first draw in the respective treatment, given the respective first draw. In the *Force* treatments, dictators have to draw at least one ball before choosing between x and y. Note that in the *Control* treatments the within treatment differences given different news are due to the asymmetric prior belief of 35% in the *Good* state.



(a) Pooled Data **(b)** *Force* Treatments **(c)** *NoForce* Treatments

These figures present the proportion of dictators who continue acquiring information after the first draw.

Figure 3.4. Proportion of Dictators Continuing after the First Draw

Discussion. Finding 3 is less prominent in the *NoForce* treatments, where the dictators can choose to draw no information. The reason might be the fact that the dictators in *NoForce* have selected themselves into the information process.

In *Tradeoff–NoForce*, almost all dictators who do not acquire information choose x directly. Had they received a further piece of good news, they would also be willing to stop immediately. Therefore, the *Tradeoff–NoForce* dictators' sorting out of the information process decreases the proportion of them who stop directly after the first good news and reduces the observed effect of the treatment. Similarly, in treatment *Control–Force*, almost all dictators who do not acquire information choose y directly. Had they received a piece of bad news, they might also stop immediately to choose y. Therefore, the self-selection out of the information process of the *Control–Force*

dictators decreases the observed proportion of them who stop right after the first bad news and hence is also against our finding.

3.2.2.2 The Entire Information Histories

Now we turn to the dictators' complete information acquisition process. Each dictator's information history evolves over time. To be able to include it in our analyses, we first split each dictator' complete information history at the unit of one draw.[2] The resulting data set consists of records at the person-draw level. For every draw of each dictator, the pseudo-observation records the dictator's information history up to that draw, whether the dictator chooses to stop or continue acquiring the information directly after that draw and time-constant characteristics of the dictator such as her identity, treatment assignment, and gender. After each draw, we can distinguish between information histories dominated in amount by good and bad news, using a binary dummy variable.

In the framework of a Cox proportional hazard model, we compare the decision to stop acquiring further information between treatments, given these two types of information histories: one dominated in amount by good news, and the other by bad news.[3]

We are interested in the dictators' hazard to stop acquiring information. The Cox proportional hazard model factors the hazard rate to stop acquiring information into a baseline hazard function $h_0(t)$ and covariates X_t that shift the baseline hazard proportionally, as in (3.1). The baseline hazard function $h_0(t)$ fully captures the time dependency of the hazard.[4]

$$h(t|X_t) = h_0(t) \cdot exp(X_t \cdot b). \tag{3.1}$$

Our model specification is as follows:

$$h(t|X) = h_0(t) \cdot exp(\beta_1 \text{Tradeoff} + \beta_2 \text{Info} + \beta_{12} \text{Tradeoff} \times \text{Info} + \alpha z_t), \tag{3.2}$$

where "Tradeoff" is a dummy variable for treatment *Tradeoff*, "Info" is a categorical variable denoting information histories that are dominated by bad news, good news, or balanced between the two, with bad news dominance as the baseline. z_t is a control variable that measures the accuracy of the individual belief after each

2. Time-varying covariates in survival analysis are often obtained by the method of splitting episodes (see Blossfeld, Rohwer, and Schneider, 2019, pp 137-152).

3. The Cox model has the advantage that the coefficient estimates are easy to interpret. We report a robustness check using the logistic model in Appendix 3.E. The results of the logistic model are in line with those of the Cox model.

4. Unlike many other regression models, the Cox model naturally includes no constant term, since the baseline hazard function already captures the hazard rate at covariate vector 0 (see for example Cleves, Gould, Gutierrez, and Marchenko, 2010).

ball drawn.[5] After controlling for the belief accuracy, the color of the balls per se appears to have no significant effect on dictators' stopping decisions, as shown later in Table 3.5. To allow for different shapes of the hazard function with respect to gender, cognitive ability (measured by Raven's matrices test) and prosocial types (categorized by SVO measure by Murphy, Ackermann, and Handgraaf, 2011), we stratify the Cox model by these variables (Allison, 2002).[6]

We are interested in the following two hazard ratios:
(i) the first one reflects the effect of the treatment on the hazard rate, given *bad* news dominance in the information history. That is, *ceteris paribus*

$$\begin{aligned} \text{HR}_{Bad} = \frac{h(t|\text{Bad, Tradeoff } = 1)}{h(t|\text{Bad, Tradeoff} = 0)} &= \frac{exp(\beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_{12} \cdot 1 \cdot 0 + \alpha z_t)}{exp(\beta_1 \cdot 0 + \beta_2 \cdot 0 + \beta_{12} \cdot 0 \cdot 0 + \alpha z_t)} \\ &= \frac{exp(\beta_1 + \alpha z_t)}{exp(\alpha z_t)} \\ &= exp(\beta_1); \end{aligned} \tag{3.3}$$

(ii) the second one reflects the effect of the treatment on the hazard rate, given *good* news dominance in the information history. That is, *ceteris paribus*

$$\begin{aligned} \text{HR}_{Good} = \frac{h(t|\text{Good, Tradeoff } = 1)}{h(t|\text{Good, Tradeoff} = 0)} &= \frac{exp(\beta_1 \cdot 1 + \beta_{2,\text{Good}} \cdot 1 + \beta_{12,\text{Good}} \cdot 1 \cdot 1 + \alpha z_t)}{exp(\beta_1 \cdot 0 + \beta_{2,\text{Good}} \cdot 1 + \beta_{12,\text{Good}} \cdot 0 \cdot 1 + \alpha z_t)} \\ &= \frac{exp(\beta_1 + \beta_{2,\text{Good}} + \beta_{12,\text{Good}} + \alpha z_t)}{exp(\beta_{2,\text{Good}} + \alpha z_t)} \\ &= exp(\beta_1 + \beta_{12,\text{Good}}). \end{aligned} \tag{3.4}$$

Our prediction suggests that HR_{Bad} is smaller than 1 and HR_{Good} is larger than 1. That is, (i) $\beta_1 < 0$; (ii) $\beta_1 + \beta_{12,\text{Good}} > 0$.

In Table 3.5, we report the Cox model results, with standard errors clustered at the individual level. Pooling all treatments, the Cox model coefficient estimates yields Finding 4.

Finding 4. *(i) Having received more bad news than good news, the dictators are more likely to* ***continue*** *acquiring information in Tradeoff than in Control; (ii) while they are more likely to* ***stop*** *in Tradeoff than in Control, having received more good news than bad news.*

5. We use the following score as a proxy for the accuracy of beliefs: $\text{belief}_{\text{Good}} \times \text{belief}^2_{\text{Bad}} + \text{belief}_{\text{Bad}} \times \text{belief}^2_{\text{Good}}$. It is a probabilistic belief's expectated Brier score (Brier, 1950). Brier score is a proper score function that measures the accuracy of probabilistic predictions.

6. As shown in Table 3.5, after the stratification, our main covariates affect the hazard to stop acquiring information proportionally. That is, the proportional hazard assumption of the Cox model is not violated.

The estimated coefficient of the treatment dummy is $\beta_1 = -.28$, and its interaction with the categorical variable indicating good news dominance is $\beta_{12} = .43$, both significant at 5 percent level. If bad news dominates the information history, the hazard to stop acquiring information in *Tradeoff* is $exp(-.28) = .76$ of that in *Control*, i.e. 24% lower in *Tradeoff*. In contrast, if good news dominates, the hazard in *Tradeoff* is $exp(-.28 + .43) = 1.16$ of that in *Control*, i.e. 16% higher in *Tradeoff*. That is, the treatment of having a selfishly preferred option makes the dictators more likely to continue acquiring information, when they have received predominantly bad news, and more likely to stop when they have predominantly good news. The estimation in the *Force* and *NoForce* treatments point in the same direction.

The Role of Cognitive Ability. When we focus exclusively on dictators above the median cognitive ability, as measured by Raven's matrices test (Table 3.6), we find that the effects in Finding 4 become *stronger* than the average effect that we report in Table 3.5. Having received more bad news, these dictators' hazard to stop acquiring information in *Tradeoff* is $exp(-.35) = .70$ of that in *Control*. Having received more good news, the hazard to stop acquiring information in *Tradeoff* is $exp(-.35 + .59) = 1.30$ of that in control. In comparison, considering all dictators, these numbers are .76 and 1.16, indicating that the tendency to acquire information strategically is more moderate averaging across dictators with all levels of cognitive ability than focusing on the ones with high cognitive ability. This finding suggests that the information acquisition behavior in Finding 4 is more likely out of strategic considerations than due to limited cognitive abilities.

Table 3.5. The Cox Proportional Hazard Model Results

		Pooling All		Force	NoForce
$\hat{\beta}_1$	treatment *Tradeoff*	-.28**	-.24*	-.38*	-.18
		(.12)	(.13)	(.21)	(.16)
$\hat{\beta}_{12}$	Tradeoff ×				
	Good news dominance	.43**	.41**	.32	.42*
		(.20)	(.21)	(.39)	(.26)
	Balanced	-.35	-.42	-.59	-.34
		(.38)	(.38)	(.69)	(.47)
$\hat{\beta}_2$	Good news dominance	-.14	-.23	-.18	-.23
		(.16)	(.12)	(.31)	(.22)
	Balanced	-.52**	-.56**	-.48	-59**
		(.24)	(.22)	(.38)	(.30)
Control Variables:					
Belief Accuracy		Yes	Yes	Yes	Yes
Gender, IQ, Prosociality FEs		Yes	Yes	Yes	Yes
Force treatment FE		No	Yes	–	–
Observations (individuals)		458	458	161	297
Chi2 p-value		.00	.00	.00	.00
Violation of PH		NO	NO	NO	NO

This table presents the estimated *coefficients* of the Cox model in (3.2), with standard errors clustered at the individual level. *, **, and *** denote significance at the 10, 5, and 1 percent level. The dependent variable is the hazard to stop acquiring information, and the key coefficients of interests are $\hat{\beta}_1$ and $\hat{\beta}_{12}$. $\exp(\hat{\beta}_1)$ reflects the treatment effect on the dictators' hazard to stop acquiring further information, given information histories dominated by bad news; and $\exp(\hat{\beta}_1 + \hat{\beta}_{12}|\text{Good news dominance})$ reflects the treatment effect on the hazard, given information histories dominated by good news (derivation see Equation (3.4)).

The fixed effects are taken into account by stratification, which allows the baseline hazard to differ according to the control variables, i.e., gender, the prosocial types (categorized by the SVO test), and the cognitive ability (measured by Raven's matrices test). We also control for the belief accuracy, measured by the Brier score of the beliefs after each draw (see Footnote 5). The reported likelihood Chi-square statistic is calculated by comparing the deviance ($-2\times$ log-likelihood) of each model specification against the model with all covariates dropped. The violation of the proportional hazard assumption of the Cox model (PH) is tested using Schoenfeld residuals. In all four cases, the PH is not violated for each covariate nor globally. We use the Breslow method to handle ties.

Table 3.6. The Cox Model Results For Above and Below Median Raven's Scores

		Above Median	Below Median
$\hat{\beta}_1$	treatment *Tradeoff*	-.35**	-.17
		(.16)	(.20)
$\hat{\beta}_{12}$	Tradeoff ×		
	Good news dominance	.59**	-.21
		(.27)	(.30)
	Balanced	.32	-1.03*
		(.54)	(.59)
$\hat{\beta}_2$	Good news dominance	-.10	-.25
		(.22)	(.27)
	Balanced	-.98**	-.21
		(.40)	(.32)
Control Variables:			
Belief Accuracy		Yes	Yes
Gender, IQ, Prosociality FEs		Yes	Yes
Force treatment FE		No	Yes
Observations (individuals)		267	191
Chi2 p-value		.00	.00
Violation of PH		NO	NO

This table presents the Cox model results for the subjects above and below median cognitive ability, measured by the number of correctly answered questions in Raven's matrices test, pooling data from all treatments. Standard errors are clustered at the individual level. The median number of correct answers to Raven's test is four out of five in our experiment. In this table, the subjects above the median have given correct answers to four or five questions in Raven's test, and the subjects below the median have correctly answered below four questions in Raven's test. We find that subjects with higher cognitive ability have a higher tendency to acquire information strategically. For a comprehensive table description, please see that of Table 3.5.

3.3 The Receiver Welfare

Do the dictators in *Tradeoff*, for whom x is self-rewarding, more often choose the option that reduces the receivers' payment, than the dictators in *Control*? This might seem to be the case, since the dictators in treatment *Tradeoff* might bias towards choosing x, whereas the dictators in *Control* are impartial between the option x and y. Indeed, our data show that in both states the dictators in *Tradeoff* are more likely to choose x than dictators in *Control* (details see below). However, we find that, despite the higher tendency to choose x, the dictators in *Tradeoff* do *not* choose the option that reduces the receivers' payments significantly more often (*Tradeoff*: 32%; *Control*: 27%; Chi-2 $p = 0.17$). These two observations seem to contradict each other. What is the missing piece of the puzzle?

An option being remunerative does not only directly affect the agent's decision between the options (*the decision effect*), but also indirectly by affecting how she

acquires information (*the information effect*). In Section 2.4, we theoretically show that, while the decision effect of the remuneration is always negative on the welfare of the other, the information effect is positive for some agent types. This information effect can sometimes offset the decision effect and leads to an overall neutral or even positive effect of the remuneration on the welfare of the other.

This counter-intuitive result arises from a moral hazard problem: when impartial between options, the agent might acquire little information. Therefore she sometimes mistakenly chooses the harmful option because she is ill-informed about the state. One option being remunerative can mitigate this moral hazard problem. Although she now more often falsely chooses x, the agent *less* often falsely chooses y because she now requires higher certainty in the innocuousness of y before choosing it.

Below, we disentangle the decision and the information effect in our experimental data. A direct between-treatment comparison of the receiver welfare confounds two effects of the self-reward on the receiver welfare: first, it directly affects their decision between x and y, given any acquired information (decision effect); second, the self-reward affects dictators' information acquisition, which in turn affects their beliefs about the unknown state and their choices between the options (information effect). Before we disentangle the decision effect and the information effect, we first present the dictators' choice of x and y given realized posterior beliefs at their decisions.

We observe in our data that, fixing the posterior belief, the dictators in *Tradeoff* decide differently between x and y than the dictators in *Control*. This difference affects the proportion of receivers whose incomes are reduced by the dictators' decisions (the decision effect). Specifically, in both treatments most dictators who have received more good news than bad news choose option x (*Tradeoff*: 91%, *Control*: 93%, chi-2 $p = 0.63$). Difference arises among those who have received more bad news than good news – a significantly higher fraction of these dictators in *Tradeoff* choose option x (*Tradeoff*: 27%, *Control*: 3%, chi-2 $p = 0.00$). Similarly, among those who have received equal number of good and bad news (final belief on x being harmless $= 0.35$), including those who acquire no information, significantly more dictators in treatment *Tradeoff* choose x than those in the *Control* treatment (*Tradeoff*: 81%, *Control*: 11%, chi-2 $p = 0.00$).

To empirically disentangle the decision effect and the information effect of the remuneration on the receivers' welfare, we construct a *Counterfactual* scenario, in which dictators acquire information as in the *Control* treatment, but decide as in the *Tradeoff* treatment given the acquired information and the final posterior beliefs (as illustrated in Table 3.7). When comparing the receiver welfare in the *Counterfactual* to the *Control* treatment, we isolate the decision effect by keeping fixed the information acquisition behavior; when comparing the receiver welfare in the *Counterfactual*

to that in the *Tradeoff* treatment, we isolate the information effect by keeping fixed the decision between x and y given beliefs.

Table 3.7. *Counterfactual* Scenario

	Control	*Tradeoff*
posterior beliefs	×	
decision given belief		×
compared to the *Counterfactual*	*decision effect*	*information effect*

Tables 3.8a and 3.8b show the decision effect and the information effect respectively. In Table 3.8a, we compare the *Counterfactual* with the *Control* and find a negative decision effect. The dictators in the *Counterfactual*, who employ the decision rules in *Tradeoff* given any posterior belief, choose x more often in both states. Overall, in the *Counterfactual*, the proportion of unharmed receivers is lower than in the *Control* treatment (62% compared to 73%). This means that the decision effect is negative: option x being self-rewarding for the dictators leads to a change of decision rule that makes the receivers worse-off.

In Table 3.8b, we compare *Tradeoff* with the *Counterfactual* and find a positive information effect. The remuneration makes a higher fraction of dictators choose x when x is harmless (81% compared to 75%), and a higher fraction of dictators to choose y when y is harmless (60% compared to 54%). Overall, in *Tradeoff*, the proportion of unharmed receivers is higher than in the *Counterfactual* (68% compared to 62%). The information effect of remuneration on the receiver welfare is hence positive: option x being self-rewarding makes the dictators acquire information strategically, and in turn, improves the receiver welfare.

As discussed before, there is a moral hazard problem when no option is remunerative – the dictators do not fully learn the state before they make a decision and hence often mistakenly choose the harmful option for the receiver. Note that in both states, the proportions of dictators who choose the harmless option for the receiver are lower in *Counterfactual* than in *Tradeoff*. This difference can only be due to different information acquisition behavior since the decision rule is the same between the *Counterfactual* and *Tradeoff*. In our experiment, in *Control*, 36% dictators who are actually in the *Good* state stop acquiring information at a belief in the Good state lower than their prior. The additional payment that the dictators can obtain by choosing x mitigates this moral hazard problem: in treatment *Tradeoff*, the proportion of dictators in the Good state who stop acquiring information below the prior is 26% – lower than in *Control*.

Aggregating both effects, the proportion of the receivers spared from harm does not significantly differ between the *Tradeoff* and the *Control* (68% compared to 73%, Chi-2 $p = 0.17$). It is decreased from 73% (*Control*) to 62% (*Counterfactual*) by more selfish decision-making, i.e. the decision effect, and is increased from 62%

(*Counterfactual*) to 68% (*Tradeoff*) by strategic information acquisition, i.e. the information effect.

Table 3.8. The Effects of Remuneration on Receiver Welfare

(a) The Decision Effect

State	Good State (x harmless)	Bad State (y harmless)	Overall
Counterfactual:			
% no harm	75%	54%	62%
(# total dictators)	(88)	(158)	(246)
Control:			
% no harm	54%	83%	73%
(# total dictators)	(88)	(158)	(246)
The decision effect:			-11%

(b) The Information Effect

State	Good state (x harmless)	Bad state (y harmless)	Overall
Tradeoff:			
% no harm	81%	60%	68%
(# total dictators)	(87)	(163)	(250)
Counterfactual:			
% no harm	75%	54%	62%
(# total dictators)	(88)	(158)	(246)
The information effect:			6%

This table presents the decision effect and the information effect of the remuneration in our experiment. The *Counterfactual* is calculated by combining the posterior beliefs from the *Control* and the mapping from beliefs to choices in the dictator game from *Tradeoff*. Comparing the *Conterfactual* to the *Control* (*Tradeoff*), we obtain the decision effect (information effect).

3.4 Concluding Remarks

This paper experimentally investigates how people acquire information about the externalities of their options before making a decision.

We present experimental evidence that when faced with a self-benefiting option that might harm others, individuals acquire noisy information strategically: they tend to carry on acquiring information when they have received mostly information suggesting that the selfish decision harms others; while they tend to stop having received information indicating the opposite. Moreover, in our experiment, individuals with higher intelligence exhibit a stronger tendency to acquire information this way, suggesting that this information acquisition behavior is more likely to be due to strategic considerations than limited cognitive ability.

This empirical finding sheds light on how people acquire information in various contexts where decisions incur unknown consequences on others, and noisy information is available for inquiry. One example is the credence goods market. The research on credence goods has been focusing on the deceptive behavior of the credence goods provider, while the psychology of them is less understood. The credence goods providers – physicians, car mechanics, taxi drivers – who care for the well-being of their customers face a dilemma between their monetary compensations and their unwillingness to harm their customers. Our finding suggests that credence goods providers might mitigate this dilemma by strategically learning about the best option for the customers. If by examining the need of a customer, they can persuade themselves that a profitable option is the right one for the customer, their dilemma is resolved.

Our findings also help to understand labor market discrimination. A discriminatory recruiter who likes to think of himself as nondiscriminatory might be able to maintain his positive self-view while hiring in a biased manner, by selectively stopping interviewing the candidate to persuade himself that a candidate of his less preferred character is disqualified. This insight has implications on the quality distribution of successful labor market candidates across ethnic groups and gender.

In other contexts, such as charitable giving and media consumption for voting, our result highlights the importance of the first pieces of information sent and received. If the potential donors' first information about a charitable organization is negative, she might readily stop learning about the charity and decide to keep her money in her pocket. The charity will then have a hard time to raise for its cause. When a voter inquires about an ethical issue with a personal cost for him (e.g., additional taxes), if the first news articles that he reads lean against it, the voter is likely to stop the inquiry and vote against it.

Our experimental data confirms the prediction in Chapter 2 that motivated information acquisition can improve the welfare of the other affected. This finding provides empirical evidence for the policy relevance that we raised – delegating the job of collecting information to an independent investigator, who is disinterested

in the decision, can sometimes lead to worse decision making and more negative externalities.

3.A Summarizing Statistics

Here we provide summarizing statistics on our data. The basic information of the subjects in each treatment is summarized in Table 3.9.

Table 3.9. Basic Information of Subjects

		no. obs.	Good State	women	student	av. age
Force	Tradeoff	82	.34	.45	.95	22
	Control	79	.37	.54	.95	22
	p value		.73	.24	.56	.50
NoForce	Tradeoff	168	.35	.66	.93	24
	Control	167	.35	.65	.92	24
	p value		.97	.79	.56	.36
Pooled	Tradeoff	250	.35	.59	.94	24
	Control	246	.36	.61	.93	24
	p value		.82	.62	.56	.25

This table presents the basic characteristics of our subjects in each treatment. The Mann-Whitney U test verifies that our randomization was successful.

3.B Number of Balls Drawn and the Posterior Beliefs

Table 3.10 summarizes the dictators' information acquisition behavior.

Table 3.10. Information Acquisition Behavior

		no. balls (median)	av. belief at decision	% stop above prior
Force	Tradeoff	7.5	.30	.33
	Control	4	.36	.37
	p value	.04	.04	.67
NoForce	Tradeoff	5	.34	.37
	Control	6	.33	.33
	p value	.92	.76	.44
Pooled	Tradeoff	6	.35	.36
	Control	5	.36	.34
	p value	.24	.82	.71

This table presents the statistics of the dictators' information acquisition behavior and the Mann-Whitney-U test p values comparing between *Tradeoff* and *Control*, respectively. In the *NoForce* treatments, only dictators who draw at least one ball are included.

3.C Dictator Game Decision

Table 3.11 summarizes the dictator game decisions.

Table 3.11. Dictator Game Decisions

		Choosing x%			Harm %
		Good	Bad	Overall	
Force	Tradeoff	.71	.43	.54	.38
	Control	.62	.14	.32	.23
	p value	.46	.00	.01	.04
NoForce	Tradeoff	.86	.38	.55	.30
	Control	.51	.18	.29	.29
	p value	.00	.00	.00	.84
Pooled	Tradeoff	.81	.40	.54	.32
	Control	.54	.16	.30	.27
	p value	.00	.00	.00	.17

The first three columns of this table presents the proportions of dictators who choose x given *Good* and *Bad* states and the treatments, together with the Mann-Whitney U test p values comparing between *Tradeoff* and *Control* respectively. In the *Good* state, x does not harm the receiver, while in the *Bad* state it does.The last column presents the percentage of dictators whose decision reduced the receivers' payoffs in the dictator game.

3.D The Optimal Belief Cutoffs

In this section, we infer the belief cutoffs $(\underline{p}, \bar{p})$ in Chapter 2, using the experimental data and compare them between treatments.

We find that the large majority of subjects behave consistently with the model (431 out of 496; *Control*: 228 out of 246; *Tradeoff*: 203 out of 250), i.e., they choose x if they stop at a posterior weakly above the prior or y if they stop at a posterior weakly below the prior.[7] For dictators who stop acquiring information at the equivalent of their prior belief of 0.35, 0.35 is interpreted as their upper cutoffs if they choose x and as their lower cutoffs if they choose y.

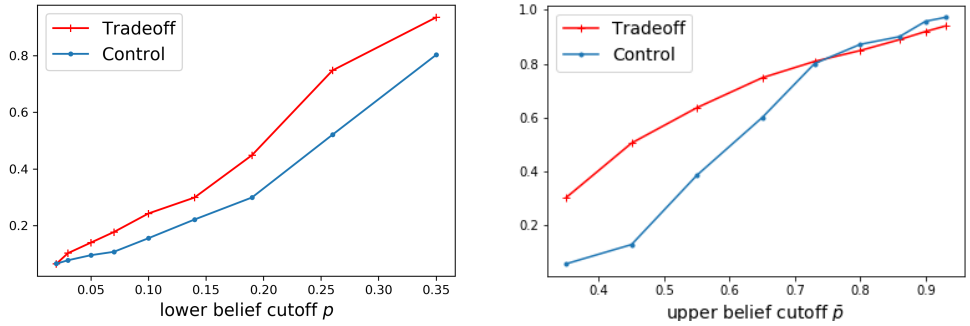
Table 3.12 summarizes the fraction of dictators who stop at their upper belief cutoffs $\bar{p}$. It reveals that the distribution of posterior belief cutoffs differ between *Tradeoff* and *Control*. In *Tradeoff*, 49% dictators stop acquiring information at a posterior belief above the prior belief, while in *Control* the fraction is only 31% (Chi square, $p = 0.00$). This finding is consistent with our theoretical model.

7. In the *Control* treatment, 14 dictators choose y after having received more good news, 4 dictator choose x after having received more bad news. In the *Tradeoff* treatment, 10 dictators choose y after having received more good news, 37 subjects choose x after having received more bad news.

Figure 3.5 shows the empirical cumulative distribution function of the upper and lower belief cutoff. Both the lower and the upper cutoff are *lower* in the Tradeoff treatment, consistent with Theorem 4.[8]

Table 3.12. Proportion of Dictators Reaching the Upper Belief Cutoff $\bar{p}$

	Overall	*Tradeoff*	*Control*	Chi-2 p
Stop at $\bar{p}$	39%	49%	31%	0.00



(a) CDf of the lower belief cutoff

(b) CDF of the upper belief cutoff

These figures show the empirical cumulative distribution functions of the lower belief cutoff (Figure 3.6a) and the upper belief cutoff (Figure 3.6b). The CDF of the lower belief cutoff reflects the data of dictators who stop information acquisition at posterior beliefs weakly below the prior and choose y. The CDF of the upper belief cutoff reflects dictators who stop weakly above the prior and choose x.

Figure 3.5. Distribution of the Observed Belief Cutoffs

8. For the interpretation of the right tail of the distribution, note that only 5% dictators stop at beliefs higher than 0.80, such that only very few observations drive the estimation of the cumulative distribution functions at very high beliefs.

3.E Robustness Check: The Logistic Regression

Using the data at the person-draw level, we estimate the following logistic model as a robustness check and find result similar to that in Section: 3.2.2.2.

$$logit\, h(X) = X \cdot b + Z \cdot a + (C + T \cdot c), \tag{3.5}$$

where $h(X)$ is the probability that the dictator stops acquiring information after that draw; X denotes the same covariates of interest as in the Cox model, i.e.

$$X \cdot b = \beta_1 \text{Tradeoff} + \beta_2 \text{Info} + \beta_{12} \text{Tradeoff} \times \text{Info}.$$

The control valuables in Z include gender, cognitive ability, prosociality and belief accuracy, all measured in the same way as in the Cox model in Section 3.2.2.2. T is a vector of time dummies, which captures the time dependency of the probability to stop acquiring information.

When interpreting the results, this logistic model can be viewed as a hazard model in which the covariates proportionally affect the *odds* of stopping acquiring information (Cox, 1975). Formally,

$$\begin{aligned} \frac{h(t)}{1-h(t)} &= \frac{h_0(t)}{1-h_0(t)} \cdot exp(X_t \cdot b + Z_t \cdot a) \\ \Rightarrow \underbrace{log\left(\frac{h(t)}{1-h(t)}\right)}_{logit\, h(X)} &= \underbrace{log\left(\frac{h_0(t)}{1-h_0(t)}\right)}_{C+T\cdot c} + X_t \cdot b + Z_t \cdot a. \end{aligned} \tag{3.6}$$

Unlike in the framework of the Cox model, the coefficients here cannot be interpreted as hazard ratios. Instead, they should be interpreted as odds ratios. Our prediction that the hazard to stop acquiring information is lower in *Tradeoff* when bad news dominates suggests a negative β_1. And the prediction that the hazard is higher when good news dominates suggests a positive $\beta_1 + \beta_{12,Good}$. Results reported in Table 3.13 support these predictions.

Table 3.13. The Logistic Model Results

		Pooling All		Force	NoForce
$\hat{\boldsymbol{\beta}}_{\mathbf{1}}$	treatment *Tradeoff*	-.25*	-.26*	-.56**	-.18
		(.15)	(.15)	(.25)	(.18)
$\hat{\boldsymbol{\beta}}_{\mathbf{12}}$	Tradeoff ×				
	Good news dominance	.35*	.37*	.71**	.34
		(.22)	(.22)	(.37)	(.26)
	Balanced	-.54	-.53	-.62	-.40
		(.40)	(.41)	(.73)	(.49)
$\hat{\beta}_2$	Good news dominance	-.21	-.21	-.14	-.26
		(.18)	(.18)	(.29)	(.24)
	Balanced	-.67**	-.68**	-.46	-78**
		(.28)	(.28)	(.46)	(.35)
Control Variables:					
Belief Accuracy		Yes	Yes	Yes	Yes
Gender, IQ, Prosociality		Yes	Yes	Yes	Yes
Time Dummies		Yes	Yes	Yes	Yes
Force Treatment Dummy		No	Yes	–	–
Observations (person-draws)		4,658	4,658	1,567	2,932
Pseudo R2		.07	.07	.09	.07

This table presents the estimated *coefficients* of the logistic model, with standard errors clustered at the individual level. *, **, and *** denote significance at the 10, 5, and 1 percent level. The dependent variable is the hazard to stop acquiring information, and the key coefficients of interests are $\hat{\beta}_1$ and $\hat{\beta}_{12}$. $\exp(\hat{\beta}_1)$ reflects the treatment effect on the dictators' odds to stop acquiring further information, given information histories dominated by bad news. And $\exp(\hat{\beta}_1 + \hat{\beta}_{12}|\text{Good news dominance})$ reflects the treatment effect on the odds, given information histories dominated by good news. We control for belief accuracy, gender, the prosocial types (categorized by the SVO test), and the cognitive ability (measured by in Raven's matrices test). The time dependency of the odds is accounted for by including a dummy for each period.

3.F Complementary Stage

After the experiment, we elicited the dictators' posterior beliefs on the state and their SVO scores. We also asked them to answer a questionnaire consisting of questions on their sociodemographics, self-reported risk preferences, time preferences, preferences for fairness, reciprocity. A selective subset of the HEXACO personality inventory (Ashton and Lee, 2009) and five items from Raven's progressive matrices intelligence test are also included.

Elicited Beliefs. In the experiment, we display the Bayesian posterior belief on the state after each draw of information on the screens of the dictators. After the dictators stop acquiring information, we elicit subjects' beliefs on the state, given all the information acquired. Figure 3.6 plots the histogram of the difference between the Bayesian posterior beliefs, and the elicited posterior beliefs at the end of the information acquisition. The majority of subjects' elicited beliefs coincide with the Bayesian posterior beliefs after the last ball they draw (299 out of 496), the elicited beliefs of the self-rewarding option x being harmless are higher than the Bayesian posterior beliefs by 2.60% (one-sample t-test $p = 0.00$). Figure 3.6 reveals no systematic bias in the elicited beliefs.

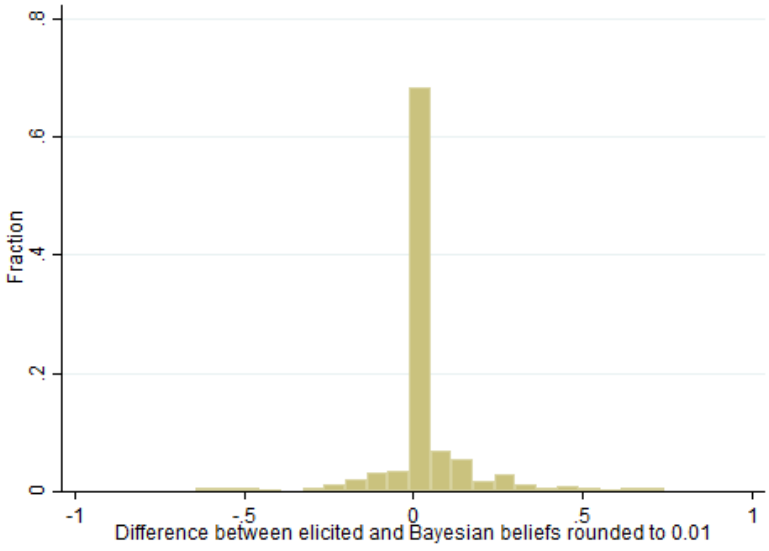


Figure 3.6. Difference between elicited posterior beliefs and Bayesian posterior beliefs

SVO Scores. The average SVO score of all the subjects is 20.49, with no significant difference between *Tradeoff* and *Control* treatments (Mann-Whitney-U test, $p = 0.84$). According to Murphy, Ackermann, and Handgraaf (2011), 48% subjects are categorized as "prosocials", 15% "individualists" and 37% "competitive type".

Cognitive Abilities. On average, the subjects answered 3.60 out of 5 questions in Raven's matrices test correctly. There is no significant difference between *Control* and *Tradeoff* treatments (Chi-square $p = 0.12$). When asked about a simple question on probability, in both treatments 92% subjects answer correctly (Mann-Whitney-U test $p = 0.85$).[9]

Preferences. To elicit risk preferences, time preferences, preferences for fairness, and reciprocity, we use survey questions in Falk, Becker, Dohmen, Huffman, and Sunde (2016). We report the exact questions in Table 3.14. All answers are given on a 0 to 10 scale.

HEXACO-60 proposed by Ashton and Lee (2009) is a personality inventory that assesses the following six personality dimensions: Honesty-Humility (HH), Emotionality (EM), Extraversion (EX), Agreeableness (AG), Conscientiousness (CO), and Openness to Experiences (OP). We select 4 questions with the highest factor loading in each dimension (as reported in Moshagen, Hilbig, and Zettler, 2014) and in addition, include 4 questions from the Altruism versus Antagonism scale (AA) proposed in Lee and Ashton (2006). Table 3.15 reports the exact questions we ask. All questions are answered on a scale from 1 to 5, where 5 means strongly agree, and 1 means strongly disagree. We use the German self-report form provided by hexaco.org.

9. We use the following question to elicit subjects' understanding of probabilities:
Imagine the following 4 bags with 100 fruits in each. One fruit will be randomly taken out. For which bag, the probability of taking a banana is 40%?
A. A bag with 20 bananas.
B. A bag with 40 bananas.
C. A bag with 0 banana.
D. A bag with 100 bananas.
The correct answer is B.

Table 3.14. Preferences Elicitation in the Questionnaire

Preferences for	Question
Risk	Please tell me, in general, how willing or unwilling you are to take risks. (10 means very willing, 0 means completely unwilling)
Time	How willing are you to give up something beneficial for your today to benefit more from that in the future? (10 means very willing, 0 means completely unwilling)
Altruism	I am always ready to help others, without expecting anything in return.
Fairness	Q1: I think it is very important to be fair. Q2: I, in general, agree that unfair behaviors should be punished.
Positive reciprocity	I am always ready to go out of my way to return a favor.
Negative reciprocity	I am always ready to take revenge if I have been treated unfairly.

Table 3.15. Selected Items From the HEXACO Personality Inventory

Dimension	Question
HH	12. If I knew that I could never get caught, I would be willing to steal a million dollars. 18. Having a lot of money is not especially important to me. 42. I would get a lot of pleasure from owning expensive luxury goods. 60. I'd be tempted to use counterfeit money if I were sure I could get away with it.
EM	17. When I suffer from a painful experience, I need someone to make me feel comfortable. 41. I can handle difficult situations without needing emotional support from anyone else. 47. I feel strong emotions when someone close to me is going away for a long time. 59. I remain unemotional even in situations where most people get very sentimental
EX	10. I rarely express my opinions in group meetings. 22. On most days, I feel cheerful and optimistic. 28. I feel that I am an unpopular person. 40. The first thing that I always do in a new place is to make friends.
AG	3. I rarely hold a grudge, even against people who have badly wronged me. 15. People sometimes tell me that I'm too stubborn. 21. People think of me as someone who has a quick temper. 45. Most people tend to get angry more quickly than I do.
CO	2. I plan and organize things, to avoid scrambling at the last minute. 26. When working, I sometimes have difficulties due to being disorganized. 44. I make a lot of mistakes because I don't think before I act. 56. I prefer to do whatever comes to mind, rather than stick to a plan.
OP	1. I would be quite bored by a visit to an art gallery. 13. I would enjoy creating a work of art, such as a novel, a song, or a painting. 25. If I had the opportunity, I would like to attend a classical music concert. 55. I find it boring to discuss philosophy.
AA	97. I have sympathy for people who are less fortunate than I am. 98. I try to give generously to those in need. 99. It wouldn't bother me to harm someone I didn't like. 100. People see me as a hard-hearted person.

References

Allison, Paul. 2002. "Bias in Fixed-Effects Cox Regression With Dummy Variables." *Manuscript, Department of Sociology, University of Pennsylvania*, [97]

Ambuehl, Sandro. 2017. "An Offer You Can't Refuse? Incentives Change What We Believe." *CESifo Working Paper Series No. 6296. Available at SSRN: https://ssrn.com/abstract=2917195*, [86]

Andreoni, James, Justin M Rao, and Hannah Trachtman. 2017. "Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable giving." *Journal of Political Economy* 125 (3): 625–653. [85]

Ashton, Michael C, and Kibeom Lee. 2009. "The HEXACO–60: A short Measure of the Major Dimensions of Personality." *Journal of Personality Assessment* 91 (4): 340–345. [112, 113]

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. "Attention Discrimination: Theory and Field experiments with Monitoring Information Acquisition." *American Economic Review* 106 (6): 1437–75. [86]

Bénabou, Roland, and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30 (3): 141–64. [83, 86]

Blossfeld, Hans-Peter, Gotz Rohwer, and Thorsten Schneider. 2019. *Event History Analysis with Stata*. Routledge. [96]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "hroot: Hamburg registration and organization online tool." *European Economic Review* 71: 117–120. [92]

Brier, Glenn W. 1950. "Verification of forecasts expressed in terms of probability." *Monthly weather review* 78 (1): 1–3. [97]

Cleves, M, W Gould, R Gutierrez, and Y Marchenko. 2010. *An Introduction to Survival Analysis Using Stata*. College Station, TX, Stata Press. [96]

Cox, David R. 1975. "Partial Likelihood." *Biometrika* 62 (2): 269–276. [110]

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33 (1): 67–80. [85, 93]

DellaVigna, Stefano, John A List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127 (1): 1–56. [85]

Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman. 2015. "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism." *American Economic Review* 105 (11): 3416–42. [85]

Ditto, Peter H, and David F Lopez. 1992. "Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions." *Journal of Personality and Social Psychology* 63 (4): 568. [85]

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): 114–38. [85]

Exley, Christine, and Judd B Kessler. 2018. "Motivated Errors." [85]

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*, [113]

Falk, Armin, and Nora Szech. 2016. "Pleasures of Skill and Moral Conduct." *CESifo Working Paper Series*, [85]

Feiler, Lauren. 2014. "Testing Models of Information Avoidance with Binary Choice Dictator Games." *Journal of Economic Psychology* 45: 253–267. [85]

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171–178. [92]

Gino, Francesca, Michael I Norton, and Roberto A Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives* 30 (3): 189–212. [83]

Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel van Veldhuizen. 2016. "Motivated Self-Deception, Identity and Unethical Behavior." *Working Paper*, [85]

Golman, Russell, David Hagmann, and George Loewenstein. 2017. "Information Avoidance." *Journal of Economic Literature* 55 (1): 96–135. [85]

Grossman, Zachary. 2014. "Strategic ignorance and the robustness of social preferences." *Management Science* 60 (11): 2659–2665. [85]

Haisley, Emily C, and Roberto A Weber. 2010. "Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior." *Games and Economic Behavior* 68 (2): 614–625. [85]

Lee, Kibeom, and Michael C Ashton. 2006. "Further Assessment of the HEXACO Personality Inventory: Two New Facet Scales and an Observer Report Form." *Psychological Assessment* 18 (2): 182. [113]

Lee, Kibeom, and Michael C Ashton. 2018. "Psychometric Properties of the HEXACO-100." *Assessment* 25 (5): 543–556. [91]

Masatlioglu, Yusufcan, A Yesim Orhun, and Collin Raymond. 2017. "Preferences for Non-Instrumental Information and Skewness." *Working Paper*, [86]

Mobius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat. 2011. "Managing Self-Confidence: Theory and Experimental Evidence." Working paper. National Bureau of Economic Research. [85]

Moshagen, Morten, Benjamin E Hilbig, and Ingo Zettler. 2014. "Faktorenstruktur, Psychometrische Eigenschaften und Messinvarianz der Deutschsprachigen Version des 60-item HEXACO Persönlichkeitsinventars." *Diagnostica*, [113]

Murphy, Ryan O, Kurt A Ackermann, and Michel Handgraaf. 2011. "Measuring Social Value Orientation." *Judgment and Decision Making* 6 (2): 771–781. [91, 97, 112]

Raven, John Carlyle et al. 1998. *Raven's progressive matrices and vocabulary scales*. Oxford pyschologists Press. [91]

Serra-Garcia, Marta, and Nora Szech. 2019. "The (in) Elasticity of Moral Ignorance." *CESifo Working Paper*, [85]

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99–118. [85]

Zimmermann, Florian. Forthcoming. "The Dynamics of Motivated Beliefs." Working paper. [85]