# Essays on Beliefs and Economic Behavior

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

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

vorgelegt von

**Thomas Wilhelm Graeber**

aus Göttingen

Bonn

2018

| | |
|---|---|
| Dekan: | Prof. Dr. Daniel Zimmer, LL. M. |
| Erstreferent: | Prof. Dr. Armin Falk |
| Zweitreferent: | Prof. Dr. Lorenz Götte |
| Tag der mündlichen Prüfung: | 27. August 2018 |

# Acknowledgements

In early 2015, I made a habit of bothering my advisor Armin Falk on a weekly basis to enthusiastically pitch new ideas, only to find out that they were not nearly as brilliant as I had thought. He could have quickly dismissed me after confronting me with the truth about (what I know now) were terrible ideas. Instead, in countless discussions, he slowly made me grasp what makes for good research. This was my primary source of learning in graduate school and made each successive idea a little less terrible. Academically, I am indebted to Armin for pushing me to not be easily satisfied but to dig deeper and *get to the bottom of things*. But perhaps more importantly, he has provided a research environment in which I felt comfortable at a personal level. The experience that research need not be a lonely enterprise but lives off personally and intellectually rewarding discussions is what kept me enthusiastic and most shaped my plans for the future. For his support at all levels I am grateful to Armin.

I thank my second advisor and coauthor, Lorenz Goette, for numerous productive, inspiring and fun meetings. I have greatly benefited from the confidence he put in me early on during my research efforts. I am grateful to the *Bonn Graduate School of Economics* and the *Institute on Behavior and Inequality* for financial and administrative support, as well as to Holger Gerhardt for his relentless help in preparing and running experiments.

I consider myself extremely lucky to be part of a close network of graduate students, who have made the past five years such an enjoyable experience inside and outside of academia. This includes, most of all, my friends Lukas Püttmann, Chris Roth, Kilian Russ, Fabian Walders and Justus Winkelmann.

I thank my parents, who have always encouraged me to pursue whatever goals I set for myself. Their unconditional love and support has provided an unchanging notion of home to rely on at all times. I dedicate this dissertation to them. Finally, my greatest thanks and admiration goes to my girlfriend, Maria. She has changed my life in a much more fundamental way than any of my research activities ever will. Our life together is my indispensable source of happiness.

# Contents

# List of Figures

# List of Tables

# 1

# Introduction

The core of economic models of human behavior are preferences and beliefs. Preferences capture how a person values different outcomes and beliefs specify what a person thinks about unobserved states of the world. Both are combined in an optimization routine that leads to a choice. Standard economic theory puts a parsimonious and tractable structure on the elements of this *as if* model of behavior: people have egoistic and stable preferences, incorporate all available information into their beliefs according to a normative updating rule, and are endowed with unconstrained cognitive resources to identify the optimal action. While these assumptions produced a powerful framework that rationalizes numerous phenomena, they also generated predictions that are systematically at odds with reality. This has motivated the literature on behavioral economics that equips the economic model with more realistic assumptions to explain what seemed to be empirical puzzles. Much of this work focuses on enriching the scope of preferences and on relaxing the notion of unconstrained optimization.

Instead, this thesis empirically investigates the systematic role of subjective beliefs in economic behavior. How do people incorporate information into their beliefs? What are the cognitive mechanisms underlying different updating rules? To what extent is observed heterogeneity in beliefs and behavior predictable? Are beliefs shaped by people's individual experience? How do beliefs translate into economic behavior?

The following four chapters revolve around these questions and explore the hypothesis that a more nuanced account of the nature of subjective beliefs improves the explanatory power of the economic model of human behavior. The unifying approach of this work is cross-disciplinary and fundamentally relies on incorporating ideas from other fields such as cognitive science, psychology, sociology and anthropology.

The empirical motivation for **Chapter 2: “Inattentive Inference”** is the pervasive miscalibration to information in practice: the dominant puzzle is *overreaction* induced by excessive belief swings (Bondt and Thaler, 1985), whereas other evidence documents belief inertia and information rigidities, leading to *underreaction*. There is a limited understanding, however, of the common processes underlying belief formation that reconcile such divergent findings.

Chapter 2 addresses this issue, starting from the simple observation that for most information structures in practice, learning in a normatively optimal (Bayesian) way is practically impossible. The reason is that any piece of news usually contains information about a multiplicity of variables. Correct inference requires taking into account all underlying variables. However, the difficulty of updating grows exponentially in the number of dimensions. At the same time, mental computing is costly and the processing capacity of the human brain is limited (Newell and Simon, 1972). To accommodate these constraints, we need to reduce the complexity of real-world contexts to manageable levels. One way of doing this is to selectively attend to the personally most relevant variables in an information structure (*signal*), while ignoring other dimensions (*noise*).

Chapter 2 studies selective attention in belief formation. In simple choice experiments, I investigated how people deal with noise in information structures. I found that the vast majority of beliefs corresponded to exactly three updating rules. The largest fraction exhibited *noise neglect*, which means a person treated a piece of information as if it were specific to the variables he was most interested in, while ignoring other dependencies. This overattribution to a subset of dimensions that guide subsequent actions generated overreaction. The second, somewhat smaller share accorded to the normatively optimal Bayesian benchmark, as if taking into account all variables. The third and smallest fraction committed *information neglect* and displayed no updating at all in the face of new information. This set of beliefs sticked to the prior and therefore underreacted.

A series of additional experiments provided three main insights. First, the origin of different updating rules were the set of elements of an information structure that a person attended to. Such selective processing of the context induced a simplified *subjective representation* of the information environment. This mental model determined how the person perceived a given situation, and formed the basis for further mental computations that resulted in a posterior belief. Second, the relative prevalence of the three updating modes emerged as if resulting from trading off the expected benefits against the cognitive costs of forming the corresponding beliefs. Third, the mental process by which people adopted a representation occurred outside their awareness. That the underlying psychological mechanisms were unconscious had far-reaching implications for, e.g., the success of different debiasing strategies such as nudging and the persistence of bias in the presence of feedback.

**Chapter 3: “Heterogeneity of Loss Aversion and Expectations-Based Reference Points”** moves the focus to how people’s beliefs affect their decisions. Specifically, this chapter considers the role of forward-looking beliefs, i.e., expectations. A seminal insight from psychology is that we tend to evaluate outcomes relative to a reference point. In the canonical model of economic behavior, however, utility is defined over final levels of consumption or wealth. In theories of reference-dependent decision-making, by contrast, people code outcomes as gains or losses relative to some reference point. Yet, the *location* of this reference point is a critical degree of freedom. A recent theoretical advance characterizes the reference point based on people’s expectations about their own future outcomes (Kőszegi and Rabin, 2006). Imagine that an employee expects to be paid five thousand euros more in the next year. At the end of this year, he learns that his salary will only increase by one thousand euros. He will partly perceive this pay raise as a loss, because he compares the outcome against his own expectations. In recent years, empirical tests of this model yielded mixed results and there remains a lack of consensus on the location – and thus the empirical relevance – of reference points.

Chapter 3 attempts to reconcile different approaches and findings. In this study that is joint work with Lorenz Goette, Charles Sprenger and Alexandre Kellogg, we developed a tightly controlled exchange experiment with two main innovations: First, the design recognizes that testing the role of expectations-based reference points requires experimental control of other plausible avenues of reference dependence, such as the status quo or personal experience. Second, it accommodates a critical confound related to the key behavioral parameter, loss aversion. Loss aversion captures that people dislike losses more than equal-sized gains. A growing body of evidence documents substantial heterogeneity in measured levels of loss aversion, with a substantial fraction of people being loss-neutral or even loss-loving. Different levels of loss aversion, however, lead to different signs of comparative statics. In our results, recognizing heterogeneity in loss aversion allowed us to reliably recover the central prediction of expectations-based reference points. Moreover, our manipulation of individual exchange experience identified a distinct effect of exchange experience on behavior, which was driven by the *subjective* perception of previous trading experience. In sum, this study sheds light on the simultaneous forces of forward-looking and backward-looking sources of reference-dependent behavior and stresses the importance of systematic considerations of heterogeneity in empirical work.

The first two chapters assume an almost mechanical perspective on beliefs by exploring information processing and the translation of beliefs into behavior. In **Chapter 4: “Breaking Trust: On the Persistent Effect of Economic Crisis Experience”**, I incorporate the specific content of beliefs and consider one particularly important belief: trust. Trust is the degree of belief in the benevolent intentions of another person. It is considered a basic foundation for human progress (Harari, 2015). In the realm of economic behavior, trust plays a central role as a prerequisite for all forms of economic exchange: without a minimal amount of trust in the counterpart, no person would be willing to sign a contract. In fact, trust has been shown to affect economic outcomes at the individual, group and societal levels. However, much less is known about the origins of trust. Recent evidence documents that levels of trust vary substantially across locations and over time, but the determinants of this geographical and temporal variation are not well understood.

In Chapter 4, which is joint work with Tom Zimmermann, we analyzed the economic implications of a *breach of trust* argument, positing that trust is not easily restored once it has been abused. Building on a nascent literature on the economic implications of people's experience, we hypothesized that trust is partially determined by the experience of catastrophic macroeconomic events. Using a variety of identification strategies in a large cross-country sample, we estimated a persistent and robust negative long-term effect of economic crisis experience on trust in other people. In line with the breach of trust hypothesis, the effect was specific to living through crises in trust-intensive domains, most of all banking crises. The effect was not driven by distrust in financial institutions but was accommodated by a lack of confidence in the political class, and operated via beliefs rather than changes in preferences.

**Chapter 5: "Negative Long-run Effects of Prosocial Behavior on Happiness"** studies happiness, a topic that has played a secondary role in standard economic analysis. Welfare theory as well as economic policy in practice have focused on objective and quantifiable measures such as output and growth. However, the correlation between these measures and a population's perceived well-being is far from perfect. In recent times, measures of subjective well-being are increasingly viewed as relevant indicators of a society's welfare, and a rising number of countries have incorporated national happiness levels as a policy objective (Layard, 2011). This has sparked scientific interest in the causes of happiness. Perhaps most prominently, recent studies contribute to a debate spanning more than two millennia on the hypothesis that *prosocial behavior* is a key to happiness. The existing causal evidence indeed confirms a positive influence of prosocial behavior on happiness, but are limited to short-term effects of an enforced prosocial or selfish act (Dunn et al., 2008).

In Chapter 5, which is joint work with Armin Falk, we reconsider this hypothesis in a behavioral experiment that extends the scope of previous studies in various dimensions. In our *Saving a Life* paradigm, every participant either saved one human life in expectation or received one hundred euros, respectively. Using a choice between two binary lotteries with different chances of saving a life, we observed subjects' intentions at the same time as creating random variation in prosocial outcomes. We repeatedly measured happiness at different time horizons after the experiment. We confirmed the previous consensus finding of a positive short-term effect, but this effect quickly faded. As time passed, the sign of the effect even *reversed*, and we recorded significantly greater happiness associated with the selfish outcome than with the prosocial outcome one month later.

These findings hint at distinct sources of happiness. On the one hand, physical consumption can generate happiness. On the other hand, people derive happiness from their memories, thoughts and expectations, i.e., they *consume their beliefs* (Ariely and Norton, 2009). Prosocial choices and outcomes create the latter type of happiness, promoting feelings of *warm glow* or *doing the right thing*. These feelings, however, plausibly occur in temporal proximity to the time of choice and fade over time. Happiness derived from physical consumption, by contrast, is linked to the time when actual consumption occurs, which can be spread out over time. Chapter 5 provides an initial piece of evidence that a comprehensive understanding of the effects of prosocial behavior on happiness requires a more nuanced view that accounts for delayed effects.

A common thread of this thesis is the focus on the sources and implications of heterogeneity in beliefs and economic behavior. First, average behavior often masks a substantial amount of underlying structure, e.g., in how people process information or react to trading experience. Second, much of this heterogeneity is predictable. Beliefs systematically respond to contextual features and individual experience. Uncovering and organizing these influences is a promising path towards a deeper understanding of the role of beliefs in economic behavior.

## References

**Ariely, Dan and Michael I. Norton (2009):** "Conceptual consumption." *Annual Review of Psychology*, 60, 475–499. [5]

**Bondt, Werner F. M. and Richard Thaler (1985):** "Does the stock market overreact?" *The Journal of Finance*, 40 (3), 793–805. [2]

**Dunn, Elizabeth W., Lara B. Aknin, and Michael I. Norton (2008):** "Spending money on others promotes happiness." *Science*, 319 (5870), 1687–1688. [4]

**Harari, Y. N. (2015):** *Sapiens: A Brief History of Humankind.* HarperCollins. [4]

**Kőszegi, Botond and Matthew Rabin (2006):** "A model of reference-dependent preferences." *The Quarterly Journal of Economics*, 121 (4), 1133–1165. [3]

**Layard, Richard (2011):** *Happiness: Lessons from a New Science.* Penguin UK. [4]

**Newell, Allen and Herbert A. Simon (1972):** *Human Problem Solving.* Prentice-Hall. [2]

# 2

# Inattentive Inference

## 2.1 Introduction

Updating beliefs in a normatively optimal way is cognitively demanding even for simple information structures. The reason is that complexity grows exponentially in the number of variables going into an information-generating structure.[1] An investor assessing an analyst recommendation, a consumer reading a customer review of a product or an economist interpreting macroeconomic indicators in principle need to account for thousands of underlying variables when making inference from these pieces of information. However, mental computing is costly and the processing capacity of the human brain is limited. To accommodate these constraints, we can selectively attend to only some elements of an information structure when forming a posterior. For example, the consumer might only care about the functional quality of a product, even though he knows that the product rating also reflects the reviewer's assessment of its aesthetics. To simplify an updating problem, we might then account for a selected few of the variables included in an information structure, while ignoring others, or we might not even incorporate a piece of information at all. This paper studies selective attention in belief formation from noisy information.

The empirical analysis builds on a simple experimental design that tightly controls the information environment. An updating problem consists of two real-valued random variables, $X$ and $Y$, that are drawn from known distributions. Subjects get paid to guess the realization of $X$, but not of $Y$. Before stating a guess, they receive a piece of information $I$ that depends on both random variables, e.g., $I = X + Y$. From the perspective of a subject, this information structure is noisy: it has a *signal* part, $X$, which is the subject's learning target, and a

[1] Assume an information structure that depends on $k$ binary variables. The number of possible realizations of these $k$ binary variables is $O(2^d)$. This rapid growth of complexity is a form of the curse of dimensionality.

*noise* part, $Y$.[2] However, to learn from $I$ about $X$, subjects need to account for the variation coming from $Y$. Intuitively, extracting a signal from noisy information requires accounting for the noise.

In a set of baseline experiments conducted in the laboratory and online, three modes of updating captured the vast majority of observed beliefs. The largest share of beliefs was formed *as if ignoring* $Y$, that means subjects interpreted the information as if it did not depend on $Y$. They overattributed the information to $X$ and consequently overreacted in their guesses of $X$, relative to the rational (Bayesian) adjustment. This mode of updating, called *noise neglect* henceforth, was not an artifact of the complexity of the updating problem. In a control condition, *Broad*, subjects were incentivized to predict both $X$ and $Y$ instead of only $X$, so that $Y$ was not noise. Both conditions featured exactly identical information structures and subjects should rationally have formed the same belief.[3] In *Broad*, however, $Y$ was not neglected and most beliefs were closely aligned with the Bayesian posterior. The second, smaller share of beliefs accorded to Bayesian updating. The resulting posteriors were well-calibrated to the informativeness of the signal. The third and smallest systematic portion of stated posteriors sticked to the prior, i.e., no updating occurred. I will refer to this mode as *information neglect*, which caused underreaction relative to the Bayesian adjustment. Together, beliefs *exactly* in line with these three modes made up between sixty and up to more than ninety percent of beliefs in each task.

The central finding of a pronounced trimodality of posterior beliefs was robust to a battery of variations of the data and information structures and modifications of the experimental procedures. It provides the point of departure for a comprehensive investigation of the heterogeneity, predictability and the underlying cognitive mechanisms of updating patterns. To organize and guide through the analyses I develop a simple conceptual framework in Section 2.2.

In this framework, belief formation proceeds in three steps. First, the agent chooses which elements of an information environment $(X, Y, I)$ to attend to. Attention to each of the three elements is all-or-nothing, i.e., a dimension is either fully processed or completely ignored. The attention vector induces a *subjective representation*, or mental model of the information environment. Second, upon observing information $i$, the agent forms a posterior belief given his represen-

[2] Noise is generally defined as an unwanted source of stochastic variation that masks the signal (Shannon and Weaver, 1949). $Y$ is noise because given an action, i.e., a guess of $X$, the subject's utility does not depend on the realization of variable $Y$.

[3] That the information structure was held fixed across conditions distinguishes this design from previous studies (Caplin et al., 2011; Dean and Neligh, 2017; Enke, 2017; Enke and Zimmermann, 2017; Khaw et al., 2017). That means the data-generating process, the signal structure, the induced prior, the Bayesian posterior and the stake size were kept exactly identical.

tation of the situation. Third, he takes an action, which can be thought of as predicting $X$. Utility directly depends on the realization of $X$, but not of $Y$.[4]

Different subjective representations lead to different posterior beliefs, but they also come at different cognitive cost. This cognitive cost captures the computational resources required to calculate a posterior given a representation. In the model, the agent pays attention so as to adopt the representation that maximizes the expected utility benefit net of the cognitive cost, as if following from a form of cost-benefit analysis. The key property of the cost function is that required mental resources are determined by the *dimensionality* of the updating problem, reflecting that complexity is above all driven by the number of variables. The dimensionality-based form of the cost function is the reason people will choose from a discrete set of updating modes and not choose mixtures. In Section 2.2.2, I argue that the framework is conceptually distinct from rational inattention theory (Caplin and Dean, 2015; Sims, 2003).[5] Importantly, inattentive inference features a mechanic link between belief formation and actions: because we tailor an action to (our beliefs about) certain variables, we attend to these – but not other – variables by default, even in the absence of new information. In the model, the agent always attends to $X$, irrespective of an additional updating problem.

The framework generates the three cases observed empirically based on subjective representations that follow from different attention strategies. Bayesian updating results from a complete representation of the environment. Information neglect means that an agent does not represent the information $I$. Noise neglect, in turn, corresponds to a representation in which the agent processes the information $I$, but removes $Y$ from the information structure. This framework provides a structured way to study inference from noisy information, using a cognitive foundation geared to the empirical findings. The primitive of different updating rules is the set of elements of a situation that an agent attends to. This representation serves as the basis for other mental operations, e.g., computations, that result in a posterior.[6]

Motivated by the conceptual framework, the paper proceeds in two steps. First, in Section 2.4, I investigate whether inattentive inference is driven by a

[4] More generally, $X$ and $Y$ can have arbitrary dimensions. Treatment *Broad* is nested as a case where $X$ has two elements and the dimensionality of $Y$ is zero.

[5] Inattentive inference is concerned with *processing* exogenously given information, rather than the *acquisition* and choice of an information structure. It also suggests a different object of attention, i.e., elements of an information structure. Moreover, it establishes a direct link between attention and actions.

[6] This distinction between representation and subsequent computations squares with the computational theory of mind in psychology (Horst, 2011) that has been invoked by other recent work on belief formation (Enke, 2017).

form of cost-benefit analysis. Second, I shed light on the underlying cognitive mechanisms in Section 2.5.

If inattentive inference is the product of a consideration of the benefits and costs of different updating modes, it should systematically respond to variations in those. The benefits depend on features of the information environment. In additional experiments I tested the effects of bias and variance introduced by noise. Treatment *Signal-to-Noise Ratio* varied the ratio between the variance of $X$, $\sigma_X^2$, and the variance of $Y$, $\sigma_Y^2$. The utility loss associated with noise neglect relative to Bayesian updating rises as the variation of $Y$ increases, while that of information neglect decreases. Intuitively, if an ever greater part of the information structure is noise, it gets more harmful to treat the information as if it were precise, and less harmful to ignore the information altogether. These predictions were borne out in the data. With higher signal-to-noise ratio, noise neglect increased at the expense of information neglect. Notably, the overall share of beliefs in line with the three updating modes remained approximately constant. The second treatment, *Directional Bias*, analyzed the effect of directional bias in information while fixing the signal-to-noise ratio. Under noise neglect, beliefs are more biased the larger the deviation between the mean information value and the mean of $X$. As this difference increased across tasks, people indeed became less likely to neglect the noise. These two treatments demonstrated that the *relative* prevalence of different updating modes varied in line with their expected utilities.

To shed light on whether inattentive inference was also shaped by cognitive costs, I analyzed the effect of between-subject differences in cognitive skills as a proxy of this cost. An incentivized measure of cognitive skill significantly predicted the propensity to form Bayesian beliefs, albeit with a moderate effect size. Moreover, there was substantial within-subject consistency in updating modes, indicating the existence of individual-specific inference styles. At the same time, a significant fraction of subjects in all experiments used at least two different updating modes. The combined findings highlight that most people were able to produce Bayesian beliefs, but did not always do so.

The evidence on the plausibility of a cost-benefit type reasoning does not sufficiently characterize the underlying processes. Building on the distinction between mental representations and subsequent computations on those representations, Section 2.5 examines the mechanisms through which the initial representation is formed. The concept of a cost-benefit analysis seemingly implies a conscious, willed activity. Yet, the process of adopting a representation need not be deliberate but could occur automatically. In consequence, the resulting neglect itself might not be perceptible to the agent's conscious mind. This idea leverages research from neuroscience and psychology showing that even complex mental operations are routinely executed outside of a person's awareness

(Dijksterhuis and Aarts, 2010). Viewing subjective representations as the result of an unconscious optimization routine has several implications.

First, the effect of interventions and nudges in the inference process should depend on whether they affect the mental representation formed by a person, or only the downstream computations while leaving the representation unaltered. In a bonus task, a tenfold increase in stake size caused subjects to try harder and spend more time on the problem, but this consciously exerted effort left performance unaffected. It apparently did not help noticing the initial neglect. This shows that inattentive inference systematically occurred under effortful solution strategies.[7] On the other hand, the conceptual framework naturally gives rise to the possibility of "lightbulb moments" that make people "wake up" after receiving hints that alter the mental representation directly. In treatment *Hint*, a simple reminder to "also think about the role of $Y$" nearly eliminated noise neglect. In fact, once nudged to attend to $Y$, subjects were able and willing to compute Bayesian beliefs.

Second, people should generally be confident in their deficient beliefs if they are unaware of the underlying neglect. In condition *Confidence*, I elicited subjects' minimum valuations for their stated beliefs.[8] Strikingly, reservation prices were unrelated to inattentiveness in beliefs, implying that subjects who committed noise neglect were nevertheless fully confident in their beliefs. This and further experimental variations consistently suggested that subjects lacked any metacognitive experience of their neglect, i.e., they were unaware of both the processes responsible for inattentive inference and the resulting discrepancy between their mental model and the external environment. They confidently computed a posterior, but employed a faulty solution strategy, i.e., representation of the problem.

A third implication of unawareness is that inattentive inference can be persistent, even in the presence of feedback. Learning that a belief was flawed should not necessarily lead to improvement if people were fundamentally unaware about the unconscious processes at the root of their mistakes. Reflecting on the source of error would instead lead people to first "blame" those steps of their solution strategies that they are consciously aware of. I developed three treatments to causally test the hypothesis that learning from feedback is less likely the more *consciously executed* steps are associated with the inference process, *holding fixed the complexity of the problem*. The results suggest that a critical reason for the persistence of biases is that subjects are unaware of *why* errors oc-

[7] This is one reason why inattentive inference is not easily reconciled with dual process models, in which a primary characterization of Type 1 reasoning is as providing effortless responses (Kahneman, 2003).

[8] I elicited the minimal certain amount subjects prefer over being paid out for their stated guesses using an incentive-compatible price-list method.

cur, which compromises targeting them adequately in the presence of surprising feedback.

Section 2.2 introduces the motivating conceptual framework. Section 2.3 discusses the baseline design and results. Section 2.4 analyzes predictability and heterogeneity of updating patterns and Section 2.5 investigates mechanisms. Section 2.6 discusses the related literature and Section 2.7 concludes.

## 2.2 Conceptual Framework

### 2.2.1 Baseline Model

This section introduces a simple framework that organizes the empirical findings. An agent observes a signal $i$. The state of the world is fully characterized by two real-valued stochastic variables, $X$ and $Y$, with known distributions described by joint density function $h$. An observed piece of information $i$ is generated by a known information structure $g$, $I = g(X,Y)$. The entire information environment is described by $R = (X,Y,I)$.

I assume the agent proceeds in three steps. First, he chooses which elements of the information environment $R$ to attend to. An attention vector $a = (a_X, a_Y, a_I)$ induces a subjective representation $\tilde{R}(a)$ of the environment. For reasons described later I will focus on attention that is all-or-nothing in each dimension. A feature of the environment is either fully attended to, or not at all, $a_i \in \{0,1\}$. E.g., attention vector $a = (1,0,1)$ corresponds to a subjective representation $\tilde{R} = (X,S)$. Second, upon observing information $i$, the agent forms a subjective posterior belief described by joint density $\tilde{h}(\tilde{R},i)$ given his representation of the situation. I will discuss below how $\tilde{h}$ is connected to each possible subjective representation. Third, he takes an action $m(\tilde{h})$ based on his belief.

This is the basic three-step structure of the framework. Utility $v(\tilde{R}) := u(m(\tilde{h}),x)$ depends on the realization of $X$, but not of $Y$. Accordingly, the optimal action depends only on the posterior belief about $X$.[9] Dimension $Y$ will sometimes be called *noise*. To fix ideas, an action can be thought of as guessing the realization of variable $X$ using information that is also affected by $Y$, and utility is determined by some incentive structure that rewards the accuracy of the guess given the true realization of $X$. Note that the agent always attends to $X$, which is imperative to take an action. That means belief formation is shaped by a mechanical relationship between action and attention: when updating beliefs, the agent by default already pays attention to those variables that he tailors his

[9] Put differently, conditional on an action, utility is independent of the realization of variable $Y$. In Appendix 2.B.1 I refine and endogenize the classification of variables as $X$ or $Y$. This extension provides a foundation of this distinction based on which variables an agent optimally tailors his action to. This in particular pertains to cases where $X$ and $Y$ have many dimensions.

actions to. To simplify notation, I sometimes omit dimension $X$ in the attention vector, $a = (a_X = 1, a_Y, a_I) = (a_Y, a_I)$.

Different subjective representations $\tilde{R}$ lead to different posterior beliefs, but they also come at different cognitive cost $C(\tilde{R})$. This cognitive cost captures on the computational resources required to calculate a posterior given a representation. I assume that the agent pays attention to select the representation $\tilde{R}$ that maximizes the expected utility benefit minus the cognitive cost:

$$\tilde{R} \;=\; \tilde{R}(a^*) \quad \text{with} \quad a^* \;=\; \underset{a\,\in\,\{0,1\}^2}{\arg\max}\, \mathbb{E}\left[v(\tilde{R}(a))\right] - C(\tilde{R}(a)) \tag{2.1}$$

This formulation purports that people engage in a form of cost-benefit analysis that systematically responds to two elements. On the hand, features of the environment change the expected benefit of different representations via $\mathbb{E}\left[v(\tilde{R}(a))\right]$. On the other hand, behavior depends on the cost structure of different updating modes, $C(\tilde{R}(a))$. The key property of the cost function is that required mental resources are determined by the *dimensionality* of the updating problem, which reflects the curse of dimensionality: complexity as measured by the joint sample space grows exponentially in the number of variables, a phenomenon called *combinatorial explosion*. For a given environment $R$, the agent's perception $\tilde{R}$ determines which features the agents cognitively represents. Specifically, I assume that accounting for a continuous variable or a signal recruits fixed amounts of mental resources:

$$C(\tilde{R}(a)) \;=\; q \cdot (a_Y + l \cdot a_S) \tag{2.2}$$

Since $X$ is always attend to, I normalized its cost to 0. The fixed cognitive cost of computation associated with an additional dimension in the updating problem is $q \geq 0$, which in practice varies between individuals. The cost of accounting for the piece of information – rather than ignoring it – is $q \cdot l$, with $l \geq 0$.

The cost function is a fundamental unknown in models of attention and provides a critical degree of flexibility (Caplin and Dean, 2015). The motivation of the present cost function is that the cost of information processing does not primarily depend on the expected informativeness of a signal, but on the richness of the information environment. The assumed step cost structure associated with the dimensionality of a problem reflects the realistic feature that each additional dimension renders the cognitive process discretely more difficult, and leads to the dimension-wise all-or-nothing attention strategy from above.

There are three possible attention strategies $a = (a_Y, a_I)$. $a = (1, 1)$ is the full-attention benchmark where are all dimensions of the environment are fully represented. $a = (0, 1)$ means he attends to the information but not to the noise $Y$, a case labeled *noise neglect*. $a = (0, 0)$ specifies the case in which the agent does not attend to the information, which further obviates the need to attend to the noise $Y$. I call this case *information neglect*.[10] Moving from info neglect over noise neglect to full attention, the process of updating beliefs on the corresponding representations is accompanied by increasing cognitive costs.

How are posterior beliefs formed on a given subjective representation? The fully attentive representation $\tilde{R}^B = \tilde{R}((1,1)) = (X, Y, I)$ induces Bayesian updating:

$$\tilde{h}\left(x|i;\tilde{R}^B\right) \ = \ h(x) \cdot \frac{h(i|x)}{h(i)} \tag{2.3}$$

An agent who commits information neglect, $\tilde{R}^{IN} = \tilde{R}((0,0)) = (X)$ builds a posterior that is equal to the prior:

$$\tilde{h}\left(x|i;\tilde{R}^{IN}\right) \ = \ h(x) \tag{2.4}$$

The most interesting case is noise neglect, $\tilde{R}^{NN} = \tilde{R}((0,1)) = (X, I)$, which requires specifying how an agent interprets information that is perceived as being generated by $X$ alone. While the empirical findings indicate a somewhat more intricate general rule, let us assume for now that the agent picks the element in the support of $X$ that is closest to the signal realization, which reasonably fits the evidence in most algebraic signal structures studied in this paper:

$$\tilde{h}\left(x|i;\tilde{R}^{NN}\right) \ = \ \mathbb{1}_{\left\{W = \underset{\text{supp}(X)}{\arg\min} \left|x - i\right|\right\}} \tag{2.5}$$

I will call belief formation that is in line with this framework *inattentive inference* because deviations from Bayesian updating are driven by inattention to specific dimensions of the information environment.

The above framework has three key features. First, it views mental representations as the primitives of belief formation. Second, cognitive cost are primarily determined by the dimensionality of an updating problem, which is motivated by the curse of dimensionality that occurs when trying to interpret a piece of information as the result of the joint realizations of many variables. This characterization of cognitive costs is why people stick to a discrete set of representations and the ensuing updating rules. They adopt representations *as if* based on a cost-benefit analysis. Third and perhaps most importantly, attention in belief formation is fundamentally determined by the nature of the action an agent takes. The framework relies on a classification of variables as being signal ($X$) or noise ($Y$) to an agent. This is the cornerstone of the belief formation framework. The relevance of $Y$ for the agent is limited to the belief formation process. That means

[10] Note that attention $a = (1,0)$ would lead to the same action as information neglect, but at a higher cost due to the unnecessary attention paid to $Y$. I will not discuss this case any further.

$Y$ is relevant to the extent that it can help form a more accurate belief about $X$, but $Y$ has no effect on utility *given a belief about X*. The framework as outlined above relies on a crude distinction into $X$ and $Y$ variables based on whether the realization of a variable affects utility given an action, or not. Appendix 2.B.1 develops a more realistic version of this distinction, in which an agent optimally chooses which variables he wants to tailor his action to. He will ignore variables that only affect the action (and thereby his expected utility) by a sufficiently small amount. This extension results again in a binary characterization of variables as being signal ($X$) or noise ($Y$). However, it provides a clearer intuition for the sources of noise neglect. If an agent takes an action based on what he thinks about $X$, he will consider $X$ even in the absence of any new information. When new information arrives, he thinks about $X$ already, though not about $Y$. In a sense, the agent can leverage the cognitive cost (incurred for acting on $X$) in the belief formation process. If the agent could only decide between bearing the combined cost of attending to all dimensions ($X, Y$ and $I$) and ignoring the information altogether, he might frequently neglect information in order to avoid the large cognitive cost. However, since he pays the cognitive cost of attending to $X$ in any case, both information neglect and Bayesian updating (based on a complete representation) might be dominated by noise neglect, which exploits the fact that he already attends to $X$ but avoids the cost of accommodating $Y$.

### 2.2.2 Relationship to Rational Inattention Theory

The framework shares a common ground with a vibrant line of research that is associated with the term *rational inattention*. All of this literature shares the notion that information is costly, and people (rationally) trade of that cost against the utility of improved decision quality. There are three key differences between inattentive inference and this line of work.

First, rational inattention is concerned with the *acquisition* of information. People are assumed to choose (some elements of) the structure of their information. In the most prominent type of model pioneered by Sims (2003), subjects have unrestricted power over the structure of their information, in Verrecchia (1982) people choose the variance of a piece of information, and in Reis (2006) they decide whether to receive a fully informative signal or no information. In inattentive inference, by contrast, information structures are exogenous and people cannot change them. The focus is instead on *processing* available information.

The second distinction from rational inattention concerns the primitive of attention and the nature of cognitive costs. In rational inattention theory, the object of attention are pieces of information, while inattentive inference considers attention to elements of a given information structure. The observed information is only one element of the information structure. There is no room in rational inattention models to not attend to individual variables included in an information structure: If people acquire a piece of noisy information, they update in a *multidimensional* way that recognizes all underlying variables. In particular, models of rational inattention cannot generate noise neglect. This difference in the object of attention is reflected in the specification of cognitive costs. The most widespread approach in rational inattention theory is to model agents as finite-capacity channels, where costs are based on the Shannon mutual information between prior and posterior beliefs (Sims, 2003). In many situations this can be thought of as putting a limit on the variance reduction from prior to posterior. Inattentive inference, by contrast, suggests that the cost of processing is not primarily driven by, e.g., the expected informativeness of an information structure, but by the dimensionality of an updating problem. Fixing the overall amount of uncertainty reduction that could be achieved by updating in a Bayesian way, inattentive inference builds on the idea that costs increase in the number of variables included in the information structure.

Third, inattentive inference is fundamentally motivated by a mechanical link between belief formation and the action an agent takes. The way we tailor our actions to certain unobserved states of the world leads us attend to some variables by default, even in the absence of newly arriving information. This predetermined attention plausibly affects how we choose to allocate additional attention in the face of new information. This idea cannot be modeled with the tools of rational inattention. Assume, for example, the case of being paid to guess only $X$ versus being paid to guess both $X$ and $Y$, fixing the absolute stake size between conditions. A compound piece of information $I$ on $X$ and $Y$ provides an identical updating problem in both conditions from the perspective of rational inattention. It does not in inattentive inference.

I will next outline the empirical setting and return to specific predictions of the framework at various stages of the empirical analysis.

## 2.3 Evidence for Inattentive Inference

I analyze two empirical settings in this paper. The laboratory provides maximum control and thus serves two purposes. First, it provides specific evidence on the existence of noise neglect. Second, I analyze the psychological mechanisms that drive observed behavior using tightly controlled settings.

Online experiments provide access to a different, more diverse sample observed under less controlled choice conditions. I complement the baseline and mechanism evidence about noise neglect with online experiments for two reasons. First, the online study uses a simpler but more general design variation to study updating from noisy information. It is less centered on the identification of noise neglect. It serves to test the robustness and generality of the laboratory results and allows precise and inexpensive replication by other researchers. Second, I study the nature of updating rules in a large variety of information structures and explore predictors of heterogeneity in belief formation. The online experiments allow to run multiple treatment variations with a large number of participants, which would be infeasible in the laboratory.

### 2.3.1 Baseline Study on Noise Neglect: Laboratory Evidence

#### 2.3.1.1 Design

Causally identifying noise neglect in belief formation requires *(i)* a fully controlled and transparent data-generating process and information structure that is known to subjects, *(ii)* an experimental manipulation of the presence of noise, *(iii)* limited complexity to minimize confusion, *(iv)* a clear prediction for the posterior under neglect of noise, and *(v)* an incentive-compatible procedure to extract beliefs. In the following I present a tightly controlled laboratory experiment that meets all of these criteria.

The crux of the design is to create an environment that allows to vary the presence of noise *without changing the information structure or data-generating process*. The simplest such setting features only two unknown states of the world, i.e., two unobserved random numbers $X$ and $Y$, generated by stochastic processes known to subjects. The numbers are independently drawn from two discrete uniform distributions, each with a size of the sample space below ten. Subjects receive an easily understandable signal $I$ on the two unknown draws, such as the sum of the two numbers, $I = X + Y$. That virtually all subjects are, in principle, capable of forming normatively optimal beliefs in this simplistic setup is confirmed in the data. I define noise as a source of variation that constitutes an unwanted modification to a signal (Shannon and Weaver, 1949). This definition directly lends itself to an interpretation in terms of incentives as discussed in Section 2.2: noise is a stochastic component in the information structure, but its realization does not directly affect the agent's utility given an action. If subjects are paid for their accuracy of guessing $X$, but have no monetary prediction incentives for $Y$, then I consider $Y$ to be noise within the information structure. However, $Y$ is not characterized as noise as soon as it is explicitly incentivized. Accordingly, there are two experimental conditions in the baseline design: In *Narrow*, subjects are paid to guess only $X$, while in *Broad*, subjects are paid to guess both $X$ and $Y$. This implementation of noise highlights its inherenlty subjective nature. Given an informational structure, what is noise to one person might be the signal for another person. Note that the induced prior, the signal structure, and the Bayesian posterior are identical in *Narrow* and *Broad*. Moreover, by randomly paying for only one of the guesses in *Broad* (about $X$ or $Y$), the incentive size is kept constant.

**Table 2.1.** Overview of baseline tasks

| Sample space *X* | Sample space *Y* | Info structure | Info value |
|---|---|---|---|
| 30, 40, 50, 60, 70 | 10, 20, 30, 40, 50, 60, 70, 80, 90 | $(X+Y)\div 2$ | 60 |
| 180, 190, 200, 210, 220 | 180, 190, 200, 210, 220 | $(X+Y)\div 2$ | 200 |
| 130, 140, 150, 160, 170 | -25, -15,-5, 0, 5, 15, 25 | $X+Y$ | 165 |
| 80, 90, 100, 110, 120 | -30, -20, -10, 0, 10, 20, 30 | $X+Y$ | 80 |
| 230, 240, 250, 260, 270 | 210, 220, 230, 240, 250, 260, 270, 280, 290 | $(X+Y)\div 2$ | 230 |

*Notes*: This table provides an overview of the five baseline belief tasks in the laboratory study. The distributions of $X$ and $Y$ as well as the signal structure are identical in both treatment conditions. $X$ and $Y$ were independently drawn from two discrete uniform distribution, i.e., every indicated outcome was equally likely.

Subjects play the five updating problems of Table 2.1 in random order without receiving feedback in between. For example, in the first task of of Table 2.1, $X$ is one of five numbers, 30, 40, 50, 60 or 70 with equal probability, while $Y$ is independently drawn as a multiple of 10 between 10 and 90. Subjects learn that the average of $X$ and $Y$ is 60 and then state their belief. To solve the problem, a subjects needs to identify all $(X, Y)$ combinations with an average of 60, that is (30, 90), (40, 80), (50, 70), (60, 60), (70, 50). Both numbers being drawn uniformly and independently, it is intuitive that each of these outcomes is equally probable. The elicitation procedure extracts the maximum amount of information about subjective beliefs by having subjects indicate the full posterior distribution instead of point predictions. At the end, one of the tasks is randomly selected to be paid out based on the Binarized Scoring Rule with a prize of 10 euros (Hossain and Okui, 2013).[11] Subjects receive extensive instructions and had to complete eight control questions that test their understanding of the instructions, the data-generating process and signal structure, as well as the elicitation protocol. In two unpaid practice tasks subjects learned how to indicate a given belief in a way that maximizes their payoff. This training stage was identical in both treatments.

A notable feature of this design is that unlike previous belief formation studies, the present experiment does not alter the updating task between conditions, significantly reducing concerns about differential complexity (Caplin et al., 2011; Dean and Neligh, 2017; Enke, 2017; Enke and Zimmermann, 2017; Khaw et al., 2017).

Beyond the tasks in Table 2.1, a number of different task specifications and additional treatment variations address robustness of the baseline results and examine the nature of updating rules (see Sections 2.3.1.6 and 2.3.1.7).

[11] The scoring rule proposed by Hossain and Okui (2013) elicits truthful beliefs even if subjects are risk averse or do not follow the expected utility hypothesis.

#### 2.3.1.2 Baseline Hypothesis: Existence of Noise Neglect

I adapt the framework presented in Section 2.2 to the experimental setting with discrete data. The Bayesian posterior belief about $X$ given the signal is characterized by a discrete probability distribution $P(X|I) = \frac{P(I|X)\cdot P(X)}{P(I)}$. This normative benchmark (i) applies independent of the decision maker's incentives and (ii) depends on $Y$ through $I$. Accordingly, the treatment manipulation is inconsequential under Bayesian updating.

Consider a person who selectively attends to the dimensions he perceives as being most important while taking into account cognitive costs. In condition *Broad*, $X$ are $Y$ are equally important for the decision maker's payoff. In condition *Narrow*, however, the realization of $Y$ neither changes the decision maker's optimal action, i.e., her optimal prediction about $X$, nor does it affect her payoff given an action. She might be (partially) inattentive to $Y$, which is noise to her.

Note that if inattention leads to a neglect of $Y$, it is a priori unclear which form this neglect takes. For example, the decision maker might underestimate the variance of $Y$, replace $Y$ with a default value, be somehow unaware of its existence, or apply a specific rule in the belief formation process. Different hypotheses about the structure of inattentive inference require different auxiliary assumptions, such as about the default value. The baseline design is agnostic about the functional form of noise neglect – if there is any – and seeks to impose minimal assumptions on its parametric structure.[12] In all baseline tasks of Table 2.1, the information structure is an unbiased estimator of the mean of $X$. Either subjects receive the average of the drawn numbers and the prior distributions of $X$ and $Y$ have an identical mean, or they see the sum of the drawn numbers and $Y$ has a mean of zero. This provides a natural way of how $Y$ can be neglected, namely by interpreting the information as an unmodified signal about $X$.[13] The inattentive posterior distribution with full noise neglect in the baseline design can be specified in accordance with Equation 2.5 as $\tilde{P}^{NN}(X|I) = \mathbb{1}_{\{X=x^*\}}$ where $x^* = \underset{x\,\in\,\mathrm{supp}(X)}{\arg\min}\,\left|x - i\right|$.

**Hypothesis 1.** *Beliefs formed in Narrow and Broad significantly differ. Subjects in condition Narrow display noise neglect.*

[12] I investigate the precise patterns in additional experimental variations, see Section 2.3.1.7.

[13] Noise neglect in condition *Narrow* is observationally equivalent to taking the observed information at face value for the unobserved $X$. I demonstrate that belief formation here is not driven by anchoring on the signal value. First, anchoring cannot explain a treatment effect because the signal is identical across treatments. Anchoring should similarly affect beliefs in condition *Broad*. Second, the additional treatment variation *Computation* explicitly rules out anchoring effects, see Section 2.3.1.6.

#### 2.3.1.3 Procedural Details

Subjects in condition *Broad* guess the joint distribution of $X$ and $Y$ and are randomly paid for their accuracy in guessing either of these. The decision screen is displayed in Appendix Figure 2.G.24. Subjects in condition *Narrow* only guess the marginal distribution of $X$ (Appendix Figure 2.G.21).[14] To reduce potential experimenter demand, the design unobtrusively obfuscates the experimental objective. Subjects received their signal in encrypted form and had to decipher it using a simple two-step decoding protocol.[15] No subject had trouble implementing the protocol. Each belief elicitation (excluding the deciphering stage) was subject to a time limit of five minutes. The findings are robust to removing both the deciphering and the time limits (see Section 2.3.1.6). The belief updating problems were followed by a questionnaire. To shed light on correlates of subject-level heterogeneity in belief formation, I measured performance on a incentivized test of cognitive ability (10 Raven matrices, 0.2 euros per correct answer) and elicited a measure of risk preferences (Falk et al., 2016) as well as two personality questionnaires, the Big 5 inventory (Rammstedt and John, 2005) and the Interpersonal Reactivity Index (Paulus, 2009).

144 student subjects (72 in condition *Narrow*, 72 in condition *Broad*) participated in six sessions of the baseline experiment run at the University of Bonn's BonnEconLab in July 2017. Treatment status was randomized within-session. I invited subjects using hroot (Bock et al., 2014) and implemented the study in oTree (Chen et al., 2016). Mean earnings amounted to 11.40 euros – including 5 euros show-up fee – for an average session duration of 57 minutes.

#### 2.3.1.4 Baseline Results: Laboratory

**Result 1.** *Beliefs in Narrow significantly differ from Broad in line with noise neglect in condition Narrow.*

Figure 2.1 shows the raw beliefs in all five baseline tasks. It displays the sample distribution of subjective belief distributions for each condition, as well as the Bayesian benchmark and the value of the observed information. The average subject in *Broad* forms belief that are closely aligned with the sophisticated

[14] While this is the natural and preferred design to test for noise neglect, note that there is treatment difference in *what* is elicited, namely $X$ and $Y$ versus only $X$. Additional treatment variations harmonize the elicitation protocol, i.e., subjects with *Narrow* incentives predict both $X$ and $Y$, and subjects with *Broad* incentives predict first the marginal of $X$, and then the marginal of $Y$ on a separate subsequent page. All main findings persist. See Section 2.3.1.6.

[15] Concretely, subjects saw a sequence of letters. First, each letter had to be translated into a digit based on a displayed table. Then the number 20 had to be added to the result. Subjects were familiarized with the deciphering process in the practice stage. See also the instructions reproduced in Appendix 2.G.

Bayesian posterior. In *Narrow*, by contrast, subjects on average assign too much probability mass to outcomes close to the signal value, as implied by inattention to $Y$.

Three key implications are that (i) there is no systematic misunderstanding of the experimental setup, since subjects in *Broad* successfully form Bayesian beliefs, (ii) in *Narrow*, beliefs *overshoot* in the direction of the information value and (iii) are *overprecise* relative to Bayesian and *Broad* beliefs. Task (3) in Figure 2.1 exemplifies the role of overprecision. Since the signal realization coincides with the mean of the Bayesian posterior distribution, subjects in *Narrow* form beliefs featuring the correct *expected value* of $X$. They are, however, far too faithful that this expected value of $X$ was actually drawn. This would be unobservable if I only elicited point predictions about the mean of $X$.

Table 2.2 provides an overview of summary statistics and non-parametric tests by task. Median beliefs in *Narrow* (column 3) and *Broad* (column 4) closely correspond to the observed information (column 1) and Bayesian benchmark (column 2), respectively. Column 7 shows that belief distribution means and belief distribution variances are significantly different between treatments at the 0.1% level (M-W $U$ tests).[16]

**Table 2.2.** Beliefs about $X$ in baseline tasks

| Observed information | Bayesian posterior distribution | Subjective posterior distribution | | Sign test of median | | M-W U test |
|---|---|---|---|---|---|---|
| | | *Narrow*<br>N=72 | *Broad*<br>N=72 | *Narrow*<br>vs. Bayesian | *Broad*<br>vs. Bayesian | *Narrow*<br>vs. *Broad* |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | *distribution mean*<br>*(distribution variance)* | *median of distribution means*<br>*(median of distribution variances)* | | *p-value: distribution of means*<br>*(p-value: distribution of variances)* | | |
| 60 | 50 | 60 | 50 | < 0.001 | 0.664 | < 0.001 |
| | (200) | (0) | (200) | (< 0.001) | (0.011) | (< 0.001) |
| 230 | 237.6 | 230 | 240 | < 0.001 | < 0.001 | < 0.001 |
| | (71.7) | (0) | (67) | (< 0.001) | (< 0.001) | (< 0.001) |
| 200 | 200 | 200 | 200 | 1.000 | 0.012 | 0.024 |
| | (200) | (0) | (200) | (< 0.001) | (0.004) | (< 0.001) |
| 80 | 95 | 80 | 95 | < 0.001 | 0.508 | < 0.001 |
| | (125) | (0) | (125) | (< 0.001) | (0.180) | (< 0.001) |
| 165 | 155 | 165 | 155 | < 0.001 | 1.000 | < 0.001 |
| | (125) | (25) | (125) | (< 0.001) | (0.180) | (< 0.001) |

*Notes*: This table displays beliefs in *Narrow* and *Broad* for each one of the five baseline tasks. An elicited belief corresponds to a full distribution, which is described here by its mean and variance. I show medians of subjective distribution means and variances in each condition and compare these to the mean and variance of the Bayesian posterior distribution. Column (7) shows treatment comparisons for the distributions of distribution means and variances. The task order was randomized.

[16] This holds for all tasks except the distribution means in task (3), in which the observed information coincides with the Bayesian posterior mean.

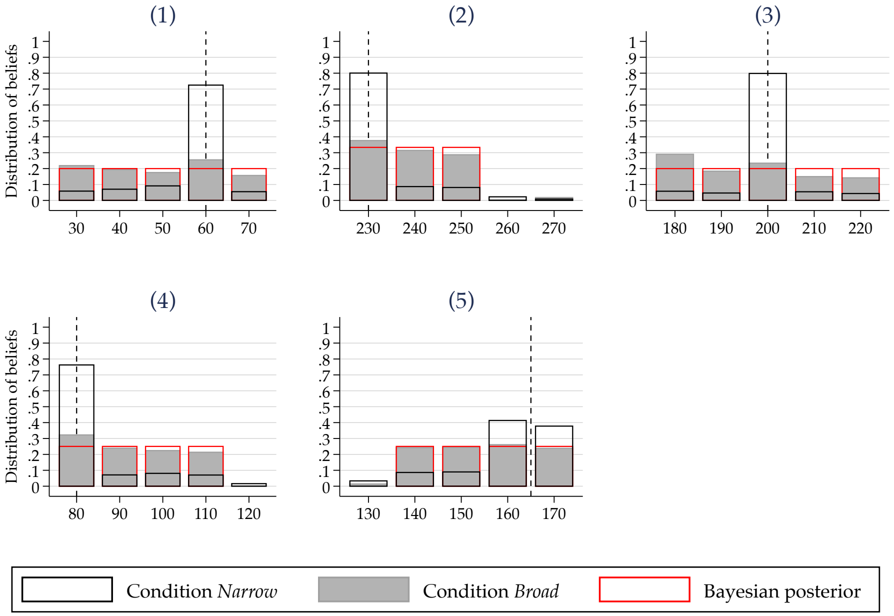


**Figure 2.1.** Distribution of elicited belief distributions about $X$ in each one of five baseline tasks. N=72 for each condition in each task. The horizontal axis shows possible outcomes of $X$. The Bayesian posterior belief is provided for reference. The observed signal is indicated by the vertical dashed line. In all five tasks, $X$ and $Y$ follow independent discrete uniform distributions that were shown to subjects. Task order was randomized. In task (1), $Y \sim \mathcal{U}\{10, 20, \ldots, 80, 90\}$, in task (2), $Y \sim \mathcal{U}\{-30, -20, \ldots, 20, 30\}$, in task (3), $Y \sim \mathcal{U}\{180, 190, 220, 210, 220\}$, in task (4), $Y \sim \mathcal{U}\{-30, -20, \ldots, 20, 30\}$ and in task (5), $Y \sim \mathcal{U}\{-25, -15, -5, 0, 5, 15, 25\}$. Subjects observed the mean of the drawn numbers in tasks (1), (2) and (3), and they saw the sum in (4) and (5).

Noise neglect comes at a sizeable cost for the decision maker. The average expected payoff for the beliefs stated in the baseline tasks was 53% higher in *Broad* than in *Narrow* (5.86 versus 3.82 euros, $p < 0.0001$, M-W $U$ test).[17]

#### 2.3.1.5 Typical Beliefs and Updating Rules

The measures of central tendency analyzed above obfuscate the presence of specific updating rules. Next I analyze what are *typical* beliefs in each condition. To this end I characterize each stated belief in terms of how close it is to the Bayesian posterior relative to noise neglect, recognizing that each observation corresponds to a full distribution rather than a single value. To obtain a single measure of distance between distributions, I first calculate the Hellinger distances (Hellinger, 1909) between the stated posterior and the Bayesian posterior:[18]

$$H_B = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k}\left(\sqrt{\tilde{P}(X_i|S)} - \sqrt{P^B(X_i|S)}\right)^2} \tag{2.6}$$

Given an analogous distance to the inattentive posterior distribution, $H_{NN}$,[19] I define a score of inattention to noise, $\theta$, that captures the distance of the subjective belief distribution to the Bayesian distribution, relative to the sum of the distances of the subjective distribution to the inattentive and the Bayesian posterior:

$$\theta = \frac{H_B}{H_B + H_N} \tag{2.8}$$

A Bayesian belief corresponds to $\theta = 0$ and noise neglect to $\theta = 1$. The parameter $\theta$ can be backed out for every stated belief, independent of the specific updating task, so I proceed with a joint analysis of the data pooled together from all tasks. Figure 2.2 is a histogram of empirical inattention parameters, split by treatment condition. More than 70% of beliefs are roughly Bayesian in

[17] Actual earnings for the baseline tasks also significantly differed across groups (means of 4.56 in *Narrow* and 2.22 euros in *Broad*, $p = 0.005$, M-W $U$ test), but these further depended on randomness induced by the binarized scoring rule as well as an additional choice by subjects that affected their payoff (see Section 2.5.2).

[18] The Hellinger distance is a bounded metric frequently used to characterize the similarity between two probability distributions (Bandyopadhyay et al., 2016). It is suited to the present purpose as it is a proper metric, unlike, e.g., the Kullback-Leibler divergence.

[19] $H_{NN}$ is calculated as:

$$H_{NN} = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k}\left(\sqrt{\tilde{P}(X_i|S)} - \sqrt{P^{NN}(X_i|S)}\right)^2} \tag{2.7}$$

*Broad*, whereas only less than 20% are in *Narrow*. Instead, about 60% in *Narrow* are characterized as close to fully inattentive, with the remaining 20% located in between the two poles. This figure indicates that in the vast majority of cases, stated beliefs are either fully sophisticated or fully inattentive to noise.

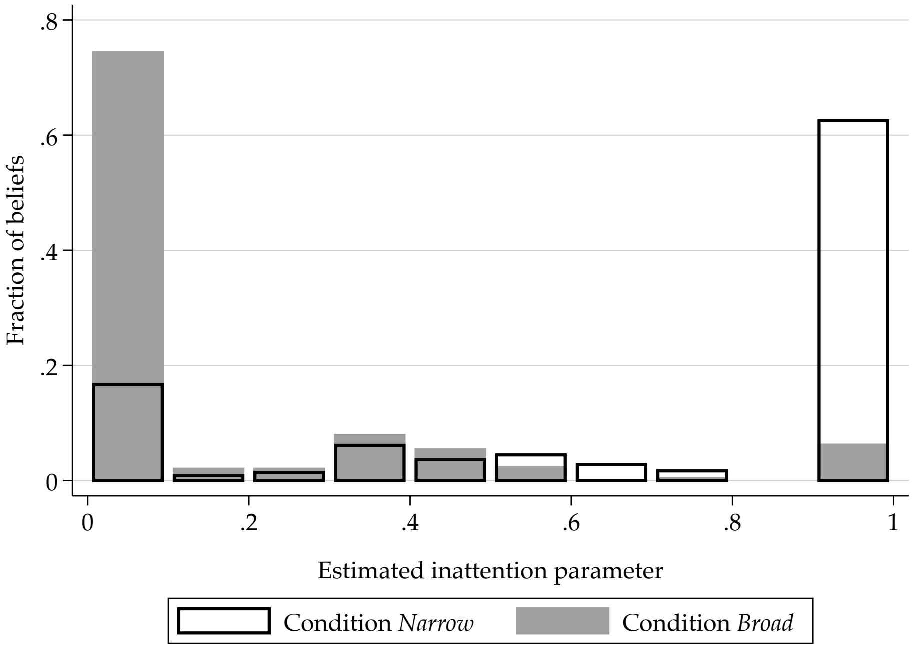


**Figure 2.2.** Inattention to $Y$ in baseline tasks. N=1135. Indicated are treatment-specific binned histograms for the implied inattention parameters from all beliefs elicited in the five baseline tasks. Inattention is calculated as $\theta = \frac{H_B}{H_B+H_N}$, where $H_B$ and $H_N$ denote the Hellinger distance of the subjective distribution to the Bayesian posterior and the inattentive posterior distribution, respectively. A parameter of $\theta = 0$ is consistent with Bayesian updating. $\theta = 1$ is a fully inattentive belief.

### 2.3.1.6 Robustness

The baseline laboratory experiment documents the occurrence of noise neglect in a specific configuration of the information environment. In additional experimental variations I examine the robustness as well as competing explanations for the findings. These extensions include *(i)* additional tasks introducing various departures from the simple discrete uniform case, *(ii)* a direct test of a signal anchoring heuristic, *(iii)* two treatments that exactly align the elicitation procedures across conditions, and *(iv)* a simplification version that removes the deciphering stage and time limits.

Four additional tasks were presented in random order after the baseline tasks.[20] First, once the numbers are correlated instead of independent, there is an additional incentive to attend to $Y$. Subjects in *Narrow* partially accommodated this incentive, as indicated by lower median inattention of 0.59. Second, moving toward a (more naturalistic) continuous data structure increased the complexity of the inference problem and pushed the median subject in *Broad* away from the Bayesian benchmark. Third, normal instead of uniform data appears to have had a similar added-complexity effect on *Broad*. Fourth, a signal realization outside of the range of $X$ "wakes up" some subjects in *Narrow* to a small extent and increases the share of Bayesian beliefs. Most importantly, highly significant treatment effects persist in all four tasks ($p < 0.001$, M-W $U$ tests).

Treatment *Computation* directly tests whether noise neglect is driven by a simple face value heuristic, whereby inattentive subjects anchor their guess of $X$ on the observed information value.[21] If this were the case, then the observed inattentiveness to $Y$ might not be the specific neglect of noise, but only an instance of a more general simplification strategy. Treatment *Computation* is identical to *Narrow*, but inserts a simple algebraic computation into the signal structure, *such that it remains equally plausible to anchor on the observed signal value*. For example, instead of $I = \frac{X+Y}{2}$, subjects receive the modified signal $I = \frac{X+Y}{2} - (2 \cdot 10) + 30$. I find very limited evidence for anchoring on the observed signal. Instead, subjects are well able and willing to invert the computations, but then still do not account for $Y$.[22] Inattention to noise implied by beliefs in *Computation* is indistinguishable from *Narrow* ($p = 0.37$, M-W $U$ test), and significantly different from *Broad* ($p < 0.001$). This suggests the baseline finding reflects a specific error in probabilistic reasoning rather than mere anchoring on the signal value.

Another block of treatments addresses the sensitivity of the baseline findings to specific experimental procedures. A central insight of the experiment is that attention can be directed using incentives, which in turn affects belief formation. The experiment, however, varies the elicitation procedure along with the incentive structure: subjects in *Narrow* only state a belief about $X$, whereas subjects in $Broad$ guess both $X$ and $Y$.[23] To better understand what portion of the treatment effect can be explained by the difference in the elicitation mechanism, two additional treatments were designed to obtain a full 2 (incentives

[20] See Appendix 2.D.1 for the robustness task specifications and detailed results.

[21] Even then, anchoring cannot explain the treatment effect without further assumptions.

[22] Further treatment details, figures and results are relegated to Appendix 2.D.2.

[23] This is a deliberate design choice. The natural setup is one in which a person only predicts the states with non-zero prediction incentives, since making a prediction in itself can provide a non-monetary incentive to pay attention. At the same time, the available information set was held exactly constant across treatments, so that subjects in *Narrow* did not have to memorize the distribution of $Y$.

*Narrow/Broad*) × 2 (elicited: only $X$ / $X$ and $Y$) between-subjects factorial design. I find that given an incentive structure, i.e., *Narrow* or *Broad*, harmonizing the elicitation protocol reduces the treatment effect by roughly one third, while all differences in estimated inattention remain highly significant (see Appendix 2.D.3). Put differently, the better part of the treatment effect is purely driven by prediction incentives.

Finally, drastically simplifying condition *Narrow* by removing the deciphering stage as well as all time limits induces a reduction in inattention ($p < 0.01$), but the treatment effect persists in a conservative comparison against the baseline condition *Broad* which includes deciphering and time limits ($p < 0.001$).

The robustness exercises substantiate the baseline findings about the prevalence and distinctness of noise neglect. All details are relegated to Appendix 2.D.

#### 2.3.1.7 The Form of Noise Neglect

Incorporating inattentive inference into models of belief formation requires understanding the form of the neglect. The term *noise neglect*, however, has no immediate formal analogue. This is because there are, in principle, many ways in which noise can be neglected in the updating process. I characterize different possibilities by whether they correspond to the (implicit) use of *(i)* a modified signal structure $g^*$, *(ii)* a modified distribution of $Y$, $h_Y^*$, or *(iii)* a non-Bayesian belief formation rule.

For example, people might update as if the information structure only depended on X, but not Y, i.e., $I = g^*(X)$. This has different implications than if people used the correct information structure $g$, but replaced the true distribution of $Y$ by something else. Different yet again is a belief formation rule that relies on the actual data and information structure, but does not comply with Bayesian updating. A candidate in this respect is a belief that ignores the prior (or base rate) and overweights the likelihood. A recent strand of the literature systematically incorporates such deviations into belief formation in the form of *diagnostic expectations* (Bordalo et al., 2017, 2018).

Since it is infeasible to identify and test every possible candidate rule, I proceed by ruling out categories of specifications based on the data. In an additional experiment, subjects faced various tasks that allow to distinguish between some of the main explanations. This evidence is reported in Appendix 2.D.5.

I make three observations. First of all, noise neglect empirically differs from likelihood-based explanations such as diagnostic expectations. People form diagnostic expectations if they overweight outcomes that become more likely in the light of new information (Bordalo et al., 2018). However, in my data people typically overweight outcomes of $X$ that are close to $i$, even if these outcomes have become less likely under $i$. For example, consider two independent, normally distributed variables $X \sim \mathcal{N}(100, 100)$ and $Y \sim \mathcal{N}(100, 100)$, and information structure $I = X + Y$. Upon observing, e.g., $i = 145$, diagnostic expectations overweight small outcomes of $X$ below 100. In the experiment, however, subjects overweight outcomes of $X$ *above* 100, as if trying to explain the information solely through $X$. Relatedly, empirical beliefs do not feature the *kernel of truth* property of diagnostic expectations, stating that beliefs generally respond to news in a directionally correct, but excessive manner. In the experiment, subjects also respond to news that is fully uninformative about $X$.

Second, I find that noise neglect is not in line with belief formation using the correct information structure, but a modified prior about $Y$. Specifically, in several tasks stated beliefs about $X$ are not in line with *any possible* belief about $Y$ defined on the union of the actual support of $Y$, the mean, median and mode of $Y$, and the number 0. This excludes any rule that replaces $Y$ by a single value in its support, its mean, etc. as well as any rule that shrinks the variance of $Y$.

Third, I find evidence for the following patterns. If $i$ is in the support of $X$, noise neglect is in line with people overweighting the outcome(s) closest to $i$. If $i$ is not in the support of $X$ but "sufficiently close", noise neglect corresponds to overweighting the outcomes in the support of $X$ that are closest to $i$. If $i$ is not in the support of $X$ and "sufficiently far" from any value with positive likelihood, noise neglect is often in line with people overweighting outcomes of $X$ whose multiple is close to $i$.

A word of caution is in order about these regularities. First, the results pertain to the specific experimental design studied in this paper, that is, algebraic signal structures in which $X$ and $Y$ are combined additively. In practice, information environments rarely have these features, let alone an explicit information structure. As such, the results above on how people deal with algebraically explicit information structure should not be over-interpreted.

The main insight from this analysis is that noise neglect is best characterized as a strong form of ignorance about the existence of $Y$. That means, people seem to apply a modified information structure $g^*(X)$ that excludes $Y$.

### 2.3.2 Baseline Study on Updating Rules: Online Evidence

The laboratory experiments provide evidence for the existence of noise neglect in a controlled experimental setting. However, the design is not well suited as a test of the prevalence of different updating modes in practice. Note that the laboratory design puts strong emphasis on the piece of information, with, e.g., an information deciphering stage that makes it practically impossible to ignore it. Moreover, subjects have to indicate a full posterior while the prior is uniform, such that entering the prior is relatively effortful. That means, while the cognitive process of neglecting a signal neglect is arguably inexpensive, the experimentally necessary procedure of having subjects state a guess to make their beliefs visible makes signal neglect practically costly. Furthermore, the baseline lab design does not allow to distinguish signal neglect from Bayesian updating in many tasks. The online task design addresses these issues, while maintaining the same basic structure.

#### 2.3.2.1 Design

There are three main modifications relative to the laboratory experiment. First, subjects do not have to indicate a full posterior distribution but are incentivized to state the mean of their posterior belief, substantially simplifying the procedure. Second, $X$ and $Y$ are not discrete with sample space size below ten, but follow (discretized) continuous distributions with a much larger sample space. The baseline tasks are displayed in Table 2.3. Third, there is no deciphering stage preceding the belief elicitation.

**Hypothesis 2.** *The heterogeneity in beliefs formed in the online experiment can be characterized by three underlying updating rules: Bayesian updating, noise neglect, and information neglect.*

#### 2.3.2.2 Procedural Details

I conducted incentivized experiments on Amazon Mechanical Turk (MTurk), an online labor marketplace frequently used by researchers. A recent study suggests that MTurk workers are more attentive to instructions than college students (Hauser and Schwarz, 2016). Participants in my online experiments had to live in the U.S. and be of legal age, have an overall approval rating of more than 95 percent, and have completed more than 100 tasks on MTurk. Workers were paid 0.5 dollars for participation and could earn up to 5 dollars for their performance on the guessing task. They played five rounds in randomized order. A decision screen is reproduced in Figure 2.G.25. One round was randomly chosen to be paid, and the payoff was determined based on a quadratic scoring rule.[24] Following the belief tasks, subjects worked on 5 Raven matrices. Correct answers were incentivized with 20 cents each. In the online experiments, all subjects were paid to predict $X$ only, analogous to condition *Narrow* in the laboratory experiments. 131 subjects participated in the online baseline experiments for an average payment of 2.7 dollars. Completion of the study took 13 minutes on average. It was implemented using oTree (Chen et al., 2016).

[24] The monetary payoff (in US dollars) was determined by the following rule:

$$\max\left\{0\ ,\ 3 - 0.1 \cdot (\text{guess of } X - \text{draw of } X)^2\right\}$$

**Table 2.3.** Online baseline tasks

| $X$ | $Y$ | $I$ |
|---|---|---|
| $\mathcal{N}(100,100)$ | $\mathcal{N}(0,100)$ | $X+Y$ |
| $\mathcal{N}(100,100)$ | $\mathcal{N}(0,400)$ | $X+Y$ |
| $\mathcal{N}(100,400)$ | $\mathcal{N}(0,100)$ | $X+Y$ |
| $\mathcal{U}[75,76,\dots,125]$ | $\mathcal{U}[-25,-24,\dots,25]$ | $X+Y$ |
| $\mathcal{U}[75,76,\dots,125]$ | $\mathcal{U}[90,91,\dots,110]$ | $\frac{X+Y}{2}$ |

*Notes*: This table provides an overview of the five baseline belief tasks in the online experiment. Note that for all normally distributed variables, the support was discretized to integers, truncated at $\mu-50$ and $\mu+50$ and then the distributions were scaled such that the they have unit probability mass.

#### 2.3.2.3 Baseline Results: Online

**Result 2.** *The distribution of subjective beliefs is trimodal. The three modes correspond to – in order of frequency – noise neglect, Bayesian updating and information neglect.*

Figure 2.3 shows all stated beliefs together with the information value received in the five baseline tasks. It further highlights which stated beliefs would correspond to noise neglect, information neglect and Bayesian updating.

There is evidence for each of those three updating rules. In fact, in each task at least 60% of stated beliefs are exactly in line with these three updating modes. Among the three modes, Bayesian updating and noise neglect are observed with roughly similar frequency, while information neglect occurs to a somewhat lesser extent. Note that the Bayesian benchmark changes across tasks, and subjects clearly respond to this change.

To illustrate the degree to which beliefs are clustered on these three updating modes, Figure 2.4 plots kernel density estimates for the task in the upper left corner of Figure 2.3. In this task, $X\sim\mathcal{N}(100,100)$, $Y\sim\mathcal{N}(0,100)$, and $I=X+Y$. The stated belief that corresponds to a Bayesian posterior in this case is $100+\lambda\cdot(i-100)$ where $\lambda=\frac{\sigma_X^2}{\sigma_X^2+\sigma_Y^2}=\frac{1}{2}$. Intuitively, since $X$ and $Y$ have equal variance, a normatively optimal guess of $X$ would attribute half of $i$'s deviation from the expected value of 100 to $X$. Information neglect, in turn, would correspond to a belief equal to the prior of $X$, $\mathbb{E}[X]=100$. This is equivalent to assigning *none* of the deviation of $i$ from its expected value to $X$. In fact, with $m$ denoting a subject's stated guess, I can back out the empirical equivalent of $\lambda$, as $\hat{\lambda}=\frac{m-100}{i-100}$. In the case of information neglect with $m=100$, $\hat{\lambda}=0$. Finally, if people commit signal neglect they state $m=i$, leading to $\hat{\lambda}=1$.

Figure 2.4 provides three insights. First, most of the probability mass is centered on the three updating modes. Second, noise neglect is relatively most fre-

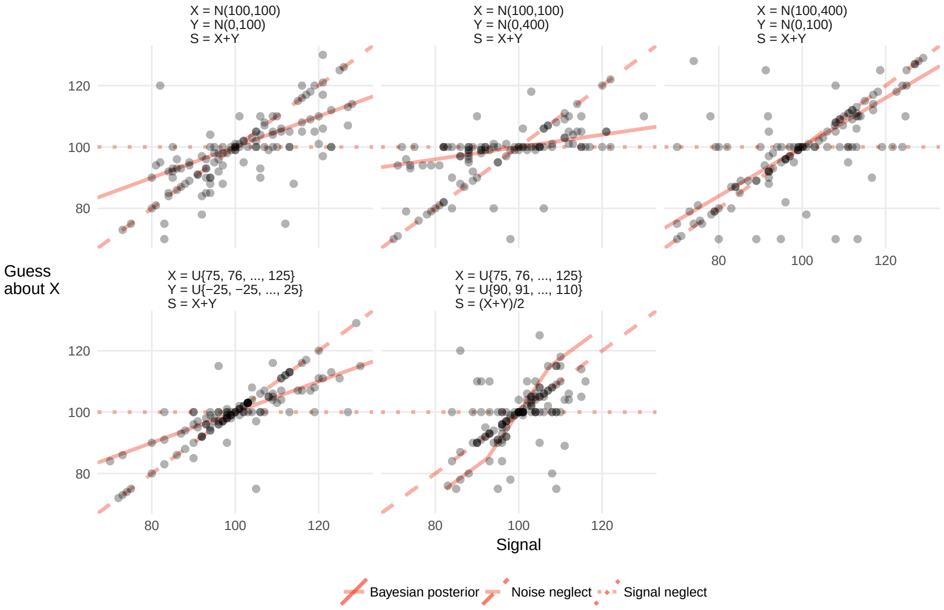


**Figure 2.3.** Beliefs in baseline tasks of online experiments. N=131 in each task. Each dot corresponds to one stated belief. The three red lines indicate the Bayesian benchmark, noise neglect, and information neglect.

quent in this task, and information neglect least frequent. Third, as indicated by the rug plot on the right, people who neglect the noise ($\hat{\lambda} = 1$) or the information ($\hat{\lambda} = 0$) do so *exactly*. By contrast, people are more dispersed around the Bayesian benchmark ($\hat{\lambda} \approx 0.5$), presumably because it is harder to compute the Bayesian posterior exactly.

In the experimental settings studied in this paper, beliefs are clearly too heterogeneous to be adequately described by a single representative updating rule. Average beliefs mask the underlying structure. At the same time, there is little randomness in stated beliefs. Instead, most beliefs accord to a discrete set of three updating modes. They align exactly with one of these modes, and there is virtually no mixing between the modes, i.e., people do not choose "combinations".

A much deeper understanding is necessary to predict the occurrence of the updating modes for different environments, and to pinpoint the underlying psychological mechanisms that can inform a formal model of belief formation from noisy information.

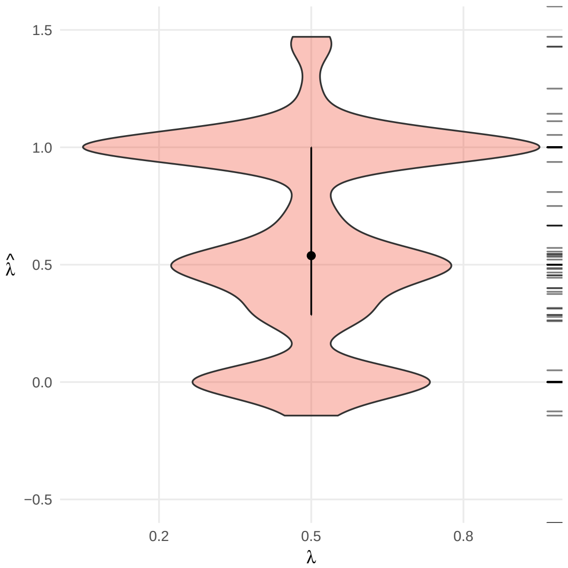


**Figure 2.4.** Kernel density plot for beliefs stated in a task where $X \sim \mathcal{N}(100, 100)$, $Y \sim \mathcal{N}(0, 100)$, and $I = X + Y$. In this task, the Bayesian belief corresponds to $100 + \lambda \cdot (i - 100)$ where $\lambda = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2} = \frac{1}{2}$. For each stated belief, the empirical counterpart of $\lambda$ is calculated as $\hat{\lambda} = \frac{m-100}{i-100}$. The plot documents three distinct clusters at $\hat{\lambda} = 0$ (information neglect), $\hat{\lambda} = 1$ (noise neglect) and around $\hat{\lambda} = \frac{1}{2}$ (Bayesian posterior.) Based on N=131. Epanechnikov kernel with bandwidth 0.07.

## 2.4 Predictability of Updating Modes

In this section I report additional evidence that sheds light on how the prevalence of different updating modes depends on features of the information environment and individual-specific factors.

In terms of the conceptual framework in Section 2.2, the former corresponds to the impact of variation in the expected utility benefit, $\mathbb{E}\left[v(\tilde{R}(a))\right]$ in Equation 2.1, while the latter refers to factors that might systematically influence the cognitive cost of updating modes, $C(\tilde{R}(a))$.

### 2.4.1 Responsiveness to the Information Environment

**Hypothesis 3.** *The relative prevalence of different updating modes responds to their expected accuracy in an information structure.*

A central comparative static concerns the effect of the signal-to-noise ratio. Relative to Bayesian updating, the expected utility of noise neglect (information neglect) is increasing (decreasing) in the signal-to-noise ratio.

I illustrate this logic in a simple example. Assume $X$ and $Y$ follow independent normal distributions, $X \sim \mathcal{N}(0,100)$ and $Y \sim \mathcal{N}(0,\sigma_Y^2)$. Information is generated as $I = X + Y$. Assume that the agent's utility of an action is described by a loss function

$$u(m,x) \; = \; -\frac{1}{2}(m-x)^2.$$

In the independent normal case with quadratic cost, the actions associated with different representations have a simple closed form as shown above. After observing information $i$, the fully attentive representation leads to the Bayesian posterior and an action equal to its mean $m = \lambda \cdot i$ with $\lambda = \frac{\sigma_X^2}{\sigma_X^2+\sigma_Y^2} = \frac{100}{100+\sigma_Y^2}$. Noise neglect means an agents takes action $m = i$, and information neglect corresponds to $m = \mathbb{E}[X] = 0$.

Assuming $q = 15$ and $\nu = 0.5$ for expositional purposes, Figure 2.5 illustrates costs and benefits of the actions associated with different representations for varying $\sigma_Y^2$. For small noise variation $\sigma_Y^2$, noise neglect is optimal, while for high $\sigma_Y^2$, signal neglect dominates. Under this parameterization, Bayesian updating prevails only for intermediate values of $\sigma_Y^2$.

In a sample with heterogeneous cognitive cost $q$, I therefore expect the frequency of noise neglect to decrease and that of information neglect to increase as $\sigma_Y^2$ rises. I directly tested this prediction in an additional online experiment with seven tasks with varying signal-to-noise ratio as shown in Table 2.4.

**Table 2.4.** Online tasks: Experiment on signal-to-noise ratio

| $X$ | $Y$ | $I$ | $\lambda = \frac{\sigma_X^2}{\sigma_X^2+\sigma_Y^2}$ |
|---|---|---|---|
| $\mathcal{N}(100,25)$ | $\mathcal{N}(0,1600)$ | $X+Y$ | 0.015 |
| $\mathcal{N}(100,100)$ | $\mathcal{N}(0,1600)$ | $X+Y$ | 0.059 |
| $\mathcal{N}(100,25)$ | $\mathcal{N}(0,100)$ | $X+Y$ | 0.25 |
| $\mathcal{N}(100,100)$ | $\mathcal{N}(0,100)$ | $X+Y$ | 0.5 |
| $\mathcal{N}(100,100)$ | $\mathcal{N}(0,25)$ | $X+Y$ | 0.75 |
| $\mathcal{N}(100,1600)$ | $\mathcal{N}(0,100)$ | $X+Y$ | 0.941 |
| $\mathcal{N}(100,1600)$ | $\mathcal{N}(0,25)$ | $X+Y$ | 0.985 |

*Notes*: This table provides an overview of the five tasks in the online experiment on the effect of the signal-to-noise ratio. Note that for all normally distributed variables, the support was discretized to integers, truncated at $\mu - 50$ and $\mu + 50$ and then the distributions were scaled such that the they have unit probability mass.

A sample of N = 209 participated in this experiment, where again task order was randomized and one task was randomly incentivized with a maximum prize of 3 dollars.

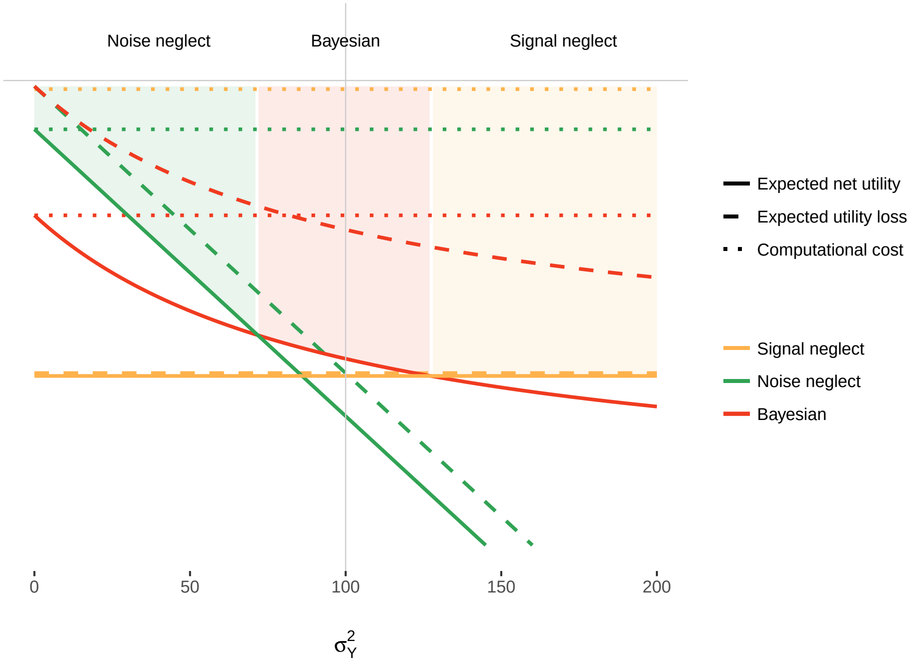


**Figure 2.5.** Illustration of the effect of the the signal-to-noise ration on updating modes. In the displayed example, $X \sim \mathcal{N}(0, 100)$ and $Y \sim \mathcal{N}(0, \sigma_Y^2)$. Information is generated as $I = X + Y$. The agent's has quadratic utility, $u(m, x) = -\frac{1}{2}(m - x)^2$. The figure assumes noise paramters $q = 15$ and $\nu = 0.5$ for expositional purposes. For small noise variation $\sigma_Y^2$, noise neglect is optimal, while for high $\sigma_Y^2$, signal neglect dominates. Under this parameterization, Bayesian updating prevails only for intermediate values of $\sigma_Y^2$.

Figure 2.6 documents the results by plotting estimated kernel densities of $\hat{\lambda} = \frac{m-100}{i-100}$ by task. In line with the previous results, there are three empirical modes in each task, corresponding to Bayesian updating, noise neglect and information neglect. Note that the value of $\lambda$ in line with Bayesian beliefs changes across tasks, as indicated by the dashed diagonal line. To support the visual analysis, I perform non-parametric test on the distributions of $\hat{\lambda}$. First, note that the summed share of beliefs in line with either one of the three updating modes (defined as being within $[\lambda - 0.05, \lambda + 0.05]$) does not significantly differ across tasks ($p > 0.1$ for all pairwise comparisons in $\chi^2$ tests). Second, for each task with $\lambda > 0.75$, the share of beliefs in line with noise neglect (again, defined as being within $[0.95, 1.05]$), is significantly higher than in all tasks with $\lambda < 0.75$ (all $p < 0.05$, pairwise $\chi^2$ tests). Third, for each task with $\lambda < 0.25$, the share of beliefs in line with info neglect (defined as being within $[-0.05, 0.05]$), is significantly higher than in all tasks with $\lambda > 0.25$ (all $p < 0.01$, pairwise $\chi^2$ tests).

This means that, in line with the hypothesis, while the overall share of beliefs in line with the three updating modes stays roughly constant, the share of noise neglect increases and the share of information neglect decreases with increasing signal-to-noise ratio $\lambda$.

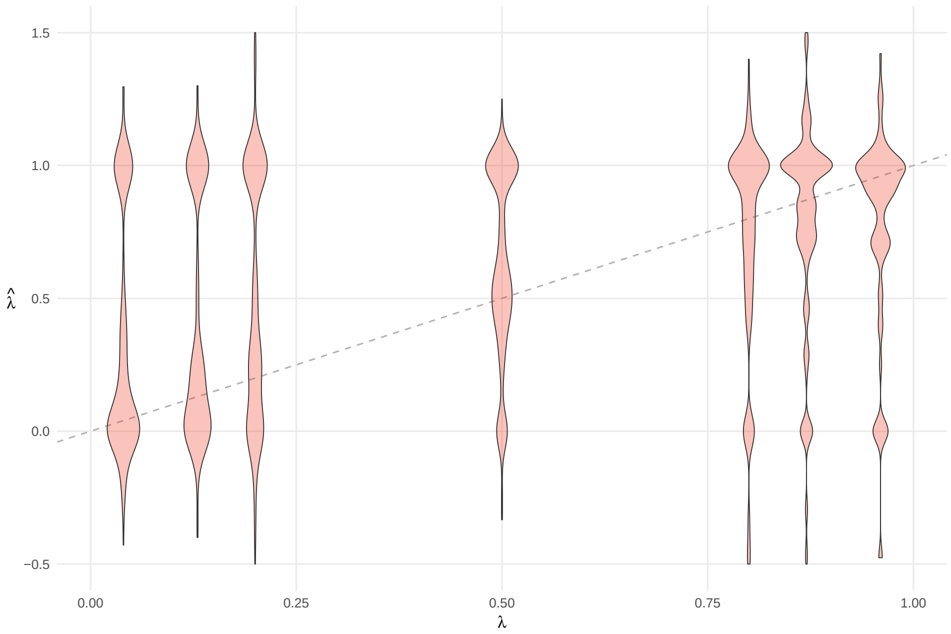


**Figure 2.6.** Kernel density estimates for seven different tasks (see Table 2.4) in an online experiment testing the effect of the signal-to-noise ratio on the prevalence of different updating modes. The horizontal axis indicates $\lambda = \frac{\sigma_X^2}{\sigma_X^2+\sigma_Y^2}$ of the task. The vertical axis shows the empirical equivalent derived from subjects guesses as $\hat{\lambda} = \frac{m-100}{i-100}$. Note that $\hat{\lambda} = 0$ indicates information neglect, $\hat{\lambda} = 1$ indicates noise neglect, and the dashed line indicates the $\hat{\lambda}$ that corresponds to Bayesian updating. Based on N=207 in each task. Epanechnikov kernel with bandwidth 0.07.

I now turn to a second implication of the conceptual framework that relates to the analysis of the *form* of noise neglect in Section 2.3.1.7. The empirical findings suggest that noise neglect induces people to overweight outcomes close to the observed information value. In the baseline laboratory experiment as well as the preceding experiment on the effect of the signal-to-noise ratio, task configurations were (purposefully) chosen such that the information is an unbiased estimator of the mean of $X$, i.e., $\mathbb{E}[I] = \mathbb{E}[X]$. In this case, noise neglect leads to overreaction without generating a directional bias on average. In information structures in which the signal does not have the same mean as $X$, however, noise neglect also induces directional bias. This, however, is reflected in the expected utility loss from neglecting noise. An interesting question is whether subjects recognize this bias, i.e., whether it makes them less likely to neglect the noise. If people in fact incorporate the expected costliness (in terms of utility losses, not cognitive costs) of each updating mode, this would be additional evidence that the prevalence of updating modes follows a form of cost-benefit analysis.

To examine this prediction I ran a variation of the online experiment in which the distribution of $X$ as well as the variance of $Y$ were kept constant across tasks, but the mean of $Y$ varied. The five task configurations are displayed in Table 2.5. Note that $\lambda = \frac{\sigma_X^2}{\sigma_X^2+\sigma_Y^2}$ is identical across tasks, setting this experiment apart form the preceding one on the signal-to-noise ratio. However, the expected value of the information, $\mu_I$, varies. To optimally learn from $I$, subjects need to account for the fact that observed values of $I$ are on average higher or lower than $X$ if $\mu_X \neq \mu_I$.

**Table 2.5.** Online tasks: Experiment on directional bias in information

| $X$ | $Y$ | $I$ | $\mu_I$ |
|---|---|---|---|
| $\mathcal{N}(100,100)$ | $\mathcal{N}(0,100)$ | $X+Y$ | 100 |
| $\mathcal{N}(100,100)$ | $\mathcal{N}(-25,100)$ | $X+Y$ | 75 |
| $\mathcal{N}(100,100)$ | $\mathcal{N}(-50,100)$ | $X+Y$ | 50 |
| $\mathcal{N}(100,100)$ | $\mathcal{N}(25,100)$ | $X+Y$ | 125 |
| $\mathcal{N}(100,100)$ | $\mathcal{N}(50,100)$ | $X+Y$ | 150 |

*Notes*: This table provides an overview of the five tasks in the online experiment on the effect of the signal-to-noise ratio. Note that for all normally distributed variables, the support was discretized to integers, truncated at $\mu - 50$ and $\mu + 50$ and then the distributions were scaled such that the they have unit probability mass.

Raw beliefs are shown in Figure 2.7. The figure suggests that subjects become less likely to commit noise neglect the larger the directional bias of the signal, i.e., the further away the mean of $Y$ is from 0. Belief in line with noise neglect have to be close to the dashed red line.

This observation is supported by non-parametric tests. The share of beliefs in line with noise neglect significantly decreases as $\mu_I$ moves away from 100.[25]

[25] That is, the share of noise neglect decreases in both directions away from 100 for adjacent tasks, e.g., both for $\mu_I = 100$ vs. $\mu_I = 75$ and $\mu_I = 75$ vs. $\mu_I = 50$. Noise neglect can be defined in different ways. I either define it as the any guess falling within a margin of 5 around the hypothetical noise neglect guess, or based on $d_{NN}^{rel} = \frac{d_{NN}}{d_{NN}+d_B+d_{IN}}$ falling within 0.05 to either side of 0, where $d$ is the distance of a stated belief $m$ to the respective benchmark belief for each of three updating modes. That means, e.g., $d_B = \left|m - m_B\right|$ is the distance to the Bayesian belief.

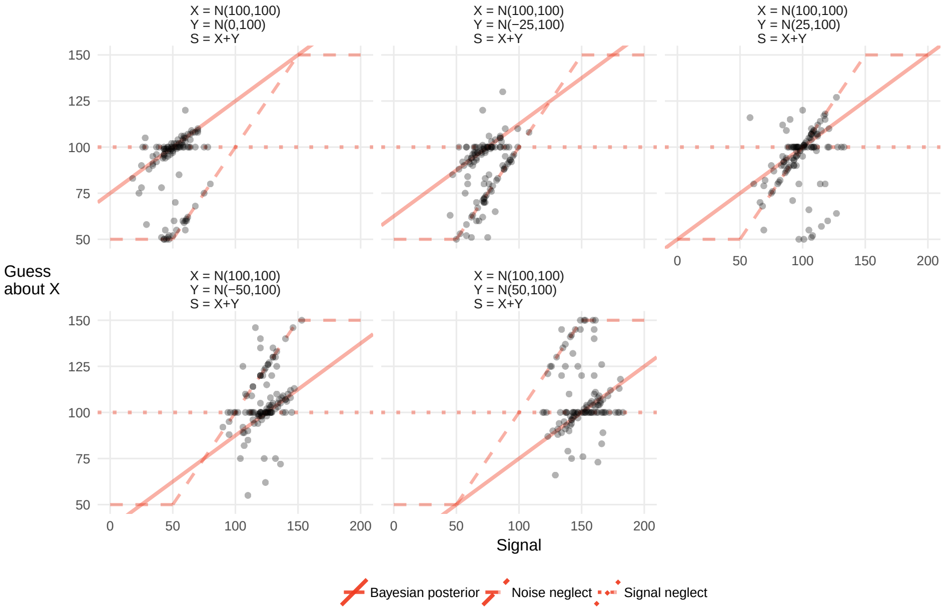


**Figure 2.7.** Beliefs in baseline tasks of online experiments. N=112 in each task. Each dot corresponds to one stated belief. The three red line indicate the Bayesian benchmark, noise neglect, and information neglect.

Notably, I find that this decrease goes hand in hand with an increase in Bayesian beliefs, rather than information neglect.[26]

Another way to illustrate this result is to characterize an observed belief by how relatively close it is to each of the three benchmarks. I do this by computing the distance to each of the three benchmarks, add up these distances, and then calculate a measure of *closeness* for each updating mode as the fraction of that mode's distance relative to the sum of distances. Specifically, I obtain three measures of relative distance for each elicited belief, all of which lie between 0, meaning the belief lies exactly on the posterior of that updating mode, and 1. These three measures sum up to 1. A ternary plot of beliefs characterized by these three distances is shown in Figure 2.8. In each triangle, the distance from the posterior under noise neglect corresponds to the vertical distance from the bottom. That means, a belief close to the horizontal axis is in line with noise neglect, while are belief further away from it indicates a larger distance from

Hence, $d_{NN}^{rel}$ is the distance of a belief to a hypothetical belief under noise neglect, *relative* to the summed distances of the elicited belief to all three updating modes. $p < 0.05$ in all pairwise $\chi^2$ tests.

[26] The share of Bayesian beliefs as defined above significantly increases with the distance of $\mu_I$ from 100, $p < 0.1$ in all pairwise $\chi^2$ tests.

noise neglect.

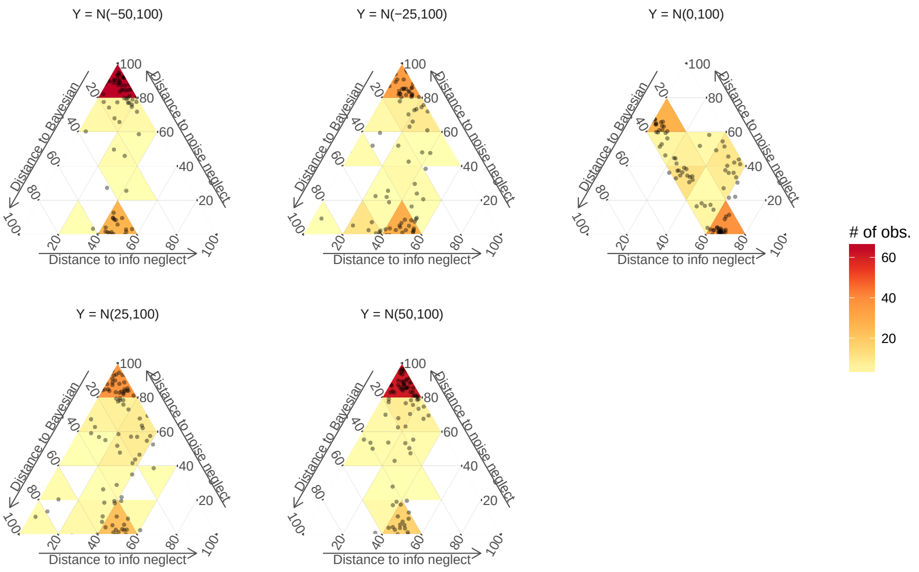


**Figure 2.8.** Beliefs in online experiments on the effect of directional bias in information. N=112 in each task. Each point corresponds to one stated belief. Red areas in the heatmap indicate regions with more stated beliefs. The displayed data is computed based on three relative measures of distance for each belief that sum to one, $d^{rel}_{NN}$, $d^{rel}_{IN}$ and $d^{rel}_{B}$. For example, $d^{rel}_{NN} = \frac{d_{NN}}{d_{NN}+d_B+d_{IN}}$ where $d$ is the distance of a stated belief $m$ to the respective benchmark belief for each of three updating modes, e.g., $d_B = \left|m - m_B\right|$ is the difference to the Bayesian belief. Hence $d^{rel}_{NN}$ is the distance of a belief to a hypothetical belief under noise neglect, *relative* to the summed distances of the elicited belief to all three updating modes.

Note, however, that as the mean of $Y$ moves away from 0, subjects are also more likely to observe information values further away from 100. With $X$ normally distributed around 100, the "plausibility of noise neglect" as judged by the probability $P(X = i)$ therefore on average decreases as the distance $\left|\mu_I - \mu_X\right|$ increases. In additional analyses reported in Appendix 2.E.1, I show that noise neglect becomes more unlikely as the directional bias of the information structure increases, *controlling for the information value observed by subjects*. That means, in two tasks with identically distributed $X$ and an identical observed information value $i$, the propensity to commit noise neglect is less likely the larger $\left|\mu_Y - 0\right|$, i.e., the larger the *expected* utility loss from noise neglect.

The evidence reported in this section suggests two things. First, subject do react in a systematic way to features in the information environment as suggested by the (as-if) cost-benefit analysis discussed in Section 2.2. Second, people respond to the structure of noise in the environment even conditional on observing the same piece of information. This implies that the process that leads to a specific updating mode such as noise neglect is not purely driven by the observed piece of information $i$, e.g., by a plausibility check of the resulting posterior, but likely starts before that, once a subject studies the information environment and allocates attention.

**Result 3.** *The empirical distribution of updating rules responded to the degree of variance and bias introduced by $Y$.*

### 2.4.2 Correlates of Inattention and Within-Subject Consistency

To understand whether attention allocation in inference problems depends on considerations about the cognitive cost of different updating modes, I will analyze between-subject differences in cognitive skills. Higher cognitive skills might be associated with a lower cost of computational effort when forming a posterior. However, there are other ways in which cognitive ability could affect performance in the updating tasks that are different from reflecting the outcome of a cost-saving choice of a mental representation that avoids certain computations. For example, subjects with lower cognitive skills might misconstrue per se the updating problem (they are not able to understand the task correctly even if they wanted to), leading to a flawed mental representation to begin with, or they might form a correct understanding of the problem but fail to execute the necessary computations. However, the baseline laboratory experiment provides evidence against both of these explanations as key determinants of inattentive inference in the experiment: Virtually all subjects in *Broad* solve the problem in Bayesian fashion. This means subjects are able to form a normatively correct representation independent of their cognitive skills, and, conditional on this correct mental representation, are both willing and able to execute the necessary contingent reasoning to arrive at an optimal belief.[27]

To shed light on whether person-fixed characteristics are associated with different updating styles, I analyze correlates of inattention to noise in the baseline laboratory experiment. Columns 1 and 2 of Table 2.6 document a significant but moderate association between cognitive skills and inattention when pooling

[27] In fact, the experiment was designed in such a way that belief formation does not boil down to performing mental arithmetic as in a test of cognitive intelligence. Instead, practically no subject is barred from "solving" the problem on cognitive capacity grounds. It is trivial, however, that this may well be the case in many information environments in practice.

both treatments together. This effect has roughly similar magnitude but seizes to be significant in a subsample restricted to condition *Narrow* (column 3). This supports the view that cognitive skills do play a role in generating noise neglect, but this is not because people with lower cognitive skills lack the capacity to understand the task correctly of to execute the necessary computations.

**Table 2.6.** Correlates of inattention in baseline laboratory experiment

| Dependent variable: | Inattention to noise $\theta$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Condition: | *Narrow* and *Broad* | | *Narrow* | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 0 if *Broad*, 1 if *Narrow* | 0.516*** | 0.517*** | | | | | | |
| | (0.042) | (0.041) | | | | | | |
| Cognitive skills (Raven) | | -0.027** | -0.022 | | -0.003 | | | |
| | | (0.014) | (0.019) | | (0.023) | | | |
| Big 5: Conscientiousness | | | | -0.007 | -0.015 | | | |
| | | | | (0.010) | (0.011) | | | |
| Willingness to take risks | | | | -0.002 | 0.027 | | | |
| | | | | (0.042) | (0.044) | | | |
| IRI: Perspective-taking | | | | 0.005 | 0.009 | | | |
| | | | | (0.014) | (0.014) | | | |
| Response time (seconds) | | | | | | -0.004*** | | -0.004*** |
| | | | | | | (0.000) | | (0.000) |
| Reading time (seconds) | | | | | | | -0.000 | 0.000 |
| | | | | | | | (0.000) | (0.000) |
| Constant | 0.154*** | 0.316*** | 0.802*** | 0.693*** | 0.572** | 0.917*** | 0.709*** | 0.614*** |
| | (0.021) | (0.089) | (0.113) | (0.241) | (0.255) | (0.026) | (0.149) | (0.107) |
| Controls | | | | | Yes | | | Yes |
| $R^2$ | 0.3778 | 0.3904 | 0.0107 | 0.0039 | 0.1222 | 0.4223 | 0.0005 | 0.4565 |
| N | 1135 | 1135 | 607 | 607 | 607 | 607 | 607 | 607 |

*Notes*: OLS regressions with implied inattention as dependent variable. Inattention is calculated as $\theta = \frac{H_B}{H_B+H_N}$, where $H_B$ and $H_N$ denote the Hellinger distance of the subjective distribution to the Bayesian posterior and the inattentive posterior distribution, respectively. Cognitive skills were measured using 10 Raven matrices and a payoff of 0.2 euros for each correct answer. Willingness to take risks is based on the survey preference module of Falk et al. (2016). The Interpersonal Reactivity Index (Davis, 1983) has a subscale on perspective-taking, defined as “the tendency to spontaneously adopt the psychological point of view of others”. Response time is the duration in seconds the subject spent on the belief elicitation page. Reading time is the duration in seconds that the subject spent on the instructions screen. Robust standard errors clustered at participant level in parentheses. Controls include gender, age, income and task-fixed effect. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Columns 4 and 5 of Table 2.6 show that inattention to noise in *Narrow* is unrelated to other selected individual-level traits. There are no significant associations with self-reported measures of conscientiousness (a Big 5 trait), willingness to take risks, and the self-reported ability to take perspective.

In columns 6 to 8, I examine the relationship with the reading time of the experimental instructions as well as response times, the latter being variably interpreted as a proxy for cognitive effort or the distinction between instinctive and contemplative modes of reasoning (e.g., Rubinstein, 2007). Noise neglect is indeed associated with lower response times.[28]

Another measure for how cognitive skills are related to inattention is subjects' consistency across tasks. Analyzing data from the baseline experiments both in the laboratory and online, I make the following observations. First, there is substantial consistency as judged by the fraction of subjects who always employ the same mode of updating. 64% of subjects in the laboratory and 43% of subjects online employ the same updating mode in all five tasks. Second, there is a non-negligible fraction of subjects (16% in the laboratory, 34% online) who used at least two different updating modes in the five tasks. Third, in the online experiment, the distribution of updating styles is systematically related to cognitive skills. When characterizing each subject by their median mode of updating, I find that on average, cognitive skill increases from the information neglect type, over the noise neglect type, to the Bayesian type.

The combined correlational evidence presented in this section allows the following conclusions. Cognitive skills are systematically related to different updating modes, but not because the complexity of the tasks exceeds the cognitive capacity of subjects with lower cognitive per se. Instead, the vast majority of participants is in principle able to form normatively correct beliefs, as shown by the condition *Broad* as well as a sizable share of online subjects who formed a Bayesian beliefs at least once. In fact, many participants *switch between modes*. An interpretation that is compatible with these results is that cognitive skills are a proxy for a person's individual cost of cognitive computations associated with belief formation, and the process that leads to the selection of a subjective representation (and resulting updating mode) at least partly accounts for this cost.

Another immediate implication of the data concerns the systematic inference patterns of groups with different average cognitive skills. The group with the lowest cognitive skills is most likely to *underreact* as here the share of information neglect is highest. Subjects with somewhat higher cognitive skill tend to disproportionately engage in noise neglect and therefore *overreact*. The group with highest cognitive skills is well calibrated and tends to perform Bayesian inference. The relationship between cognitive skills and the potential for overreaction to information might thus be non-linear.

[28] Note that inattention to noise mechanically leads to lower response times, because the fully inattentive belief in the experimental setting is characterized by lower variance, requiring fewer inputs on the elicitation screen.

## 2.5 Mechanisms: The Role of Awareness

The preceding section revealed that the presence of different updating modes is not random but systematically related to features of the information environment and person-fixed characteristics. The corresponding evidence from various experiments indicates that the process by which updating modes are selected can be described as if following a form of cost-benefit analysis. The empirical evidence reported up to this point, however, provides no direct evidence on the nature of the underlying processes. This section has the purpose to develop a deeper understanding of the cognitive mechanisms that drive inattention in belief formation. To this end I will examine people's solution strategies in fixed information environments, their confidence (or metacognition) in their beliefs, the dynamics of learning, and the effect of different external interventions in the belief formation process.

To anticipate the analysis, this section revolves around the notion of awareness. I will show that the agent's awareness about specific elements of the updating process is key to understanding inattention in inference. Note that the notion of cost-benefit analysis as modeled in the conceptual framework seemingly implies a conscious, willed activity. I stress that this is not necessarily the case. Instead, the entire process of selecting a representation need not be deliberate but might be automatic, and the relationship between the resulting, potentially reduced internal representation and the external environment can be inaccessible to the conscious mind. The subsequent computations, by contrast, can be accompanied by some form of awareness, or metacognitive experience, and can be deliberate. This squares with neuroscientific, cognitive and psychological research on the interaction between goals, attention, and consciousness. There is substantial evidence that "cost-benefit analysis" type processes operate in unconscious reasoning, e.g., that "unconscious goal pursuit is supported by attention that operates on higher cognitive processes according to principles of executive control and working memory. And these processes (and the information on which they operate) seem to run below the threshold of consciousness" (Dijksterhuis and Aarts, 2010). More concretely, this research indicates that there are *unconscious* processes of attention and inhibition that serve to interpret and manipulate information in line with a person's goals.

Understanding the framework as the formalization of an unconscious, quick optimization routine has the following implications. First, the effect of different interventions or nudges depends on whether they affect the mental representation formed by a person, or only the cognitive operations that rely on that representation. Second, people will generally be confident in beliefs formed according to information or noise neglect, because the underlying simplifications in the internal representation are inaccessible to the conscious mind. Third, inattentive representations are persistent, even in the presence of feedback. Learning that a belief was deficient does not necessarily lead to improvement if people are fundamentally unaware about the unconscious processes at the root of their mistakes.

**Hypothesis 4.** *Different updating modes are driven by different subjective representations of the information environment. These representations are formed unconsciously.*

### 2.5.1 Overcoming Inattention: Effort versus Hints

If belief formation conforms to the distinction between an initial choice of representation that occurs outside awareness, and a subsequent deliberate optimization withing that representation, then some, but not other interventions should help reduce inattention. Consciously exerted effort induced, e.g., by higher incentives, might make subjects try harder when computing a posterior, but it may not affect the unconsciously formed solution strategy. On the other hand, the framework naturally gives room for *waking up* after receiving hints that alter the agent's representation. A hint at the neglected dimension can extend the agent's representation and lead to a different posterior. Note that such hints induce a *broader* representation by pointing at the signal or the noise.

I provide evidence on both types of interventions. The *High Stakes* stage of the baseline laboratory experiment investigates the role of effort more directly. This stage consists of one surprise bonus round following. Within each condition, I re-randomized whether this round would be incentivized with the same expected payoff as the preceding tasks, or a tenfold increase thereof. Regression results in column 1 of Table 2.E.7 of Appendix 2.E.2 document that subjects significantly increased effort based on observed response times. At the same time, inattention to noise was completely unaffected by the manipulation of stakes (columns 2 and 3). Put differently, under high incentives, subjects apparently tried harder, but that did not make them more Bayesian.

Treatment *Hint* was run using a separate set of 46 subjects in the laboratory. The belief tasks were identical to those in the baseline laboratory experiment (see Table 2.1) except that on every elicitation screen, subjects saw a hint stating "Also think about the role of $Y$". People only guessed $X$, so that condition *Hint* was exactly identical to *Narrow* in the baseline experiment except for the hint.

The hint significantly reduced inattention to noise compared to *Narrow* ($p < 0.001$, M-W $U$ test). As illustrated in Appendix Figure 2.E.9, the distribution of inattention is in fact much closer to that in *Broad*, with a majority of subjects forming Bayesian beliefs.[29]

[29] Still, inattention in *Hint* is significantly higher than in *Broad* ($p < 0.001$, M-W $U$ test).

Taking stock, I find that in line with the notion of unawareness about the mental representation, only interventions that directly alter this representation affect belief formation. If people are paid more, they try harder by putting in more effort into the consciously executed steps of their updating strategy. However, this does not affect the resulting beliefs because people operate within a narrow mental representation to begin with. That deliberately exerted effort affects how hard people work on the tasks but does not change their performance strongly suggests that the neglect at the root of inattentive inference occurs outside of awareness. The hint, by contrast, broadens subjects' mental representation directly. People then perform the subsequent mental operation in the correct mental frame and predominantly arrive at Bayesian beliefs.

### 2.5.2 The Role of Confidence

I next study people's awareness about the inference process directly by examining their confidence in their own beliefs. If people had some form of awareness about noise neglect, they should be less confident in the resulting belief. If instead this neglect occurs outside of awareness, people deliberately and purposefully execute the subsequent computations and they will be confident in the result.

**Design.** Two additional experimental variations in the laboratory directly examine subjects metacognition of inference, i.e., what they think about their own solution strategy in the belief updating tasks. In stage *Confidence* following the baseline belief tasks, subjects indicate their willingness-to-accept for each previously stated belief. To this end they again see each individual updating task in combination with their own stated belief. They then indicate whether they prefer to be paid out for their belief based on the scoring rule or receive a fixed monetary amount. Subjects make this binary decision for different fixed amounts ranging from 0 euros to 6 euros, presented in a multiple-price list format. In case this task would be chosen for payoff, their decision in one of the rows of the list would be implemented. Note that the *Confidence* tasks *(i)* had no time limit such that subjects could freely rethink their stated belief, and *(ii)* the subjective valuation in each task provides a measure of confidence in the belief itself, beyond the variance implied by the stated belief distribution.

In stage *Switch-role* at the end of the laboratory baseline experiment, each subject played two bonus rounds in the opposite condition. Are participants in *Broad*, who have previously formed Bayesian beliefs in these tasks, able to transfer their successful solution strategy to an updating problem with narrow incentives that is otherwise identical? This requires a metacognitive understanding of one's own previously implemented solution.

**Results.** Columns 1 to 4 of Table 2.7 present results from the *Confidence* tasks using regressions in which the dependent variable is the subjective valuation of a stated believe, i.e., the minimal certain amount preferred over a having the stated belief paid out. A higher value corresponds to more confidence. Strikingly, more inattentive beliefs are not significantly associated with lower reservation prices. Even after reconsidering the updating problem and their own belief, subjects fail to recognize the necessity to account for $Y$ and are equally confident in their own guess. Reassuringly, the variance of the indicated belief distribution negatively affects confidence. Restricting the sample to beliefs stated in *Narrow*, there is again no relationship between the valuation of a stated belief and implied inattention. These results suggest that inattentive inference relies on processes outside of a person's awareness.

In columns 5 to 7 of Table 2.7 I analyze scores of inattention to noise on the pooled sample of beliefs from the baseline and *Switch-role* tasks. I find that *(i)* unsurprisingly, *Narrow* subjects almost immediately improve when facing the broad setup, and display a similar level of inattention as *Broad* subjects in the baseline ($p > 0.7$, see footer of Table 2.7), *(ii) Broad* subjects do transfer their experience in forming Bayesian beliefs, as indicated by a significant improvement relative to *Narrow* subjects in baseline ($p < 0.05$), *(iii)* this transfer, however, is far from perfect and a significant treatment effect between *Narrow* and *Broad* persists in the *Switch-role* tasks, albeit now with the reverse sign. In fact, mean inattention in *Broad* is 0.59 in *Switch-role*, compared to baseline means of 0.11 in *Broad* and 0.73 in *Narrow*. Put differently, the improvement in *Broad* is marginal and subjects effectively commit inattentive inference to a roughly similar extent as if they had not had the baseline experience.

In sum, the combined evidence clearly suggests that the psychological mechanism responsible for inattentive inference operates outside of people's awareness. Selective processing of the features of an information environment shaped by the structure of the prediction incentives leads to a narrow representation of the problem. The subsequent cognitive computations are executed deliberately and confidently – but rely on a flawed mental model of the environment.

**Table 2.7.** Mechanisms underlying inattentive updating: Awareness about the problem structure

| Dependent variable: | **Confidence**: Valuation for stated belief | | | | **Inattention** $\theta$ | | |
|---|---|---|---|---|---|---|---|
| Condition: | *Narrow* and *Broad* | | | *Narrow* | *Narrow* and *Broad* | | |
| Tasks: | Baseline and robustness | | | | Baseline and switch-role | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| 0 if *Broad*, 1 if *Narrow* | -0.497 | -0.499 | -0.104 | | 0.617*** | 0.616*** | 0.616*** |
| | (0.316) | (0.317) | (0.300) | | (0.047) | (0.047) | (0.046) |
| Inattention $\theta$ | -0.808 | -0.801 | -0.369 | -0.487 | | | |
| | (0.509) | (0.508) | (0.512) | (0.436) | | | |
| Treatment dummy * Inattention $\theta$ | 0.714 | 0.705 | 0.006 | | | | |
| | (0.620) | (0.619) | (0.612) | | | | |
| Variance of belief distribution | | -0.000*** | -0.000* | -0.000 | | | |
| | | (0.000) | (0.000) | (0.002) | | | |
| Willingness to take risks | | | 0.555*** | 0.645*** | | | |
| | | | (0.134) | (0.175) | | | |
| 0 if main task, 1 if reverse task | | | | | 0.484*** | 0.442*** | 0.443*** |
| | | | | | (0.052) | (0.056) | (0.056) |
| (1 if *Narrow*) * (1 if switch-role task) | | | | | -1.045*** | -1.043*** | -1.045*** |
| | | | | | (0.069) | (0.069) | (0.069) |
| Constant | 4.550*** | 4.555*** | 3.694*** | 3.011*** | 0.110*** | 0.126*** | -0.070 |
| | (0.180) | (0.181) | (0.637) | (0.613) | (0.024) | (0.031) | (0.054) |
| Task fixed effects | | | Yes | Yes | | Yes | Yes |
| Additional controls | | | Yes | Yes | | | Yes |
| i) Mean inattention *Broad*, baseline | | | | | .11 | .13 | -.07 |
| ii) Mean inattention *Narrow*, switch-role | | | | | .17 | .14 | -.06 |
| i) vs. ii): $F_{1,141}$ | | | | | 1.98 | .09 | .1 |
| iii) Mean inattention *Narrow*, baseline | | | | | .73 | .74 | .55 |
| iv) Mean inattention *Broad*, switch-role | | | | | .59 | .57 | .37 |
| iii) vs. iv): $F_{1,141}$ | | | | | 4.28** | 6.99*** | 6.71** |
| R$^2$ | 0.02 | 0.02 | 0.13 | 0.15 | 0.40 | 0.41 | 0.42 |
| # Observations | 1135 | 1135 | 1135 | 607 | 942 | 942 | 942 |

*Notes*: OLS regressionse. Inattention to noise is calculated as $\theta = \frac{H_B}{H_B+H_N}$, where $H_B$ and $H_N$ denote the Hellinger distance of the subjective distribution to the Bayesian posterior and the inattentive posterior distribution, respectively. Robust standard errors clustered at participant level in parentheses. The *switch-role* task is the final, additional belief task in which we switched experimental conditions, i.e., subjects in *Broad* had to guess only $X$ and subjects in *Narrow* guess both $X$ and $Y$. The additional controls include gender, age, income and task-fixed effect. Group means of inattention and tests of the differences in group means are reported in the footer. Robust standard errors clustered at participant level in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

### 2.5.3 Limits to Learning and the Persistence of Biases

Prima facie, it may seem puzzling that the inferential errors documented in this paper arise despite the substantial costs associated with it. After all, being permanently exposed to noisy information, why do we not learn to adequately incorporate them into our beliefs? There is no consensus about the origins for the persistence of the large catalog of heuristics and biases. A prominent view holds that they should not even be considered errors to begin with, but useful and efficient behavioral strategies in practice. Researchers reproduce these as artifacts in ecologically invalid, stylized experimental contexts, applying false normative views (e.g., Gigerenzer, 1991; Stanovich and West, 2000). This argument, however, does not square well with evidence amassed over several decades on the robustness and ubiquity of many thinking errors. There are apparent limits of learning, the reasons for which are not yet well understood.

The preceding discussions suggests a channel for limits of learning based on awareness: if people learn from surprising feedback by reflecting on their own solution strategy, they will first – and perhaps exclusively – address those elements that are available to introspection. In fact, learning in most context involves feedback, defined as the outcome of an action that is captured by the senses (Luft, 2014). Types of feedback vary, e.g.,performance information or motivational cues such as reward or punishment, as does the learning process itself, which can be, e.g., implicit and procedural or explicit and hypothesis-driven. All learning shares the feature of a person linking her action to a consequence of the action, and then modifying her action next time she encounters a similar situation. A critical feature of inattentive inference is that the processes responsible for the initial neglect appear to be implicit and inaccessible to the conscious mind. They are unavailable to introspection and recall. This is why they cannot be actively targeted by the learner. Other components of the solution strategy, such as controlled thought and even intuitive reasoning accompanied by some metacognitive experience or accessible result (Ackerman and Thompson, 2017), can come to mind when receiving surprising feedback. These processes are more likely to be targeted in the learning process. But they are not the source of the error in the case of inattentive inference. Under this hypothesis, errors due to unconscious processes can survive even in the presence of feedback.[30] A comparative static coming out of this line of argument is that learning to overcome inattentive inference given feedback varies with the presence of conscious steps of reasoning involved in the solution strategy. If inattentive inference is oriented toward the consciously accessible elements of forming a response, then more of these will make it less likely to get to the point of challenging one's internal representation, holding fixed the complexity of the problem. I designed additional treatments to directly test this hypothesis.

**Design.** Treatment *Feedback* is akin to *Narrow*, but further provides the most natural unspecific type of feedback in this setting. In each of the five baseline tasks, after guessing $X$, subjects see the true value of $X$. Under fully inattentive inference, the true value can subjectively be a zero probability event. Still, I expect limited improvement across tasks, because subjects may not get to the point of reflecting on their neglect of $Y$.

Next, to create exogenous variation in the extent of consciously accessible reasoning, condition *Computation with Feedback* inserts a simple algebraic com-

[30] By feedback I mean all types of feedback that are unspecific to the specific neglect committed. An explicit hint to $Y$ allows to exert executive attention and directly influence the percept.

putation to the signal structure, which is identical to that in *Computation*.[31] Recall that these computations are extremely simple, e.g.,"+20 − 30". Notably, standard accounts of dual processing consider these simple algebraic problems as recruiting Type 1 reasoning, because the answer suggests itself without intention (Evans and Stanovich, 2013; Thompson, 2013). A key distinction to the nature of the processes driving inattentive inference is that it creates a metacognitive experience, that is we are aware of somehow having produced a result. The results in *Computation* (Section 2.3.1.6) showed that the computation is inconsequential for the guesses about *X* that subjects actually submit. Presented with surprising feedback about the actually drawn number, however, subjects in *Computation with Feedback* might recall the conscious part of their inference strategy, i.e., undoing the calculation. The computation provides an obvious – albeit unlikely – source of error. I hypothesize that this significantly reduces learning relative to the *Feedback* baseline.

Note that reduced learning when computations are added can also be the result of increased complexity. Adding more sources of error can reduce the likelihood to question each individual one of them in the learning process. I therefore design a third treatment, *Computational Feedback*, in which *(i)* subjects receive the same feedback as before, *(ii)* the inference problem including the signal structure is identical to the no-computations case in *Feedback*, and *(iii)* the setting features identical computational complexity to *Computation with Feedback*. Now, if reduced learning in *Computation with Feedback* is solely the result of increased complexity, one should expect the same in *Computational Feedback*. If, however, learning is only compromised by computations performed *when doing inference*, we would not expect to see reduced learning here, since the inference problem is identical to *Feedback*. I predict the latter.

In *Computational Feedback*, subjects receive a signal on *X* and *Y* without additional computations, i.e., the mean or sum as before. This time, however, these same computations are added at the feedback stage. That means, instead of observing the true value of *X*, subjects see a different value on which they first need to perform the computations to arrive at the true value of *X*. Here, the computations are clearly executed *after* stating a guess about *X*, i.e., after inference. Seeing a surprising true value of *X*, now, subjects presumably recall that they performed the calculations when provided with the feedback, and that they earlier on indicated a guess, which itself was independent of these computations. That is, the computations are not directly associated with the inference process. Since the algebraic calculations are extremely simple, I expect subjects are instead somewhat more likely to reflect on the inference stage.

[31] That is, *Computation with Feedback* is identical to condition *Computation*, except for the feedback; and it is identical to condition *Feedback*, except for the simple computation that needs to be undone.

**Results.** In the first round before receiving feedback for the first time, inattention is expectedly indistinguishable across the feedback treatments (see Appendix Figure 2.F.12).[32] I now analyze inattention scores of beliefs stated after feedback has been received in preceding rounds. In what follows, I restrict my attention to the fifth and last round, since learning effects should be highest after several rounds of feedback.[33] For ease of exposition, Figure 2.9 depicts mean inattention by treatment condition. All statistical analyses, however, are based on empirical distributions of inattention.[34] Inattention in the three relevant no-feedback conditions is displayed above the dashed horizontal line for comparison. I document three key findings. First, the provision of feedback alone (condi-

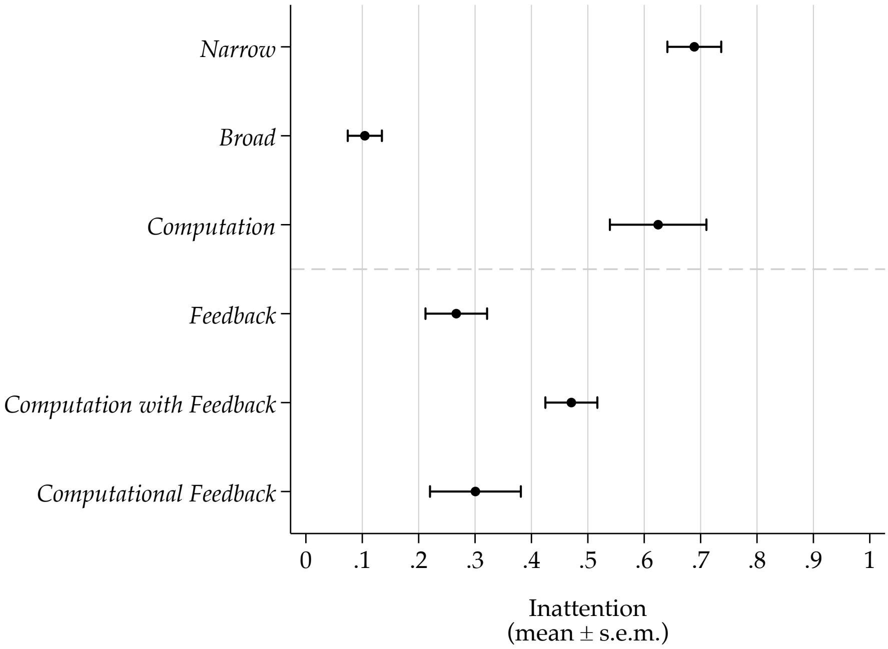


**Figure 2.9.** Treatment means of inattention to $Y$. The three treatments above the dashed line show conditions without feedback for reference. Feedback is about the truly drawn value of $X$. Displayed are implied inattention scores in the final baseline round, after having received feedback in the four preceding rounds. Sample sizes are N = 72 in both *Narrow* and *Broad*, N = 48 in *Feedback*, and N = 24 in each of Computation, Computation with Feedback and Computational Feedback.

[32] This again shows that the computation added to the signal does not influence stated beliefs.

[33] The following findings persist if beliefs from rounds two to five are pooled together. See further results in Appendix 2.F.

[34] See Appendix 2.F for distribution plots.

tion *Feedback*) significantly decreases inattention relative to *Narrow* ($p < 0.001$, M-W $U$ test). At the same time, learning is far from perfect, as indicated by a remaining treatment effect between *Feedback* and the *Broad* benchmark without feedback ($p < 0.001$). That means, feedback about the actually drawn $X$ generates substantial improvements, but it is no guarantee that subjects figure out their neglect of $Y$.

Second, I find clear evidence for the hypothesis that additional computations in the inference process reduce learning. Inattention in *Feedback* and *Computation with Feedback* significantly differ ($p = 0.008$). This effect presumably operates by diminishing subjects' propensity to realize that there are parts of the problem *that they have not attended to*. Notably, the addition of simple algebraic computation virtually eliminated learning. Inattention in *Computation with Feedback* is not significantly lower than in *Computation* ($p = 0.21$).

Third, the documented reduction in learning is not driven by an increase in complexity. In fact, inattention in *Computational Feedback* is indistinguishable from *Feedback* ($p = 0.48$), but significantly differs from *Computation with Feedback* ($p = 0.005$) in the predicted direction.

Taking stock, the data clearly suggest that subjects learn to account for $Y$ only if they cannot attribute prediction mistakes to any steps in the reasoning process that are available to introspection. In an attempt to do gather direct evidence of this hypothesis, the design includes an additional choice in all feedback treatments. On the feedback screen that informs about the actual draw, subjects could choose to be reminded of up to exactly one aspect of the preceding belief task: the distribution of $X$, the distribution of $Y$, or the signal structure. Revealing such details can help figure out the source of an erroneous guess and make better subsequent guesses. In the first round, i.e., upon receiving feedback for the first time, subjects are indeed more likely to reveal the distribution of $Y$ in *Feedback* than in *Computation with Feedback* ($p = 0.044$). This effect, however, is not robust and loses significance when pooling all rounds. Procedural details and further results are relegated to Appendix 2.F.[35]

## 2.6 Related Literature

Early work on updating patterns found that beliefs sometimes underreact to new information as if overweighting the base rate (Edwards, 1982), and sometimes overreact in line with a neglect of the base rate (Bar-Hillel, 1980). As

[35] Finally, in practice available signals are usually noisy or imprecise. The possibility that observed feedback is not exactly right might provide another obvious way for subjects resolve the conflict between their subjective belief and feedback received, again reducing learning. In condition *Imperfect Feedback*, subjects are given feedback about the true $X$ that is correct only with 80% probability, but would see a value of $X$ which is not the true one with 20% probability. Again, learning is reduced in a similar way as by adding the computation, see Appendix 2.F.4.

discussed in Section 2.3.1.7, over-sensitivity to likelihoods is not generally compatible with the form of noise neglect. Information neglect, on the other hand, is observationally equivalent to a strong form of conservatism. This paper focuses on the cognitive primitives of updating behavior, and identifies selective attention to elements of an information environment, which hinges on origins of over- and underreaction that are conceptually distinct from miscalibration to likelihoods or base rates.

An increasing number of recent empirical studies examines the role of attention in economic contexts (Chetty et al., 2009; De los Santos et al., 2012; Hirshleifer and Teoh, 2003; Lacetera et al., 2012). That certain aspects of the environment are not properly represented by agents is the central finding of recent work on the neglect of correlations (Enke and Zimmermann, 2017) and unobserved signals (Enke, 2017). Unlike in these studies where agents face multiple pieces of information with possible interrelations, inattentive inference is about how a single signal is processed, or interpreted, given that the information structure is attended to. At the same time, my analyses on the underlying mechanisms are inspired by – and the findings are largely consistent with – the evidence in these papers. Moreover, Enke (2017) argues that neglect is due to a flawed representation of a problem, modulated by a problem's complexity. This paper builds on this idea and focuses on how exactly representations are formed.

Inattentive inference is further connected to research on other patterns of misreading information, such as confirmatory bias (Rabin and Schrag, 1999) or misattribution in social learning (Eyster and Rabin, 2005, 2014). An additional implication of my results is that two persons with standard preferences and identical prior beliefs can draw predictably different conclusions after seeing the same evidence.[36]

On the theory side, Gabaix (2014) formulates a model in which agents first choose a sparse attention vector given fixed psychic costs of attending to each dimension. They rationally trade off the cost of not tailoring an action to some unobserved state of the world against the cost of attending to that dimension. The intuition built into Gabaix (2014) is that of agents maximizing within a sparse representation of the world. This sparse model excludes the dimensions that are least costly to ignore in a given optimization problem, and replaces these with default values. In the model, the agent finds out the exact realization of a

[36] Blackwell and Dubins (1962) show that beliefs of people with identical priors should converge if they are Bayesian. Here I show that deviations from Bayesian updating induce disagreements in a predictable manner. A long literature on the origins of belief polarization partly invokes non-Bayesian updating as a source of different opinions, including confirmatory bias (Rabin and Schrag, 1999) and ambiguity aversion (Baliga et al., 2013).

variable upon attending to it.[37] While the notion of sparsity in attention is closely related to the motivation of inattentive inference, Gabaix (2014) abstracts from the implications for inference from noisy information in itself. The present paper shows that attention to a variable is beneficial even if the optimal action does not depend on it.[38]

Another strand of literature on rational inattention focuses on the *acquisition* of information when the amount of available sources of information exceeds the decisionmaker's processing capabilities or processing is costly (see also Section 2.2.2). These models assume rational inference from those signals that an agent actually attends to (e.g., Matejka and McKay, 2014; Sims, 2003). In *learning though noticing* models (Gagnon-Bartsch et al., 2017; Hanna et al., 2014; Schwartzstein, 2014), people face uncertainty about which dimensions are important for predicting a relevant variable. While learning through noticing models explicitly describe a failure to *process* given data, as is the case in my experiments, a crucial difference is that inattentive inference is not driven by uncertainty about the structure of the world but occurs even in the absence of such uncertainty.

## 2.7 Conclusion

This paper reports causal evidence from more than twenty different experiments on how people update from noisy information. In my view, it provides four core messages.

First, belief formation is not sufficiently homogenous to be described by a single rule or some "representative agent". At the same time, updating rules are drawn from a discrete set that can be precisely characterized. There is little evidence for the use of mixtures of these inference modes.

Second, the primitive of different updating modes are subjective representations of a context. These representations correspond to the set of elements of an information structure that an agent attends to, i.e., that he processes into a mental model. This mental model forms the basis for subsequent computations that result in a posterior. The distinction between mental representations and computations on those representations leverage insights from the computational theory of mind.

Third, the relative prevalence of updating modes is not random but systematic and predictable. People's inference strategies respond to variations in the

[37] In Proposition 16 and Online Appendix, Sextion XI, Gabaix (2014) considers an extension where the agent perceives the realization of a variable with noise and decides on a signal precision. Inference from a noisy signal is assumed to be optimal.

[38] This is because attending to a dimension can have a positive externality for learning about other variables that directly affect actions.

expected benefits and costs. This paper avoids making general statements about whether people over- or underreact to information *on average*. Instead, I stress that average patterns are likely to vary systematically with features of the context. Inattentive inference is in line with overreaction in some situations and underreaction in others, yet does not give arbitrary flexibility but provides testable comparative statics. One interesting example is the hypothesized relationship to average cognitive skills in a group of people. As cognitive skills increase, the propensity to neglect noise increases at the expense of information neglect, implying a switch from underreaction to overreaction. As cognitive skills further increase, people become increasingly Bayesian and thus well-calibrated to information. These and other predictions provided by the conceptual framework could be an interesting avenue for further research.

Fourth, this paper highlights the role of awareness for understanding belief formation. The processes that lead to the adoption of a representation, as well as the representation itself, appear to live below people's threshold of awareness. This has far-reaching implications, some of which are explored in this paper. More generally, I suggest that the characterization of behavioral mechanisms in terms of associated awareness can help better understand and potentially unify the diverse corpus of behavioral deviations from rationality. In fact, the awareness result and other findings on the psychological mechanisms are not easily accounted for by the dominant approach to belief formation in economics that focuses on higher-level cognitive processes such as intuition and reasoning (Kahneman, 2003). Instead, the combined evidence on inattentive inference might be much better captured by the mechanisms of *perception*. Perception is the processing of sensory information to make sense of a situation by creating an internal representation of it (Bernstein, 2013). Future work can help to clearly draw out the implications of this distinction.

## References

**Ackerman, Rakefet and Valerie Thompson (2017):** "Meta-reasoning: Shedding metacognitive light on reasoning research." In. *International Handbook of Thinking & Reasoning.* Psychology Press. [46]

**Baliga, Sandeep, Eran Hanany, and Peter Klibanoff (2013):** "Polarization and ambiguity." *The American Economic Review*, 103 (7), 3071–3083. [50]

**Bandyopadhyay, Prasanta, Gordon Brittan, and Mark Taper (2016):** *Belief, Evidence, and Uncertainty: Problems of Epistemic Inference*. Springer International. [23]

**Bar-Hillel, Maya (1980):** "The base-rate fallacy in probability judgments." *Acta Psychologica*, 44 (3), 211–233. [49]

**Bernstein, D. (2013):** *Essentials of Psychology*. Cengage Learning. [52]

**Blackwell, David and Lester Dubins (1962):** "Merging of opinions with increasing information." *The Annals of Mathematical Statistics*, 33 (3), 882–886. [50]

**Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014):** "hroot: Hamburg registration and organization online tool." *European Economic Review*, 71, 117–120. [20]

**Bordalo, Pedro, Nicola Gennaioli, Rafael LaPorta, and Andrei Shleifer (2017):** "Diagnostic expectations and stock returns." Tech. rep. National Bureau of Economic Research. [26]

**Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2018):** "Diagnostic expectations and credit cycles." *The Journal of Finance*, 73 (1), 199–227. [26, 69]

**Caplin, Andrew and Mark Dean (2015):** "Revealed preference, rational inattention, and costly information acquisition." *The American Economic Review*, 105 (7), 2183–2203. [9, 13]

**Caplin, Andrew, Mark Dean, and Daniel Martin (2011):** "Search and satisficing." *The American Economic Review*, 101 (7), 2899–2922. [8, 18]

**Chen, Daniel L., Martin Schonger, and Chris Wickens (2016):** "oTree—An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance*, 9, 88–97. [20, 28]

**Chetty, Raj, Adam Looney, and Kory Kroft (2009):** "Salience and taxation: Theory and evidence." *The American Economic Review*, 99 (4), 1145. [50]

**Davis, Mark H. (1983):** "Measuring individual differences in empathy: Evidence for a multidimensional approach." *Journal of Personality and Social psychology*, 44 (1), 113–126. [39]

**De los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest (2012):** "Testing models of consumer search using data on web browsing and purchasing behavior." *The American Economic Review*, 102 (6), 2955–2980. [50]

**Dean, Mark and Nathaniel Neligh (2017):** "Experimental tests of rational inattention." Tech. rep. Columbia University. [8, 18]

**Dijksterhuis, Ap and Henk Aarts (2010):** "Goals, attention, and (un) consciousness." *Annual Review of Psychology*, 61, 467–490. [11, 41]

**Edwards, Ward (1982):** "Conservatism in human information processing." In. *Judgment under Uncertainty: Heuristics and Biases*. Ed. by Daniel Kahneman, Paul Slovic, and Amos Tversky. Cambridge University Press, 359–369. [49]

**Enke, Benjamin (2017):** "What you see is all there is." Mimeo. [8, 9, 18, 50]

**Enke, Benjamin and Florian Zimmermann (2017):** "Correlation Neglect in Belief Formation." *The Review of Economic Studies*. [8, 18, 50]

**Evans, Jonathan and Keith E. Stanovich (2013):** "Dual-process theories of higher cognition: Advancing the debate." *Perspectives on Psychological Science*, 8 (3), 223–241. [47]

**Eyster, Erik and Matthew Rabin (2005):** "Cursed equilibrium." *Econometrica*, 73 (5), 1623–1672. [50]

**Eyster, Erik and Matthew Rabin (2014):** "Extensive imitation is irrational and harmful." *The Quarterly Journal of Economics*, 129 (4), 1861–1898. [50]

**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." Mimeo. [20, 39]

**Gabaix, Xavier (2014):** "A sparsity-based model of bounded rationality." *The Quarterly Journal of Economics*, 129 (4), 1661–1710. [50, 51, 61]

**Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein (2017):** "Channeled attention and stable errors." Mimeo. [51]

**Gigerenzer, Gerd (1991):** "How to make cognitive illusions disappear: Beyond 'heuristics and biases'." *European Review of Social Psychology*, 2 (1), 83–115. [45]

**Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein (2014):** "Learning through noticing: Theory and evidence from a field experiment." *The Quarterly Journal of Economics*, 129 (3), 1311–1353. [51]

**Hauser, David J. and Norbert Schwarz (2016):** "Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants." *Behavior Research Methods*, 48 (1), 400–407. [28]

**Hellinger, Ernst (1909):** "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen." *Journal für die reine und angewandte Mathematik*, 136, 210–271. [23]

**Hirshleifer, David and Siew Hong Teoh (2003):** "Limited attention, information disclosure, and financial reporting." *Journal of Acounting and Economics*, 36 (1), 337–386. [50]

**Horst, Steven (2011):** "The computational theory of mind." *Stanford Encyclopedia of Philosophy*. [9]

**Hossain, Tanjim and Ryo Okui (2013):** "The binarized scoring rule." *The Review of Economic Studies*, 80 (3), 984–1001. [18]

**Kahneman, Daniel (2003):** "Maps of bounded rationality: Psychology for behavioral economics." *The American Economic Review*, 93 (5), 1449–1475. [11, 52]

**Khaw, Mel Win, Luminita Stevens, and Michael Woodford (2017):** "Discrete adjustment to a changing environment: Experimental evidence." *Journal of Monetary Economics*, 91, 88–103. [8, 18]

**Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor (2012):** "Heuristic thinking and limited attention in the car market." *The American Economic Review*, 102 (5), 2206–2236. [50]

**Luft, Caroline Di Bernardi (2014):** "Learning from feedback: The neural mechanisms of feedback processing facilitating better performance." *Behavioural Brain Research*, 261, 356–368. [46]

**Matejka, Filip and Alisdair McKay (2014):** "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *The American Economic Review*, 105 (1), 272–298. [51]

**Paulus, Christoph (2009):** "Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index." [20]

**Rabin, Matthew and Joel L. Schrag (1999):** "First impressions matter: A model of confirmatory bias." *The Quarterly Journal of Economics*, 114 (1), 37–82. [50]

**Rammstedt, Beatrice and Oliver P. John (2005):** "Kurzversion des big five inventory (BFI-K)." *Diagnostica*, 51 (4), 195–206. [20]

**Reis, Ricardo (2006):** "Inattentive consumers." *Journal of Monetary Economics*, 53 (8), 1761–1800. [15]

**Rubinstein, Ariel (2007):** "Instinctive and cognitive reasoning: A study of response times." *The Economic Journal*, 117 (523), 1243–1259. [40]

**Schwartzstein, Joshua (2014):** “Selective attention and learning.” *Journal of the European Economic Association*, 12 (6), 1423–1452. [51]

**Shannon, Claude E. and Warren Weaver (1949):** *The mathematical theory of communication*. University of Illinois Press. [8, 17]

**Sims, Christopher A. (2003):** “Implications of rational inattention.” *Journal of Monetary Economics*, 50 (3), 665–690. [9, 15, 16, 51]

**Stanovich, Keith E. and Richard F. West (2000):** “Individual differences in reasoning: Implications for the rationality debate?” *Behavioral and Brain Sciences*, 23 (5), 645–665. [45]

**Thompson, Valerie (2013):** “Why it matters: The implications of autonomous processes for dual process theories—Commentary on Evans & Stanovich (2013).” *Perspectives on Psychological Science*, 8 (3), 253–256. [47, 65]

**Verrecchia, Robert E. (1982):** “Information acquisition in a noisy rational expectations economy.” *Econometrica*, 1415–1430. [15]

## Appendix 2.A Treatment Overview

**Table 2.A.1.** Overview of laboratory treatments

| Condition | Description | Covered in |
|---|---|---|
| *Baseline experiment: Narrow and Broad* | (Elements of baseline experiment in respective order below) | |
| Baseline Tasks | 5 updating tasks in random order. $X$ and $Y$ follow independent discrete uniform distributions with outcome spaces smaller than 10. The information is the mean or the sum of the draws. | Appendix 2.C |
| Robustness Tasks | 5 updating tasks in random order. Data are correlated, drawn from a larger sample space, discretely normally distributed, or the information is outside of the range of $X$. | Appendix 2.D.1 |
| Bonus Task | 1 surprise task with similar configuration to baseline. Within each condition, subjects are re-randomized and face either the same expected incentive size as before, or tenfold incentives. | Main text |
| Confidence | For each baseline and robustness problem, subjects indicate their valuation for their stated belief using a multiple-price list method. | Appendix 2.E.4 |
| Switch-role | 2 tasks with similar configuration as baseline, but subjects face incentives from opposite treatment condition. That means *Narrow* is paid for $X$ and $Y$, while *Broad* paid for $X$ only. | Appendix 2.E.5 |
| *Computation* | Identical to *Narrow* baseline, except that a simple, task-varying algebraic calculation is added to the information structure (e.g., "the mean $+20-30$"). | Appendix 2.D.2 |
| *Simplification* | Identical to *Narrow* baseline, but deciphering stage and all time limits removed. | Appendix 2.D.4 |
| *Narrow with joint elicitation* | Identical to *Narrow* baseline, but subjects indicate the joint distribution of $X$ and $Y$ (while only $X$ is paid for). | Appendix 2.D.3 |
| *Broad with sequential elicitation* | Identical to *Broad* baseline, but subjects indicate the marginal distributions of $X$ and $Y$ in sequential order, such that the first screen is identical to *Narrow* baseline. | Appendix 2.D.3 |
| *Hint* | Identical to *Narrow* baseline, but subjects receive a reminder on the elicitation screen, stating "Also think about the role of $Y$". | Appendix 2.E.3 |
| *Feedback* | Identical to *Narrow* baseline, but subjects observe the actual draw of $X$ after stating their guess. | Appendix 2.F.1 |
| *Computation with feedback* | Identical to *Computation*, but subjects observe the actual draw of $X$ after stating their guess. | Appendix 2.F.2 |
| *Computational feedback* | Identical to *Computation with feedback*, except that the computation is added to the feedback instead of the information. | Appendix 2.F.3 |
| *Imperfect feedback* | Identical to *Feedback*, but subjects receive the true draw as feedback only with 80% probability, while seeing another value with 20% probability. | Appendix 2.F.4 |

**Table 2.A.2.** Overview of online treatments

| Condition | Description | Covered in |
|---|---|---|
| *Baseline experiment (Narrow only)* | 5 updating tasks in random order (Table 2.3). $X$ and $Y$ follow independent distributions. Subjects only state a mean posterior belief about $X$. No deciphering stage. | Main text |
| *Form of Noise Neglect* | 10 updating tasks in random order (Table 2.D.5). Identical to baseline online experiment but different information structures to analyze different candidates for the belief formation rule under noise neglect. | Appendix 2.D.5 |
| *Signal-to-Noise Ratio* | 7 updating tasks in random order. Identical to baseline online experiment but different information structures (Table 2.4). The signal-to-noise ratio is varied between tasks. | Main text |
| *Directional Bias* | 5 updating tasks in random order. Identical to baseline online experiment but different information structures (Table 2.5). All elements of the information structure are kept fixed across tasks except the mean of $Y$. | Main text |

## Appendix 2.B Conceptual Framework

### 2.B.1 An Endogenous Characterization of Signal and Noise

The basic framework presented in Section 2.2 focused on the belief formation problem and took the classification of variables as signal ($X$) or noise ($Y$) as given. In the simplest case, both $X$ and $Y$ were one-dimensional. $X$ was characterized as affecting the agent's utility *given an action*, while $Y$ did not. This is to say that $Y$ was only relevant for the agent in the belief formation process, i.e., to form a more accurate belief about $X$, but not the in selection of an action *given a belief about X*.

This crude distinction can be refined to better suit information structures in practice. $X$ is usually multi-dimensional, i.e., there are more than one and often thousands of variables that affect utility given an action, even if to a minimal degree. In that case, it is impossible for the agent to form beliefs about and tailor his actions to all of those variables. The reason is again that tailoring an action optimally to a belief about a variable is cognitively costly. Instead, agents can restrict their optimization to a subset of the dimensions in $X$, accounting for those that change actions most and have the biggest impact on utility, while ignoring those that only change actions and utility by a little bit.

In the following I endogenize the characterization of variables as belonging to $X$ or $Y$ by modeling an initial step in which the agent selects the variables that he will tailor his actions to. This initial step is a crucial input for any subsequent belief formation processes, which has been left out from the framework presented in the main text.

Concretely, this initial stage classifies each stochastic variable in the world into the following three types:

Category 1 Variables that do affect utility given an action, and which are considered by the agent in selecting an action. These variables form the *consideration vector*. Only variables included in the consideration vector can play the role of $X$ in the belief formation framework.

Category 2 Variables that do affect utility given an action, but which are *not* considered by the agent in selecting an action. This type is a subset of $Y$ in the belief formation framework.

Category 3 Variables that do *not* affect utility given an action, and which are *not* considered by the agent in selecting an action. This type corresponds to $Y$ in the belief formation framework.

The only new class is Category 2, which are variables that only change actions by a little bit, so that accounting for them would not affect utility by much. The agent ignores these variables in his optimization. Moreover, when new information is received prior to taking an action, the agent would treat these variables exactly like $Y$, i.e., as noise. All variables in an information structure that do not form part of a consideration vector classify as noise, or $Y$, even if they affect utility given an action.

To illustrate, assume a consumer wants to buy a printer. Printers are fully characterized by three unobservable attributes: durability, ease of handling and maintenance cost, $Z = (Z_{durability}, Z_{handling}, Z_{maintenance})$. The utility derived from the consumer's action, a choice of printer, depends on the durability and ease of handling, but not on maintenance cost. It also depends much more on the durability than on ease of handling. Tailoring the subjective assessment of a printer to each additional dimension comes at a cognitive cost, which can be thought of as forming an explicit belief about the dimension and selecting an optimal action based on that belief. Similar to the belief formation framework, I assume that the agent can either fully account for a variable or not at all, and there is a fixed psychic cost per dimension, $d$.[39] The agent can now choose the dimensions that he will base his action on. For example, he might reasonably choose to asses a given printer based only on his expectations about its durability, while ignoring both ease of handling and maintenance cost, since he cares little or not at all about those. Crucially, this process occurs *before* learning about any information structures and is entirely unconnected to considerations about belief formation. Instead, this is purely about the "importance" of an attribute in the agent's

[39] Cost $d$ might be different from the cost of forming a posterior $q$.

utility function. However, this selection process is an input to any following belief updating. Assume the agent now receives a relevant piece of information about a printer that is influenced by the unobserved levels of all three variables, e.g., a customer rating. He will then treat durability as signal, $X = Z_{durability}$, and the other two variables as noise, $Y = (Z_{handling}, Z_{maintenance})$.

More generally, I conceptualize the process of selecting Category 1 variables, or his consideration vector, as follows:

1. For each candidate consideration vector, the agent determines his optimal actions given his prior belief about the distribution of variables in the consideration vector. He solves the problem of finding the optimal action based on the (false) assumption that his utility not does depend on any variable that is excluded from the consideration vector.

2. Given these hypothetical actions, he calculates his expected benefit from each consideration vector, based on his true utility function. He selects the consideration vector with highest expected benefit net of cognitive cost.

Let $Z \in \mathbb{R}^n$ be a variable with dimension corresponding to the number of variables that affect utility given an action, i.e., all variables except those in Category 3. The agent's prior belief is a distribution $h \in \Delta(Z_1 \times Z_2 \times \cdots \times Z_n)$. Action $m(h) \in M$ has arbitrary dimension. Utility is $u(m,z) = u(m, z_1, z_2, \ldots, z_n)$. The consideration vector $Z_{consid}$ consists of some, but not necessarily all dimensions of $Z$. Which elements are contained in $Z_{consid}$ is specified by an $n$-dimensional vector $b$ of zeros and ones that indicates the dimensions of $Z$ that are accommodated when selecting an action. Let $u^b$ a modified utility function in which utility only depends on the variables in the consideration vector, $u^b(m, z_{consid})$.[40] I further call $h^b \in \Delta(Z_{consid})$ the belief about variables in the consideration set and $m^b(h^b)$ the action that optimizes $u^b$.

The agent selects a consideration vector as follows, where $d$ is a fixed psychic cost per dimension:

$$b = \underset{b \,\in\, \{0,1\}^n \text{ s.th. } \sum_{i=1}^n b_i \geq 1}{\arg\min} \mathbb{E}\left[u\left(m^b(h^b), z_1, z_2, \ldots, z_n\right)\right] - \sum_{i=1}^n b_i \cdot d \quad (2.B.1)$$

Intuitively, the agent determines optimal actions in lower-dimensional space, ignoring some dimensions as dictated by a candidate consideration vector, and then evaluates these actions based on his true utility function. Changing the consideration vector does not change the range of available actions, e.g., "buy" and "don't buy" a given printer model. The consideration vector always needs to include at least one dimension.

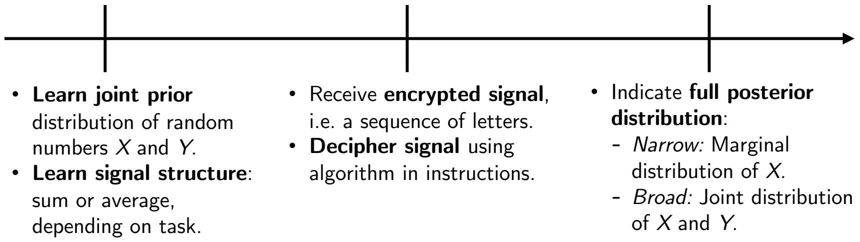
[40] $u^b$ is the projection of $u$ onto the space $(M \times Z_{consid})$.

This notion of how people focus on “more relevant" dimensions is closely related to and inspired by the “sparse max” of Gabaix (2014). In fact, $b$ is a sparse vector which has similarities to Gabaix’ (2014) attention vector $m$.[41]

What happens if the agent receives information before taking an action? The consideration vector specifies the dimensions in an information structure that the agent always attends to. The intuition is that because the agent tailors his actions to a given variable, he forms a belief about that variable in any case and thereby incurs a fixed cognitive cost. One example relating to the experimental paradigm is the action of guessing a realization of $X$ directly: it is impossible to not think about $X$ when stating a guess even if subjects would not receive an additional piece of information. This is one justification for noise neglect that is absent in all other work on attention such as rational inattention. Because people attune their action to their beliefs about some attributes, but not others, they give attentional priority to those dimensions even in the absence of new information. When new information arrives, they benefit from already thinking about $X$ which makes it less costly to relate the new information to that variable. This can generate overreaction where rational inattention models would predict inertia.

## Appendix 2.C Baseline Experiments

### 2.C.1 Procedure of Updating Tasks: Laboratory Experiment



- **Learn joint prior** distribution of random numbers $X$ and $Y$.
- **Learn signal structure**: sum or average, depending on task.

- Receive **encrypted signal**, i.e. a sequence of letters.
- **Decipher signal** using algorithm in instructions.

- Indicate **full posterior distribution**:
  - *Narrow:* Marginal distribution of $X$.
  - *Broad:* Joint distribution of $X$ and $Y$.

**Figure 2.C.1.** Timeline of updating task in laboratory experiment.

[41] There are some differences. The sparse max stresses the importance of *defaults* for each variable, which are assumed to be their respective means. The above formalization abstracts from explicit default values, suggesting that people truly *ignore* dimensions outside of the consideration vector, rather than using defaults. Gabaix (2014) also allows for continuously chosen attention and closes the model by approximating the expected utility losses relative to the full attention case, while I focus on the simpler, binary case. More generally, the sparse max has a different goal than the present paper and is not specifically geared to the specific role of attention for belief formation.

### 2.C.2 Consistency of Attention Across Tasks

In this section I examine how consistently inattentive or consistently Bayesian subjects behave across tasks. Figure 2.2 in the main text includes five beliefs per subjects. But does each subject exhibit a stable level of attention? Figure 2.C.2 shows kernel density estimates of subject-level mean inattention. While there is a strong clustering of subjects in the *Broad* condition who always form beliefs that are close to an implied inattention of zero, there are no such two clusterings in the *Narrow* treatment – one at each end of the attention spectrum – as could be expected from Figure 2.2. Instead, there is a smaller clustering at mean inattention values of between 0.8 and 1. Indeed, I find that many subjects in *Narrow* condition formed close to Bayesian beliefs in some tasks, and close to fully inattentive beliefs in other tasks. In fact, 15.5% of subjects in *Narrow* indicated both a fully Bayesian and a fully inattentive belief at least once. This may suggest that a subject's degree of attention to $Y$ varies across situations to some extent, even for largely identical updating contexts.

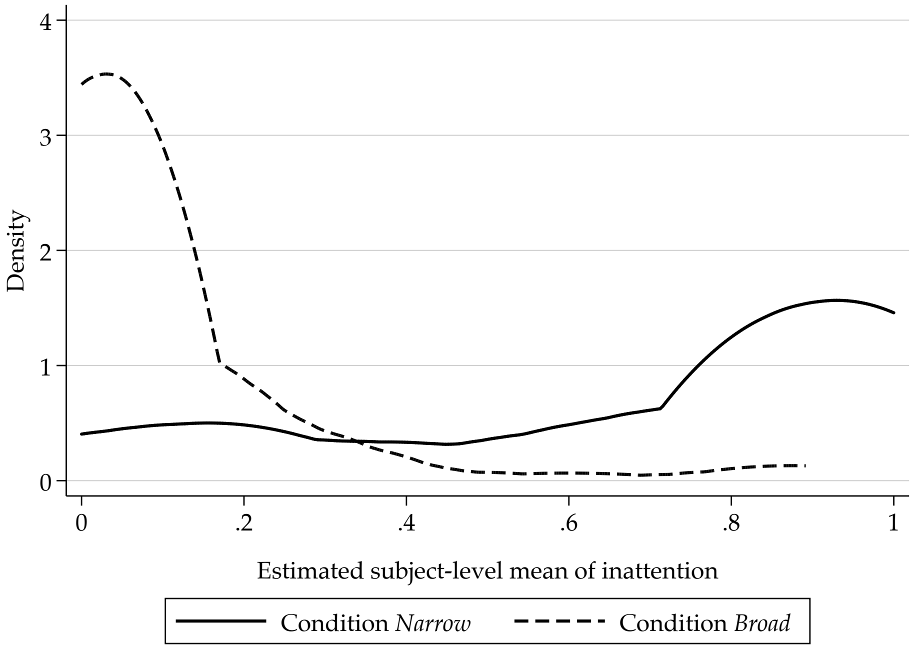


**Figure 2.C.2.** Subject-level mean of inattention to $Y$. N=144. For each subject I calculate the mean inattention in the five baseline tasks. The curves show kernel density estimates for each treatment (both N=72). A parameter of $\theta = 0$ is consistent with Bayesian updating. $\theta = 1$ means complete inattention. Epanechnikov kernel with bandwith 0.1.

# Appendix 2.D Robustness Treatments

## 2.D.1 Task Variations

**Table 2.D.3.** Overview of robustness tasks

| Task | Sample space *X* | Sample space *Y* | Signal type | Signal value |
|---|---|---|---|---|
| Correlated data (r=0.7) | $\{95, 96, \ldots, 104, 105\}$ | $\{-15, -14, \ldots, 14, 15\}$ | $(X+Y) \div 2$ | 104 |
| Larger sample space ($>$ 10) | $\{190, 191, \ldots, 209, 210\}$ | $\{180, 181, \ldots, 219, 220\}$ | $(X+Y) \div 2$ | 208 |
| Discrete normally distributed numbers | $\{170, 180, \ldots 220, 230\}$ | $\{-50, -40, \ldots 40, 50\}$ | $X+Y$ | 220 |
| Signal out of X range | $\{240, 241, \ldots, 259, 260\}$ | $\{-15, -14, \ldots, 14, 15\}$ | $X+Y$ | 230 |

*Notes*: This table provides an overview of the four robustness belief tasks. The distributions of $X$ and $Y$ as well as the signal structure are identical in both treatment conditions. $X$ and $Y$ were independently drawn from two discrete uniform distribution, i.e., every indicated outcome was equally likely.

**Table 2.D.4.** Median inattention in robustness tasks

| Task | Median inattention $\theta_{Mdn}$ | | Mann-Whitney *U* test |
|---|---|---|---|
| | *Narrow*<br>N=72 | *Broad*<br>N=72 | ($p$-value) |
| Correlated data (r=0.7) | 0.59 | 0.00 | $< 0.001$ |
| Larger sample space ($>$ 10) | 1.00 | 0.33 | $< 0.001$ |
| Discrete normally distributed numbers | 0.44 | 0.27 | $< 0.001$ |
| Signal out of X range | 0.49 | 0.17 | $< 0.001$ |

*Notes*: This table displays group medians of implied inattention parameters by treatment condition for four additional belief formation tasks. Inattention is calculated as $\theta = \frac{H_B}{H_B + H_N}$, where $H_B$ and $H_N$ denote the Hellinger distance of the subjective distribution to the Bayesian posterior and the inattentive posterior distribution, respectively. Task order was randomized within each of the two blocks. 72 subjects participated in each condition.

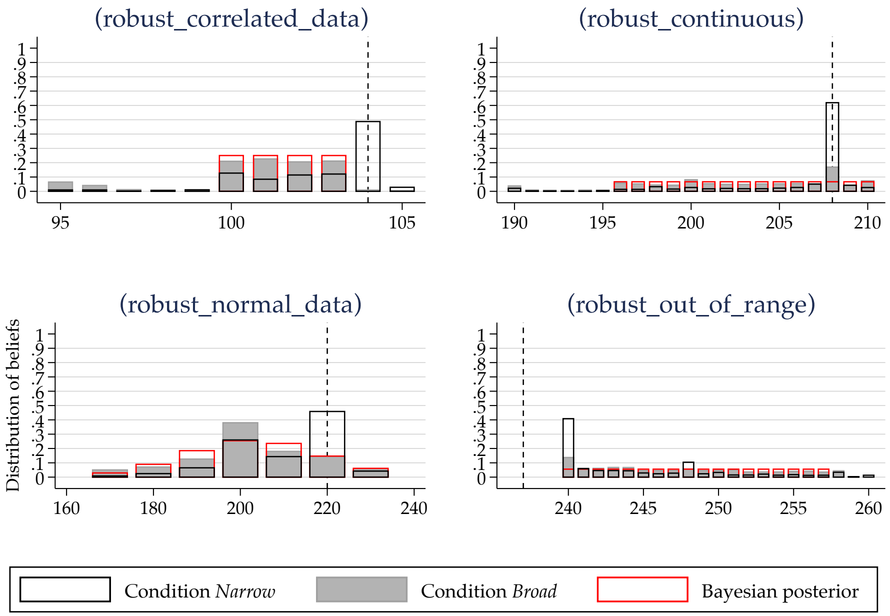


**Figure 2.D.3.** Distribution of elicited belief distributions about $X$ in each one of four robustness tasks. N=72 for each condition in each task. The horizontal axis shows possible outcomes of $X$. The Bayesian posterior belief is provided for reference. The observed signal is indicated by the vertical dashed line. The conrresponding task configurations are shown in Table 2.D.3.

### 2.D.2 Face Value Heuristic and Anchoring

In treatment *Computation*, a simple algebraic computation was added on top of the signal structure. The resulting signals provided in the five baseline tasks were "average of $X$ and $Y$ minus (3·5) plus 35", "sum of $X$ and $Y$ plus (2·10) minus 30", "sum of $X$ and $Y$ plus 40 minus (4·5)", "average of $X$ and $Y$ minus (8·5) plus 10", and "average of $X$ and $Y$ plus (3·5) plus 10". Note that given the simplicity of these calculations, it is possible that subjects did not have to execute these computations effortfully but the results automatically came to mind. This is suggested by research on dual processing (Thompson, 2013).

The computations were chosen such anchoring on the signal value remains equally plausible. If subjects apply a simple face value heuristic, they should ignore both the the computation and the variation of Y. Figure 2.D.4 shows raw beliefs in condition *Computation*, including the signal value and the signal value *after accounting for the computation*. There is limited evidence for anchoring on the signal value. Subjects do not simply take the signal at face value, but they take into account the computation and still neglect $Y$.

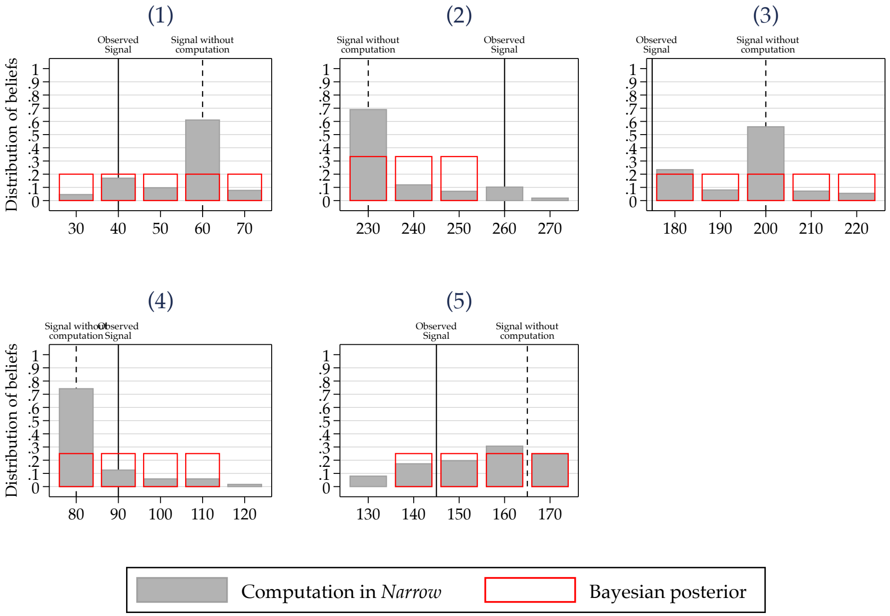


**Figure 2.D.4.** Distribution of elicited belief distributions about $X$ in condition *Computation*. N=24 in each task. The horizontal axis shows possible outcomes of $X$. The Bayesian posterior belief is provided for reference. The observed signal is indicated by the solid dashed line, and the signal value after undoing the computation is shown by the dashed line. In all five tasks, $X$ and $Y$ follow independent discrete uniform distributions that were shown to subjects. Task order was randomized.

### 2.D.3 Elicitation Procedure

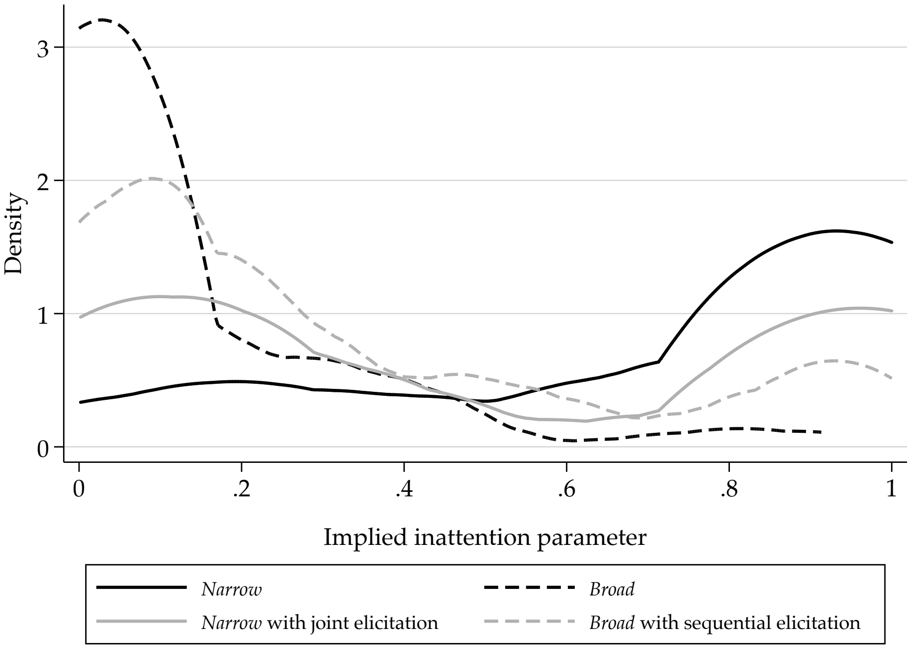


**Figure 2.D.5.** Subject-level mean of inattention to $Y$ in four conditions. Based on N=216. For each subject I calculate the mean inattention in the five baseline tasks. The curves show kernel density estimates for each treatment (*Narrow* N=72, *Broad* N=72, *Narrow with joint elicitation* N=24, *Broad with sequential elicitation* N=48). A parameter of $\theta = 0$ is consistent with Bayesian updating. $\theta = 1$ means complete inattention. Epanechnikov kernel with bandwith 0.1.

In treatments *Narrow* and *Broad*, prediction incentives are different, but the elicitation method also differs. In *Narrow*, subjects only indicate the marginal of X, while in Broad, subjects indicate the joint distribution of X and Y. To rule out that treatment effects are driven by this difference in what is elicited, I designed two additional treatments. In *Narrow with joint elicitation*, only $X$ is paid for (as in *Narrow*) but the joint distribution is elicited exactly as in *Broad*. In *Broad with sequential elicitation*, $X$ and $Y$ are paid for (as in *Broad*) but now the subject first indicates the marginal of $X$, and then indicates the marginal of $Y$ on a separate screen. This way, the first screen (for the marginal of $X$) is exactly identical to *Narrow*. Figure 2.D.5 plots kernel density estimates of the within-subject mean of inattention in the five belief tasks for all four treatments. Mean inattention in the four treatments is 0.25 (*Broad*), 0.34 (*Broad with sequential elicitation*), 0.47 (*Narrow with joint elicitation*), and 0.57 (*Narrow*). These findings imply that the treatment effect is not an artifact of different elicitation methods. Harmonizing the elicitation procedure somewhat reduces the effect in the predicted direction, but prediction incentives as such have a unique effect.

### 2.D.4 Simplification

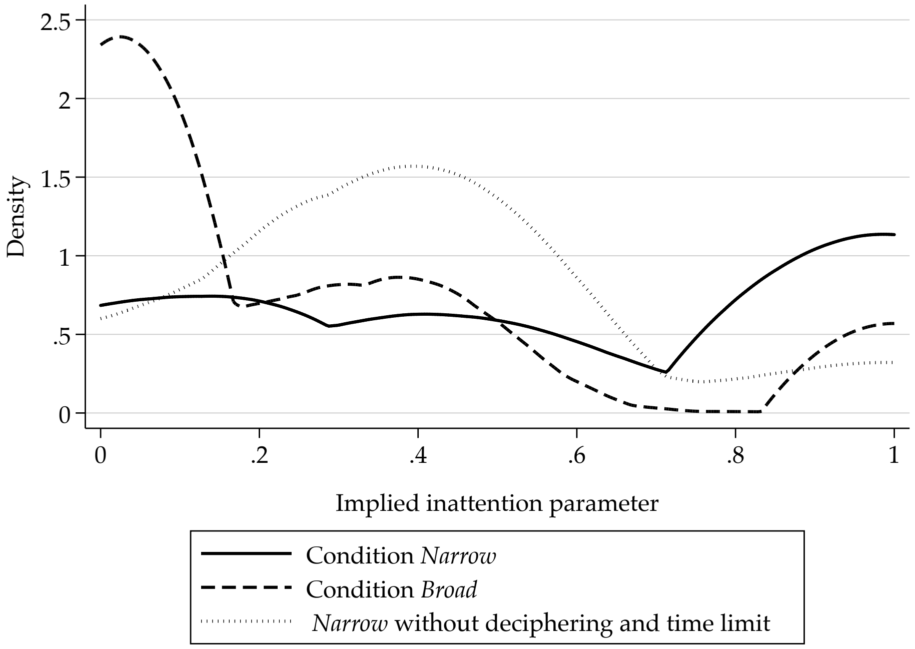


**Figure 2.D.6.** Implied inattention to $Y$ in three conditions. Based on 1,944 stated beliefs. The curves show kernel density estimates for each treatment (*Narrow* N=864, *Broad* N=864, *Simplification* N=216). A parameter of $\theta = 0$ is consistent with Bayesian updating. $\theta = 1$ means complete inattention. Epanechnikov kernel with bandwith 0.1.

To study the role of complexity in the experimental setting, an additional condition drastically simplifies the experimental procedure by removing the deciphering stage as well as all time constraints. In this treatment, subjects are paid to predict $X$ as in Narrow, but they do not have to decipher the signal and have unlimited time to indicate their guess. Effectively, they are given the distributions of $X$ and $Y$, and directly see the value of the signal. There is a statistically significant reduction in inattention relative to Narrow in this case ($p = 0.00$). At the same time, inattention remains far higher than in Broad ($p = 0.00$). Mean inattention is 0.57 in *Narrow*, 0.40 in *Simplification*, and 0.25 in *Broad*. Also, there is somewhat reduced bunching at fully inattentive and fully Bayesian beliefs. Considerable simplifications improve predictions, but do not eliminate the effect of

Narrow incentives. Figure 2.D.6 plots kernel density estimates of the distribution of inattention parameters in *Simplification* together with *Narrow* and *Broad* for reference.

### 2.D.5 The Form of Noise Neglect

**Table 2.D.5.** Online experiment on form of noise neglect

| $X$ | $Y$ | Info structure $I$ | Observed info $i$ |
|---|---|---|---|
| $\mathcal{U}\{75,76,\ldots,125\}$ | $\mathcal{U}\{90,91,\ldots,110\}$ | $\frac{X+Y}{2}$ | Individual draw |
| $\mathcal{U}\{75,76,\ldots,125\}$ | $\mathcal{U}\{-25,-24,\ldots,25\}$ | $X+Y$ | Individual draw |
| $\mathcal{N}(100,400)$ | $\mathcal{N}(50,100)$ | $\frac{X+Y}{2}$ | Individual draw |
| $\mathcal{N}(100,400)$ | $\mathcal{N}(100,100)$ | $\frac{X+Y}{2}$ | Individual draw |
| $\mathcal{N}(100,400)$ | $\mathcal{N}(100,100)$ | $X+Y$ | Individual draw |
| $\mathcal{U}\{50,51,\ldots,150\}$ | $\mathcal{U}\{50,51,\ldots,150\}$ | $X+Y$ | 145 |
| $\mathcal{U}\{75,76,\ldots,125\}$ | $\mathcal{U}\{90,91,\ldots,110\}$ | $\frac{X+Y}{2}$ | 116 |
| $\mathcal{N}(100,400)$ | $\mathcal{N}(100,100)$ | $2\cdot X+2\cdot Y$ | 412 |
| $\mathcal{N}(100,400)$ | $\mathcal{N}(100,100)$ | $2\cdot X+Y$ | 266 |
| $\mathcal{N}(100,400)$ | $\mathcal{N}(100,100)$ | $X+Y$ | 110 |

*Notes*: This table provides an overview of the ten belief tasks in the online experiment on the form of noise neglect. Note that for all normally distributed variables, the support was discretized to integers, truncated at $\mu-50$ and $\mu+50$ and then the distributions were scaled such that the they have unit probability mass.

Table 2.D.5 displays the ten tasks used in an online experiment on the form of noise neglect with 79 subjects recruited from Mturk. In five of those tasks, information values were drawn individually for each subject, while in the remaining tasks one information value was drawn jointly for all subjects to obtain higher power for a specific realization.

Figures 2.D.7 and 2.D.8 illustrate the corresponding results, which are also discussed in the main text in Section 2.3.1.7. In each of the tasks in 2.D.7, the solid reference line corresponds to Bayesian posteriors while the dashed line indicates reference beliefs under noise neglect.

Figure 2.D.8 demonstrates that the form of noise neglect is not generally in line with people using a modified distribution of $Y$. To see this, the green line indicates a corresponding threshold: all belief on the opposite side of the Bayesian posterior are not compatible with *any* possible implied distribution of $Y$ on the actual support of $Y$. At the same time, these tasks indicate that noise neglect is not easily reconciled with oversensitivity to the likelihood (or neglect of base rates), as would be in line with, e.g., diagnostic expectations (Bordalo et al., 2018). Consider for example the task displayed in the upper right corner of Figure 2.D.8, where $X\sim\mathcal{U}\{50,51,\ldots,150\}$, $Y\sim\mathcal{U}\{50,51,\ldots,150\}$, $I=X+Y$ and $i=145$. Here, an information value of 145 indicates that a relatively small value of $X$, i.e., $x < 100$, has been drawn, and the likelihood increase is greatest for values of $X$ below 100. However, people predominantly choose values above 100, close to 145.

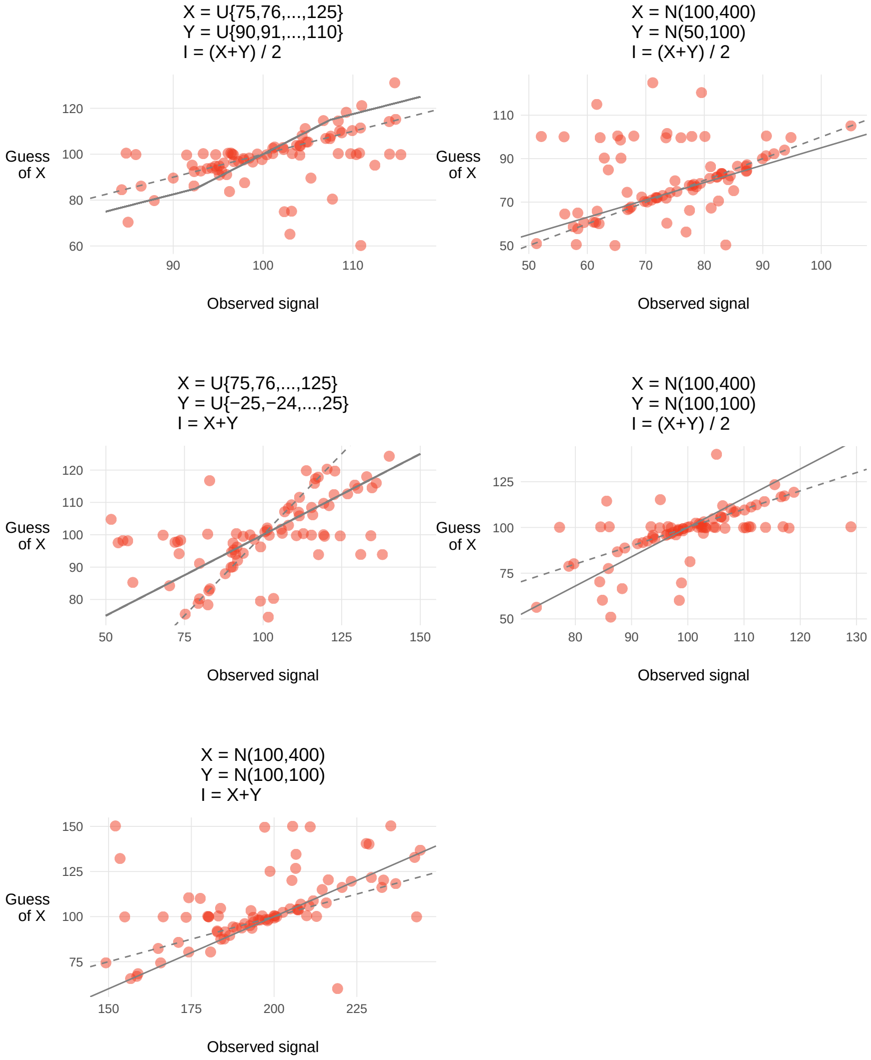


**Figure 2.D.7.** Raw beliefs in online experiment on the form of noise neglect. The solid reference line indicates the Bayesian posterior, the dashed line shows noise neglect. N = 79 in each task. Displayed are the five out of ten tasks in which the information value was individually drawn for each subject. The task order (of all ten tasks) was randomized at the individual level.

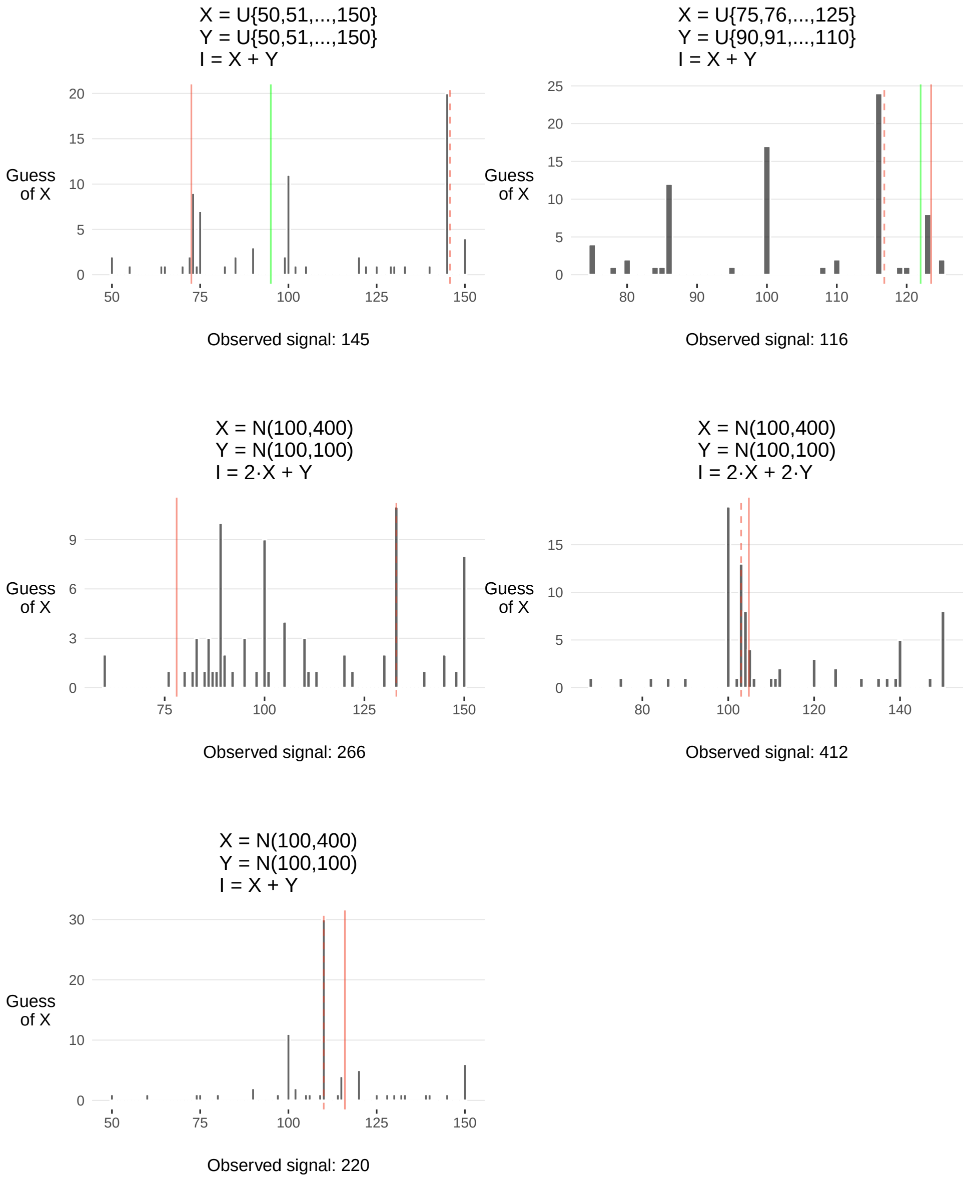


**Figure 2.D.8.** Raw beliefs in online experiment on the form of noise neglect. The solid red reference line indicates the Bayesian posterior, the dashed red line shows noise neglect. The green line indicates a threshold. All belief on the opposite side of the Bayesian posterior are not compatible with any possible implied distribution of $Y$ on the actual support of $Y$. These tasks therefore provide evidence against the idea that noise neglect is in line with people using a modified distribution of $Y$. N = 79 in each task. Displayed are the five out of ten tasks in which all subjects observed the same information value. The task order (of all ten tasks) was randomized at the individual level.

# Appendix 2.E Mechanism Treatments

## 2.E.1 Directional Bias

**Table 2.E.6.** Directional bias

| Dependent variable: | Rel. distance from noise neglect | 1 if rel. distance from noise neglect < 0.1 |
|---|---|---|
| | (1) | (2) |
| $Y \sim \mathcal{N}(-50, 100)$ | 0.088** | -0.128** |
| | (0.044) | (0.061) |
| $Y \sim \mathcal{N}(-25, 100)$ | 0.142** | -0.176** |
| | (0.058) | (0.084) |
| $Y \sim \mathcal{N}(25, 100)$ | 0.154*** | -0.211*** |
| | (0.043) | (0.064) |
| $Y \sim \mathcal{N}(50, 100)$ | 0.169*** | -0.238*** |
| | (0.056) | (0.084) |
| Absolute difference of signal from mean | 0.005*** | -0.006*** |
| | (0.001) | (0.002) |
| Constant | 0.262*** | 0.744*** |
| | (0.031) | (0.051) |
| R$^2$ | 0.15 | 0.12 |
| # Observations | 548 | 549 |

*Notes*: The dependent variables are computed based on $\frac{d_{NN}}{d_{NN}+d_B+d_{IN}}$ where $d_.$ is the distance of a stated belief $m$ to the respective benchmark belief for each of three updating modes, e.g., $d_B = \left|m - m_B\right|$ is the difference to the Bayesian belief. Hence the dependent variable in (1) is the distance of a belief to a hypothetical belief under noise neglect, *relative* to the summed distances of the elicited belief to all three updating modes. The dependent variable in (2) is a dummy for whether this relative distance is smaller than 0.1, such that a belief is plausibly classified as noise neglect. In all tasks, $X \sim \mathcal{N}(100, 100)$ and $I = X + Y$. OLS regressions. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

## 2.E.2 The Role of Effort: Manipulation of Stake Size

**Table 2.E.7.** Inattentive inference and effort

| Tasks: | Bonus round (variation of stakes) | | |
|---|---|---|---|
| Conditions: | *Narrow* and *Broad* | | *Narrow* |
| Dependent variable: | **Response time** | **Inattention $\theta$** | |
| | (1) | (2) | (3) |
| High stakes in bonus task | 19.778** | -0.017 | 0.001 |
| | (8.711) | (0.054) | (0.100) |
| 0 if *Broad*, 1 if *Narrow* | -32.444*** | 0.502*** | |
| | (7.538) | (0.084) | |
| Treatment dummy * High stakes | -3.830 | 0.019 | |
| | (11.795) | (0.113) | |
| Constant | 66.111*** | 0.101** | 0.603*** |
| | (5.671) | (0.042) | (0.073) |
| $R^2$ | 0.24 | 0.37 | 0.00 |
| # Observations | 144 | 144 | 72 |

*Notes*: OLS regressions. In the bonus round I randomly vary within each treatment whether incentives are 1 euro or 10 euros. Response time is the duration in seconds the subject spent on the belief elicitation page. Inattention is calculated as $\theta = \frac{H_B}{H_B+H_N}$, where $H_B$ and $H_N$ denote the Hellinger distance of the subjective distribution to the Bayesian posterior and the inattentive posterior distribution, respectively. Robust standard errors clustered at participant level in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

### 2.E.3 Hint Treatment

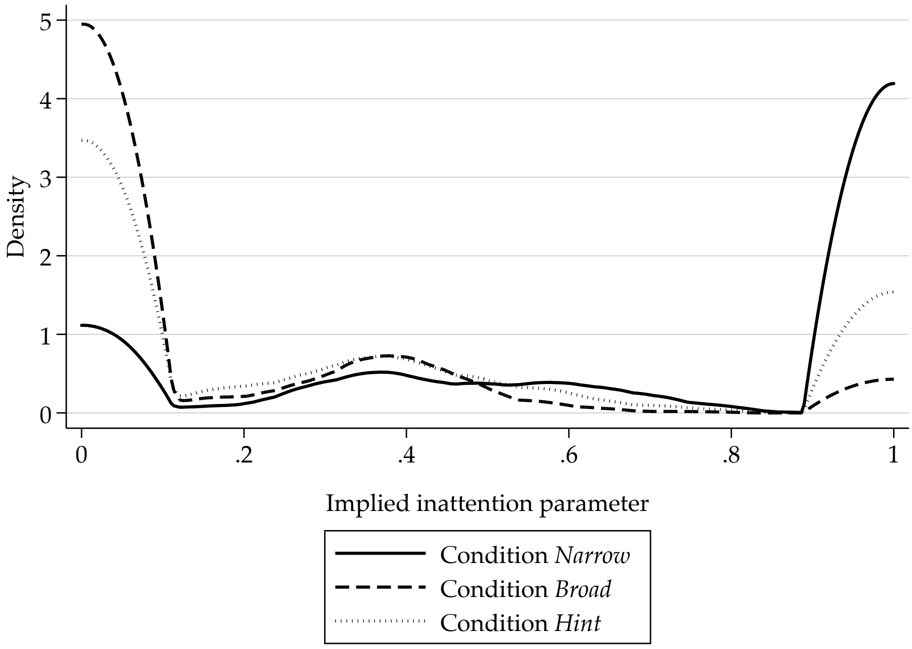


**Figure 2.E.9.** Implied inattention to $Y$ in three conditions. Based on 950 stated beliefs. The curves show kernel density estimates for each treatment (*Narrow* N=360, *Broad* N=360, *Hint* N=230). A parameter of $\theta = 0$ is consistent with Bayesian updating. $\theta = 1$ means complete inattention to noise. Epanechnikov kernel with bandwith 0.05.

### 2.E.4 Confidence Ratings

After finishing the baseline, robustness and bonus belief tasks in the laboratory, each of the tasks was again presented successively including all previously shown information as well as the subject's stated guess. In a list with fixed monetary amounts from 0 euros to 6.25 euros in steps of 0.25 euros, subjects then indicated whether they prefer to be paid out for their stated belief, or receive this fixed amount, in case this belief task would be selected to count. Single switching was enforced. Figure 2.E.10 shows that implied inattention of the belief and stated valuations for the belief are virtually unrelated.

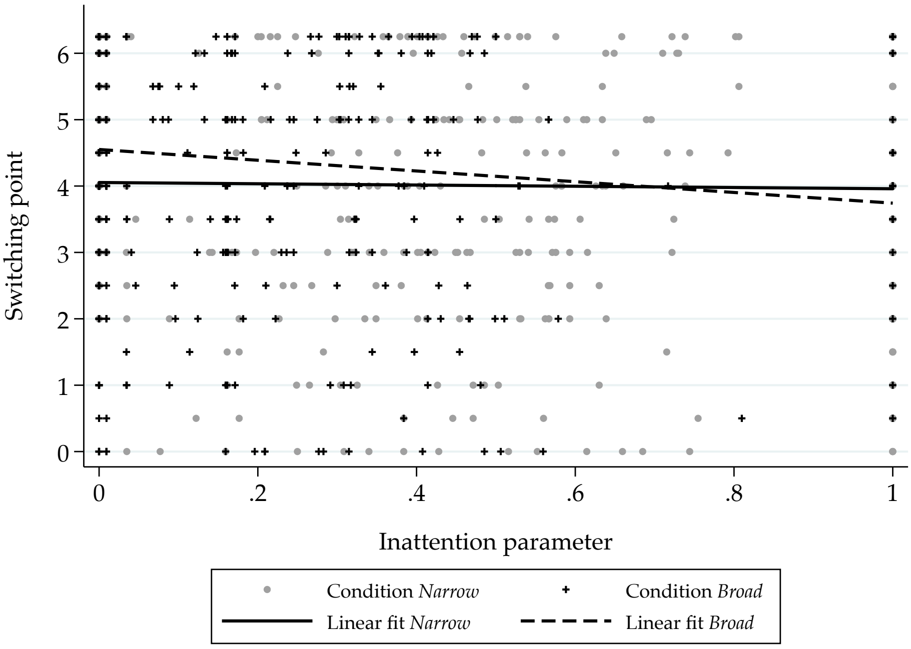


**Figure 2.E.10.** Scatterplot and linear regression fits for valuations of stated beliefs and implied inattention by condition. Based on N=360 each for condition *Narrow* and condition *Broad*.

### 2.E.5 Switch-Role Tasks

As the last part of the main baseline experiment, i.e., following the confidence tasks, subjects were (unexpectedly) presented with two additional tasks in which roles were switched with the respective other condition. The switch-role task configurations were comparable to those of the baseline tasks. Figure 2.E.11 displays group means of inattention for each of the blocks of tasks by condition. Having previously predicted $X$ and $Y$ in condition *Broad* makes subjects somewhat less inattentive than in the *Narrow* baseline, but not by much. A highly significant reverse treatment effect persists in teh switch-role tasks.

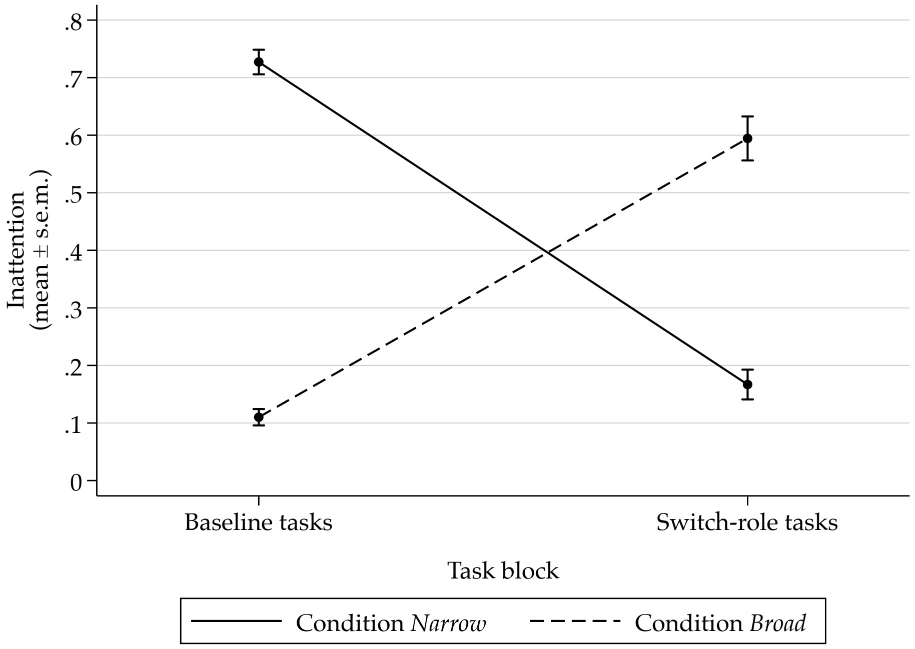


**Figure 2.E.11.** Group means of inattention by task block and condition. Based on N=360 baseline beliefs and N=144 switch-role beliefs each for condition *Narrow* and condition *Broad* .

## Appendix 2.F Learning Treatments

In the first baseline round, i.e., before receiving feedback for the first time, inattention scores do not significantly differ between the four learning treatments, as expected.

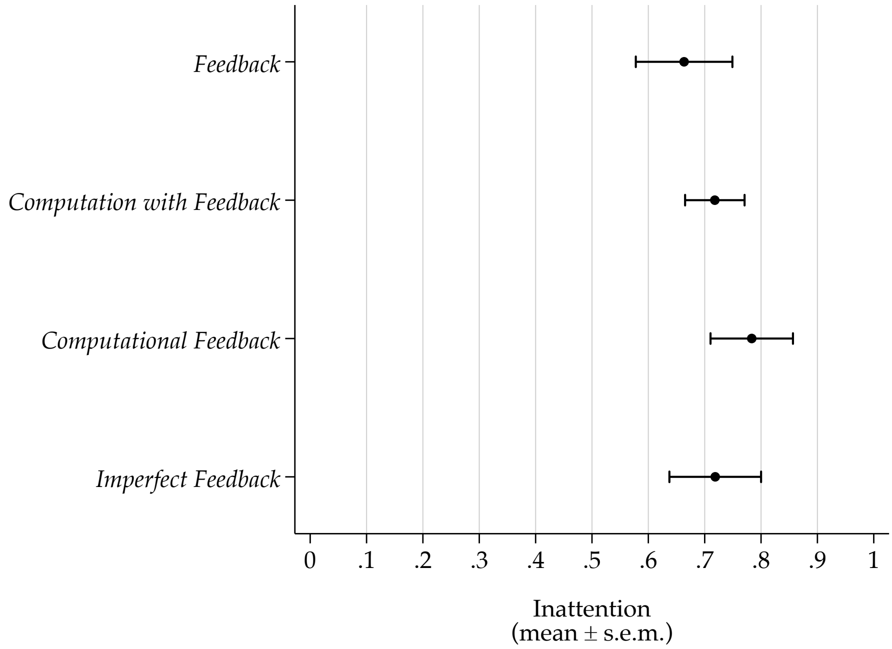


**Figure 2.F.12.** Treatment means of inattention to $Y$ in the first round. Displayed are implied inattention scores in the initial baseline round. Subjects have not previously received feedback when stating these guesses. Sample sizes are N = 48 in both *Feedback* N = 24 each in all other three conditions.

### 2.F.1 Feedback

From the initial experiments we know that the neglect of $Y$ is typically confident and occurs outside subjects' awareness. The key hypothesis motivating the feedback treatments is that people fail to reflect on steps of their solution strategy that are not available to introspection or recall, interfering with learning even in the presence of surprising feedback. Condition *Feedback* is akin to *Narrow*, but also shows the actually drawn number of $X$ after guessing it. Relative to the no-feedback benchmark (condition *Narrow*), there is marginally significant learning after receiving feedback for the first time (p=0.06, in a regression of inattention in the second round on a treatment dummy and including task-fixed effects). After having received feedback four times, mean inattention is .27 as compared to .69 in the no-feedback baseline. Despite this sizable improvement, inattention is still significantly greater than in the fifth round of the no-feedback setting with Broad incentives (mean inattention 0.10, p=0.00). Figure 2.F.13 shows a histogram of inattention parameters, and Figure 2.F.14 histograms of the raw beliefs in condition *Feedback*.

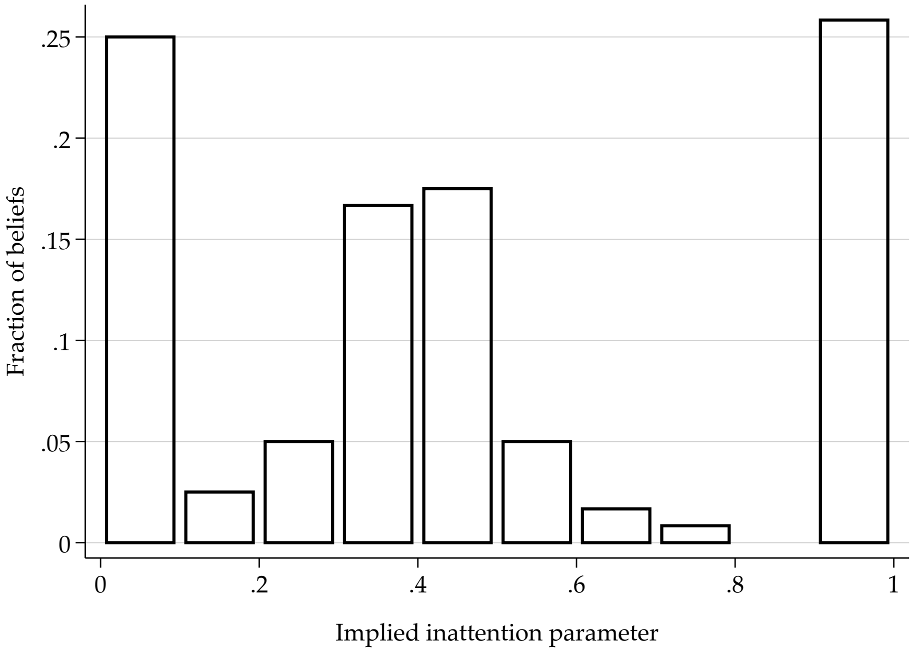


**Figure 2.F.13.** Histogram of implied inattention to $Y$ in condition *Feedback*. Based on 216 stated beliefs. A parameter of $\theta = 0$ is consistent with Bayesian updating. $\theta = 1$ means complete inattention.

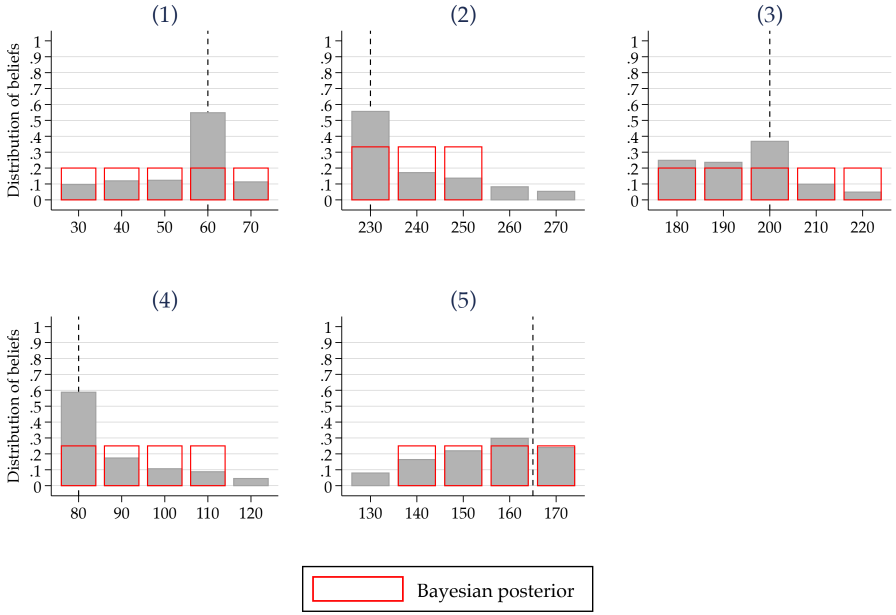


**Figure 2.F.14.** Distribution of elicited belief distributions about $X$ in each one of five baseline tasks of condition *Feedback*. N=24 for each condition in each task. The horizontal axis shows possible outcomes of $X$. The Bayesian posterior belief is provided for reference. The observed signal is indicated by the vertical dashed line. Task order was randomized.

### 2.F.2 Computation with Feedback

To directly test the hypothesis that people fail to reflect on the non-accessible elements of their solution strategy, *Computation with Feedback* provides feedback that is identical to *Feedback*, but the initial signal about $X$ and $Y$ is modified by a simple algebraic computation. This condition is identical to the anchoring treatment *Computation*, but including the feedback stage. As found in the *Computation* condition and confirmed here, the additional computation is inconsequential for the guesses about X that subjects submit (see also Figure 2.F.12). Virtually every subject correctly accounts for the computation but then tends to forget about $Y$. Presented with surprising feedback about the actually drawn number, however, subjects might now first remember the conscious part of their inference strategy, i.e., undoing the calculations. The computations provide them with "a place to hang their coat" in the sense of an obvious – albeit unlikely – source of error. This is what I find: Adding the computation virtually eliminates learning. Figure 2.F.15 shows a histogram of inattention parameters, and Figure 2.F.16 histograms of the raw beliefs in condition *Computation with Feedback*.

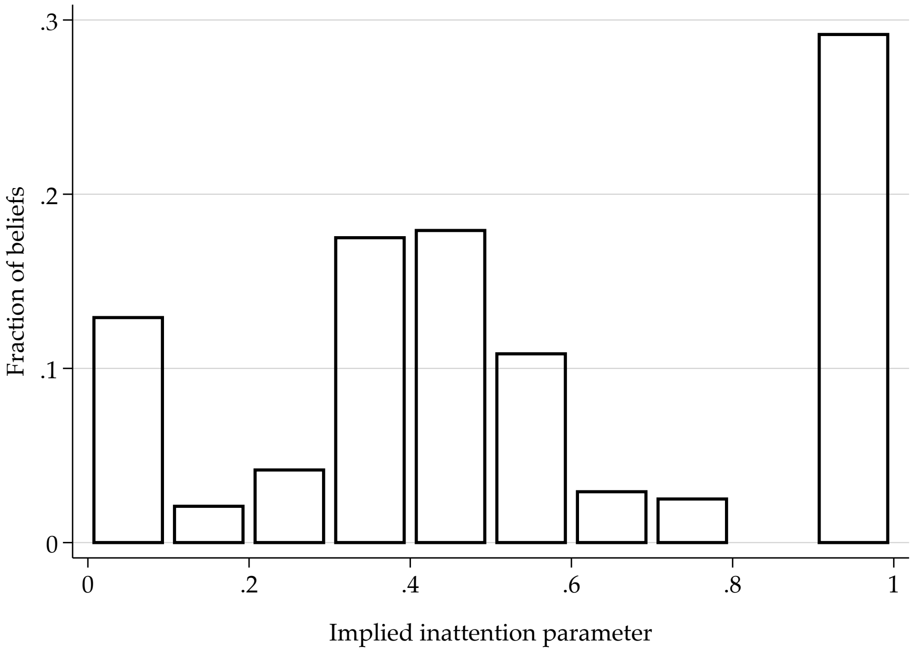


**Figure 2.F.15.** Histogram of implied inattention to $Y$ in condition *Computation with Feedback*. Based on 216 stated beliefs. A parameter of $\theta = 0$ is consistent with Bayesian updating. $\theta = 1$ means complete inattention.

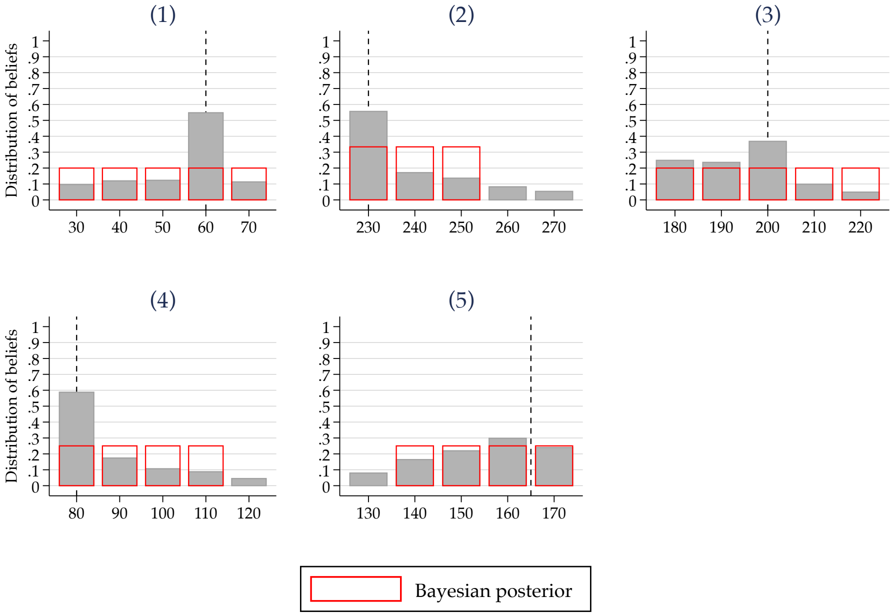


**Figure 2.F.16.** Distribution of elicited belief distributions about $X$ in each one of five baseline tasks of condition *Computation with Feedback*. N=24 for each condition in each task. The horizontal axis shows possible outcomes of $X$. The Bayesian posterior belief is provided for reference. The observed signal is indicated by the vertical dashed line. Task order was randomized.

### 2.F.3 Computational Feedback

Reduced learning when algebra is added could result from increased complexity. In condition *Computation Feedback*, therefore, subjects have narrow incentives and receive a signal on $X$ and $Y$ without additional computations, i.e., the mean or sum as before. This time however, the same computations as in *Computation with Feedback* are added at the feedback stage. That means, instead of seeing the true value of $X$, subjects see a different value on which they first perform the computations and then arrive at the true value of $X$. The results suggest it is not computational complexity of a problem per se that reduces learning form feedback. Instead, it is precisely the consciously accessible steps of reasoning performed *when doing inference* that interfere with reflecting on the role of $Y$. Figure 2.F.17 shows a histogram of inattention parameters, and Figure 2.F.18 histograms of the raw beliefs in condition *Computational Feedback*.

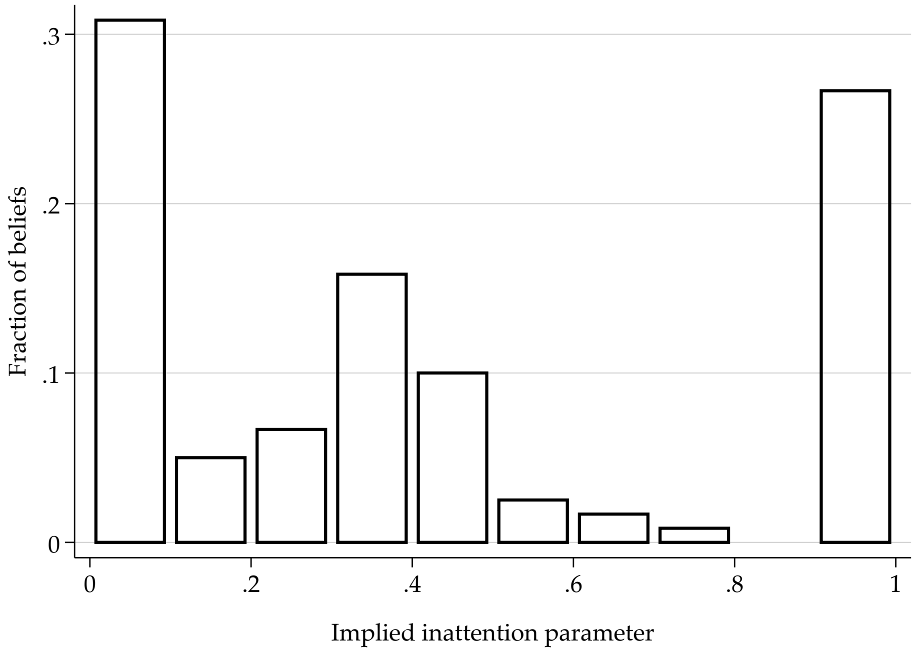


**Figure 2.F.17.** Histogram of implied inattention to $Y$ in condition *Computational Feedback*. Based on 216 stated beliefs. A parameter of $\theta = 0$ is consistent with Bayesian updating. $\theta = 1$ means complete inattention.

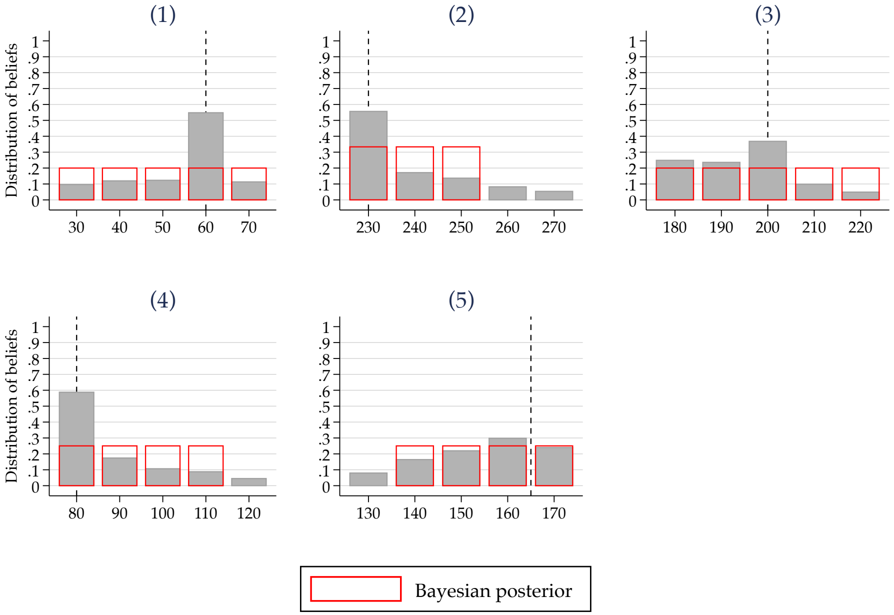


**Figure 2.F.18.** Distribution of elicited belief distributions about $X$ in each one of five baseline tasks of condition *Computational Feedback*. N=24 for each condition in each task. The horizontal axis shows possible outcomes of $X$. The Bayesian posterior belief is provided for reference. The observed signal is indicated by the vertical dashed line. Task order was randomized.

### 2.F.4 Imperfect Feedback

Learning in practice is often based on imprecise signals. The possibility that observed feedback is not exactly right might provide another obvious way for subjects to explain a conflict between their stated belief and received feedback, reducing learning. In an additional treatment, feedback about the true $X$ was only correct with 80% probability, and the remaining 20% subjects would see a value of $X$ which is not the true one. Pooling beliefs following the first four rounds of feedback, there is only a small and marginally significant positive effect of receiving this feedback on inattention relative to receiving no feedback at all ($p = 0.09$). As predicted, simple solutions for why beliefs conflict with the feedback compromise the ability to reflect on the role of Y. Figure 2.F.19 shows a histogram of inattention parameters, and Figure 2.F.20 histograms of the raw beliefs in condition *Imperfect Feedback*.

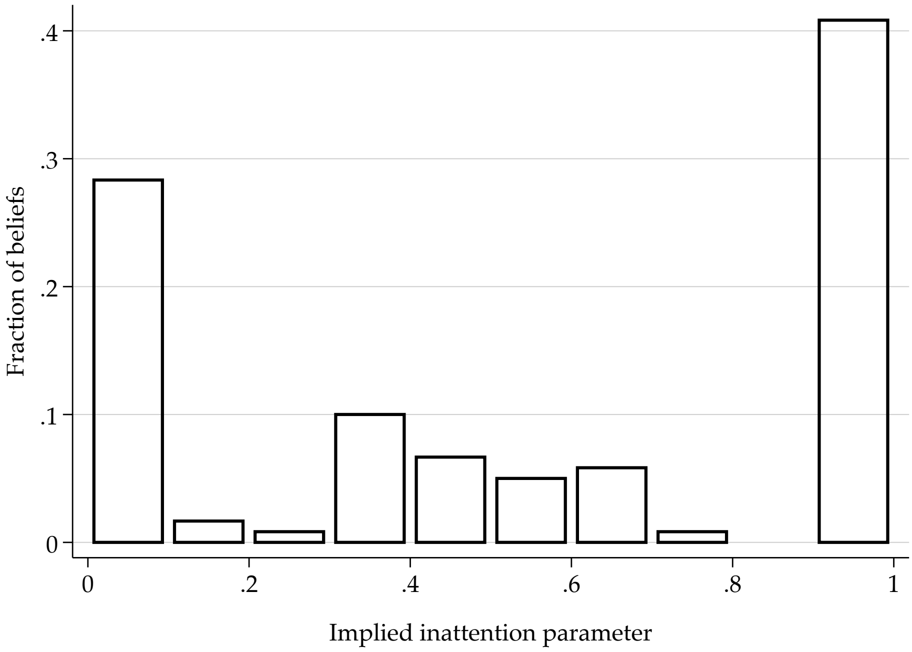


**Figure 2.F.19.** Histogram of implied inattention to $Y$ in condition *Imperfect Feedback*. Based on 216 stated beliefs. A parameter of $\theta = 0$ is consistent with Bayesian updating. $\theta = 1$ means complete inattention.

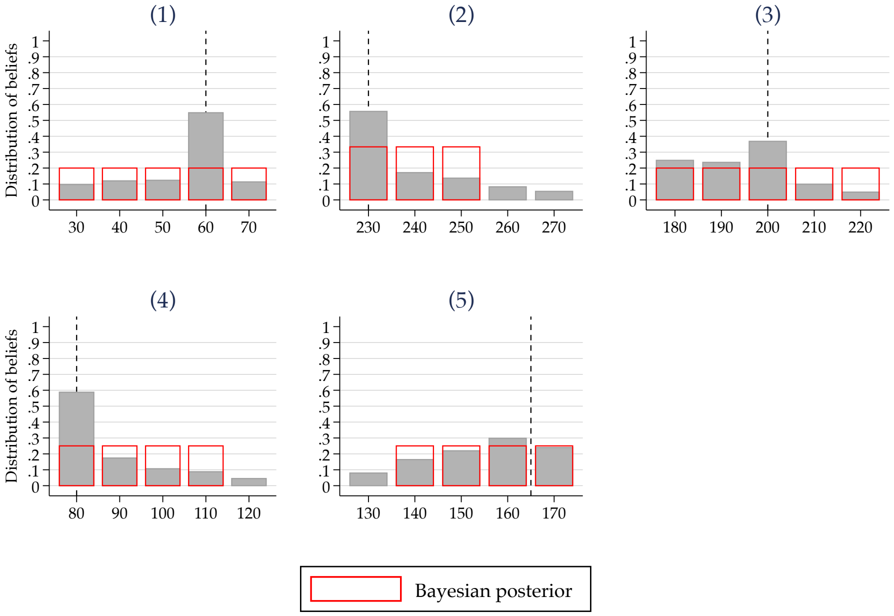


**Figure 2.F.20.** Distribution of elicited belief distributions about $X$ in each one of five baseline tasks of condition *Imperfect Feedback*. N=24 for each condition in each task. The horizontal axis shows possible outcomes of $X$. The Bayesian posterior belief is provided for reference. The observed signal is indicated by the vertical dashed line. Task order was randomized.

## Appendix 2.G Experimental Instructions

### 2.G.1 Main Instructions in *Narrow* and *Broad*

*All instructions were computerized. Translated from German into English.*

Welcome. For your participation you will receive a fixed payment of 10.00 € , which will be paid to you in cash at the end. In this study you will take decisions on the computer. Depending on how you decide you can earn additional money. **During the study it is not allowed to communicate with other participants. Note also that the curtain of your cubicle must be closed during the entire study.** Please turn off your mobile phone now, so that other participants will not be disturbed. Please only use the designated functions on the computer and make your entries using the keyboard and the mouse. If you have questions, please make a hand signal. Your question will be answered at your seat. To proceed click "Next".

**Your Task**
You will successively receive 9 different guessing tasks. The guessing tasks are about guessing numbers that are randomly drawn. The better your guess, the more money you can earn. In each guessing task there is a random number X. The computer randomly picks X from a range of possible numbers. You will receive an encrypted hint about which number was actually drawn, and you can then indicate your guess about X. There are 9 rounds in total. In each round you receive a new guessing task. That means, in each round the computer again determines a number X independently of the other rounds. Your payoff depends on how precisely you guess, that means how accurate your guess is. At the end of the study, one of the 9 rounds is picked at random and you will be paid according to the precision of your guess in that round.

**The Guessing Tasks**
Example. Imagine there are exactly 3 balls. These 3 balls have the following numbers on them: 10, 20, 30. In this example, the number X is determined as follows: The computer randomly draws one of these three balls. Each ball is drawn with equal probability. It is equally likely that the "10" will be drawn, that the "20" will be drawn, or that the "30" will be drawn. The number X is then the number of the ball that was randomly drawn by the computer. However, you will not be told which number X was drawn. Instead you receive an additional hint. You can look at this hint, before you guess the number X. Please note:

- For each guessing you will be informed about which numbers can be drawn. In different guessing tasks, different numbers can be drawn. Sometimes the numbers repeatedly occur across rounds. However, the draws in these rounds are completely independent of one another.

- The additional hint can give you different types of information in different rounds. In each round you will learn anew, what the additional hint means. Therefore you should pay attention in every new guessing task to which information the hint indicates.

Your guess. You can state your guess by allocating 100 percentage points to the different numbers. The *more certain* you are, that a particular number was drawn, the *more* points you should allocated to this numbers. Similarly, the more certain you are, that a particular was *not* drawn, the *fewer* points you should allocate to this numbers. The sum of your allocated points must be exactly 100. In the example above, if after receiving the additional note you are, for example, sure that X = 30, then you should allocate 100 points to the number 30, and 0 points to both the numbers 10 and 20. In the example above, if after receiving the additional note you are, for example, sure that X = 20, then you should allocate 100 points to the number 20, and 0 points to both the numbers 10 and 30. In the example above, if after receiving the additional note you think, for example, that the number 30 have definitely not been drawn, but the 10 and 20 have been drawn with equal probability, then you should allocate 50 points each to the number 10 and 20, and 0 points to the number 30. You can arbitrarily allocate the points. However you can only allocate full points, that means for example that you cannot allocate half points. For instance, you could allocate 21 points to number 30, 47 points to number 20, and 32 points to number 10. **The more points you assign to the number that was actually drawn, the more money you can earn. Similarly, the fewer points you allocate to those numbers, that are not equal to X, the more money you can earn.** The calculation of your payoff will be explained in greater detail in the following section.

**Your payment**

In addition to your show up fee you will be paid based on how precisely you guessed. To this end one of the 9 rounds will randomly be picked and you will be paid according to the precision of your guess in that round. This means for you that *each one* of your guesses is potentially relevant for your payment and accordingly you should carefully think through every guess. You can either earn and additional 10 € or 0 € from your guessing task. While the following explanation might look difficult, the basic principle is very simple: **the better your guess, that is the more percentage points your guess assigns to the actually drawn number and the fewer percentage points it allocates to every wrong number, the more likely it is that you receive the 10€ .** Concretely this means the following: In expectation you will earn most money if you allocate your points according to how probable you find it that the respective numbers was drawn (with 1 point = 1 percent). If you have understood this, it is not necessary for the maximization of your earnings to read the following section on the details of the calculation of your additional payment. You can then directly click on "Next."

For your information: Details on the calculation of your additional earnings. For working on the guessing tasks it is *not necessary* that you read and fully understand the following section on the calculation of your payoff. you can also skip this part. After you have stated your guess, the computer will randomly draw another number $k$j This number is between 0 and 20,000. (More precisely, this numbers is drawn from a discrete uniform distribution on the interval from 0 to 20,000.) You will then receive the 10 € if the sum $S$ is smaller or equal to $k$. $S$ is the sum of the following elements:

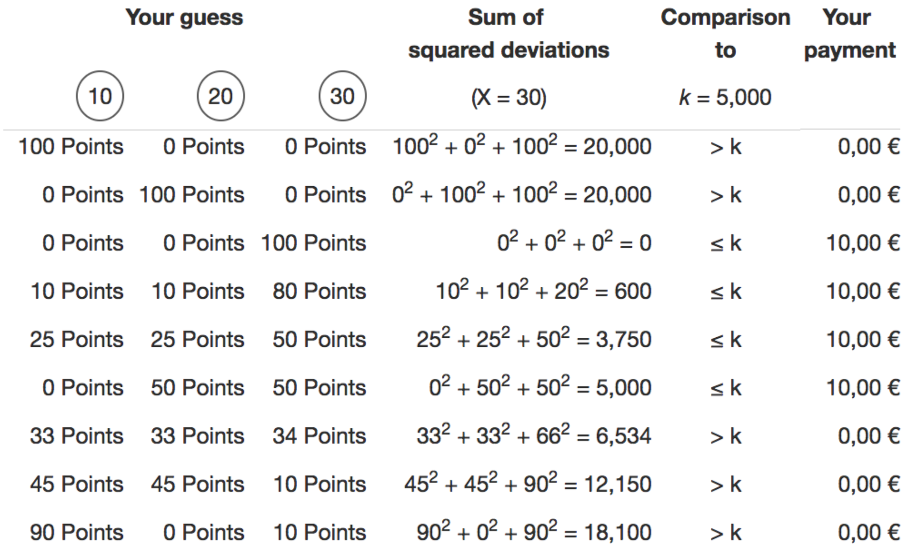
- The squared deviation between the number of points that you allocated to the actually drawn numbers X, and 100 points.
- For *each* possible number, that has not been drawn (i.e., every other number than X): The squared deviation between 0 points and the number of points that you allocated to this numbers.

An exact mathematical formula of the sum $S$ is displayed in the footnote.[42] If the sum $S$ is bigger than $k$ you will receive 0 € . Accordingly, the payoff rule is as follows:

Payment = 10.00 € , if $S \leq k$
Payment = 0.00 € , if $S > k$

This means the following: If the sum of the squared deviations exceeds a particular value $k$, you will receive 0 € . If, however, the sum of the squared deviations is smaller than $k$, you will receive 10 € in addition. You can notice here that it should be your goal a) to keep the difference between the points allocated to X and 100 points as low as possible, that is to allocate as many points as possible to X, and b) to allocate as few points as possible to ever other number than X. An *example*: Let us assume that the computer has randomly drawn the number X = 30, while the numbers 10, 20 and 30 could have been drawn with equal probability. Also the number $k = 5,000$ For the following guesses you would receive the indicated payments.

[42] *Footnote text:* Exact mathematical formulation: There are $N$ possible number from which X is drawn. In the example, $N = 3$. The number of points that you allocate to the $i$th of the $N$ numbers is $p_i$. The indicator function $\mathbb{1}_i$ takes the value 1, if X is the $i$th number, and 0 otherwise. The sum $S$ is calculated as follows: $S = \sum_{i=1}^{N}(\mathbb{1}_i - p_i)^2$. The expected payoff amount is maximized by indicating the probability distribution of the numbers after receiving the additional hint.



| Your guess | | | Sum of squared deviations | Comparison to | Your payment |
|---|---|---|---|---|---|
| 10 | 20 | 30 | (X = 30) | $k$ = 5,000 | |
| 100 Points | 0 Points | 0 Points | $100^2 + 0^2 + 100^2 = 20,000$ | $> k$ | 0,00 € |
| 0 Points | 100 Points | 0 Points | $0^2 + 100^2 + 100^2 = 20,000$ | $> k$ | 0,00 € |
| 0 Points | 0 Points | 100 Points | $0^2 + 0^2 + 0^2 = 0$ | $\leq k$ | 10,00 € |
| 10 Points | 10 Points | 80 Points | $10^2 + 10^2 + 20^2 = 600$ | $\leq k$ | 10,00 € |
| 25 Points | 25 Points | 50 Points | $25^2 + 25^2 + 50^2 = 3,750$ | $\leq k$ | 10,00 € |
| 0 Points | 50 Points | 50 Points | $0^2 + 50^2 + 50^2 = 5,000$ | $\leq k$ | 10,00 € |
| 33 Points | 33 Points | 34 Points | $33^2 + 33^2 + 66^2 = 6,534$ | $> k$ | 0,00 € |
| 45 Points | 45 Points | 10 Points | $45^2 + 45^2 + 90^2 = 12,150$ | $> k$ | 0,00 € |
| 90 Points | 0 Points | 10 Points | $90^2 + 0^2 + 90^2 = 18,100$ | $> k$ | 0,00 € |

In particular this means the following: If you allocate all 100 points to the right number X, you will received the 10 € **in any case**. However, you will also receive 10 € in many cases in which you allocate less than 100 points to X. The more points you allocate to the right number X, the more likely it is, that you receive the 10 € . **In expectation, you will earn the most money if you allocate the points according to how probable you think it is that the respective number was drawn.** Please note:

- It is not necessary, to allocate 100 points to the number that you think is most likely. As you can see in the examples of the table, you can also win 10 € if you have allocated less than 100 points to the right number X. Your earnings depend on the randomly drawn number $k$.

- Your guess in *one* randomly picked round will be paid. The guessing task that is payoff relevant for you is determined by the computer at the end of the study. Therefore you should indicate your best guess in each guessing task, independent of all other guessing tasks.

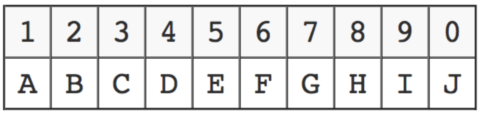
**Summary**

**In each round it is your task to state a guess about the number that was randomly drawn by the computer.** Before this, you will get a computer-generated, **encrypted hint**. For each guessing task you will see this additional hint and you can subsequently indicate your guess. Which hint you will receive, and how this hint is encrypted will be explained in the following. For the deciphering of the hint and your subsequent guess there is a time limit. You will previously be informed about how much time you have. The remaining time will be displayed while working on the tasks.

**Encryption of Hints**
You receive additional hints that have been encrypted by an **encryption device**. The encryption device transforms each hint (a number) into a letter code. You first need to decrypt the letter code back into a number in order to use the hint.

Decryption of the additional hint. When you get an encrypted sequence of letters as hint, you can decipher this hint by following these steps:

a. Transform the sequence of letters into a number using the code table.



| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
|---|---|---|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G | H | I | J |

b. Add 20 to the number

Before every guessing task you will receive an encrypted hint that you can decipher before stating your own guess. Whenever you receive a hint, you will see the code table as well as the decryption instructions. **That means you don not have to remember the decryption procedure.** You will soon get the opportunity to practice the decryption on an example hint.

**Control Questions**
Please notify one of the experiments now if you have questions about the instructions so far. If there is something that is unclear to you, please re-read the respective information carefully. You can return to the previous pages by clicking "Back". If you click on "Proceed to control questions", you will receive several control questions, which ensure your understanding of the instructions. You will not get paid for the control questions. However, you have to correctly answer all control questions to proceed to the guessing tasks. After you have correctly answered all control questions, you will be presented with the first guessing task.

## 2.G.2 Control Questions

**Control Question 1 of 9**
What is your main task in this study?

- There are several number from which X can be drawn. I need to add these numbers up to a sum.
- I guess the drawn number X.

**Control Question 2 of 9**

The numbers from which X is drawn vary across rounds. Sometimes the numbers occur in different rounds. For example, it could be that in two different rounds, the number X is randomly drawn from the number “10”, “20”, and “30”. Please evaluate the following statement: “In both round, each of the 3 numbers is drawn with equal probability.”

- **Wrong.** If, for example, the “10” was drawn in the first round, it is more probable that “10” will not be drawn in the next round.
- **Correct.** Both rounds are completely independent. The draw in the first round has no influence on which number is drawn in the second round.

**Control Question 3 of 9**

In guessing X, how can you make most money?

- By allocating the points to the numbers as precisely as possible based on how certain I am, that the respective number is X.
- By varying my guess and allocating by instinct sometimes more points to high numbers and sometimes more points to low numbers.

**Control Question 4 of 9**

After you have read the description of the guessing task and received the additional hint, you think that the number “20” is the most likely drawn number among the numbers. However you are not certain that it is the “20”. Assess the following statement: “To maximize my payoff I have to put all 100 points on the number “10”.”

- **Correct.** It is only this way that I can earn the 10 euros.
- **Wrong.** While I should put more points on the “20” than on all other numbers, I should not putt all points on the “20” , because I am not certain. If for example i am 60% sure that X = 20, I should put exactly 60 points on the “20”, and allocate the remaining 40 points to the other numbers. This way it is most probably that I earn the 10 euros.

**Control Question 5 of 9**

Which of your guess is payoff-relevant?

- Every guess is paid out.
- No guess is paid out.

- A randomly picked guess is paid out.

**Control Question 6 of 9**

Imagine the number X is drawn with equal probability from the following four numbers: 50, 60, 70, 80. You have no additional information. Please indicate how in this case you should allocate the 100 points to the four numbers such that you make winning the 10 euros as likely as possible. Start by picking a number in the selection box to the left and assign a number of percentage points in the input field to the right. Use further input rows if you want to assign percentage points to other numbers.

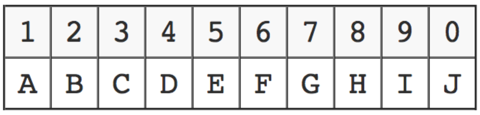
**Control Question 7 of 9**

As before the number X is drawn with equal probability from the following four numbers: 50, 60, 70, 80. Please imagine now that after deciphering the hint you are certain that the "70" was drawn. Please indicate how in this case you should allocate the 100 points to the four numbers such that you make winning the 10 euros as likely as possible. If you want to allocate 0 percentage points to a number then you do not have to enter this into an extra row, but you can simply skip this number (0 points will automatically be allocated).

**Control Question 8 of 9**

Imagine you receive the hint: AJ. Please decipher the hint and enter your result below.

a. Transform the sequence of letters into a number using the code table.



| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
|---|---|---|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G | H | I | J |

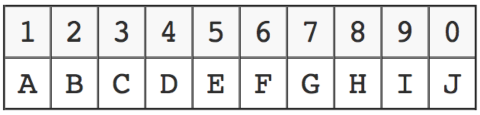
b. Add 20 to the number

**The decrypted hint reads:**

**Control Question 9 of 9**

Imagine now you receive the hint: ACJ. Please decipher the hint and enter your result below.

a. Transform the sequence of letters into a number using the code table.



| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
|---|---|---|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G | H | I | J |

b. Add 20 to the number

**The decrypted hint reads:**

### 2.G.3 Task Instructions

Next to X another number was drawn by the computer, Y. Whether a participant has to guess Y as well was randomly determined at the beginning of the study and has no impact on the size of possible earnings.

[ *Treatment Narrow:* **To you applies the following**: You indicate a guess **only about X** and will be paid for your guess of X as described. ]

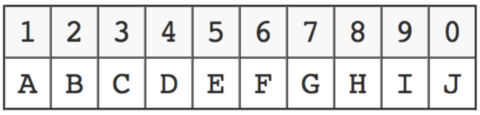
[ *Treatment Broad:* **To you applies the following**: You guess **both drawn numbers**, X and Y. One of the numbers will later be picked and you will be paid for your guess of this number as described. ]

[ **The following description varies by task** ]
**X** was randomly drawn from the following 5 numbers between 80 and 120, where each number was equally likely: 80, 90, 100, 110, 120.
**Y** was randomly drawn from the following 7 numbers between -30 and 30, were each number was equally likely: -30, -20, -10, 0, 10, 20, 30.
X and Y were drawn independently.

[ *Treatment Broad:* You will guess X and Y *simultaneously*, that means in each entry row you have to pick both a number for X and a number for Y and indicate a percentage alongside, which is your guess that these two numbers were drawn together. ]
When you click “Next”, you will first receive your *additional hint* on the following page. You have 5 minutes time to decipher the hint. Then you have another 5 minutes of time to indicate you guess. The remaining time will be displayed on the upper right corner of the pages.

**Your Additional Hint**
Your additional hint for the guess of X [ X and Y ] is: FJ. The completely decrypted hint indicates the sum of the 2 drawn numbers, i.e., X + Y.
Decryption Instructions.

a. Transform the sequence of letters into a number using the code table.

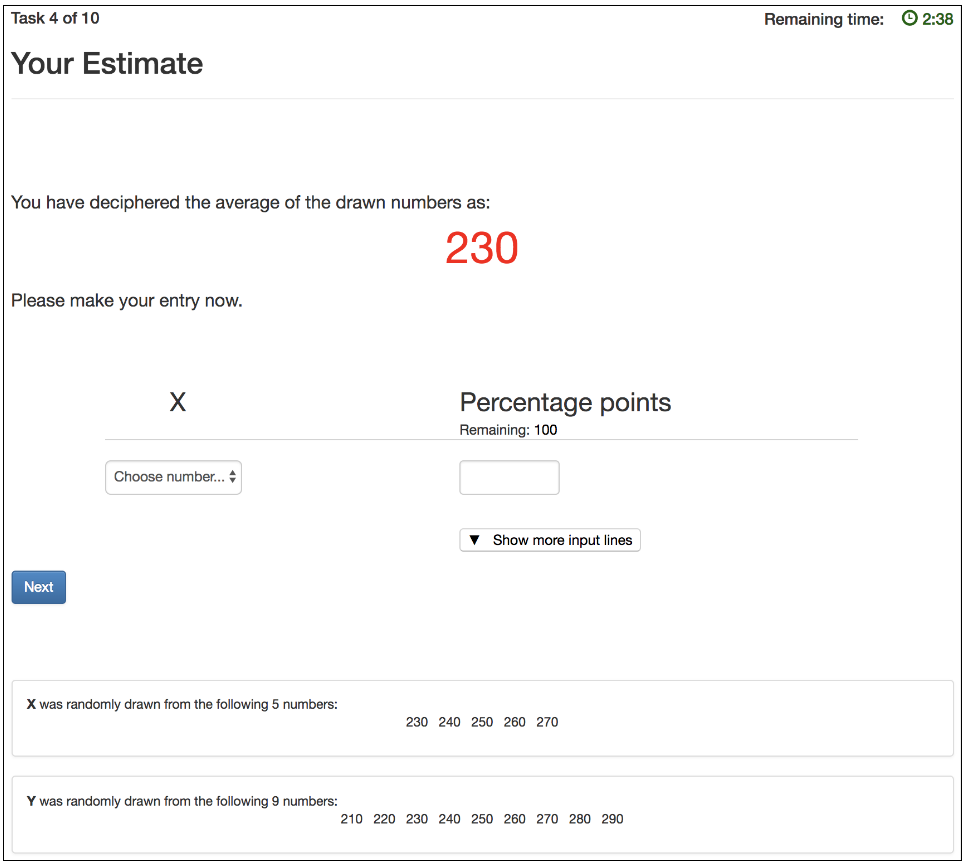


| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
| A | B | C | D | E | F | G | H | I | J |

b. Add 20 to the number

[ *Calculator provided.* ] On the next page you will see the entry fields for your guess. You can now enter your decrypted additional hint below, then it will be displayed again on the next page.**Your deciphered additional hint reads: ...** Once you click on "Next" you have 5 minutes time to indicate your guess.

### 2.G.4 Decision Screens: Laboratory, Baseline Study



Task 4 of 10

Remaining time: 2:38

Your Estimate

You have deciphered the average of the drawn numbers as:

230

Please make your entry now.

X

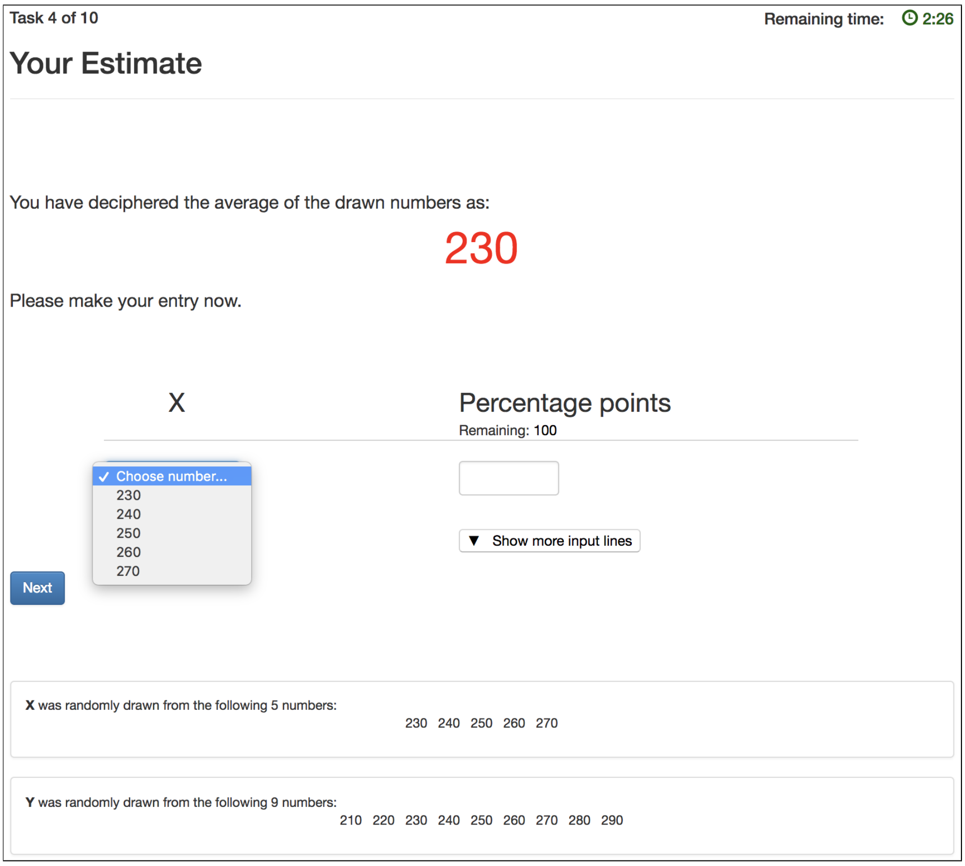
Percentage points

Remaining: 100

**X** was randomly drawn from the following 5 numbers:
230 240 250 260 270

**Y** was randomly drawn from the following 9 numbers:
210 220 230 240 250 260 270 280 290

**Figure 2.G.21.** Exemplary decision screen in condition *Narrow* (translated from German). The number 230 indicates the average of X and Y. Subjects state their belief by indicating a full posterior distribution for X. They have to select values for X using the dropdown menu and enter a number of percentage points in the fields on the right. They can use arbitrary many entry lines. The current sum of percentage points is indicated and has to equal exactly 100 before one can proceed. On the bottom of the screen, the distributions of X and Y are indicated as a reminder.



Task 4 of 10

Remaining time: 2:26

Your Estimate

You have deciphered the average of the drawn numbers as:

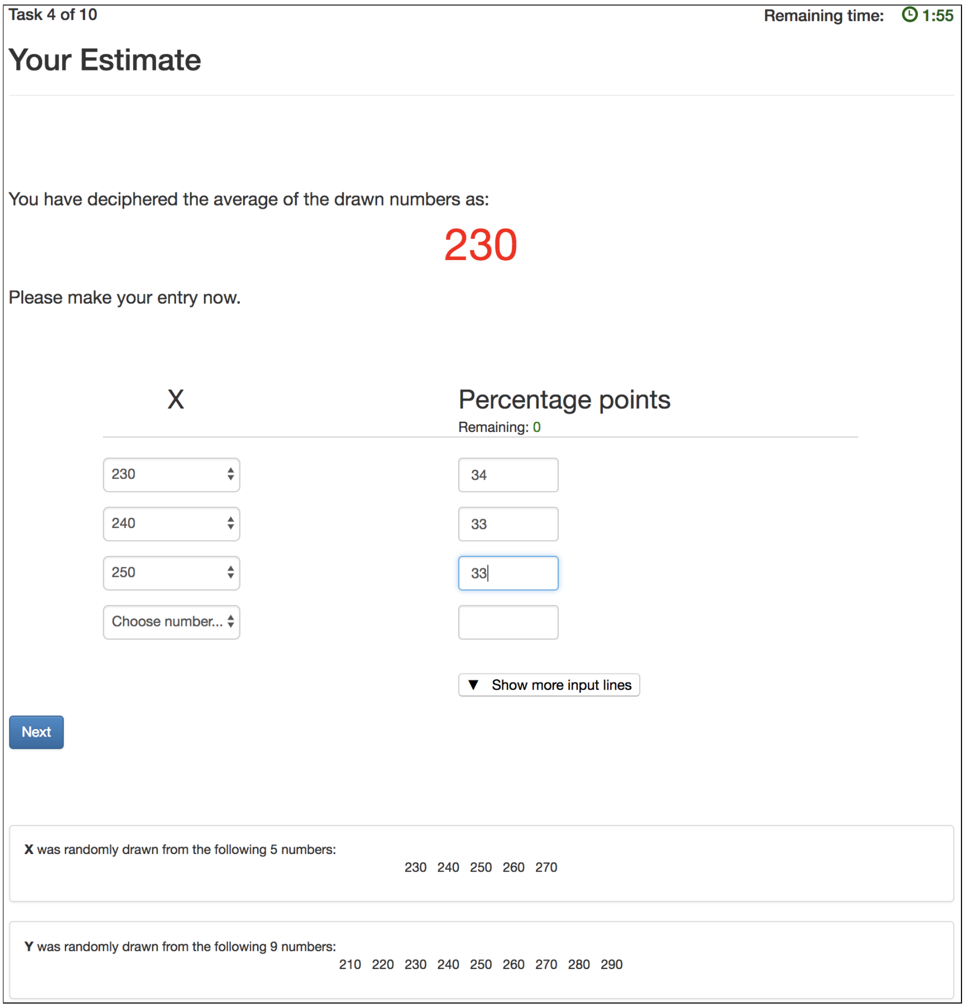
230

Please make your entry now.

**X** was randomly drawn from the following 5 numbers:
230 240 250 260 270

**Y** was randomly drawn from the following 9 numbers:
210 220 230 240 250 260 270 280 290

**Figure 2.G.22.** Exemplary decision screen in condition *Narrow* (translated from German). Use of dropdown menu.



Task 4 of 10

Remaining time: 1:55

Your Estimate

You have deciphered the average of the drawn numbers as:

230

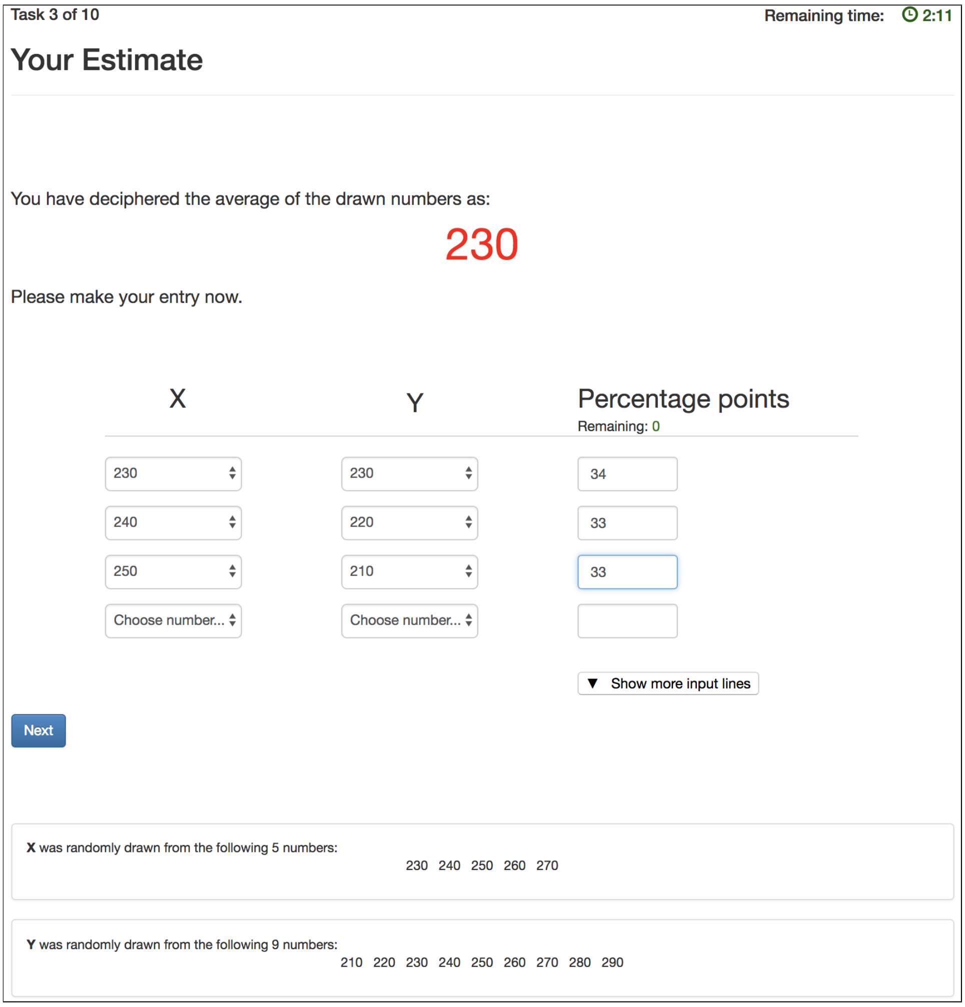
Please make your entry now.

| X | Percentage points<br>Remaining: 0 |
|---|---|
| 230 | 34 |
| 240 | 33 |
| 250 | 33 |
| Choose number... | |

**X** was randomly drawn from the following 5 numbers:
230 240 250 260 270

**Y** was randomly drawn from the following 9 numbers:
210 220 230 240 250 260 270 280 290

**Figure 2.G.23.** Exemplary decision screen in condition *Narrow* (translated from German). Use of multiple entry rows to indicate the full subjective distribution.



Task 3 of 10

Remaining time: 2:11

Your Estimate

You have deciphered the average of the drawn numbers as:

230

Please make your entry now.

| X | Y | Percentage points<br>Remaining: 0 |
|---|---|---|
| 230 | 230 | 34 |
| 240 | 220 | 33 |
| 250 | 210 | 33 |
| Choose number... | Choose number... | |

**X** was randomly drawn from the following 5 numbers:
230 240 250 260 270

**Y** was randomly drawn from the following 9 numbers:
210 220 230 240 250 260 270 280 290

**Figure 2.G.24.** Exemplary decision screen in condition *Broad* (translated from German). The number 230 indicates the average of X and Y. Subjects state their belief by indicating a full posterior distribution for X and Y. They have to select values using the dropdown menu and enter a number of percentage points in the fields on the right. They can use arbitrary many entry lines. The current sum of percentage points is indicated and has to equal exactly 100 before one can proceed. On the bottom of the screen, the distributions of X and Y are indicated as a reminder.

### 2.G.5 Decision Screens: Online, Baseline Study

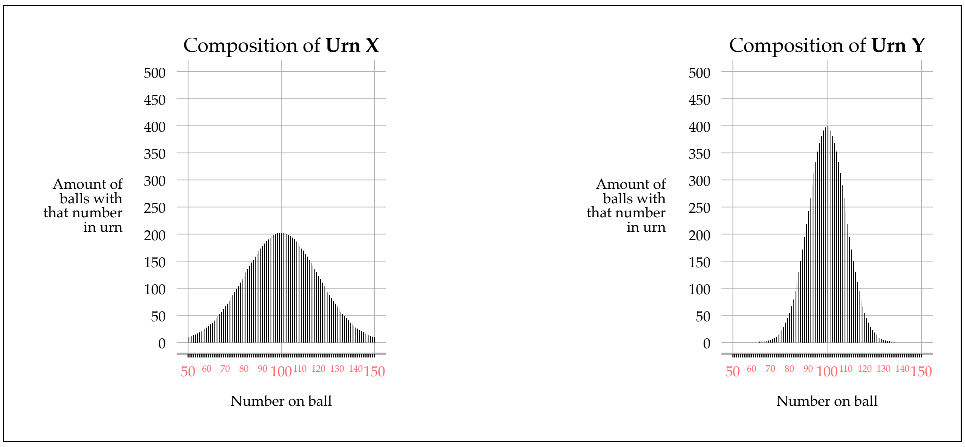


**Figure 2.G.25.** Exemplary task in online experiment. One number each would be drawn from Urn X and Urn Y. In this example, subjects would also learn the average of the drawn numbers. Subjects state their belief by indicating the mean of their posterior belief about the number drawn from Urn X.

# 3

# Heterogeneity of Loss Aversion and Expectations-Based Reference Points

*Joint with Lorenz Goette, Alexandre Kellogg and Charles Sprenger*

## 3.1 Introduction

Models of reference-dependent preferences are regarded as a major advance in behavioral economics, rationalizing a range of observations at odds with the canonical model of expected utility over final wealth (Camerer et al., 1997; Kahneman et al., 1990; Odean, 1998; Rabin, 2000). Critical to such applications is the formulation of the reference point around which gains and losses are encoded. Given flexibility in this reference point, the model is endowed with a powerful degree of freedom, limiting its ability to be coherently applied across contexts.

Several prominent views on the source of reference points have been articulated, from the status quo formulation of Kahneman and Tversky (1979), to adaptive formulations based on "customary" wealth or endowments (Baucells and Sarin, 2007; DellaVigna et al., 2017; Lattin and Bucklin, 1989; Markowitz, 1952; Putler, 1992; Shefrin and Statman, 1985; Wathieu, 1997). More recently, a literature has examined forward-looking characterizations of the reference point, based on rational expectations of potential outcomes (Kőszegi and Rabin, 2006, 2007) (henceforth KR).[1] Expectations-based models have the promise to

[1] Our analysis will focus on the formulations of KR. An earlier literature also provided formulations of reference dependence grounded in rational expectations, but without the equilibrium concepts we analyze (Bell, 1985; Loomes and Sugden, 1986).

be readily and broadly applicable, closing the degree of freedom with a foundation to which economic tools are already adapted.

Our study begins with an observation: even with reference points grounded in rational expectations, heterogeneity in the model's key behavioral parameter, loss aversion, can confound inference. In the case of KR preferences, heterogeneity in loss aversion breaks down comparative static tests that have been used to test the model (see Section 3.2 for detail). Given both the number of contradictory empirical studies for expectations-based models (see, e.g., Abeler et al., 2011; Ericson and Fuster, 2011; Gneezy et al., 2017; Goette et al., 2016; Heffetz and List, 2014), and the recognized variation in individual measures of loss aversion (see, e.g., Erev et al., 2008; Harinck et al., 2007; Nicolau, 2012; Sokol-Hessner et al., 2009; Sprenger, 2015), heterogeneity is potentially an issue of first order importance.

We design an exchange experiment with the objective of examining the force of expectations-based models while recognizing heterogeneity in loss aversion. Our central treatment plausibly alters expectations of exchange for a given object, and we experimentally control the prior experiences of agents. The manipulation of experience allows us to collect, and validate, a measure of loss aversion for an alternate object, providing an assessment of heterogeneity. Our objective is achieved through between-subjects variation and a purposeful parsimony of choices, with a single binary decision per subject.

We implement our study in a sample of 607 subjects. In a first stage, subjects are randomly endowed with one of two objects. Though no choices are made, subjects are asked to provide ratings of both objects, and their initial mood is measured using standard psychological scales. Subsequently, based on a randomization device, the endowed object is taken away for half of the subjects and replaced with the alternative object, after which mood is measured again. The initial ratings allow us to form a taxonomy of types, constructed from a simple structural model of rating statements.[2] The randomized confiscation and corresponding changes in mood measures provide for an initial validation of

[2] We also provide reduced form evidence based only on the ratings themselves. The structural model assumes ratings are driven by consumption utilities and loss aversion. Though no choices are made, the core assumption is that subjects rate the object truthfully. The measure of loss aversion estimated is consistent with rational expectations as subjects were not told in advance that their their endowed object may be taken away when the ratings data was collected. An alternative design would attempt to precisely measure loss aversion either through statements of small stakes risk aversion (Fehr and Goette, 2007; Sprenger, 2015) or some other choice. Such tests would require both additional assumptions (e.g., about the correlation between consumption utility and loss aversion) and additional experimental choices. Recognizing both the polluting potential of such choices and the challenge of modeling the full body of experimental behavior through the lens of the KR model (for discussion, see Sprenger, 2015), we opted for this more broad categorization. Failure to correctly categorize types should lead to a lack of predictive validity in Stage 2 of the experiment, working against our identified results.

our taxonomy, ensuring that people who are classified as loss averse actually do experience sensations of loss in their measured mood.

In a second stage, subjects are again endowed with one of two objects. The second stage objects have no plausible complementarities with either object in the first stage, eliminating the desire to construct bundles of objects across the two stages. In this second stage, subjects make their only choice in the experiment. Forty percent of subjects are asked a baseline endowment effect question of whether they would like to trade their object for the alternative. The other sixty percent of subjects are asked whether they would like to trade their object under a probabilistic forced exchange mechanism akin to Goette et al. (2016). With probability 0.5, regardless of their decision, exchange will be forced. Under the KR model, individuals who are loss averse should grow more willing to exchange relative to baseline when probabilistically forced to do so, while those who are loss loving should grow less willing to exchange (statements which we formalize in Section 3.2).

The second stage provides for two central analyses. First, in the baseline condition we further validate our taxonomy of loss averse types by examining whether individuals coded as loss averse in Stage 1 are also unwilling to exchange for a completely different object in Stage 2. Second, we study expectations-based forces by examining sensitivity of behavior to probabilistic forced exchange. Given these predictions depend upon the heterogeneity in loss aversion, this exercise is conducted separately for the different types identified in Stage 1.

We document three key findings. First, on average subjects do appear to prefer their randomly endowed object in Stage 1, indicating an endowment effect in ratings.[3] Correspondingly, we estimate loss aversion on aggregate. At the individual level 36% of subjects are classified as loss averse in our structural analysis, 40% as potentially loss neutral, and 25% as loss loving. Our relative proportion of loss averse and loss loving is comparable to other recent findings on the heterogeneity of loss aversion (Chapman et al., 2017).[4]

Second, the taxonomy of loss aversion is respected in the responsiveness of mood to randomized experience. Loss averse types have significantly larger decreases in mood than loss loving types if their Stage 1 object is confiscated. More compellingly, this taxonomy is reflected in Stage 2 behavior. In the Stage 2 base-

[3] 47% of subjects report a higher rating for their endowed object, 22% report the same rating, and 31% percent report a higher rating for the alternative object.

[4] Based on willingness to pay and willingness to accept data for lotteries from a representative sample, Chapman et al. (2017) find an endowment effect for 60% of the respondents, no endowment effect for 10% of the subjects, and a reverse endowment effect for approximately 30% of the sample. Loss averse types in our data are 1.44 times more likely than loss loving types. In their data loss averse types are twice as frequent.

line condition, loss averse types are less willing to trade than others, delivering a substantial endowment effect for a different, randomly-assigned object.

Third, the comparative statics of expectations-based models are decisively supported in Stage 2. Following KR predictions, loss-averse types grow significantly more willing to trade under probabilistic forced exchange, while loss-loving types grow significantly less-so. Recognizing and accounting for the heterogeneity in types is critical as the aggregate data reproduce the null findings of Goette et al. (2016) for a similar forced exchange mechanism.

We believe our results add to the discussion of reference-dependent preferences and exchange anomalies in general. First, recognizing and accounting for heterogeneity in loss aversion allows for more nuanced tests of expectations-based reference dependence. Given different findings across prior studies (Ericson and Fuster, 2011; Heffetz and List, 2014), the null aggregate effects here and in Goette et al. (2016), and our theoretical development demonstrating that KR comparative statics change sign for different types of loss aversion, heterogeneity appears to be a confound of first order importance. We show, in a simple setting, that the forces of expectations-based models are reliably recovered once heterogeneity in loss aversion is accounted for.

Second, a body of research has questioned the generality of exchange anomalies such as the endowment effect. One line in particular has argued that trading experience can increase the willingness to engage in exchange (List, 2003, 2004), with the implication that the endowment effect should be "selected" out in markets. Indeed, Engelmann and Hollard (2010), show that even a very minute body of experience can eliminate the endowment effect. In Section 3.4.3, we link the experiences of Stage 1, and the subjective perception thereof, to exchange behavior in Stage 2. Even accounting for heterogeneity in types, there remains a marked distaste for exchange in Stage 2. This "residual" endowment effect is related to experience in Stage 1. Interestingly, the effects of experience are not reflected in the objective outcome of keeping or losing one's object, but rather in the subjective perception of this experience.[5] Individuals with a negative perception of their Stage 1 experience are less willing to exchange in Stage 2. Such an observation may help to explain the findings of Engelmann and Hollard (2010). In their study, experience is induced through trading rounds, in which subjects must make an exchange in order to keep *any* object. Making such an explicit connection between exchange and positive experience should indeed lead to more willingness to trade. This also suggests a path by which exchange anomalies may persist: negative experiences (both for exchanging and not exchanging) can lead to less willingness to exchange subsequently. As such,

[5] Subjectively, the experience could be positive or negative depending on the subject's loss aversion.

the endowment effect need not be selected quickly out of the market through trading experience alone if perceptions thereof are not uniformly positive.

The paper proceeds as follows. In Section 3.2, we set the theoretical background and derive behavioral predictions. Section 3.3 and 3.4 present the experimental design and results, respectively. Section 3.5 concludes.

## 3.2 Theoretical Considerations and Design Guidance

We examine the forces of expectations-based reference-dependent preferences in simple exchange settings with two goods, recognizing heterogeneity of loss aversion. The theoretical development hues closely to our experimental design, providing motivation for our analyses.

Consider a two-dimensional utility function over the two objects of interest, good X and good Y. Let $\mathbf{c} = (m_X, m_Y)$ and $\mathbf{r} = (r_X, r_Y)$ represent vectors of consumption utility and reference utility, respectively. The KR model specifies a utility function with two components, consumption utility, $m(\mathbf{c}) \equiv m_X + m_Y$, and gain-loss utility, $n(\mathbf{c}|\mathbf{r}) \equiv n_X(m_X|r_X) + n_Y(m_Y|r_Y) \equiv \mu(m_X - r_X) + \mu(m_Y - r_Y)$, with separability across consumption dimensions. Let $m_X \in \{0,X\}$ and $m_Y \in \{0,Y\}$ stand for both the outcome and the corresponding consumption utility of owning no or one unit of good X, and no or one unit of good Y, respectively. Overall utility is described by

$$\begin{aligned} u(\mathbf{c}|\mathbf{r}) \;=\; u(m_X, m_Y|r_X, r_Y) \;&=\; m_X + n_X(m_X|r_X) + m_Y + n_Y(m_Y|r_Y) \\ &=\; m_X + \mu(m_X - r_X) + m_Y + \mu(m_Y - r_Y), \end{aligned}$$

where

$$\mu(z) = \begin{cases} \eta z & \text{if } z \geq 0 \\ \eta\lambda z & \text{if } z < 0. \end{cases}$$

In this piece-wise linear gain-loss function, the parameter $\eta$ captures the magnitude of changes relative to the reference point, and $\lambda$ is the degree of loss aversion.

### 3.2.1 Determination of the Reference Point

For the KR model, the vector $\mathbf{r}$ is determined as part of a consistent forward-looking plan for behavior. The KR model posits a reference-dependent expected utility function $U(F|G)$, taking as input a distribution $F$ over consumption outcomes, $\mathbf{c}$, which are valued relative to a distribution $G$ of reference points, $\mathbf{r}$. That is

$$U(F|G) = \int\int u(\mathbf{c}|\mathbf{r})dF(\mathbf{c})dG(\mathbf{r}).$$

A *Personal Equilibrium* is a situation where, given that the decision-maker expects as a reference some distribution $F$, she indeed prefers $F$ as a consumption distribution over all alternative consumption distributions, $F'$. Ex-ante optimal behavior has to accord with expectations of that behavior. Formally, given a choice set, $\mathscr{D}$, of lotteries, $F$, over consumption outcomes $\mathbf{c} = (m_X, m_Y)$, *Personal Equilibrium* states the following:
*Personal Equilibrium (PE):* A choice $F \in \mathscr{D}$, is a personal equilibrium if

$$U(F|F) \;\geq\; U(F'|F) \; \forall \; F' \;\in\; \mathscr{D}.$$

Regardless of endowment, if good X is to be chosen in a PE, then $\mathbf{r} = (X, 0)$ and if good Y is to be chosen in a PE then $\mathbf{r} = (0, Y)$.

#### 3.2.1.1 Manipulating r: Probabilistic Forced Exchange

As noted above, the PE concept requires a consistency between **c** and **r**. In a simple exchange experiment over two objects, potential PE selections are $[\mathbf{c}, \mathbf{r}] = [(X, 0), (X, 0)]$ and $[\mathbf{c}, \mathbf{r}] = [(0, Y), (0, Y)]$. Depending on the endowment of X or Y, only one of these choices represents an unwillingness to trade. Assuming an endowment of X, the individual can support not exchanging $[\mathbf{c}, \mathbf{r}] = [(X, 0), (X, 0)]$ in a PE if

$$U(X, 0|X, 0) \;>\; U(0, Y|X, 0),$$

or

$$X \;>\; \frac{1 + \eta}{1 + \eta\lambda} Y. \tag{3.1}$$

Note that the smallest value of $X$ at which the individual can support not exchanging, $\underline{X} = \frac{1+\eta}{1+\eta\lambda} Y$, is inferior to $Y$ if $\lambda > 1$. As such, loss averse individuals can support not exchanging $X$ for $Y$ even if $Y$ would be preferred on the basis of consumption utility alone. This describes the mechanism by which the KR model generates an endowment effect. Figure 3.1 graphs $\underline{X}$ against $\lambda$ for $Y = 1$, $\eta = 1$, showing that as $\lambda$ increases, the lowest value of $X$ at which the agent can support not exchanging decreases following a simple inverse relationship.

Also graphed in Figure 3.1 is the alternate PE cutoff value corresponding to an agent who fulfills an expectation to exchange their endowed object X for Y.

$$U(0, Y|0, Y) \;>\; U(X, 0|0, Y),$$

or

$$X \;<\; \frac{1 + \eta\lambda}{1 + \eta} Y.$$

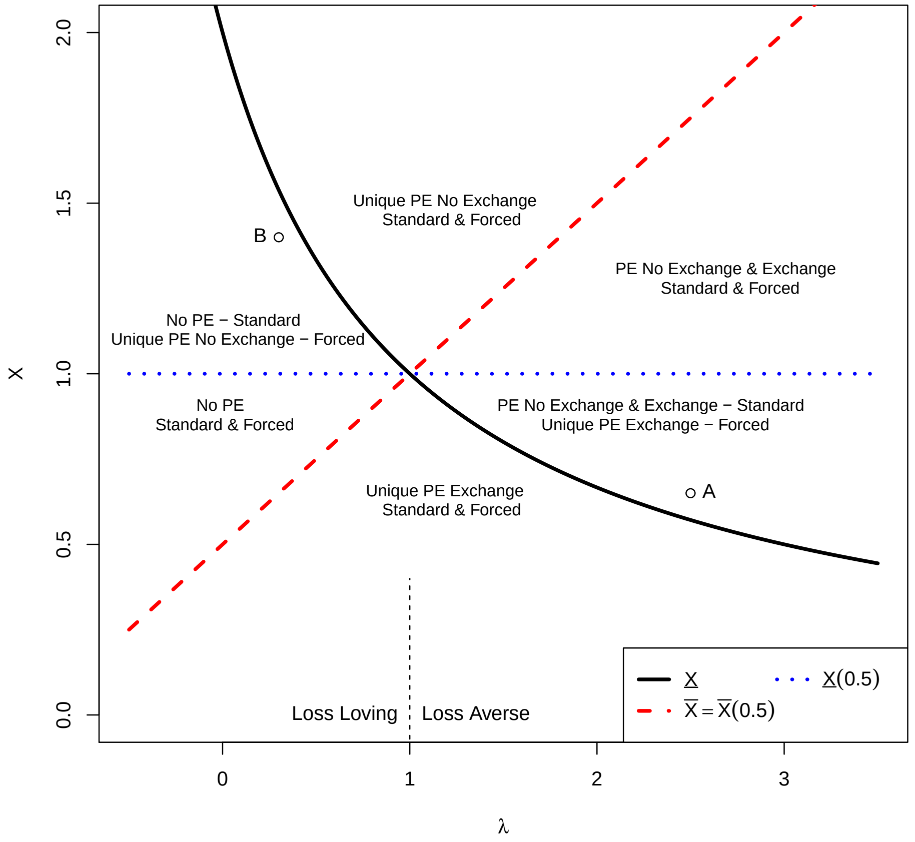


**Figure 3.1.** Loss aversion and personal equilibrium values. PE cutoff values for an agent endowed with good X, having consumption utility $Y = 1$ and $\eta = 1$. For $X > \underline{X} = \frac{1+\eta}{1+\eta\lambda}Y$, agents can support not exchanging as a PE in standard exchange environment. For $X < \overline{X} = \frac{1+\eta\lambda}{1+\eta}Y$, agents can support exchanging as a PE in a standard exchange environment. With forced exchange probability of 0.5 $\underline{X}(0.5) = Y$ and $\overline{X}(0.5) = \overline{X} = \frac{1+\eta\lambda}{1+\eta}Y$. Endowed with X, loss averse agents with $\lambda > 1$ (as in point A) can support not exchanging in PE in standard exchange environment, but cannot with probabilistic forced exchange. Loss loving agents with $\lambda < 1$ (as in point B) cannot support not exchanging in PE in standard exchange environment, but can with probabilistic forced exchange.

The highest value of X at which the agent can support exchanging, $\overline{X} = \frac{1+\eta\lambda}{1+\eta}Y$, increases linearly with $\lambda$. Note that for $\underline{X} < X < \overline{X}$, there will be multiple equilibria, with the agent able to support both exchanging and not exchanging as a PE. The KR model is constructed with a notion of equilibrium refinement, *Preferred*

*Personal Equilibrium* (PPE), in which ex-ante utility is used as a basis for selection and, hence, for making more narrow predictions. We provide our results without appeal to equilibrium selection, assuming only that actions are more likely to be taken if they are PE than if they are not.[6]

Now, consider a setting of probabilistic forced exchange. With probability 0.5 the agent, assumed endowed with X, will be forced to exchange X for Y regardless of their choice. If the individual wishes to retain her object, she is subject to a stochastic reference point, as with probability 0.5 it will be confiscated. She can support attempting not to exchange if

$$\begin{aligned} U(0.5(X,0) + 0.5(0,Y)|0.5(X,0) + 0.5(0,Y)) \ > \\ U(0,Y|0.5(X,0) + 0.5(0,Y)), \end{aligned}$$

or

$$X \ > \ Y. \tag{3.2}$$

The attempt to retain X is only supported on the basis of consumption utility values, regardless of the level of loss aversion. The manipulation of probabilistic forced exchange changes the PE cutoff for $\underline{X}$ from $\underline{X} = \frac{1+\eta}{1+\eta\lambda}Y$ to $\underline{X}(0.5) = Y$. Figure 3.1 illustrates the changing PE cutoff values associated with not exchanging. Loss averse agents can no longer support not exchanging in PE at values of X lower than Y.

Though probabilistic forced exchange alters the PE considerations associated with not exchanging, it leaves unchanged the PE considerations associated with exchanging. The agent can support exchanging in PE if

$$U(0,Y|0,Y) \ > \ U(0.5(X,0) + 0.5(0,Y)|0,Y),$$

which as before is

$$X \ < \ \frac{1+\eta\lambda}{1+\eta}Y.$$

$\overline{X}(0.5) = \overline{X}$ is noted in Figure 3.1.

Manipulating forced-exchange probability carries clear value for testing the KR model. Under the standard assumption of loss aversion, $\lambda > 1$, agents can support not exchanging in PE for values of $X < Y$ in a standard exchange experiment, but cannot do so with forced exchange probability of 0.5. The intuition is simple: attempting to retain the object exposes the agent to potential losses under forced exchange. She cannot support accepting these losses. Under the

[6] Goette et al. (2016) discuss PPE considerations with probabilistic forced exchange, ensuring that the core comparative statics associated with probabilistic forced exchange are maintained under equilibrium refinement.

assumption that actions are more likely to be taken if they are PE than if they are not, agents' willingness to exchange should increase with forced exchange. This is a unique prediction of expectations-based models not shared by prior formulations of the reference point. Importantly, this comparative static prediction hinges on agents being loss averse. In the next subsection, we investigate heterogeneity in loss aversion, showing that the comparative static associated with probabilistic forced exchange can reverse sign if individuals have $\lambda < 1$.

### 3.2.2 Heterogeneity in Loss Aversion

A number of recent studies have questioned the universality of loss aversion (see, e.g., Chapman et al., 2017; Erev et al., 2008; Harinck et al., 2007; Nicolau, 2012; Sprenger, 2015)).[7] Heterogeneity in loss aversion can confound the identification of expectations-based models. Under KR preferences, different values of $\lambda$ can lead to different directional predictions for the effects of forced exchange. Figure 3.1 illustrates the logic, graphing the PE cutoff values for not exchanging, $\underline{X} = \frac{1+\eta}{1+\eta\lambda}Y$ and $\underline{X}(0.5) = Y$, and for exchanging, $\overline{X} = \overline{X}(0.5) = \frac{1+\eta\lambda}{1+\eta}Y$.

Consider the case of a point like $A$, with $\lambda > 1$ and a valuation $X$ slightly below $Y = 1$. In the standard exchange experiment this individual can support not exchanging even though $X < Y$ as $X > \underline{X}$. With forced exchange probability 0.5, this individual can no longer support not exchanging as $X < \underline{X}(0.5)$. Assuming that actions are more likely to be taken when they are PE than when they are not leads to the Goette et al. (2016) comparative static prediction: individuals should grow more willing to exchange with probabilistic forced exchanged.

Now, consider a point like $B$ with $\lambda < 1$ and a value of $X$ slightly above $Y = 1$. Such an individual cannot support not exchanging as a PE in the standard exchange experiment even though $X > Y$ as $X < \underline{X}$.[8] With forced exchange probability of 0.5 this individual can now support not exchanging as a PE as $X > \underline{X}(0.5)$. Again, assuming that actions are more likely to be taken when they are PE than when they are not leads to the opposite prediction from the prior case. An agent with $\lambda < 1$ grows less willing to exchange with probabilistic forced exchange, reversing the sign of the previously described comparative static.[9]

[7] Though $\lambda > 1$ obtains for the majority of subjects, a substantial fraction are found to be close to loss neutral, $\lambda = 1$, and loss loving, $\lambda < 1$.For example, in the individual estimates of Sprenger (2015), 27% of the sample has $\lambda < 1$ within the 95% confidence interval of their estimated $\lambda$, while the remaining 73% are significantly loss averse.

[8] This individual can also not support exchanging as a PE given his loss-lovingness as $X > \overline{X}$. That is, no PE selections exist for this individual. KR note the possibility of multiplicity and absence of equilibria in their theoretical development.

[9] Note that the example provided relied on both differences in loss aversion, $\lambda$, and consumption utility, $X$, between points $A$ and $B$. This is only for illustrative purposes. If two agents instead had the same value of $X$, either above or below $Y$, with one being loss averse and the other loss

Taken together the analysis of probabilistic forced exchange and heterogeneity give insights for our experimental design. Our study adapts Goette et al.'s (2016) central manipulation of probabilistic forced exchange to a binary exchange situation with two objects, and also manipulates prior experiences to deliver and validate measures of loss aversion.[10]

## 3.3 Experimental Design and Procedures

Our design is comprised of two stages. In Stage 1, a taxonomy of loss averse types is created, exchange experience is manipulated via random confiscation, and the effects of this experience on mood are measured. In Stage 2, subjects are assigned to either a standard exchange study or one with probabilistic forced exchange, making their only choice in the experiment. Stage 1 experiences and measures of loss aversion can then be connected to Stage 2 behavior. Figure 3.2 illustrates the experimental order of events.

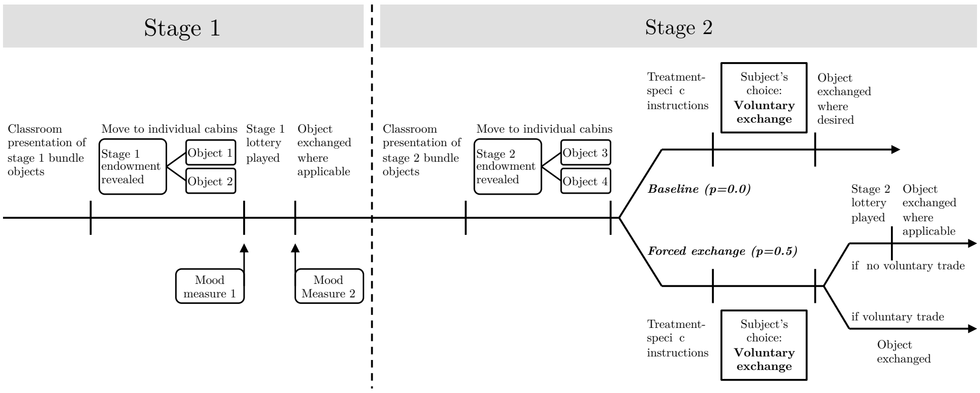

Stage 1
Stage 2
Classroom presentation of stage 1 bundle objects
Move to individual cabins
Stage 1 endowment revealed
Object 1
Object 2
Stage 1 lottery played
Object exchanged where applicable
Mood measure 1
Mood Measure 2
Classroom presentation of stage 2 bundle objects
Move to individual cabins
Stage 2 endowment revealed
Object 3
Object 4
Treatment-specific instructions
Subject's choice: Voluntary exchange
Object exchanged where desired
Baseline (p=0.0)
Forced exchange (p=0.5)
Stage 2 lottery played
Object exchanged where applicable
if no voluntary trade
if voluntary trade
Object exchanged
Treatment-specific instructions
Subject's choice: Voluntary exchange

**Figure 3.2.** Timeline of laboratory experiment. The figure displays the course of events in both treatment conditions, baseline ($p = 0.0$) and forced exchange ($p = 0.5$).

loving, then one of them would be affected by probabilistic forced exchange (either positively or negatively) and the other would not. This implies that if $X$ is symmetrically distributed around $Y$, and $X$ and $\lambda$ are independent, the sign of comparative statics can differ depending on whether $\lambda > 1$ or $\lambda < 1$. Loss averse agents will grow more willing to exchange on average while loss loving agents will grow less willing to exchange on average as exchange is probabilistically forced.

[10] Unlike Goette et al. (2016), we also study a direct exchange mechanism that does not require eliciting the willingness to pay or willingness to accept in monetary terms using price lists.

### 3.3.1 Stage 1: Measures of Loss Aversion and Manipulation of Experience

**Procedures.** The experimenter welcomed the participants in a presentation room and informed them that the study would consist of two stages. At each seat there was a card with a number (placed face down). Then, without further explanation, the experimenter projected on the wall two equally-sized pictures of the respective Stage 1 objects for that session along with the description and two short bullet points on the characteristics of the product. The exact information presented to subjects is reproduced in Appendix 3.C.

After allowing sufficient time (three minutes) to study the projected information, the experimenter asked subjects to turn the card in front of them over and move to the cubicle with the corresponding number in the adjacent computer laboratory. In their private cubicle, which was separated and not visible from the outside, subjects would find one of the two presented goods. Computer instructions then informed the subject that she possesses the object in front of her, and that she is free to inspect it more closely.

After three minutes allotted for inspection of the good, we asked subjects how much they liked and wanted each one of the goods. Specifically, for each object we asked “How much do you like this product?” and “How much would you want to have this product?” with response scales ranging from 0=“Not at all” to 8=“Very much”. These ratings data are used to construct our measures of loss aversion, notably collected without experimental choice. These ratings are collected before any further instructions are given, including instructions related to confiscation.

Next, the computer instructions announced that the experimenter would randomly draw a number between 1 and 20 using a rotating lottery device placed on a table in the middle of the room. Half of the subjects learned that they would lose their current good and receive the other one in return in case a number between 1 and 10 is drawn. Instructions for the other half read that this exchange would only take place if a number between 11 and 20 is drawn.[11] The experimenter drew the number in a way that both the lotto device containing the 20 balls and the drawn number was visible from every cabin. The exchange was executed after the draw by the experimenter, who, without further comment, replaced the object for subjects who had lost their good due to the drawn number. Subsequent instructions informed subjects that they would keep their current object and asked them to return to the lecture room for the second stage.

Immediately before and immediately after the random confiscation was conducted, we elicited subjects’ mood using standard psychological scales (Bradley

[11] This *loss condition* was counterbalanced within each subsample endowed with the same good, such that irrespective of the draw, exchange would take place for exactly half of the subjects initially endowed with either good.

and Lang, 1994). Subjects answered the question “Please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?” by positioning a slider on an 11-point response scale. The lower end (0) was labeled using the words “Unhappy, Angry, Unsatisfied, Sad, Desperate” and the upper end (10) was labeled “Happy, Thrilled, Satisfied, Content, Hopeful”. The individual change in these scores are used to provide an initial validation of our taxonomy of types.

### 3.3.2 Stage 2: Probabilistic Forced Exchange, Heterogeneity and Prior Experience

**Procedures.** The basic procedures in the second stage were deliberately kept exactly identical to those in the first stage. Upon their return to the lecture room, the experimenter projected another page onto the wall, this time presenting the objects of the Stage 2 goods bundle of that session. In the meantime, a second experimenter allocated objects to the cubicles in the computer laboratory next door in a pre-specified order. Subjects were ushered back to their cubicle where again they found their second object, learned that it belonged to them and were allowed sufficient time for inspection. We studied two conditions.

**Baseline treatment.** In the baseline condition, subjects received an opportunity to voluntarily exchange their endowed good for the other one. Whichever way they chose, they would keep or receive their desired object and there would be no further exchange. The baseline condition is a standard exchange setting.

**Forced exchange treatment.** The second condition implemented an exchange study with probabilistic forced exchange. The instructions specified that irrespective of their choice of exchanging their endowed object, exchange would take place anyway with a probability of 50% based on a draw from the lotto device as in the first stage. This means that for a subject who decided to trade voluntarily, the forced exchange did not bear any consequences. However, for a subject who chose to keep her object, there was an additional chance of losing it.

Several noted issues with experimental investigations of market exchange motivated our purposefully simple design (Plott and Zeiler, 2005, 2007). First, subjects take a simple binary choice, alleviating potential concerns related to the use of multiple-price lists in exchange experiments. Specifically, we do not need to elicit a willingness to pay or willingness to accept in monetary terms, but simply ask whether the subject is willing to trade the endowed good for the other one. As such, mistaken perceptions of market power do not play a role, nor do income effects. Second, unlike previous market exchange experiments, we create a private environment that limits confounds from social interaction. In particular, subjects take their decisions anonymously in a private cabin; they find their endowment placed in front of them when entering the cabin instead of receiving it personally through the hands of the experimenter (which has been criticized for triggering the misperception of the endowment as a gift (see, e.g., Plott and Zeiler, 2005, 2007); and subjects do not interact with other subjects at any stage during the experiment.

### 3.3.3 Sample Details

A sample of 607 students from the University of Bonn participated in the experiment which was conducted using the software z-Tree (Fischbacher, 2007) in June and July 2015 at the BonnEconLab. We conducted 31 sessions with 17 to 20 participants each. Table 3.1 provides an overview of the subject pool by treatment conditions.

**Table 3.1.** Summary statistics and treatment assignment

| | **Stage 1** | | | |
|---|---|---|---|---|
| | Bundle 1 | | Bundle 2 | |
| | USB stick | Pen set | Picnic mat | Thermos |
| *A) Initial Endowment* | 160 | 152 | 150 | 145 |
| *– in % of subject pool* | 26.36% | 25.04% | 24.71% | 23.89% |
| *B) Lost Endowment* | 80 | 76 | 75 | 72 |
| *– in % of A)* | 50.00% | 50.00% | 50.00% | 49.66% |
| | **Stage 2** | | | |
| | Bundle 1 | | Bundle 2 | |
| | USB stick | Pen set | Picnic mat | Thermos |
| *C) Initial Endowment* | 150 | 145 | 160 | 152 |
| *– in % of subject pool* | 24.71% | 23.89% | 26.36% | 25.04% |
| *D) Baseline Condition* | 60 | 58 | 60 | 55 |
| *– in % of C)* | 40.00% | 40.00% | 37.50% | 36.18% |
| *E) Probability 0.5 Condition* | 90 | 87 | 100 | 97 |
| *– in % of C)* | 60.00% | 60.00% | 62.50% | 63.82% |
| Total number of observations | 607 | | | |

*Notes*: Stage 2 condition (baseline or probability 0.5 of forced exchange) is randomized within each session. The use of each bundle as the Stage 1 bundle was counterbalanced at the session level.

The objects used for the exchange experiment included a USB stick, a set of three erasable pens, a picnic mat and a thermos.[12] We selected these four objects on the basis of a pre-experimental survey evaluation of 12 candidate goods to ensure that all items were of approximately equal value to potential participants. We put particular emphasis on ruling out complementarities between items across rounds. The former two (USB stick and pens) and the latter two objects (picnic mat and thermos) each constituted a bundle. Every subject faced exactly one exchange situation with each bundle of objects. The use of each bundle as Stage 1 bundle was counterbalanced at the session level, with the respective other bundle used in Stage 2. Within each session, the endowments of one of the two objects within the bundle was counterbalanced in both stages.[13]

## 3.4 Experimental Results

We present the results in three subsections. First, we examine stated good ratings and the effect of experience in Stage 1, providing our taxonomies of loss averse types and validating these taxonomies with evidence on the change in mood induced by forced exchange. Second, we examine behavior in Stage 2, linking heterogeneity in loss aversion to probabilistic forced exchange. A third subsection is dedicated to the effects of subjective experience on exchange behavior.

### 3.4.1 Stage 1: Loss Aversion, Experience, and Mood

Though no choices were made in Stage 1, we collect two pieces of evidence. First, subjects provide their ratings for both objects. Second, subjects provide a measure of mood once before being informed about the randomized confiscation procedure and once after they learned their random outcome and the exchange was carried out where applicable.

Figure 3.3 provides histograms of subject's liking of their endowed and the alternative object. Given random assignment of endowed objects and the counterbalanced design, the distributions of ratings should be identical. Instead, the distribution of ratings for subjects' own object skews higher than the alternative, yielding a statistically significant stated preference for the endowed good (Wilcoxon signed-rank test, $z = 4.57, p < 0.01$).

[12] Pictures and information presented to subjects are reproduced in Appendix 3.C.

[13] That is, if for a given session the USB stick and pens bundle constituted the first stage bundle, the picnic mat and thermos bundle would be the second stage bundle. Half of the subjects were initially endowed with the USB stick in the first stage. Among this half of the session participants, again half would initially receive the picnic mat and the other half the thermos at the beginning of the second stage.

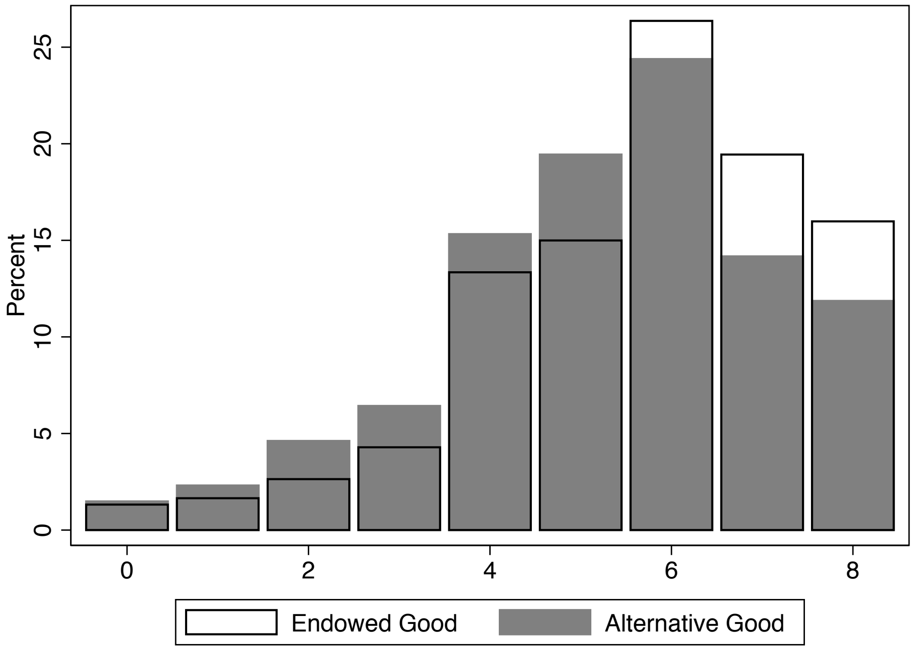


**Figure 3.3.** Preferences and endowments. Self-reported scores of liking for the endowed and alternative goods. (Wilcoxon signed-rank statistic $z = 4.57, p < 0.01, N = 607$).

Within subject we also find a tendency towards preferring the endowed object relative to the alternative. 47% of subjects report a higher liking score for their endowed object, 22% report the same score, and 31% report a higher score for the alternate object.[14]

The liking scores for the endowed and alternative object provide a basis for measuring loss aversion at the aggregate and individual level. We construct a simple structural model of these ratings based upon standard random utility methods (McFadden, 1974). Consider an individual endowed with $X$ that is asked to provide ratings statements for both $X$ and $Y$ prior to being informed of the random confiscation implemented in Stage 1. Through the lens of the KR model such an individual evaluates $X$ based upon $U(X,0|X,0)$. Given that the agent is endowed with $X$ and is uninformed of the possibility of confiscation at the time of the ratings, she plausibly evaluates $Y$ based upon $U(0,Y|X,0)$. With standard logit shocks, $\epsilon_X$ and $\epsilon_Y$, the parameters associated with these utilities are easily estimated. Unlike choice data, agents may provide the same rating

[14] Our design also collects a score for "wanting" each object. The corresponding percentage shares for wanting scores are virtually identical (48%, 23%, 29%, respectively). For analysis using these wanting scores as the basis of analysis see Figure 3.A.1, Table 3.A.2, and Table 3.A.3.

score for both objects. As such, the estimator must account for identical ratings, something to which standard logit techniques are also already well adapted (see, e.g., Cantillo et al., 2010). We assume agents will provide a higher rating for their endowed object, $X$, if

$$U(X,0|X,0)+\epsilon_X \;>\; U(0,Y|X,0)+\epsilon_Y+\delta,$$

where $\delta$ is a discernibility parameter to be estimated. Similarly agents provide a higher rating for the alternative object, $Y$, if

$$U(0,Y|X,0)+\epsilon_Y \;>\; U(X,0|X,0)+\epsilon_X+\delta,$$

and provide the same rating if

$$|U(X,0|X,0)+\epsilon_X-(U(0,Y|X,0)+\epsilon_Y)| \;\le\; \delta.$$

Under the functional form assumptions of $\eta=1$ and $m_X=X, m_Y=Y$, for someone given object $X$, we obtain familiar probabilities for the ranking of ratings $R(X)$ and $R(Y)$,

$$\begin{aligned}
P(R(X)\;>\;R(Y)) &= \frac{\exp(U(X,0|X,0))}{\exp(U(X,0|X,0))+\exp(U(0,Y|X,0)+\delta)} = \frac{\exp(X)}{\exp(X)+\exp(2Y-\lambda X+\delta)}\\
P(R(Y)\;>\;R(X)) &= \frac{\exp(U(0,Y|X,0))}{\exp(U(0,Y|X,0))+\exp(U(X,0|X,0)+\delta)} = \frac{\exp(2Y-\lambda X)}{\exp(X+\delta)+\exp(2Y-\lambda X)}\\
P(R(X)\;=\;R(Y)) &= 1-P(R(X)\;>\;R(Y))-P(R(Y)\;>\;R(X)),
\end{aligned}$$

where the consumption utilities values, $X$ and $Y$, the discernibility parameter $\delta$, and the loss aversion parameter, $\lambda$, are the desired estimands. For someone endowed with object $Y$, these same ratings probabilities are

$$\begin{aligned}
P(R(X)\;>\;R(Y)) &= \frac{\exp(U(X,0|0,Y))}{\exp(U(X,0|0,Y))+\exp(U(0,Y|0,Y)+\delta)} = \frac{\exp(2X-\lambda Y)}{\exp(Y+\delta)+\exp(2X-\lambda Y)}\\
P(R(Y)\;>\;R(X)) &= \frac{\exp(u(0,Y|0,Y))}{\exp(U(0,Y|0,Y))+\exp(U(X,0|0,Y)+\delta)} = \frac{\exp(Y)}{\exp(Y)+\exp(2X-\lambda Y+\delta)}\\
P(R(X)\;=\;R(Y)) &= 1-P(R(X)\;>\;R(Y))-P(R(Y)\;>\;R(X)).
\end{aligned}$$

The likelihood contribution of someone endowed with $X$ or $Y$ follows precisely the formulations above. It will not generally be possible to estimate both utility values, $X$ and $Y$, separately. So we normalize one of the goods values to be $Y=1$ and estimate the remaining parameters via maximum likelihood.

Table 3.2 provides aggregate estimates of consumption utilities, $\lambda$ and $\delta$, separately for each bundle of goods. For Bundle 1, we restrict the utility value of USB sticks to be $Y=1$ and for Bundle 2 we restrict the utility value of the thermos to be $Y=1$. Quite similar results obtain across the two bundles. For Bundle 1, $\lambda$ is estimated to be 1.559 (robust s.e. = 0.139), while for Bundle 2 it is estimates to be 1.289 (0.121). For both bundles we reject the null hypothesis of no loss aversion $\lambda = 1$, consistent with the reduced form ratings results.[15] The utility of pen sets and picnic mats are estimated to be lower than those of USB sticks and Thermoses, respectively. And, discernibility is estimated close to $\delta = 0.5$ in both cases.

**Table 3.2.** Aggregate parameter estimates

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Estimate | (Std. Error) | Estimate | (Std. Error) |
| | *Bundle 1* | | *Bundle 2* | |
| *Loss Aversion:* | | | | |
| $\hat{\lambda}$ | 1.559 | (0.139) | 1.289 | (0.121) |
| *Utility Values:* | | | | |
| $\hat{X}_1$ *(Pen Set)* | 0.632 | (0.049) | | |
| $\hat{Y}_1$ *(USB Stick)* | 1 | - | | |
| $\hat{X}_2$ *(Picnic Mat)* | | | 0.837 | (0.051) |
| $\hat{Y}_2$ *(Thermos)* | | | 1 | - |
| *Discernibility:* | | | | |
| $\hat{\delta}$ | 0.549 | (0.061) | 0.446 | (0.052) |

*Notes*: Maximum likelihood estimates. Robust standard errors in parentheses.

The aggregate estimates show evidence of loss aversion. To construct bounds for estimates of individual loss aversion, we evaluate individual choices assuming average utility and discernibility values. For example, consider an individual endowed with the pen set in Bundle 1. At the aggregate estimates of $\delta$ and $X$ for Bundle 1, if this individual were to state a higher ranking for the pen set than for the USB stick, it would imply a loss aversion parameter of $\hat{\lambda} > 3.03$.[16] Similarly, stating a higher ranking for the USB stick would imply $\hat{\lambda} < 1.30$,[17] and stating the same ranking implies $\hat{\lambda} \in [1.30, 3.03]$. Of these three possible cases, two demonstrate evidence of loss aversion $\hat{\lambda} > 1$, while the other case is plausibly loss neutral as $\hat{\lambda} = 1$ can rationalize the rankings.[18] In total, there exist twelve cases of endowments and rank orders. Table 3.3 enumerates the

[15] For Bundle 1, the null hypothesis of $\lambda = 1$ is rejected, $\chi^2(1) = 16.13$ ($p < 0.01$). For Bundle 2, the null hypothesis of $\lambda = 1$ is also rejected, $\chi^2(1) = 5.73$ ($p < 0.05$).

[16] To state a higher ranking for the pen set implies $0.632 > 2 - \hat{\lambda} \cdot 0.632 + 0.549$ or $\hat{\lambda} > 3.03$.

[17] To state a higher ranking for the USB implies $2 - \lambda \cdot 0.632 > 0.632 + 0.549$ or $\lambda < 1.30$.

[18] It may seem prima-facie surprising that providing the same ranking in this case is consistent with loss aversion. The logic is simple: given that the pen set has substantially lower consumption utility than the USB stick, one must be loss averse to rank them equally.

cases and the corresponding categorization into loss averse, loss neutral, and loss loving types. Overall 217 subjects (35.7%) are categorized as loss averse, 240 (39.5%) are categorized as loss neutral, and 150 (24.7%) are categorized as loss loving. This is the taxonomy of individual types used in our analysis.

**Table 3.3.** Individual classifications

| Case | #Obs | *Structural Bounds Taxonomy* | | | *Reduced Form Taxonomy* | | |
|---|---|---|---|---|---|---|---|
| | | *Loss Averse* | *Loss Neutral* | *Loss Loving* | *Loss Averse* | *Loss Neutral* | *Loss Loving* |
| Bundle 1 | | | | | | | |
| Endowed Pen Set | | | | | | | |
| R(Pen Set) > R(USB Stick) | 42 | $\hat{\lambda} > 3.03$ | | | X | | |
| R(USB Stick) > R(Pen Set) | 69 | | $\hat{\lambda} < 1.30$ | | | | X |
| R(USB Stick) = R(Pen Set) | 41 | $1.30 \leq \hat{\lambda} \leq 3.03$ | | | | X | |
| Endowed USB Stick | | | | | | | |
| R(USB Stick) >R(Pen Set) | 109 | | $\hat{\lambda} > 0.81$ | | X | | |
| R(Pen Set) > R(USB Stick) | 23 | | | $\hat{\lambda} < -0.29$ | | | X |
| R(USB Stick) = R(Pen Set) | 28 | | | $-0.29 \leq \hat{\lambda} \leq 0.81$ | | X | |
| Bundle 2 | | | | | | | |
| Endowed Picnic Mat | | | | | | | |
| R(Picnic Mat) > R(Thermos) | 55 | $\hat{\lambda} > 1.92$ | | | X | | |
| R(Thermos) > R(Picnic Mat) | 61 | | | $\hat{\lambda} < 0.86$ | | | X |
| R(Thermos) = R(Picnic Mat) | 34 | | $0.86 \leq \hat{\lambda} \leq 1.92$ | | | X | |
| Endowed USB Stick | | | | | | | |
| R(Thermos) > R(Picnic Mat) | 79 | $\hat{\lambda} > 1.12$ | | | X | | |
| R(Picnic Mat) > R(Thermos) | 38 | | | $\hat{\lambda} < 0.23$ | | | X |
| R(Thermos) = R(Picnic Mat) | 28 | | $0.23 \leq \hat{\lambda} \leq 1.12$ | | | X | |
| Totals: | 607 | 217 | 240 | 150 | 285 | 131 | 191 |

*Notes:* Structural bounds taxonomy of types based on individual rankings at estimated aggregate utility values and discernibility parameters from Table 3.2. Reduced form taxonomy derived from whether subject exhibits higher, lower, or equal rankings for their endowed good relative to the alternative.

Also presented in Table 3.3 is an alternate taxonomy based only on raw ranking information. Ignoring utility values and discernibility, this reduced form taxonomy codes someone as loss averse, neutral, or loving depending only on whether the individual provides a higher, equal, or lower ranking for their endowed object. Based on this reduced form taxonomy, 285 subjects (47%) are categorized as loss averse, 131 (22%) are classified as loss neutral, and 191 (31%) are classified as loss loving.[19] Though the structural taxonomy provides a more conservative classification of types, there is broad agreement between the structural and reduced form taxonomies (Pearson's $\chi^2(4) = 315.2,\ p < 0.01$). For completeness, we provide all our analysis with both the structural and reduced form bounds provided in Table 3.3.

A minimal validation to eschew random response and lend credence to our two classifications is provided by our mood measures. Table 3.4 provides a summary of reported mood regressing the change in Stage 1 Happiness, *Δ Stage 1 Happiness*, on the objective experience of losing one's object for the full sample and our different identified types. Panel A presents the results based on the structural taxonomy, while Panel B uses the reduced form classification of types.

[19] This follows exactly the relative rankings data noted above.

**Table 3.4.** Preference types and subjective experience

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | *Dependent Variable: $\Delta$ Stage 1 Happiness* | | | |
| | Full Sample | Loss Averse | Loss Neutral | Loss Loving |
| *Panel A: Structural Bounds Taxonomy* | | | | |
| Lost Stage 1 Endowment | -0.826*** | -2.679*** | -0.715** | 1.560*** |
| | (0.210) | (0.385) | (0.291) | (0.403) |
| Constant | 0.582*** | 1.198*** | 0.715*** | -0.818** |
| | (0.159) | (0.254) | (0.231) | (0.337) |
| R-Squared | 0.0249 | 0.252 | 0.0106 | 0.129 |
| # Observations | 607 | 217 | 240 | 150 |
| *Panel B: Reduced Form Taxonomy* | | | | |
| Lost Stage 1 Endowment | -0.826*** | -3.169*** | -0.443 | 2.454*** |
| | (0.210) | (0.282) | (0.420) | (0.303) |
| Constant | 0.582*** | 1.841*** | 0.226 | -1.264*** |
| | (0.159) | (0.194) | (0.319) | (0.256) |
| R-Squared | 0.025 | 0.308 | 0.009 | 0.265 |
| # Observations | 607 | 285 | 131 | 191 |

*Notes*: Ordinary least squares or two-stage least squares regression. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Panel A: taxonomy of types based on structural bounds from Table 3.3. Panel B: taxonomy of types based on reduced form rating statements from Table 3.3.

For both taxonomies, individuals categorized as loss averse have substantial deterioration in mood if they lose their endowment, while individuals categorized as loss loving grow happier. Individuals categorized as loss neutral experience intermediate effects, growing somewhat less happy when their endowment is lost. This initial validation indicates that random response is unlikely to be driving our ratings statements, and provides evidence supporting our structural classification.

### 3.4.2 Stage 2: Heterogeneous Treatment Effects

Our Stage 2 design relies on between subjects variation. Forty percent of subjects participate in a baseline standard exchange study, choosing whether to keep their endowed object or exchange for the alternative. The other sixty percent make the same choice but with probability 0.5 exchange is forced. Table 3.5 presents the choices of subjects across these two conditions with linear probability models for the effect of treatment assignment on an indicator, $Exchange(=1)$.[20]

Before turning to the effects of probabilistic forced exchange, we examine behavior in our baseline design. Baseline results are conveyed as the estimated constants in least squares regression of exchange behavior in Table 3.5. Overall 36.5 percent of subjects choose to exchange, demonstrating a significant endowment effect relative to the null hypothesis of fifty percent exchange, $F_{1,605} = 18.32$ ($p < 0.01$). A second validation of our taxonomies is derived from examining differential baseline behavior across types. Panel A of Table 3.5 shows that 33 percent of subjects coded as loss averse according to our structural model choose to exchange, yielding a significant endowment effect relative to 50 percent exchange, $F_{1,215} = 12.21$ ($p < 0.01$). The fraction of subjects exchanging increases monotonically from loss averse to loss loving types. 42.9 percent of subjects who are coded as loss loving choose to exchange, which cannot be differentiated from the 50 percent benchmark, $F_{1,148} = 1.15$ ($p = 0.29$). Similar conclusions are reached in Panel B of Table 3.5, based only on the reduced form classification of types. These qualitative differences in Stage 2 baseline exchange behavior are closely in line with theoretical predictions — loss averse agents are unwilling to exchange, while loss loving types are more eager too — further validating the Stage 1 taxonomies. It must be noted, however, that though the groups differentially deviate from the 50 percent benchmark, the difference-in-differences does not fall within standard measures for statistical significance for either the structural, $p = 0.23$, or reduced form, $p = 0.05$, taxonomies.

Behavior in conditions with probabilistic forced exchange is also reported in Table 3.5, separately for the different types of agents. Probabilistic forced exchange yields substantially different effects across types of loss aversion. Panel A documents that subjects who are coded as loss averse increase their exchange probability by nearly 16%-age points ($\sim$ 50 percent), under probabilistic forced exchange, $F_{1,215} = 5.64$, ($p < 0.05$). The sizable endowment effect from baseline is eliminated, such that exchange can no longer be differentiated from the 50 percent benchmark, $F_{1,215} = 0.07$ ($p = 0.78$).

The positive treatment effect for loss averse types is mirrored by a significant negative treatment effect for loss loving types. Subjects coded as loss loving decrease their exchange probability by nearly 25%-age points ($\sim$ 60 percent),

[20] The analysis of Table 3.5 is conducted with robust standard errors. Table 3.A.1 repeats this analysis with standard errors clustered at the session level. The statistical conclusions are unchanged. The results based on the structural taxonomy of types increase in statistical significance, while the results based on the reduced for taxonomy decrease in significance when clustering at the session level. Given our focus on the structural taxonomy, Table 3.5 represents the more conservative set of conclusions.

**Table 3.5.** Exchange behavior and probabilistic forced exchange

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | *Dependent Variable: Exchange (=1)* | | | |
| | Full Sample | Loss Averse | Loss Neutral | Loss Loving |
| *Panel A: Structural Bounds Taxonomy* | | | | |
| Forced Exchange | 0.004 | 0.158** | 0.0271 | -0.248*** |
| | (0.040) | (0.067) | (0.066) | (0.078) |
| Baseline Exchange (Constant) | 0.365 | 0.330 | 0.361 | 0.429 |
| | (0.032) | (0.049) | (0.053) | (0.067) |
| R-Squared | 0.000 | 0.025 | 0.001 | 0.072 |
| # Observations | 607 | 217 | 240 | 150 |
| $H_0$: No Baseline Endowment Effect | $F_{1,605}$=18.32 | $F_{1,215}$=12.21 | $F_{1,238}$=6.85 | $F_{1,148}$=1.15 |
| | $(p < 0.01)$ | $(p < 0.01)$ | $(p < 0.01)$ | $(p = 0.29)$ |
| $H_0$: No Forced Ex. Endowment Effect | $F_{1,605}$=27.48 | $F_{1,215}$=0.07 | $F_{1,238}$=8.14 | $F_{1,148} = 63.77$ |
| | $(p < 0.01)$ | $(p = 0.78)$ | $(p < 0.01)$ | $(p < 0.01)$ |
| $H_0$: Baseline (col. 2) = Baseline (col. 4) | | | | $\chi^2(1) = 1.45$ |
| | | | | $(p = 0.23)$ |
| $H_0$: Forced Ex. (col. 2) =Forced Ex(col. 4) | | | | $\chi^2(1) = 15.89$ |
| | | | | $(p < 0.01)$ |
| *Panel B: Reduced Form Taxonomy* | | | | |
| Forced Exchange | 0.004 | 0.119** | -0.030 | -0.149** |
| | (0.040) | (0.057) | (0.088) | (0.074) |
| Baseline Exchange (Constant) | 0.365 | 0.304 | 0.392 | 0.448 |
| | (0.032) | (0.043) | (0.069) | (0.061) |
| R-Squared | 0.000 | 0.015 | 0.001 | 0.022 |
| # Observations | 607 | 285 | 131 | 191 |
| $H_0$: No Baseline Endowment Effect | $F_{1,605}$=18.32 | $F_{1,283}$=20.65 | $F_{1,129}$=2.45 | $F_{1,189}$=0.73 |
| | $(p < 0.01)$ | $(p < 0.01)$ | $(p = 0.12)$ | $(p = 0.39)$ |
| $H_0$: No Forced Ex. Endowment Effect | $F_{1,605}$=27.48 | $F_{1,283}$=4.04 | $F_{1,129}$=6.45 | $F_{1,189} = 23.82$ |
| | $(p < 0.01)$ | $(p = 0.045)$ | $(p = 0.012)$ | $(p < 0.01)$ |
| $H_0$: Baseline (col. 2) = Baseline (col. 4) | | | | $\chi^2(1) = 3.71$ |
| | | | | $(p = 0.054)$ |
| $H_0$: Forced Ex. (col. 2) =Forced Ex(col. 4) | | | | $\chi^2(1) = 8.32$ |
| | | | | $(p < 0.01)$ |

*Notes*: Ordinary least squares regression. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Null hypotheses tested for 1) zero baseline endowment effect, regression (Constant = 0.5); 2) zero forced exchange endowment effect (Constant + Forced Exchange = 0.5); 3) Identical baseline behavior across loss averse and loss loving agents (Constant (col. 2) = Constant (col. 4)); 4) Identical treatment effects of forced exchange across loss averse and loss loving agents (Forced Exchange (col. 2) = Forced Exchange (col. 4)). Hypotheses 3 and 4 tested via seemingly unrelated regression. Panel A: taxonomy of types based on structural bounds from Table 3.3. Panel B: taxonomy of types based on reduced form rating statements from Table 3.3.

under probabilistic forced exchange, $F_{1,148} = 10.18$ ($p < 0.01$).[21] The heterogeneous treatment effect over types closely follows our theoretical development on the sign of comparative statics, and is significant at all conventional levels, $\chi^2(1) = 15.89$ ($p < 0.01$). Quite similar results are found in Panel B, basing the analysis only on the reduced form taxonomy. Loss averse agents respond to forced exchange by exchanging more often while loss loving agents respond by exchanging less often.

Figure 3.4 presents more granular analysis of treatment. For each of the twelve structural types identified in Table 3.3, we take as a measure of loss aversion the lower bound (upper bound, midpoint) for those subjects who provided a higher (lower, equal) ranking for their endowed object relative to the alternative in Stage 1. Figure 3.4 graphs these values of loss aversion against each group's treatment effect noting the classification of type. The size of each point corresponds to the number of observations. An effectively monotonic pattern of treatment effects is observed. All four groups coded as loss loving exhibit negative treatment effects, all four groups coded as loss neutral deliver effectively zero treatment effect, and all four groups coded as loss averse exhibit positive treatment effects. Even within loss averse and loss loving groups, subjects coded as more loss averse respond more positively to forced exchange.

Also graphed in Figure 3.4 are predicted treatment effects for each group. At the corresponding values of $\lambda$ and aggregate utility values, we predict the probability of exchange following closely the logit formulation elaborated in Section

[21] Given these worsened attitudes towards exchange, loss loving agents in the forced exchange condition deliver a substantial endowment effect relative to the 50 percent benchmark, $F_{1,148} = 63.77$ ($p < 0.01$).

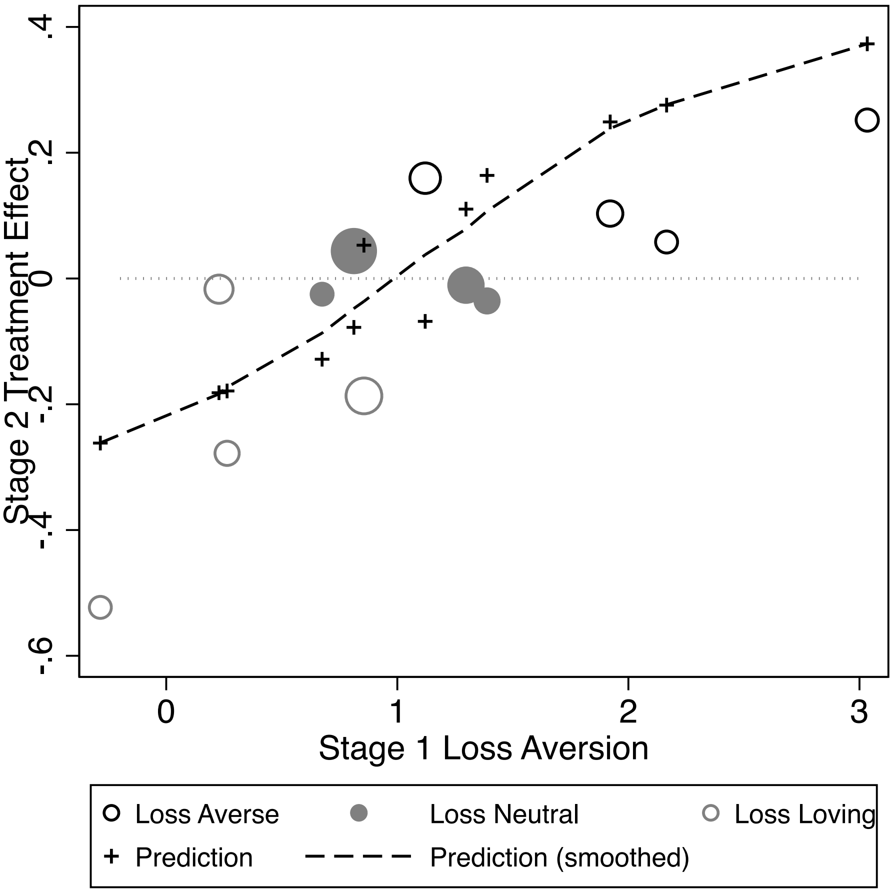


**Figure 3.4.** Loss aversion and treatment effects. Stage 1 Loss Aversion corresponds to the lower bound (upper bound) of $\hat{\lambda}$ for those who prefer their own endowment (prefer the other good), and the midpoint for those who rate the goods the same. Treatment Effect refers to the difference in the probability of exchange between forced exchange and baseline. Circles represents the treatment effect for each group in the data, with size of circle corresponding to number of observations. *Prediction* uses the estimated structural parameters (at the relevant bounds and midpoints) to calculate the logit probability of exchange.

3.4.1.[22] Given that these treatment effects will depend on Stage 2 assignment, and the corresponding aggregate utility values, Figure 3.4 also provides a locally

[22] This formulation maps PE values to choices via the assumption that an agent will exchange based on the probability that they cannot support not exchanging as a PE. For someone endowed with good $X$ in the baseline condition, the exchange probability is thus calculated as the probability of choosing $Y$:

weighted smoothed prediction. The magnitude of observed treatment effects are broadly in line with those predicted from the structural analysis of Stage 1 behavior.

### 3.4.3 Additional Results: Subjective Experience and Exchange

The comparative statics and magnitude of treatment effects match well the theoretical developments. The taxonomy of loss aversion is respected in the responsiveness of behavior to forced exchange. Nonetheless, Table 3.5 highlights a marked reticence to exchange in the forced exchange condition overall. This “residual” endowment effect falls outside the narrow predictions of the KR model, which, under our assumptions, would predict 50 percent exchange in this condition. In Table 3.6, we explore the effects of experience on subsequent exchange behavior by linking the variation in experience in Stage 1 to Stage 2 exchanges controlling for interactions between treatment and loss aversion type. Columns (1) and (4) of Table 3.6 show that actual experience of having their endowed object confiscated and replaced with the alternative in Stage 1 is not statistically related to exchange behavior in Stage 2. Controlling for the interaction of treatment and type, simply experiencing exchange via confiscation and replacement does not engender a greater willingness to exchange. If anything, the effects are directionally negative, with the experience of confiscation and replacement in Stage 1 leading to lower trading probabilities in Stage 2.

In columns (2) and (5) of Table 3.6, we examine the correlation between subjective experience in Stage 1 and exchange behavior in Stage 2. Subjects with more positive subjective Stage 1 experiences are significantly more willing to exchange in Stage 2 controlling for type and treatment assignment. Columns (3) and (6) ensure that it is the subjective evaluation of this experience, rather than the objective event of confiscation that leads to changes in exchange behavior.

$$P(Choice\ =\ Y)_{Baseline} \quad = \quad \frac{\exp(U(0,\hat{Y}|X,0))}{\exp(U(0,\hat{Y}|X,0)) + \exp(U(X,\hat{0}|X,0))} = \frac{\exp(2\hat{Y} - \hat{\lambda}\hat{X})}{\exp(\hat{X}) + \exp(2\hat{Y} - \hat{\lambda}\hat{X})},$$

where $X$ and $Y$ are the aggregate utility values for the goods in question estimate in Table 3.2 and $\hat{\lambda}$ is the measure of loss aversion taken as the lower bound (upper bound, mid point) of the range of parameters for the relevant group in Table 3.3. Similarly, in the forced exchange condition, where not exchanging can be supported in PE based only on utility values, this is

$$P(Choice\ =\ Y)_{Forced} \quad = \quad \frac{\exp(\hat{Y})}{\exp(\hat{X}) + \exp(\hat{Y})}.$$

And the predicted treatment effect is calculated as

$$Prediction\ =\ P(Choice\ =\ Y)_{Forced} - P(Choice\ =\ Y)_{Baseline}.$$

**Table 3.6.** Stage 1 experience and stage 2 exchange behavior

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | | | *Dependent Variable: Exchange (=1)* | | | |
| Lost Stage 1 Endowment | -0.049 | | -0.033 | -0.053 | | -0.038 |
| | (0.039) | | (0.039) | (0.039) | | (0.040) |
| $\Delta$ Happiness (Stage 1) | | 0.020*** | 0.019** | | 0.020*** | 0.019** |
| | | (0.007) | (0.007) | | (0.007) | (0.007) |
| Constant | 0.453 | 0.428 | 0.445 | 0.475 | 0.453 | 0.472 |
| | (0.069) | (0.067) | (0.070) | (0.065) | (0.061) | (0.065) |
| Treatment X Structural Taxonomy | Yes | Yes | Yes | No | No | No |
| Treatment X Reduced Form Taxonomy | No | No | No | Yes | Yes | Yes |
| R-Squared | 0.041 | 0.051 | 0.052 | 0.018 | 0.026 | 0.027 |
| # Observations | 607 | 607 | 607 | 607 | 607 | 607 |

*Notes*: Ordinary least squares or two-stage least squares regression. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The findings of Table 3.6 highlight the importance of the subjective perception of experience. Objectively being forced to exchange seems less critical than the subjective representation of this experience for fostering future exchange. An understanding of the subjective perception of experience helps to evaluate research on the persistence of exchange anomalies like the endowment effect (Engelmann and Hollard, 2010; List, 2003, 2004). The view from this research indicates that the endowment effect is reduced by experiences of exchange, and even a minute body of experience (over the course of one experimental session in Engelmann and Hollard (2010)) can eliminate the phenomena. Our data show that it is not the objective experience of exchanging one item for another which fosters market participation, but rather its subjective evaluation. Importantly, our results should not be read as inconsistent with those of Engelmann and Hollard (2010). Their design makes an explicit connection between exchange and positive experience as subjects must trade their endowed item in order to keep anything. As such trade is very likely to be viewed as subjectively positive and naturally leads to increased trading behavior.[23] Beyond such short-term experiments, our data also help to contextualize longer-term results such as List (2003, 2004), who shows that more experienced traders are less likely to exhibit an endowment effect. Though exchange should, on average, be a positive experience, it need not be uniformly so. Our data indicate that negative subjective

[23] We discuss further differences in the implementation of trading experience between our design and the previous examination of Engelmann and Hollard (2010) in Appendix 3.B.1.

evaluations of exchange may slow the speed at which the endowment effect is eliminated by market experience.

## 3.5 Discussion and Conclusion

Expectations-based reference-dependent preferences (Kőszegi and Rabin, 2006) (KR) represent a key advance in behavioral economics, but a host of conflicting evidence for the theory exists. In this paper we aimed to reconcile this conflicting evidence by explicitly recognizing and evaluating heterogeneity in loss aversion. Heterogeneity is critical both because the model's comparative statics can change sign depending on the level of loss aversion, and because prior work has noted that loss aversion is, by no means, a universal characteristic.

We measure loss aversion by evaluating ranking statements for a first bundle of goods without choice, and then place subjects in an exchange environment where they make choices over a second, different bundle of goods. We show that explicitly accounting for the heterogeneity in loss aversion by and large restores behavior in line with KR predictions. Individuals that are measured to be loss averse for the first bundle of goods deliver a substantial endowment effect for the second bundle, validating our taxonomy of types. Using a mechanism of probabilistic forced exchange, we then show that individuals who are measured to be loss averse grow more willing to exchange when probabilistically forced to do so; and individuals who are measured as loss loving grow less willing to exchange. These findings, and the magnitudes of the observed treatment effects are closely in line with the predictions of the KR model.

Our results help to reconcile conflicting results in the empirical study of the KR model (Ericson and Fuster, 2011; Goette et al., 2016; Heffetz and List, 2014) and follow naturally from the broad recognition of heterogeneity in loss aversion (Chapman et al., 2017; Erev et al., 2008; Harinck et al., 2007; Knetsch and Wong, 2009; Nicolau, 2012; Sokol-Hessner et al., 2009; Sprenger, 2015). If we are to recognize that loss aversion is not a universal trait, we must also recognize it as a confound of first-order importance for the KR model.

The conclusions drawn from this work rely on ex-ante measurement of the taxonomy of loss aversion. Though predicted and actual treatment effects generally coincide, our measures of loss aversion are admittedly broad. Future work could tighten the prediction using more refined measurements. Of course, more refined measurements come with potential challenges. If measurement is based on subject choices (e.g., for willingness to pay for lotteries), these choices themselves must be evaluated as part of a rational expectations equilibrium plan. Overcoming this joint challenge would represent a helpful advance over the current work.

Even accounting for KR forces, our data show a residual endowment effect of subjects being generally unwilling to exchange. Our results shed light on the mechanisms underlying such behavior. We show that unwillingness to exchange is related to prior experience, particularly the subjective perception thereof. Negative experience, regardless of objective outcome, leads to decreased exchange. This result may helpfully add to the literature on experience effects and exchange anomalies (Engelmann and Hollard, 2010; List, 2003, 2004), showing that exchange experience, even short-lived, can reduce the endowment effect. If the perception of experience influences subsequent exchange, it is possible for exchange anomalies to persist. Though exchange should generally be viewed as a positive event, with both parties gaining from trade, negative ex-post perceptions may still engender a hesitance to trade. Experiments which make explicit connection between trade and positive experience may thus be overstating the speed at which the endowment effect dissipates. And, provided that exchange is not a uniformly positive event, exchange anomalies may indeed persist with experience.

## References

**Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman (2011):** "Reference points and effort provision." *The American Economic Review*, 470–492. [100]

**Baucells, Manel and Rakesh K. Sarin (2007):** "Satiation in discounted utility." *Operations Research*, 55 (1), 170–181. [99]

**Bell, David E. (1985):** "Disappointment in decision making under uncertainty." *Operations Research*, 33 (1), 1–27. [99]

**Bradley, Margaret and Peter Lang (1994):** "Measuring emotion: the self-assessment manikin and the semantic differential." *Journal of Behavior Therapy and Experimental Psychiatry*, 25 (1), 49–59. [109]

**Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler (1997):** "Labor supply of New York City cabdrivers: One day at a time." *The Quarterly Journal of Economics*, 112 (2), 407–441. [99]

**Cantillo, Víctor, Johanna Amaya, and Juan de Dios Ortúzar (2010):** "Thresholds and indifference in stated choice surveys." *Transportation Research Part B*, 44, 753–763. [114]

**Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer (2017):** "Willingness to pay and willingness to accept are probably less correlated than you think." Tech. rep. National Bureau of Economic Research. [101, 107, 124]

**DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F. Schmieder (2017):** "Reference-dependent job search: evidence from Hungary." *The Quarterly Journal of Economics*, 132 (4), 1969–2018. [99]

**Engelmann, Dirk and Guillaume Hollard (2010):** "Reconsidering the effect of market experience on the "endowment effect"." *Econometrica*, 78 (6), 2005–2109. [102, 123, 125, 128, 129]

**Erev, Ido, Eyal Ert, and Eldad Yechiam (2008):** "Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions." *Journal of Behavioral Decision Making*, 21 (5), 575–597. [100, 107, 124]

**Ericson, Keith M. Marzilli and Andreas Fuster (2011):** "Expectations as endowments: evidence on reference-dependent preferences from exchange and valuation experiments." *The Quarterly Journal of Economics*, 126 (4), 1879–1907. [100, 102, 124]

**Fehr, Ernst and Lorenz Goette (2007):** "Do workers work more when wages are high? Evidence from a randomized field experiment." *The American Economic Review*, 97 (1), 298–317. [100]

**Fischbacher, Urs (2007):** "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental economics*, 10 (2), 171–178. [111]

**Gneezy, Uri, Lorenz Goette, Charles Sprenger, and Florian Zimmermann (2017):** "The limits of expectations-based reference dependence." *Journal of the European Economic Association*, 15 (4), 861–876. [100]

**Goette, Lorenz, Anette Harms, and Charles Sprenger (2016):** "Randomizing endowments: an experimental study of rational expectations and reference-dependent preferences." IZA Discussion Paper. [100–102, 106–108, 124]

**Harinck, Fieke, Eric Van Dijk, Ilja Van Beest, and Paul Mersmann (2007):** "When gains loom larger than losses reversed loss aversion for small amounts of money." *Psychological Science*, 18 (12), 1099–1105. [100, 107, 124]

**Heffetz, Ori and John A. List (2014):** "Is the endowment effect an expectations effect?" *Journal of the European Economic Association*, 12 (5), 1396–1422. [100, 102, 124]

**Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1990):** "Experimental tests of the endowment effect and the coase theorem." *Journal of Political Economy*, 98 (6), 1325–1348. [99]

**Kahneman, Daniel and Amos Tversky (1979):** "Prospect theory: an analysis of decision under risk." *Econometrica*, 47 (2), 263–291. [99]

**Knetsch, Jack and Wei-Kang Wong (2009):** "The endowment effect and the reference state: Evidence and manipulations." *Journal of Economic Behavior & Organization*, 71 (2), 407–413. [124]

**Kőszegi, Botond and Matthew Rabin (2006):** "A model of reference-dependent preferences." *The Quarterly Journal of Economics*, 121 (4), 1133–1165. [99, 124]

**Kőszegi, Botond and Matthew Rabin (2007):** "Reference-dependent risk attitudes." *The American Economic Review*, 97 (4), 1047–1073. [99]

**Lattin, J. M. and R. E. Bucklin (1989):** "Reference effects of price and promotion on brand choice behavior." *Journal of Marketing Research*, 26 (3), 299–310. [99]

**List, John A. (2003):** "Does market experience eliminate market anomalies?" *The Quarterly Journal of Economics*, 118 (1), 41–71. [102, 123, 125]

**List, John A. (2004):** "Neoclassical theory versus prospect theory: Evidence from the marketplace." *Econometrica*, 72 (2), 615–625. [102, 123, 125]

**Loomes, Graham and Robert Sugden (1986):** "Disappointment and dynamic consistency in choice under uncertainty." *The Review of Economic Studies*, 53 (2), 271–82. [99]

**Markowitz, Harry (1952):** "The utility of wealth." *Journal of Political Economy*, 60 (2), 151–158. [99]

**McFadden, Daniel (1974):** "Conditional logit analysis of qualitative choice behavior." In *Frontiers in Econometrics*. Ed. by Paul Zarembka. New York: Academic Press. Chap. 4. [113]

**Nicolau, Juan L. (2012):** "Asymmetric tourist response to price: loss aversion segmentation." *Journal of Travel Research*, 51 (5), 568–676. [100, 107, 124]

**Odean, Terrance (1998):** "Are investors reluctant to realize their losses?" *The Journal of Finance*, 53 (5), 177–1798. [99]

**Plott, Charles R. and Kathryn Zeiler (2005):** "The willingness to pay-willingness to accept gap, the "endowment effect," subject misconceptions and experimental procedures for eliciting valuations." *The American Economic Review*, 95 (3), 530–545. [110, 111, 129]

**Plott, Charles R. and Kathryn Zeiler (2007):** "Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory?" *The American Economic Review*, 97 (4). [110, 111, 129]

**Putler, D. S. (1992):** "Incorporating reference price effects into a theory of consumer choice." *Marketing Science*, 11 (3), 287–309. [99]

**Rabin, Matthew (2000):** "Risk aversion and expected utility theory: a calibration theorem." *Econometrica*, 68 (5), 1281–1292. [99]

**Shefrin, Hersh M. and M. Statman (1985):** "The disposition to sell winners too early and ride losers too long: Theory and evidence." *The Journal of Finance*, 40 (3), 777–790. [99]

**Sokol-Hessner, Peter, Ming Hsu, Nina G. Curley, Mauricio R. Delgado, Colin F. Camerer, and Elizabeth A. Phelps (2009):** "Thinking like a trader selectively reduces individuals' loss aversion." *Proceedings of the National Academy of Sciences*, 106 (13), 5035–5040. [100, 124]

**Sprenger, Charles (2015):** "An endowment effect for risk: Experimental tests of stochastic reference points." *Journal of Political Economy*, 123 (6), 1456–1499. [100, 107, 124]

**Wathieu, Luc (1997):** "Habits and the anomalies in intertemporal choice." *Management Science*, 43 (11), 1552–1563. [99]

## Appendix 3.A Additional Analyses and Robustness Tests

### 3.A.1 Robustness Specifications

Below we display regression results with standard errors clustered at the session level. Table 3.A.1 corresponds to Table 3.5 in the main text.

All analyses in the main text are based on a taxonomy of preference types based on liking scores. Below (Figure 3.A.1, Table 3.A.2, and Table 3.A.3) we report corresponding analyses for a categorization using the wanting scores elicited for the endowed and alternative good in Stage 1.

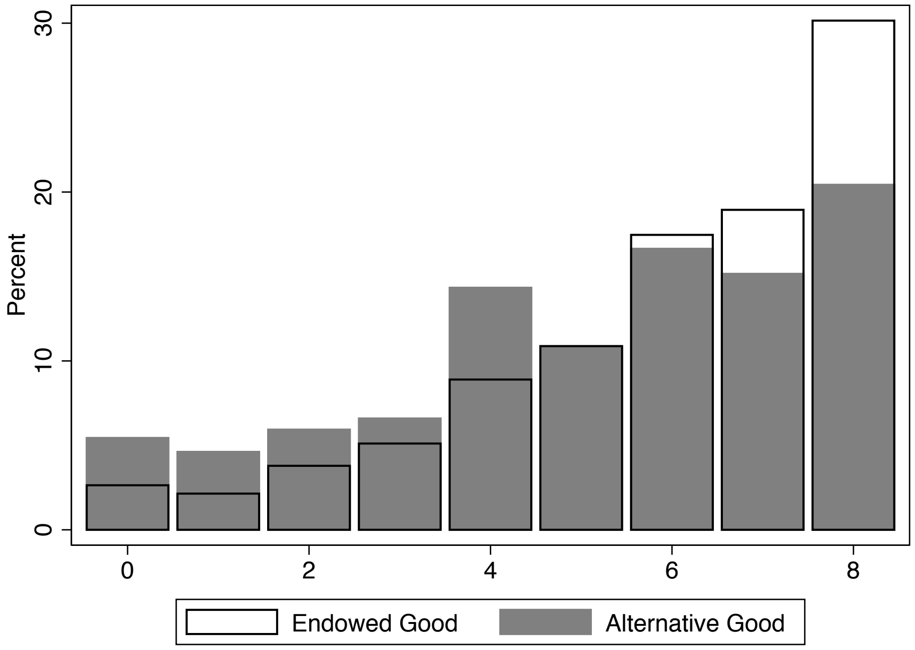


**Figure 3.A.1.** Preferences and endowments. Self-reported scores of wanting for the endowed and alternative goods. (Wilcoxon signed-rank statistic $z = 5.86$ ($p < 0.01$), N=607).

## Appendix 3.B Comments on Related Literature

### 3.B.1 Engelmann and Hollard (2010)

In the laboratory study of Engelmann and Hollard (2010), subjects play three trading rounds prior to a final trading situation with the experimenter. They show that the endowment effect in the final (voluntary) exchange vanishes for those who have been forced to trade their endowed good in the training rounds, but persists for those who were allowed to voluntarily exchange during the training round. The implementation of forced trading in Engelmann and Hollard (2010) differs from ours in important ways. In the training rounds, subjects in the treatment group are forced to exchange in the sense that otherwise, they lose their endowed good and do not receive anything in return. If endowed with good X, subjects choose between a situation of exchanging X for good Y, and losing X without receiving Y either. This way, the training rounds, in a broad sense, "condition" subjects to perceive exchange favorably by exposing them to the threat of leaving empty-handed. One explanation that reconciles their findings with our observation that the *subjective* perception of experience, i.e. its valence, determines subsequent willingness to trade, is that subjects who were forced to trade three times in a row in Engelmann and Hollard (2010) precisely grew more willing to trade because they learned to associate the no trade choice with not getting anything. Our notion of exchange – be it forced or voluntary – is motivated by typical trading situations and involves giving up the endowment *in return for something else*, instead of sacrificing the endowment for nothing in return. In any case, participants in the trade-it-or-lose-it design of Engelmann and Hollard (2010) were indeed more likely to make a subjectively positive experience than in our setting.[24]

Conceptually, Engelmann and Hollard (2010) attribute exchange asymmetries to "trade uncertainty" about market procedures, specifically that individuals misperceive and exaggerate the costs or risks of market transactions absent experience. Even in the simple, stylized and short-lived experimental exchange setting, they suggest, people grow accustomed to trading in that the perceived risks or costs of trading decrease significantly.[25] While the notion of trade uncertainty is recognizably broad and potentially incorporates our valence finding, our experimental design attempts to eliminate potential sources of uncertainty about the trading mechanism by giving it a precise, transparent and simple structure, as well as by limiting social interaction.

[24] A more critical interpretation is that training people to trade under the threat of losing leaving empty-handed otherwise is susceptible to experimenter demand effects because subjects could infer that the experimenter wants them to trade in the subsequent exchange situation.

[25] The training and second stages of Engelmann and Hollard (2010) still differ in important dimensions, however: The training rounds take place in a setting "without any restriction" where subjects can "interact, bargain, move, and so on" (Engelmann and Hollard, 2010, p.2008), while the final round is set in an isolated room facing the experimenter alone. The general setup is subject to a methodological criticism of laboratory exchange situations (Plott and Zeiler, 2005, 2007), because exchange is implemented as a direct social interaction that triggers, e.g., social comparison processes.

**Table 3.A.1.** Exchange behavior and probabilistic forced exchange: Clustered standard errors

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | *Dependent Variable: Exchange (=1)* | | | |
| | Full Sample | Loss Averse | Loss Neutral | Loss Loving |
| *Panel A: Structural Bounds Taxonomy* | | | | |
| Probabilistic Forced Exchange | 0.004 | 0.158*** | 0.027 | -0.248*** |
| | (0.034) | (0.050) | (0.071) | (0.060) |
| Baseline Exchange (Constant) | 0.365 | 0.330 | 0.361 | 0.429 |
| | (0.028) | (0.042) | (0.055) | (0.049) |
| R-Squared | 0.000 | 0.025 | 0.001 | 0.072 |
| # Observations | 607 | 217 | 240 | 150 |
| $H_0$: No Baseline Endowment Effect | $F_{1,605}$=23.85 | $F_{1,215}$=16.44 | $F_{1,238}$=6.30 | $F_{1,148}$=2.13 |
| | $(p < 0.01)$ | $(p < 0.01)$ | $(p < 0.05)$ | $(p = 0.16)$ |
| $H_0$: No Forced Ex. Endowment Effect | $F_{1,605}$=40.85 | $F_{1,215}$=0.19 | $F_{1,238}$=6.23 | $F_{1,148} = 83.39$ |
| | $(p < 0.01)$ | $(p = 0.67)$ | $(p < 0.05)$ | $(p < 0.01)$ |
| $H_0$: Baseline (col. 2) = Baseline (col. 4) | | | | $\chi^2(1) = 1.95$ |
| | | | | $(p = 0.16)$ |
| $H_0$: Forced Ex. (col. 2) =Forced Ex(col. 4) | | | | $\chi^2(1) = 25.61$ |
| | | | | $(p < 0.01)$ |
| *Panel B: Reduced Form Taxonomy* | | | | |
| Probabilistic Forced Exchange | 0.004 | 0.119** | -0.030 | -0.149* |
| | (0.034) | (0.052) | (0.096) | (0.087) |
| Baseline Exchange (Constant) | 0.365 | 0.304 | 0.392 | 0.448 |
| | (0.028) | (0.032) | (0.070) | (0.076) |
| R-Squared | 0.000 | 0.015 | 0.001 | 0.022 |
| # Observations | 607 | 285 | 131 | 191 |
| $H_0$: No Baseline Endowment Effect | $F_{1,605}$=23.85 | $F_{1,283}$=36.46 | $F_{1,129}$=2.39 | $F_{1,189}$=0.47 |
| | $(p < 0.01)$ | $(p < 0.01)$ | $(p = 0.13)$ | $(p = 0.50)$ |
| $H_0$: No Forced Ex. Endowment Effect | $F_{1,605}$=40.85 | $F_{1,283}$=3.46 | $F_{1,129}$=4.25 | $F_{1,189} = 22.79$ |
| | $(p < 0.01)$ | $(p = 0.073)$ | $(p = 0.048)$ | $(p < 0.01)$ |
| $H_0$: Baseline (col. 2) = Baseline (col. 4) | | | | $\chi^2(1) = 2.88$ |
| | | | | $(p = 0.090)$ |
| $H_0$: Forced Ex. (col. 2) =Forced Ex(col. 4) | | | | $\chi^2(1) = 6.22$ |
| | | | | $(p < 0.05)$ |

*Notes*: Ordinary least square regression. Standard errors clustered at session level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Null hypotheses tested for 1) zero baseline endowment effect, regression (Constant = 0.5); 2) zero forced exchange endowment effect (Constant + Forced Exchange = 0.5); 3) Identical baseline behavior across loss averse and loss loving agents (Constant (col. 2) = Constant (col. 4)); 4) Identical treatment effects of forced exchange across loss averse and loss loving agents (Forced Exchange (col. 2) = Forced Exchange (col. 4)). Hypotheses 3 and 4 tested via seemingly unrelated regression. Panel A: taxonomy of types based on structural bounds from Table 3.3. Panel B: taxonomy of types based on reduced form rating statements from Table 3.3.

**Table 3.A.2.** Aggregate parameter estimates: Based on wanting scores

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Estimate | (Std. Error) | Estimate | (Std. Error) |
| | *Bundle 1* | | *Bundle 2* | |
| *Loss Aversion:* | | | | |
| $\hat{\lambda}$ | 1.617 | (0.132) | 1.346 | (0.113) |
| *Utility Values:* | | | | |
| $\hat{X}_1$ *(Pen Set)* | 0.674 | (0.049) | | |
| $\hat{Y}_1$ *(USB Stick)* | 1 | - | | |
| $\hat{X}_2$ *(Picnic Mat)* | | | 0.927 | (0.050) |
| $\hat{Y}_2$ *(Thermos)* | | | 1 | - |
| *Discernibility:* | | | | |
| $\hat{\delta}$ | 0.557 | (0.060) | 0.478 | (0.053) |

*Notes*: Maximum likelihood estimates. Robust standard errors in parentheses.

**Table 3.A.3.** Exchange behavior and probabilistic forced exchange: Type categorization based on wanting scores

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | *Dependent Variable: Exchange (=1)* | | | |
| | Full Sample | Loss Averse | Loss Neutral | Loss Loving |
| *Panel A: Structural Bounds Taxonomy* | | | | |
| Probabilistic Forced Exchange | 0.004 | 0.129** | -0.003 | -0.177** |
| | (0.040) | (0.065) | (0.067) | (0.081) |
| Baseline Exchange (Constant) | 0.365 | 0.327 | 0.383 | 0.407 |
| | (0.032) | (0.048) | (0.054) | (0.067) |
| R-Squared | 0.000 | 0.017 | 0.000 | 0.035 |
| # Observations | 607 | 223 | 239 | 145 |
| $H_0$: No Baseline Endowment Effect | $F_{1,605}$=18.32 | $F_{1,221}$=13.29 | $F_{1,237}$=4.68 | $F_{1,143}$=1.89 |
| | $(p < 0.01)$ | $(p < 0.01)$ | $(p < 0.05)$ | $(p = 0.17)$ |
| $H_0$: No Forced Ex. Endowment Effect | $F_{1,605}$=27.48 | $F_{1,221}$=0.97 | $F_{1,237}$=9.62 | $F_{1,143} = 36.65$ |
| | $(p < 0.01)$ | $(p = 0.33)$ | $(p < 0.01)$ | $(p < 0.01)$ |
| $H_0$: Baseline (col. 2) = Baseline (col. 4) | | | | $\chi^2(1) = 0.97$ |
| | | | | $(p = 0.32)$ |
| $H_0$: Forced Ex. (col. 2) =Forced Ex(col. 4) | | | | $\chi^2(1) = 8.78$ |
| | | | | $(p < 0.01)$ |
| *Panel B: Reduced Form Taxonomy* | | | | |
| Probabilistic Forced Exchange | 0.004 | 0.103* | -0.093 | -0.092 |
| | (0.040) | (0.056) | (0.085) | (0.079) |
| Baseline Exchange (Constant) | 0.365 | 0.297 | 0.439 | 0.431 |
| | (0.032) | (0.042) | (0.066) | (0.065) |
| R-Squared | 0.000 | 0.011 | 0.009 | 0.008 |
| # Observations | 607 | 293 | 138 | 176 |
| $H_0$: No Baseline Endowment Effect | $F_{1,605}$=18.32 | $F_{1,291}$=23.24 | $F_{1,136}$=0.86 | $F_{1,187}$=1.11 |
| | $(p < 0.01)$ | $(p < 0.01)$ | $(p = 0.36)$ | $(p = 0.29)$ |
| $H_0$: No Forced Ex. Endowment Effect | $F_{1,605}$=27.48 | $F_{1,291}$=7.24 | $F_{1,136}$=8.40 | $F_{1,187} = 13.50$ |
| | $(p < 0.01)$ | $(p < 0.01)$ | $(p < 0.01)$ | $(p < 0.01)$ |
| $H_0$: Baseline (col. 2) = Baseline (col. 4) | | | | $\chi^2(1) = 3.01$ |
| | | | | $(p = 0.08)$ |
| $H_0$: Forced Ex. (col. 2) =Forced Ex(col. 4) | | | | $\chi^2(1) = 4.12$ |
| | | | | $(p < 0.05)$ |

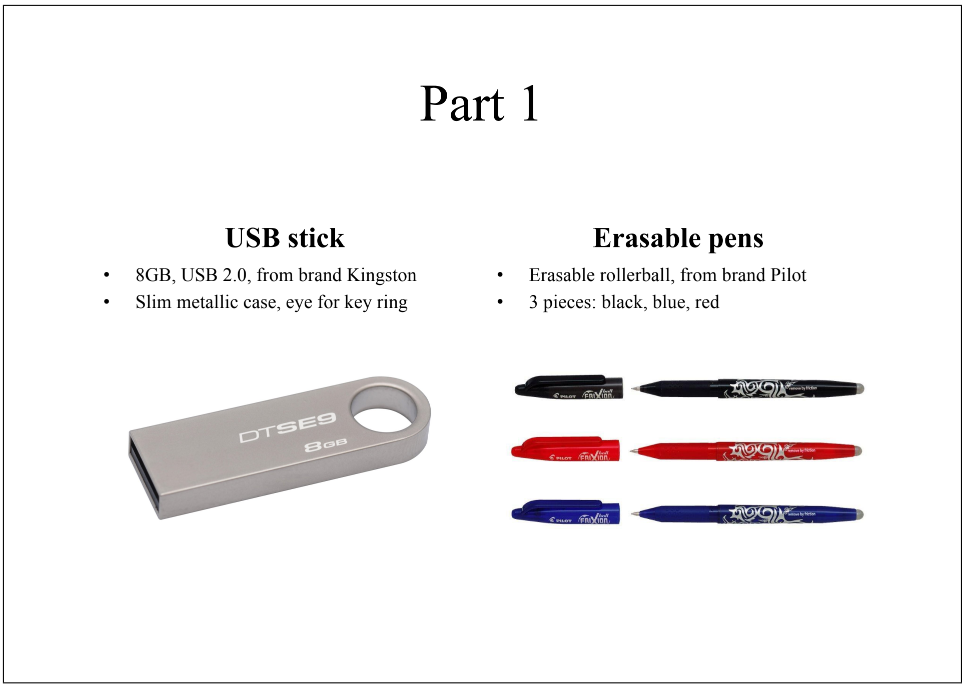
*Notes*: Ordinary least square regression. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Null hypotheses tested for 1) zero baseline endowment effect, regression (Constant = 0.5); 2) zero forced exchange endowment effect (Constant + Forced Exchange = 0.5); 3) Identical baseline behavior across loss averse and loss loving agents (Constant (col. 2) = Constant (col. 4)); 4) Identical treatment effects of forced exchange across loss averse and loss loving agents (Forced Exchange (col. 2) = Forced Exchange (col. 4)). Hypotheses 3 and 4 tested via seemingly unrelated regression. Panel A: taxonomy of types based on structural bounds from wanting scores. Panel B: taxonomy of types based on reduced form rating statements from wanting scores.

# Appendix 3.C Instructions and Material Presented to Participants

All instructions and information presented to participants have been translated from German to English.

## 3.C.1 Images of Objects Presented to Participants

The following images (Figure 3.C.2 and Figure 3.C.3) were projected to the wall of the lecture room at the beginning of the respective stage. For the displayed example, the Stage 1 bundle consisted of the USB stick and erasable pens, but this was counter-balanced at the session level.



Part 1

**USB stick**

- 8GB, USB 2.0, from brand Kingston
- Slim metallic case, eye for key ring

**Erasable pens**

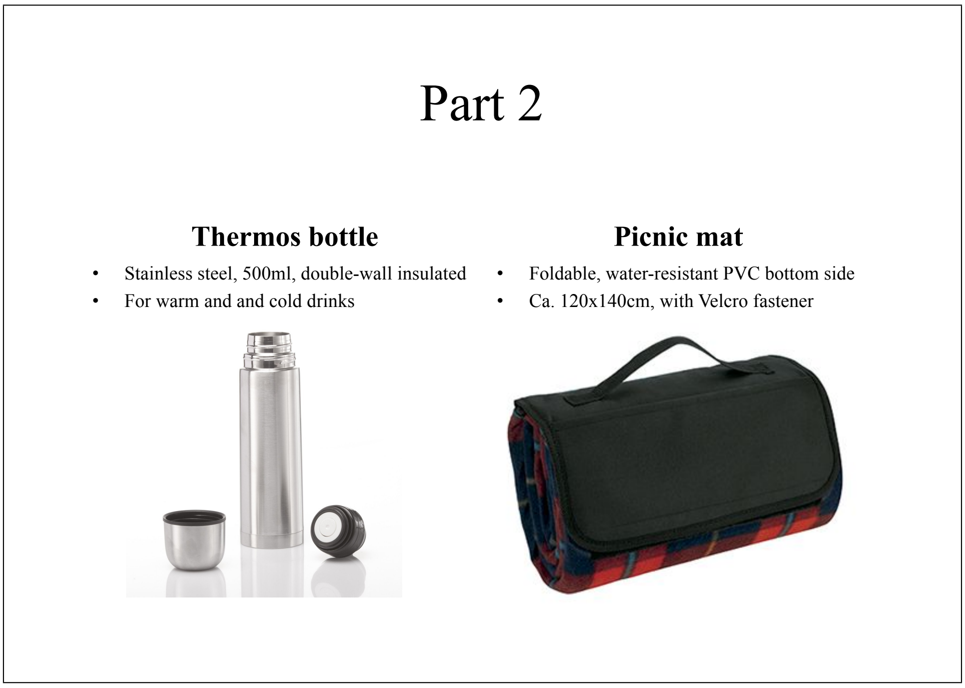
- Erasable rollerball, from brand Pilot
- 3 pieces: black, blue, red

**Figure 3.C.2.** Image 1 projected on the wall to present objects. For Stage 1 with goods bundle consisting of USB stick and erasable pens.

## 3.C.2 Instructions (Computer-Based)

**Welcome to part 1 of 2 in this experiment!**

Please close the curtain of you cabin and read the following information. All computer entries that you make in this experiment are fully anonymous and cannot be traced back to you. Speed is not important at any point in this experiment. Please always take sufficient time to read and understand the



Part 2

**Thermos bottle**

- Stainless steel, 500ml, double-wall insulated
- For warm and and cold drinks

**Picnic mat**

- Foldable, water-resistant PVC bottom side
- Ca. 120x140cm, with Velcro fastener

**Figure 3.C.3.** Image 2 projected on the wall to present objects. For Stage 2 with goods bundle consisting of thermos and picnic mat.

instructions.

The [ USB stick / erasable pens / thermos / picnic mat ] now belongs to you. You can touch and inspect it at any time. However, please do not yet open the packaging and do not use the object yet. The two objects presented to you ( [ USB stick and erasable pens / thermos and picnic mat ] ) have been randomly allocated to the cabins in equal quantities. Your cabin number was also randomly determined based on your choice of seat in the presentation room.
Please click on OK when you have read these information. If you have questions, please call an experimenter.

**Please answer the questions.**
[ USB stick / thermos ]
How much do you like this product?
How much would you want to have this product?

[ Erasable pens / picnic mat ]
How much do you like this product?
How much would you want to have this product?

**Please read the following information carefully.**

The experimenter will soon draw a random number between 1 and 20 using a lotto drum. The drawn number will then be announced loudly. If the drawn number is a number [ from 11 to 20 / from 1 to 10 ], your [ USB stick / erasable pens / thermos / picnic mat ] will be taken away from you and you instead receive [ USB stick / erasable pens / thermos / picnic mat ]. If the drawn number is a number [ from 1 to 10 / from 11 to 20 ], you will keep your [ USB stick / erasable pens / thermos / picnic mat ] and nothing happens. After the number has been drawn and the exchange of objects has taken place (if applicable), nothing else happens in this part of the experiment. You can then keep your object for good.

Please only confirm below once you have understood everything. If you have questions, please call the experimenter and wait until he comes to your cabin.

**[ Mood elicitation 1 ]**

Please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?

"Unhappy, Angry, Unsatisfied, Sad, Desperate" – "Happy, Thrilled, Satisfied, Content, Hopeful"

**The time has come. Please wait until the number has been drawn.**

Remember: If the drawn number is a number [ from 11 to 20 / from 1 to 10 ], your [ USB stick / erasable pens / thermos / picnic mat ] will be taken away from you and you instead receive [ USB stick / erasable pens / thermos / picnic mat ]. If the drawn number is a number [ from 1 to 10 / from 11 to 20 ], you will keep your [ USB stick / erasable pens / thermos / picnic mat ].

**The drawn number is [ 1 / 2 / ... / 20 ].**

This number is a number [ from 1 to 10 / from 11 to 20 ]. Therefore [ you can keep your [ USB stick / erasable pens / thermos / picnic mat ] / your [ USB stick / erasable pens / thermos / picnic mat ] will be taken away from you and you instead receive [ USB stick / erasable pens / thermos / picnic mat ] ].

Please wait while the experimenter carries out the exchange in all cabins.

**[ Mood elicitation 2 and control question. ]**

Please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?

"Unhappy, Angry, Unsatisfied, Sad, Desperate" – "Happy, Thrilled, Satisfied, Content, Hopeful”

Regarding the lottery draw, that has just taken place: What was the probability (in percent) that you would lose your initial object? Please enter a number between 0 and 100.

**Part 1 of the experiment is over!**
Please follow the instructions.

- Memorize your cabin number.
- You can no go back to the presentation room.
- Please leave your [ USB stick / erasable pens / thermos / picnic mat ] in the cabin. You will be back in the same cabin in a few minutes.
- Remember: The object now belongs to you for good and you will take it away from this experiment.

**Welcome to part 2 in this experiment!**
Please close the curtain of you cabin and read the following information. The [ USB stick / erasable pens / thermos / picnic mat ] now also belongs to you. You can touch and inspect it at any time. However, please do not yet open the packaging and do not use the object yet. The two objects presented to you for part 2 ( [ USB stick and erasable pens / thermos and picnic mat ] ) have again been randomly allocated to the cabins in equal quantities.
Please click on OK when you have read these information. If you have questions, please call an experimenter.

**[ Instructions Stage 2 – ONLY BASELINE (p=0.0) ]**
Please read the following information carefully. The [ USB stick / erasable pens / thermos / picnic mat ] from part 2 of the experiment now belongs to you and you can keep it for good. If you like, you can exchange your [ USB stick / erasable pens / thermos / picnic mat ] voluntarily for [ USB stick / erasable pens / thermos / picnic mat ]. Whichever way you decide, your choice is final and you will take your selected object with you from this experiment.
Please only confirm below once you have understood everything. If you have questions, please call the experimenter and wait until he comes to your cabin.

**[ Instructions Stage 2 – ONLY FORCED EXCHANGE (p=0.5) ]**
Please read the following information carefully. You have received a new object

in part 2 of the experiment ( [ USB stick / erasable pens / thermos / picnic mat ] ). You will soon get the opportunity to exchange your [ USB stick / erasable pens / thermos / picnic mat ] voluntarily for [ USB stick / erasable pens / thermos / picnic mat ].

If you decide to exchange, you will receive [ USB stick / erasable pens / thermos / picnic mat ] as requested for your [ USB stick / erasable pens / thermos / picnic mat ] and you can then keep your [ USB stick / erasable pens / thermos / picnic mat ] for good. The experiment is then finished.

If you decide against an exchange, there will be a probability of 50% that the exchange will be forced anyways and you have to exchange nevertheless.

Concretely, the following happens in the case that you decide against a voluntary exchange: The experimenter will draw a random number between 1 and 20 using a lotto drum (as in part 1 of the experiment). The drawn number will then be announced loudly. If the drawn number is a number [ from 11 to 20 / from 1 to 10 ], your [ USB stick / erasable pens / thermos / picnic mat ] will be taken away from you and you instead receive [ USB stick / erasable pens / thermos / picnic mat ]. If the drawn number is a number [ from 1 to 10 / from 11 to 20 ], you will keep your [ USB stick / erasable pens / thermos / picnic mat ] and nothing happens. After the number has been drawn and the exchange of objects has taken place (if applicable), nothing else happens in this part of the experiment. You can then keep your object for good.

Please only confirm below once you have understood everything. If you have questions, please call the experimenter and wait until he comes to your cabin.

**[ Mood elicitation 3 ]**

Before you get the opportunity to exchange your object, please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?

“Unhappy, Angry, Unsatisfied, Sad, Desperate” – “Happy, Thrilled, Satisfied, Content, Hopeful”

**Do you want to exchange your [ USB stick / erasable pens / thermos / picnic mat ] for a [ USB stick / erasable pens / thermos / picnic mat ]?**

Yes, I want to exchange.

No, I do not want to exchange.

**[ ONLY BASELINE (p=0.0) ]**

You have decided [ for / against ] a voluntary exchange. Please wait while the experimenter carries out the exchange in all cabins.

**[ ONLY FORCED EXCHANGE (p=0.5) ]**

You have decided [ for / against ] a voluntary exchange. Please wait while the experimenter carries out the exchange in all cabins.
[ ONLY NON-TRADERS ] After this, it will be determined whether you have to exchange anyways.
[ ONLY TRADERS ] Please wait until the experiment continues. A random number will now be drawn for those who decided against a voluntary exchange. After that the experiment continues for you.
[ ONLY NON-TRADERS ] Remember: If the drawn number is a number [ from 11 to 20 / from 1 to 10 ], your [ USB stick / erasable pens / thermos / picnic mat ] will be taken away from you and you instead receive [ USB stick / erasable pens / thermos / picnic mat ]. If the drawn number is a number [ from 1 to 10 / from 11 to 20 ], you will keep your [ USB stick / erasable pens / thermos / picnic mat ].
[ ONLY NON-TRADERS ]
The drawn number is [ 1 / 2 / ... / 20 ]
This number is a number [ from 1 to 10 / from 11 to 20 ]. Therefore [ you can keep you [ USB stick / erasable pens / thermos / picnic mat ] / your [ USB stick / erasable pens / thermos / picnic mat ] will be taken away from you and you instead receive [ USB stick / erasable pens / thermos / picnic mat ]. Please wait while the experimenter carries out the exchange in all cabins.

**[ Mood elicitation 4 ]**

Please answer the following questions about how you currently feel. Which expressions better apply to you at the moment?
“Unhappy, Angry, Unsatisfied, Sad, Desperate” – “Happy, Thrilled, Satisfied, Content, Hopeful”

**The experiment is over!**

You can keep both your objects. You will also receive a show-up fee of 4 euros. Please wait shortly in you cabin until the experimenter calls you out. Thank you for your participation!

# 4

# Breaking Trust: On the Persistent Effect of Economic Crisis Experience

*Joint with Tom Zimmermann*

## 4.1 Introduction

Trust is a fundamental prerequisite for economic exchange. A mounting body of evidence shows that measured levels of trust vary substantially across geographical locations and over time (Falk et al., 2017; Nannestad, 2008; Robinson and Jackson, 2001). This affects economic outcomes at the individual, group and societal levels.[1] But what are the determinants of the geographical and temporal variation of trust? This paper provides evidence that trust is partially determined by individuals' experience of catastrophic macroeconomic events that are jointly made by people living at the same place and at the same time.

Exploiting cohort, time and regional variation in different datasets, we find a persistent negative effect of living through times of financial crises on interpersonal trust. Our analysis is motivated by a *trust breach* argument, positing that trust is not easily restored once it has been violated (Lewicki and Bunker, 1995, 1996; Lewicki and Wiethoff, 2000; Slovic, 1993). In support of this hypothesis, the effect is only observed for crises in domains that strongly rely on trust. Banking crises and bank failures serve as a proxy for the increased likelihood of people in affected regions making adverse experience in trust-dependent

[1] Trust supports cooperative behavior (Gambetta, 1988), determines the performance of large organizations (Kramer and Tyler, 1996; Porta et al., 1997), reduces transaction costs and contributes to differences in growth between countries (Algan and Cahuc, 2010; Beugelsdijk et al., 2004; Knack and Keefer, 1997).

economic interactions.[2] We show that the aggregate consequences of financial crises operating through a population's tendency to distrust others are sizable and long-lived.

Identifying the effect of experience on trust comes with both theoretical and methodological challenges. First, there are different ways of how experience might accumulate over time, and each possibility implies a different definition for aggregated personal crisis experience. For example, a person's stock of personal crisis experience could depend both on the time distance to a past event and the person's age at that time, but there is little ex ante conceptual guidance to favor one specific formalization over another.

Second, macroeconomic crises are no ceteris paribus events, but have both concurrent and delayed effects experienced in conjunction with the crises. If financial crises were closely followed by recessions, then experience effects based on indicators for financial crises may falsely attribute the effect of living through recessions to the occurrence of financial crises. Therefore, disentangling different types of experience is essential.

Third, previous evidence on experience effects typically comes from either cross-country comparisons or within-country regional variation. These levels of comparisons require different identifying assumptions and findings in one domain are not easily generalized to the other.

Our empirical strategy addresses these concerns. First, we allow for different characterizations of accumulated crisis experience. Second, instead of estimating the effect of one specific type of macroeconomic crisis in isolation, we accommodate the fact that crises occur jointly with other crises and have repercussions in other economic domains. Our analyses account for the co-movement of different crises and measures of economic activity.

Third, we exploit two distinct individual-level data sets, one exploiting cross-country variation, and the other using within-country regional variation. In the first part of our analysis, we estimate the effect of personal experience of country-level financial and economics crises using trust data spanning over 30 years from a large-scale individual-level data set, the *World Values Survey*. In the second part we test the effect of local banking crises in the United States using detailed regional information on bank failures and more than 40 years of trust data from the *General Social Survey*. Together, our data allow to examine the *breach of trust* hypothesis about the origins of systematic heterogeneity in trust levels, using an empirical strategy that is more conservative than in previous experience studies, addressing potential confounds in the identification of experience effects.

[2] Our empirical approach assumes a broad notion of what constitutes such adverse experience. Both being directly affected – e.g., by losing assets due to a bank failure or having increased financial concerns – and being indirectly affected – e.g., by seeing friends or acquaintances lose money – potentially bears on trusting behavior.

We document two key findings from our cross-country study. First, we estimate a highly persistent effect of living through financial crises on trust in other people. Intriguingly, this effect is specific to those types of financial crises — in particular, banking crises – that occur in spheres in which individual behavior has been shown to strongly depend on trust, but not for crises in other, less trust-intensive areas, such as inflation or public debt crises. These findings hold in a broad range of specifications. Second, zooming in on specific channels, we identify two potential mechanisms: For one thing, individuals with relatively more banking crisis experience are *not* less likely to trust financial institutions, but they do tend to distrust political institutions. Giuliano and Spilimbergo (2014) showed that experiencing recessions, by contrast, increases trust in such institutions, which is replicated in our data. We document that banking crises that often occur simultaneously with recessions work in the other direction to offset the positive effect of recessions on confidence in the political system. For another thing, financial crisis experience does not shift measures of risk preferences. Instead, the effect on trust appears to operate via changes in beliefs.

Our benchmark estimate implies that experiencing a banking crisis in the previous year decreases trust by 1.6 percentage points. To put the effect size and horizon into perspective, a crisis that happened 20 years ago reduces trust today by about as much as one-third of the effect of a personal traumatic experience, e.g., a divorce or disease, in the preceding year.[3]

Our subsequent within-country study uses U.S. survey data from the *General Social Survey* in combination with historical information on bank failures provided by the *Federal Deposit Insurance Corporation* (FDIC). We reports two findings. First, our main results from the cross-country study are strongly supported using only within-country variation. Again, we control for a battery of cohort, age, region and survey year fixed effects, household-level characteristics, and recognize alternative ways for how experience accumulates. This replication in a different data set, at a different level of regional aggregation, using different measures to construct experience stocks speaks to the robustness of the identified relationship and provides external validation. Second, an advantage in this setting over the cross-country study is that we can construct a continuous measure of the severity of a local banking crisis, using information about the estimated losses from bank defaults. Our data confirm that experience stocks based on estimated losses are also associated with lower interpersonal trust levels.

We believe this paper makes three contributions. First, we provide evidence that sheds light on the determinants of trust over time and locations at the aggregate level. The literature on trust mostly focuses on its role as a determinant of

[3] Based on estimates of Alesina and La Ferrara (2002), who find that recent traumas decrease trust by 2-3 percentage points, only a bit more than the effect of a banking crisis experience in our benchmark specification, *but* the effects of banking crises are long-lasting.

individual or aggregate economic outcomes (Algan and Cahuc, 2010; Beugelsdijk et al., 2004; Knack and Keefer, 1997), or on the determinants of trusting behavior in personal interaction (Schwerter and Zimmermann, 2016). Our findings suggest that shared adverse experience in trust-sensitive areas permanently affects public trust levels. They hint at a potential feedback effect between trust and economic crises.

Second, this study points out a specific dimension of the welfare costs of economic disturbances that received little previous attention. For example, it has been argued that distrust is costly in personal interactions, organizations, and at the community level (Kosfeld and Falk, 2006; Kramer, 1999; Malul et al., 2010). As such, a persistent erosion of trust potentially increases the welfare burden of crashes such as the 2008/09 global financial crisis. Moreover, interpersonal trust has been linked to political preferences, in particular preferences for redistribution (Hetherington and Husser, 2012; Jaime-Castillo, 2016; Yamamura, 2014). By breaking interpersonal trust, macroeconomic crises could contribute to a diminished taste for redistribution in a society.

Third, we add to an emerging line of research devoted to studying how experiences shape preferences and beliefs. Previous work documents effects of macroeconomic experiences in the domains of risk taking (Malmendier and Nagel, 2011), inflation expectations (Malmendier and Nagel, 2015; Malmendier et al., 2017) and political preferences (Giuliano and Spilimbergo, 2014). Our paper shifts the focus to social preferences. Relatedly, Alesina and La Ferrara (2002) analyze correlates of trust in U.S. data. They report a negative relationship between trust and self-reported traumatic events within the previous year. Among other adverse experiences such as a divorce, diseases and financial misfortune exhibit the strongest negative correlation with trust. Moreover, our study complements work examining trust as a determinant of societal outcomes. Guiso et al. (2004) identify trust as a driver of the financial development of societies. In their view, a financial contract constitutes “the ultimate trust-intensive contract” (Guiso et al., 2004, p. 527). We analyze the reverse direction and show that trust levels are partly driven by the shared experience of financial meltdowns in the first place.

Our findings on the persistent negative effects of macroeconomic crises are remarkable in light of previous studies on the implications of other, non-economic forms of experience for trust. For example, Cassar et al. (2011) find a positive effect of natural disaster experience on trust and trustworthiness in their study of the 2004 tsunami in Thai villages. Looking at armed conflict in Uganda, De Luca and Verpoorten (2015) estimate a short-term negative effect of experiencing violence on trust, but find a rapid recovery in trust levels thereafter.

Section 4.2 sets out our empirical strategy. Section 4.3 discusses the first part of our analysis using cross-country data, and Section 4.4 present the second part that focuses on U.S. data. Section 4.5 concludes.

## 4.2 Empirical Strategy and Identification

We present the model specification and identification strategy, and discuss measurement of variables and data in the subsequent sections. Our baseline specification is akin to the approach taken in recent work on the *experience hypothesis*. We model trust as

$$\begin{aligned} trust_{ijt} &= \beta_0 + \beta_1 C_{ijt} + \beta_2 `X_{ijt} + crisis_{jt} + \gamma_a + \mu_t \\ &\quad + \nu_b + \alpha_j + \alpha_j \cdot age + \epsilon_{ijt}, \end{aligned} \tag{4.1}$$

where $trust_{ijt}$ is a self-reported measure of trust of individual $i$ in year $t$ and country $j$.

The parameter of interest is $\beta_1$, the effect of individual-level crisis experience $C_{ijt}$, which we model as an *experience stock* that an individual has accumulated over their lifetime. Our experience stock variable nests the main approaches used in the literature. Our preferred measure, which we call *delta experience,* is adapted from Fuchs-Schündeln and Schündeln (2015) and based on a simple depreciation interpretation of past experience. Delta experience assumes that the effect of crises fade over time and that past years are discounted exponentially based on their time distance relative to the survey year,

$$C_{it}^{delta}(\delta) = \sum_{s=t-age_{it}}^{t} \delta^{t-s} \mathbb{1}_{\{crisis_s\}}. \tag{4.2}$$

Here, $i$ denotes the individual and $t$ is the survey year. Crises are summed over an individual's lifetime up to the survey year, with weighting factor $\delta$. The indicator function takes value 1 if the individual experienced a crisis in year $s$ and 0 otherwise.

In this view, the time distance to a past crisis matters irrespective of age. This specification captures the notion of an *experience capital stock* that depreciates as time passes and is governed by a single parameter $\delta$. We favor this specification because of its transparency and simplicity while capturing the central insight that past events are forgotten over time ($\delta < 1$). At the same time, delta experience also nests the trivial case of no differential weighting of past events ($\delta = 1$).[4]

An alternative approach that puts the age range during which an event is experienced center stage but does not account for differential time distance to a past event is suggested in Giuliano and Spilimbergo (2014), who look at the

[4] Even a case with $\delta > 1$ is conceivable in this setup when, e.g., crisis experience becomes gradually less important the later in life it occurs. See below for an alternative model that explicitly considers age ranges to account for this idea.

effect of experience in specific age intervals.[5] Both considerations, time distance and age-specific effects, can be combined in a joint specification of the experience stock as in Malmendier and Nagel (2011). Malmendier and Nagel (2011) introduce a single parameter that simultaneously regulates how weights for past years depend on an individual's age back then and the time elapsed since that event, suggesting that a 20-year-old person with a relatively short life history may be swayed much more by an abnormal event than a 60-year-old who has a richer life experience. Table 4.1 provides an overview of alternative experience specifications considered in this paper. We investigate different plausible specifications to shed light on the importance of different modeling assumptions for our results.

**Table 4.1.** Overview of experience specifications

| Specification of experience | Approach recognizes… | |
|---|---|---|
| | time distance to an event at survey date | respondent's age at time of event |
| Delta experience<br>(Fuchs-Schündeln and Schündeln, 2015) | ✓ | |
| Age range experience<br>(Giuliano and Spilimbergo, 2014) | | ✓ |
| Lambda experience<br>(Malmendier and Nagel, 2011) | ✓ | ✓ |

*Notes.* Displayed are different definitions from the previous literature for how individual experience can be aggregated into a single measure.

We relegate further details and report results for our alternative specifications, i.e. *age range experience* and *lambda experience* to the appendix. To anticipate our results, we will show that the effect of some types of macroeconomic crises are sensitive to the choice of experience specification. Our main finding on the effect of banking crises or local banking failures on trust, however, is comparable in both magnitude and significance with those in the main text, regardless of how the experience stock is defined. To our knowledge, our paper is the first in the experience literature that considers multiple specifications for the experience stock and shows robustness of results across several such specifications.

Our identification is based on the inclusion of a myriad of fixed effects to absorb confounding variation. In particular, equation (4.1) includes unobserved age fixed effects ($\gamma_a$), time fixed effects ($\mu_t$), country fixed effects ($\alpha_j$),

[5] This formulation assumes that experiences at a certain age matter independently of elapsed time, i.e., experience does not fade. The effect of having experienced some event of interest during their 20s is the same for a person of age 31 and a person of age 65.

birth-year-cohort fixed effects ($\nu_b$), as well as country-specific linear age trends ($\alpha_j \cdot age$). Age fixed effects remove systematic life cycle effects in the evolution of trust. Survey-year fixed effects eliminate aggregate effects such as time-varying trust. Country fixed effects deal with unobserved heterogeneity in trust between countries. Because our experience measure varies not only between but also within cohorts over time, we can accommodate birth-year-cohort fixed effects that absorb unobserved cohort-level variation in trust.[6] To avoid attributing other time-varying country-specific factors to experience, we include country-specific age trends. Moreover, we use a full set of crisis-in-survey-year dummies ($crisis_{jt}$) to capture contemporaneous effects of living through a crisis on trust.

We further include observable individual-level characteristics, $X_{ijt}$, to isolate the effect of crisis experience. In some specifications, we additionally include experience stocks of other macroeconomic variables to rule out spurious attributions to specific types of crises. The latter approach is more conservative than the specifications in the extant literature that focus on a sole crisis type and do not control for simultaneous occurrences of other crises.

Our identification of $\beta_1$ comes from individual-level cross-sectional differences in trust and experience stocks and on changes in these cross-sectional differences over time. The variation in $C_{ijt}$ has two main components at the cross-sectional level: different geographical locations and age differences. The identifying assumption of our analysis is that there are no unobserved variables varying at the country-year-age-birth-year level that correlate with both our constructed individual experience measure and trust.

## 4.3 Cross-Country Evidence

Part 1 of our empirical analysis uses cross-country data to examine the effect of country-level macroeconomic shocks on trust.

### 4.3.1 Data and Measurement

Our data come from three sources. We use the *World Values Survey* (WVS) for individual data on trust and other individual characteristics. The WVS is a compilation of repeated cross-sections of national surveys at the individual level that covers a broad range of topics. Most of the questions are consistent across countries and survey dates. Six waves of the survey are currently available, with survey years spanning 1981 to 2012.

The WVS includes a standard item on generalized trust, which is the most

[6] Since we observe every cohort in our sample at least twice, the full set of cohort dummies is not perfectly collinear to the age and survey year effects.

widely studied survey question to measure trust (cf. Naeff and Schupp, 2009).[7] The trust measure is included throughout all years of the survey. The exact wording is as follows:

*Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?*

- *Most people can be trusted.*
- *Can't be too careful.*

In addition, we use individual-level demographic variables from the WVS as controls in equation (4.1). We employ information on education, income, marital status, the number of children, gender, religion and a self-assessment of an individual's social class.

Our central objective is to distinguish macroeconomic crises by type and investigate potential differences in their respective effects on trust. The main source of cross-country data on crises is the data set provided by Reinhart and Rogoff (2009),[8] who date financial crises and classify them as inflation, currency, banking, stock market, domestic debt or foreign debt crises using objective criteria. The data cover 70 countries and start in 1800. Crises are coded as binary events for each country-year.

We complement these data on financial crises with yearly GDP data collected by Barro and Ursua (2008) covering a wide range of countries starting in 1870. As in Barro and Ursua (2008), we define a GDP crisis as a year-on-year drop of GDP by more than 10%.

Figure 4.A.1 in the Appendix depicts the distribution of crisis events over countries and years. Macroeconomic crises occur frequently over time and in different countries, and some countries are much more frequently affected than others. Such heterogeneity makes it impossible to identify effects of macroeconomic crises on individual outcomes from a single cross-section of country data alone.

We calculate each individual's experience stock for each crisis type following equation (4.2), joining information on the timing of crises with age and survey

[7] The question elicits people's expectations about the trustworthiness of people in general, in line with the interpretation of trust as a belief rather than a preference (Sapienza et al., 2013).

[8] Bordo et al. (2001) and Laeven and Valencia (2013) provide alternative datasets on financial crises, but with a less detailed breakdown of types. The former is largely similar to Reinhart and Rogoff (2009) but covers fewer countries and a shorter period. The latter only starts in 1976, which is too late for our analysis. Moreover, C. D. Romer and D. H. Romer (2017) create continuous measures for crises, which would be desirable for our analysis. However, their data only covers OECD countries since the 1970s. For estimations using a continuous measure see also Section 4.4.

year information from the WVS. Alternative experience stock specifications are considered in the Appendix.

It is important to note that a systematic concurrence of different crisis types undermines the analysis of single crisis types in isolation. Figure 4.1 illustrates this point for banking and GDP crises. Banking crises were frequent in the pre-1950 period and then again post-1980. GDP crises are somewhat more evenly distributed over time (but still with only a few countries experiencing GDP crises between 1950 and 1970). Note that banking crises and GDP crises sometimes happen concurrently. When constructing experience stocks for banking crises as described above, we may falsely attribute any observed effect of such an experience to a banking crisis, when it really is due to the experience of events that typically go hand in hand with banking crises.

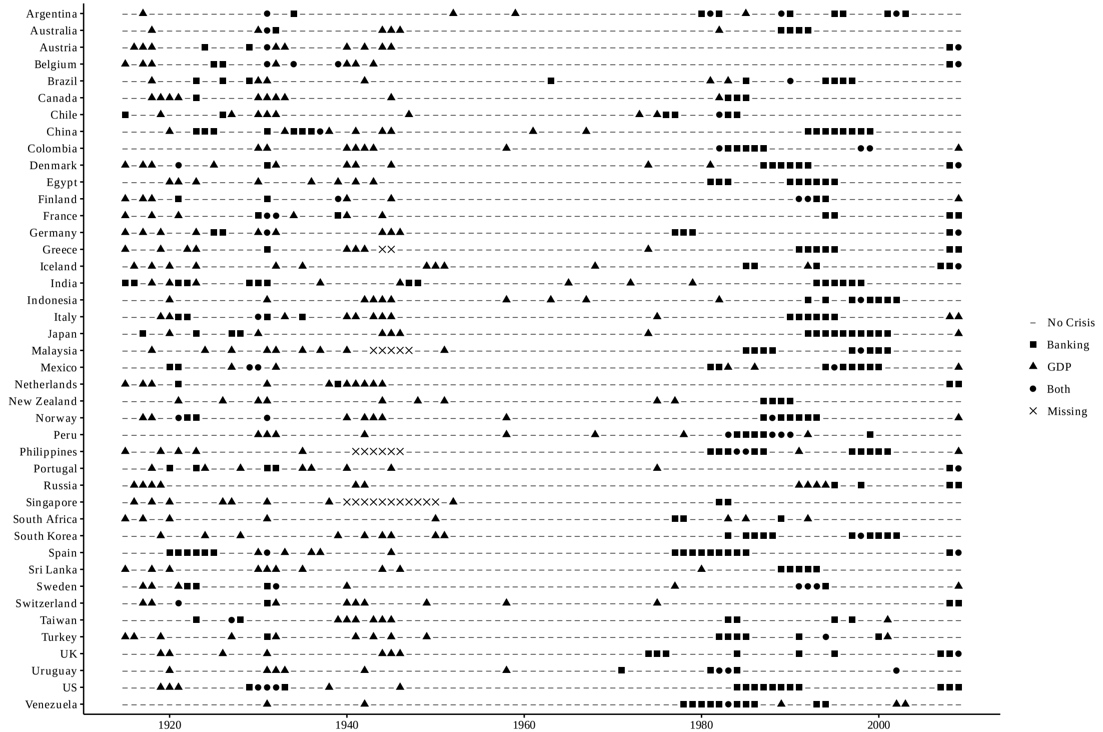


**Figure 4.1.** Map of banking crises and GDP crises. The figure shows substantial heterogeneity in the occurrence of banking and GDP crises across countries and time. There were relatively few crises from 1950 to 1970. Banking crises became much morge common from the 1980s on. Concurrent banking and GDP crisese are present but rare in our sample. The heterogeneous crisis patterns provide the basis for our analysis of crisis experience on trust. The crisis data are taken from Barro and Ursua (2008) and Reinhart and Rogoff (2009).

The problem of concurrent events is typically not addressed in research about macroeconomic experiences on individual outcomes. One way to deal with this problem is to simultaneously include experience stocks for other experiences. If different types of events are usually experienced in combination, this is reflected in a co-movement of their respective experience stocks. Some of our specifications will therefore control for experience stocks of a range of other macroeconomic events to rule out spurious relationships, reduce omitted variable bias and give a sense of the robustness of the results.

### 4.3.2 Results: Effects of Macroeconomic Crises on Trust

Table 4.2 considers each type of macroeconomic crisis available in our data in isolation. In line with Fuchs-Schündeln and Schündeln (2015), we initially set the depreciation rate of the experience stock at 2% per year ($\delta = .98$), but later show that the results are robust to varying that parameter. Table 4.2 shows two specifications for each experience stock: a baseline specification that controls for age, country, year, and cohort effects, as well as crisis-in-survey-year dummies and a baseline set of contemporaneous individual-level controls, and a more demanding specification that adds country-age trends and an additional, extensive set of controls. Table 4.2 displays least squares regressions for linear probability models. Probit estimates for all tables in the main text are provided in Appendix 4.C and yield similar results.

Note first that most crisis types are unrelated to trust, with very small coefficients that are statistically insignificant. Banking crises in columns (5) and (6) are a pronounced exception and exhibit a strong and persistent negative relation with trust. The coefficient of -.013 (or -.016 in column (6)) suggests that individuals who experienced a banking crisis 20 years ago are about 1% less likely to say that most people can be trusted.

**Table 4.2.** World Values Survey: Experience of different types of crises and the effect on trust

| Dependent variable: 1 if trusting | Delta definition of experience stock with $\delta = 0.98$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
| Experience stock of output crises | 0.011 | 0.020 | | | | | | | | | | | | |
| | (0.014) | (0.019) | | | | | | | | | | | | |
| Experience stock of inflation crises | | | 0.003 | 0.003 | | | | | | | | | | |
| | | | (0.002) | (0.004) | | | | | | | | | | |
| Experience stock of banking crises | | | | | -0.013*** | -0.016*** | | | | | | | | |
| | | | | | (0.005) | (0.006) | | | | | | | | |
| Experience stock of stock market crises | | | | | | | -0.002 | 0.003 | | | | | | |
| | | | | | | | (0.002) | (0.005) | | | | | | |
| Experience stock of currency crises | | | | | | | | | 0.001 | 0.001 | | | | |
| | | | | | | | | | (0.002) | (0.005) | | | | |
| Experience stock of domestic debt crises | | | | | | | | | | | 0.004* | 0.008 | | |
| | | | | | | | | | | | (0.002) | (0.008) | | |
| Experience stock of foreign debt crises | | | | | | | | | | | | | 0.001 | 0.005 |
| | | | | | | | | | | | | | (0.002) | (0.004) |
| Constant | 0.038 | 0.038 | 0.097* | -0.023 | 0.205*** | 0.093* | 0.167*** | -0.090 | 0.116*** | -0.011 | 0.137*** | 0.008 | 0.132*** | -0.012 |
| | (0.044) | (0.031) | (0.049) | (0.081) | (0.038) | (0.050) | (0.050) | (0.075) | (0.042) | (0.075) | (0.027) | (0.037) | (0.027) | (0.035) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Crisis at time of survey | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional controls | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes |
| Country-age trends | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes |
| $R^2$ | 0.130 | 0.140 | 0.121 | 0.130 | 0.122 | 0.131 | 0.127 | 0.136 | 0.121 | 0.130 | 0.122 | 0.131 | 0.122 | 0.131 |
| N | 97702 | 81713 | 120427 | 100405 | 120427 | 100405 | 108827 | 89658 | 120427 | 100405 | 120427 | 100405 | 120427 | 100405 |
| # countries | 33 | 32 | 44 | 43 | 44 | 43 | 39 | 38 | 44 | 43 | 44 | 43 | 44 | 43 |
| # survey years | 15 | 14 | 17 | 16 | 17 | 16 | 15 | 14 | 17 | 16 | 17 | 16 | 17 | 16 |

*Notes.* Ordinary least squares regression. Robust standard errors clustered at the country level in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Crisis at the time of survey is a dummy for whether there was a respective type of crisis in the year in which the respondent answered the survey. Baseline controls include gender, level of education and (subjective) social class. The additional controls include income decile, indicators for unemployment and marital status, subjectively assessed religiousness and number of children.

To put the effect size in perspective, consider Alesina and La Ferrara's (2002) estimate of how personal traumatic experiences (e.g., divorce or disease) in the preceding year affect trust. They find that recent traumas decrease trust by 2 to 3 percentage points and at the same time argue that traumas are among the best predictors of trust in their data. Our estimate implies that a banking crisis in the previous year decreases trust by 1.6 percentage points, only somewhat lower than the effect of a personal trauma. But the effects of banking crises are long-lasting: Even if the crisis happened 20 years ago, its effect on trust today is about one-third of the effect of a recent personal traumatic experience.

Exploiting the main insight from Table 4.2, Table 4.3 reports additional analyses on the effects of banking crises. Columns (1) to (4) show that including various controls and fixed effects has negligible effects on the size and significance of the estimate. Moreover, in column (5) we also control for other macroeconomic experience stocks which leaves the coefficient estimate virtually unchanged.

Columns (6) and (7) of Table 4.3 replicate the effect for a faster depreciation rate in the experience specification. A delta of .95 (.90) implies that experience depreciates at a rate of 5% (10%) per year, much faster than the 2% annual depreciation in our baseline specification. The effect of banking crises remains highly significant in these alternative specifications, with a similar magnitude for crises experienced recently and, owing to the faster depreciation, somewhat smaller magnitudes for crises experienced further in the past.

Figure 4.2 illustrates the effect size for two different depreciation rates and for individuals of two different ages. A recent crisis reduces trust by around 2 percentage points on average. Owing to faster depreciation of the experience stock, past crises have less of an effect on today's trust when depreciation is 5% rather than 2%. Still, our estimates imply that a banking crisis that occurred 15 years ago is associated with about 1 percentage point lower trust today, even assuming faster depreciation.

Note that age effects are not apparent in this specification: The effect of a crisis 10 years ago is the same for a 25-year-old as for a 45-year-old. The reason is that delta experience by construction only depends on the time that has passed since the crises occurred.

Our specifications are demanding and more conservative than identification strategies employed in other examinations of experience effects. The latter typically do not consider the role of concurrent experiences and presume a single specification for the accumulation of experience. We performed analogous analyses for alternative definitions of the experience stock in Appendix 4.B. All specifications strongly support the negative relationship between banking crisis experience and trust. In accordance with previous research, our results also indicate a potential age effect: Younger individuals are relatively more affected by recent crises than older people with a longer life history.

**Table 4.3.** World Values Survey: Trust and banking crisis experience using delta definition of experience stock

| Dependent variable: 1 if trusting | $\delta = 0.98$ | | | | | $\delta = 0.95$ | $\delta = 0.90$ |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Experience stock of banking crises | -0.013*** | -0.013*** | -0.015** | -0.016*** | -0.017*** | -0.020*** | -0.022** |
| | (0.004) | (0.005) | (0.006) | (0.006) | (0.006) | (0.007) | (0.008) |
| Constant | 0.186*** | 0.205*** | 0.133** | 0.093* | 0.027 | 0.039 | -0.005 |
| | (0.038) | (0.038) | (0.050) | (0.050) | (0.109) | (0.128) | (0.118) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Banking crisis at time of survey | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional controls | | | Yes | Yes | Yes | Yes | Yes |
| Country-age trends | | | | Yes | Yes | Yes | Yes |
| Other macro crisis experience | | | | | Yes | Yes | Yes |
| $R^2$ | 0.121 | 0.122 | 0.127 | 0.131 | 0.131 | 0.131 | 0.132 |
| N | 120427 | 120427 | 100405 | 100405 | 100405 | 100405 | 100405 |
| # countries | 44 | 44 | 43 | 43 | 43 | 43 | 43 |
| # survey years | 17 | 17 | 16 | 16 | 16 | 16 | 16 |

*Notes.* Ordinary least squares regression. Robust standard errors clustered at the country level in parentheses. Significance levels: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Other macroeconomic experiences include average percentage change in GDP, as well as inflation, stock market and domestic debt crises. Baseline controls include gender, level of education and (subjective) social class. The additional controls include income decile, indicators for unemployment and marital status, subjectively assessed religiousness and number of children.

### 4.3.3 Results: Confidence in Institutions

One interpretation of our result is that banking crises erode trust in institutions, particularly banks. Indeed, there is empirical evidence that aggregate levels of trust in banks dropped in the aftermath of the 2008/09 financial crisis (Sapienza and Zingales, 2012). The observed long-term effect on interpersonal trust could be a byproduct of a breach of institutional trust. This idea encapsulates the hypothesis that banking crises induce a loss of trust in a certain group of people who are blamed, or viewed as responsible, for the crises, such as bankers. Individuals might then extrapolate this attitude to other people more generally.

We test this reasoning by examining the effect of different types of domestic macroeconomic crises on measures of trust in different institutions. The specifications presented in Table 4.4 closely conform to our baseline analyses in Tables 4.2 and 4.3.

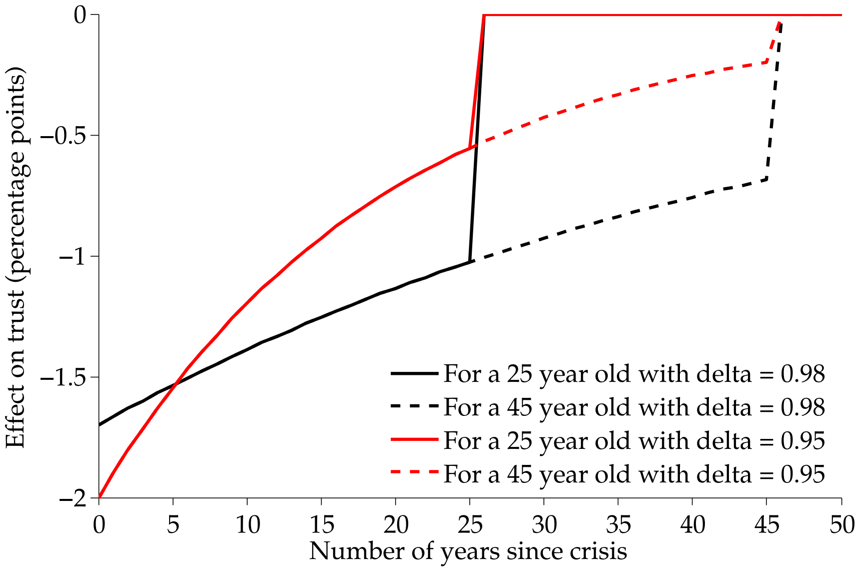


**Figure 4.2.** Estimated effect of experiencing a banking crisis X years ago on trust today, for different parameter values. Calculated effects sizes are based on our estimation results using the Fuchs-Schündeln and Schündeln (2015) experience specification with a fixed depreciation rate. The red curves assume a depreciation rate of 5%, the dark curves assume depreciation at 2%. Even at the faster depreciation rate, our estimates imply a decrease in the likelihood to be trusting due to a banking crisis 15 years ago of about 1 percentage point.

We make the following observations. First, there is no persistent effect of banking crises on trust in banks (columns (7) and (8)).[9] Second, note that banking crises have a strong and consistent negative effect on trust in the political class, represented by the government, parliament and political parties. This effect is notable given that we replicate the reverse effect of output crises on trust in the political class, which was put forward by Giuliano and Spilimbergo (2014). Other crises do not produce consistent effects across specifications. Therefore, we reconfirm Giuliano and Spilimbergo (2014) in estimating a permanently improved attitude toward the government after depressions, but further add that banking crises have a striking effect in the opposite direction. Our findings are in line with evidence reported by Rainer and Siedler (2009). In their study on the relationship between democracy and trust after the reunification of Germany, they both find that social distrust is driven by negative economic outcomes in

[9] However, note that the question on trust in banks is only available for a reduced sample of the WVS. While the effect goes in the expected direction in column (8), statistical power is much lower due to the limited number of available observations.

**Table 4.4.** World Values Survey: Confidence in institutions using delta definition of experience stock with $\delta = 0.98$

| | Confidence in institutions | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dependent variable: | Government | | Parliament | | Political parties | | Banks | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Experience stock of banking crises | -0.033*** | -0.059*** | -0.035** | -0.060** | -0.022** | -0.039** | 0.010 | 0.019 |
| | (0.010) | (0.018) | (0.015) | (0.023) | (0.010) | (0.017) | (0.013) | (0.019) |
| Experience stock of output crises | 0.129** | 0.180** | 0.087** | 0.135*** | 0.062** | 0.092** | 0.017 | -0.365 |
| | (0.063) | (0.067) | (0.033) | (0.046) | (0.031) | (0.035) | (0.045) | (0.171) |
| Experience stock of inflation crises | 0.016** | 0.018* | 0.012*** | 0.009 | 0.009** | 0.006 | 0.016*** | 0.080 |
| | (0.006) | (0.010) | (0.004) | (0.008) | (0.003) | (0.007) | (0.005) | (0.028) |
| Experience stock of stock market crises | -0.003 | 0.000 | -0.000 | 0.013 | 0.001 | 0.012 | -0.006 | 0.029 |
| | (0.005) | (0.011) | (0.005) | (0.010) | (0.004) | (0.009) | (0.008) | (0.058) |
| Experience stock of domestic debt crises | 0.044** | 0.043 | 0.030*** | 0.015 | 0.025*** | -0.017 | -0.031 | -0.123 |
| | (0.020) | (0.035) | (0.011) | (0.027) | (0.009) | (0.017) | (0.020) | (0.123) |
| Constant | 2.252*** | 3.241*** | 1.876*** | 1.702*** | 1.758*** | 1.400*** | 2.562*** | 2.153** |
| | (0.144) | (0.154) | (0.113) | (0.287) | (0.097) | (0.193) | (0.157) | (0.284) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional controls | | Yes | | Yes | | Yes | | |
| Any crisis at time of survey | | Yes | | Yes | | Yes | | Yes |
| Country-age trends | | Yes | | Yes | | Yes | | Yes |
| $R^2$ | 0.124 | 0.134 | 0.162 | 0.153 | 0.147 | 0.129 | 0.176 | 0.228 |
| N | 145673 | 86754 | 147606 | 88492 | 146911 | 88815 | 37495 | 4118 |
| # countries | 44 | 36 | 44 | 37 | 44 | 37 | 27 | 3 |
| # waves | 4 | 4 | 5 | 4 | 4 | 4 | 2 | 2 |
| # survey years | 17 | 13 | 18 | 13 | 17 | 13 | 4 | 2 |

*Notes.* Ordinary least squares regression. Robust standard errors clustered at the country level in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The dependent variables have four levels: 1="None at all", 2="Not very much", 3="Quite a lot", and 4="A great deal". Alternative specifications with ordinal regressions are reported in Appendix 4.C. Baseline controls include gender, level of education and (subjective) social class. The additional controls include income decile, indicators for unemployment and marital status, subjectively assessed religiousness and number of children. We do not use additional controls in specification (8) as this reduces the available sample to a single wave.

the past – rather than living in a democracy per se –, and that "in sharp contrast to social trust," levels of institutional trust converge over time (Rainer and Siedler, 2009, p. 251).

Another corresponding question on the underlying behavioral channels is whether the effect on trust is driven by changes in beliefs versus preferences. To shed light on this issue, we performed additional analyses, which similarly support the interpretation as a belief-based mechanism. These analyses are reported in Appendix 4.D.1.

Taken together, our cross-country analyses provides three insights. First, only banking crises, but not other types of crises, have a significant and persistent effect on trust in our data. We interpret the robustness of this effect to numerous different specifications and alternative plausible approaches to formalize experience as compelling evidence for the strength of the relationship. It is worth noting that other types of experience are occasionally at the margin of significance, but these effects are unstable and disappear in the more demanding specifications. This observation suggests care when drawing conclusions about experience effects from observational data. Second, the estimated magnitude is sizable, suggesting that banking crises can have substantial repercussions on societies. Third, while we replicate the key finding of Giuliano and Spilimbergo (2014), our data suggest that the erosion of interpersonal trust due to banking crises is not an artifact of reduced trust in financial institutions, but goes hand in hand with distrust in political institutions.

## 4.4 Evidence from the U.S.: Regional Bank Failures

Part 2 of our empirical analysis investigates the relationship between regional bank failures and trust in U.S data. This section serves three objectives. First, the examination of local bank failures has the potential to corroborate the specific relationship between trust and adverse *breach of trust* experience in a strongly trust-dependent domain, the banking sector, which we observe at the country level. It that sense, it may lend credence to the behavioral foundation for the estimated effect of macro-level events. Second, it shifts the analysis to a more granular level, exploiting variation across regional populations with much lower heterogeneity than in a cross-country study. Third, unlike in our cross-country analysis, we can construct a continuous banking crisis variable that measures the severity of a local banking crisis, using information on the estimated losses from bank defaults.

### 4.4.1 Data and Measurement

Our main source of data is the sensitive geocoded data of the *General Social Survey* (GSS). The GSS has been conducted annually from 1972 to 1994 (with few exceptions) and every other year afterwards. Each sample typically comprises around 3000 individuals and the sample is aimed to be representative of the United States. For our purposes, it is important to note that the trust question in the GSS is identical to the question in the WVS. The GSS contains a large number of individual demographic variables, similar in scope to the WVS. In addition, we obtained confidential geographic information about the locations of survey respondents at the time of the survey and at age 16 (state-level).

Our second source of data is a list of banks that failed between 1934 and 2010, provided by the *Federal Deposit Insurance Corporation* (FDIC). Our sample contains more than 3800 bank failures including bank names, locations, assets, deposits and dates of failure, as well as estimated losses. The majority of our sample of failures is made up of small and medium-sized local banks: Median deposits of the banks in our sample at the time of their failure was around USD

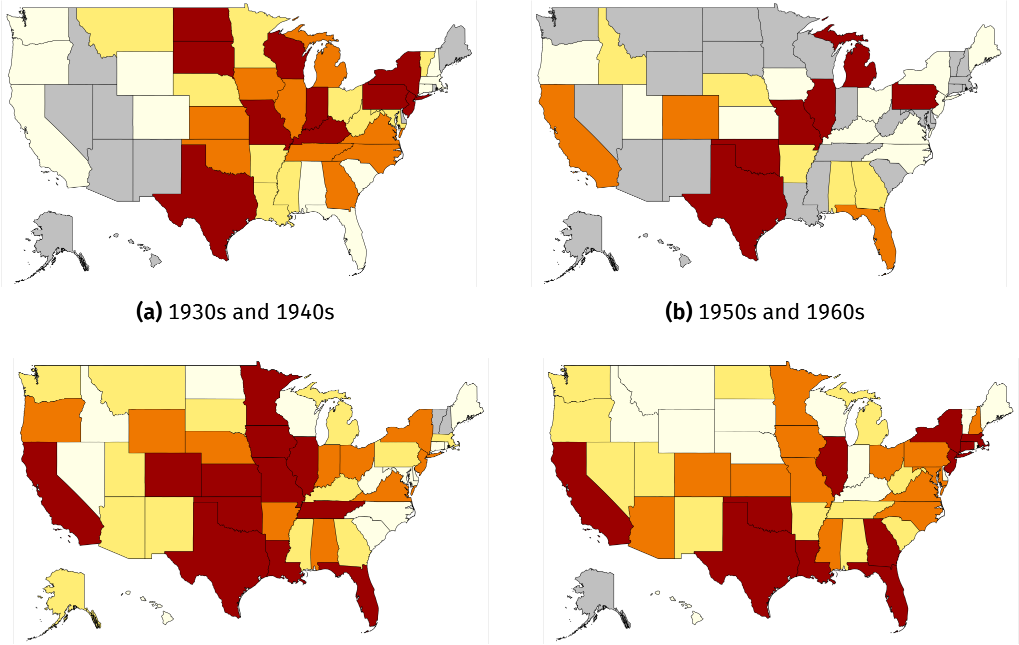
65 million, and median losses are estimated at USD 15 million. This is crucial for the interpretation of estimated effect magnitudes. Figure 4.3 displays the heterogeneity in bank failures over states and time.[10] Such geo-temporal variation forms the basis for our analysis. We aggregate the data on historical bank failures into two binary variables: An annual binary banking crisis indicator at the state level, and an annual continuous measure of the estimated losses from bank failures at the state level. Individual-level experience stocks of bank failures and losses from bank failures are constructed exactly as in section 4.2 based on these two variables. Since estimated losses are provided in nominal dollars, we harmonized historical losses by first converting these figures to real dollars based on yearly CPI (U.S. Bureau of Labor Statistics, 2017). That respondents move their place of residence between U.S. states during their lifetime is common, so we cannot generally infer that a person actually experienced the the bank failures that previously occurred at her place of residence when the survey was conducted. To limit this concern, we restrict our attention to subjects who live in the same states at the time of survey as they did at the age of 16. If anything, any residual migration should bias our estimates downwards.

### 4.4.2 Results: Effects of Regional Bank Failures on Trust

We report results for experience stocks built from both crisis event measures, the binary one and the continuous one. First, columns (1) to (4) of Table 4.5 display results for the delta experience stock regression specifications of the same form as in equation (4.1) and Table 4.3. Here, crisis experience is computed based on the binary crisis indicator as before. The results are qualitatively similar to those in the the cross-country analysis above, indicating a strongly significant negative effect of experiencing bank failures on trust. The magnitude of coefficients, however, are not directly comparable to those of section 4.3.[11] Given our scaling of experience stocks, the coefficients in columns (1) to (4) imply that the experience of 100 small bank failures 20 years ago makes individuals about 1.3 percentage points less likely to answer our trust question in the affirmative. This effect size for the experience of 100 small bank failures is about the same as experiencing one country-level banking crisis based on our estimates in section 4.3. To put this into perspective, recall the median estimated loss in our sample of failures of only about USD 15 million.

[10] The figures show variation in states' exposure to the absolute number of bank failures. We also used relative figures such as the number of bank failures per population and obtained similar degrees of heterogeneity.

[11] We scaled experience stocks of bank failures such that the experience of one country-level crises as coded in section 4.3 would result in the same experience stock as the experience of 100 small bank failures at the same time.



**(a)** 1930s and 1940s

**(b)** 1950s and 1960s

**(c)** 1970s and 1980s

**(d)** 1990s and 2000s

**Figure 4.3.** Bank failures over time and states. Displayed is the frequency of bank failures in U.S. states for two-decade intervals. The maps are based on quartile splits for each period of 20 years. Darker shades imply more bank failures. The combined figure shows strong heterogeneity over time as to which states have been affected relatively more by bank failures. For example, residents of California or Florida have experienced an increasing exposure to bank failures over time – relative to other states –, whereas Texas has seen a comparably high number of bank failures throughout.

Our second measure is based directly on estimated losses for each failure, aggregated as for the binary measure.[12] Regressions results for this continuous crisis measure are reported in columns (5) to (8) in Table 4.5. The sign and significance and sign of estimated coefficients provide additional support for the negative effect of bank failures on trust.

**Table 4.5.** General Social Survey: Trust and banking crisis experience

| Dependent variable: 1 if trusting | Delta definition of experience stock with $\delta = 0.98$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Experience stock of bank failures | -0.012*** | -0.013*** | -0.014*** | -0.013*** | | | | |
| | (0.002) | (0.002) | (0.002) | (0.002) | | | | |
| Estimated losses | | | | | -6.853*** | -7.077*** | -7.345*** | -6.361*** |
| | | | | | (1.300) | (1.255) | (1.281) | (1.659) |
| Constant | 1.333*** | 1.295*** | 1.274*** | 1.782*** | 1.329*** | 1.290*** | 1.269*** | 1.776*** |
| | (0.044) | (0.080) | (0.083) | (0.091) | (0.044) | (0.079) | (0.082) | (0.091) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bank failures at time of survey | Yes | Yes | Yes | Yes | | | | |
| Estimated losses in year of survey | | | | | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | | Yes | Yes | Yes | | Yes | Yes | Yes |
| Additional controls | | | Yes | Yes | | | Yes | Yes |
| State-age trends | | | | Yes | | | | Yes |
| $R^2$ | 0.136 | 0.139 | 0.141 | 0.145 | 0.136 | 0.139 | 0.141 | 0.144 |
| N | 14951 | 14950 | 14911 | 14911 | 14951 | 14950 | 14911 | 14911 |
| # states | 51 | 51 | 51 | 51 | 51 | 51 | 51 | 51 |
| # survey years | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 |

Ordinary least squares regression. Robust standard errors clustered at the country level in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Baseline controls include gender, age, income bracket and level of education. The additional controls further include unemployment and marital status, religion and number of children.

Similar to the WVS, the GSS results are robust to alternative definitions of the experience stock and to using alternative values for the depreciation rate. Moreover, we again find evidence that interpersonal trust is affected more if banking crises are experienced at a younger age (Table 4.B.5 in the Appendix). In line with Giuliano and Spilimbergo (2014), our data indicate that the *formative period* from the age of 16 to the age of 25 deserves particular attention.[13] Traumatic experience in a trust-intensive domain during that age interval has the potential to persistently erode interpersonal trust.

[12] The crisis measure is computed similar to equation (4.2), as

$$C_{it}^{delta}(\delta) = \sum_{s=t-age_{it}}^{t} \delta^{t-s} EstimatedLoss_s, \tag{4.3}$$

with $EstimatedLoss_s$ denoting the cumulated (real) losses from bank failures in a given region in year $s$.

[13] Note that Giuliano and Spilimbergo (2014) in fact run their analyses on the same individual-level data set, the GSS.

## 4.5 Conclusion

Recent work attests to a substantial variation in trust levels across geographical locations and over time. These differences affect outcomes such as financial development, stock market participation, and economic growth. However, the sources of the heterogeneity in trust levels across locations and over time are not yet well understood. In this paper we provide field evidence that trust is not an exogenous and stable characteristic of individuals or societies, but is malleable and systematically shaped by economic experiences during individuals' lifetime.

Our empirical analysis of cross-country data establishes that, among a set of different economic crisis types, only banking crises have a robust, sizeable and long-lasting negative effect on interpersonal trust. Notably, this relationship is not an artifact of an erosion of trust in financial institutions. The effect is more pronounced for experiences made during adulthood and in the 20s. Our findings in cross-country data suggests that drastic, adverse personal experience in the domain of banking operates like a persistent *breach of trust*. This intuition is directly confirmed in separate state-level data from the U.S. We exploit regional variation in the experience of small bank failures across U.S. states and replicate a sizable effect on trust.

This study makes three contributions. First, our results have clear implications for understanding the determinants of trust. Why do adverse macroeconomic events undermine generalized trust in strangers? An emerging literature on lay perceptions of economic events shows that people tend to rationalize macroeconomic crises as man-made rather than as the result of bad luck or systemic reasons – i.e., they tend to attribute crises to the misbehaving of other people. For example, Leiser et al. (2010) provide evidence that people favor personal (*blaming*) over impersonal (*analyzing*) explanations of the 2008/09 financial crisis and display a tendency toward intentional rather than causal accounts.[14] In ongoing companion work, we examine the behavioral mechanisms underlying the observed erosion of interpersonal trust more directly. Exploiting a rich individual-level panel data set, we identify financial stress and financial worries as key driving forces of broken trust. Second, our results bear on the welfare costs of macroeconomic crises. Distrust is regarded as costly (Kosfeld and Falk, 2006; Kramer, 1999; Malul et al., 2010), but such indirect effects are notoriously hard to quantify. Recent work examines the repercussions of financial

[14] Leiser et al. (2016, p.156) describe this as people assuming a "malevolent hand behind every negative outcome". Similarly, Kumagai et al. (2006) show that victims of natural disasters attribute an overly high share of responsibility for such disasters to human agency. This also relates to the more fundamental human inclination to judge events as being intentional rather than due to chance or simply causally determined; see, for example, Kelemen and Rosset (2009) and Rosset (2008).

crises on political preferences (Funke et al., 2016). In light of this research, our findings suggest that the erosion of trust can add to such channels, for example via the link between trust and preferences for redistribution (Hetherington and Husser, 2012; Yamamura, 2014).

Third, our data confirms that robustness should be a concern in observational studies on experience effects. We design a conservative empirical strategy based on more demanding specifications that recognize the multiplicity of plausible formal characterizations of experience. Caution in drawing conclusions from experience regressions is warranted: In our data, some types of experience show unsteady patterns across specifications. This stands in contrast to the effect of experiencing banking crises or bank failures on trust, which is robustly supported throughout our analyses.

## References

**Alesina, Alberto and Eliana La Ferrara (2002):** “Who trusts others?” *Journal of Public Economics*, 85 (2), 207–234. [141, 142, 151]

**Algan, Yann and Pierre Cahuc (2010):** “Inherited trust and growth.” *The American Economic Review*, 100 (5), 2060–2092. [139, 142]

**Barro, Robert J. and Jose F. Ursua (2008):** “Macroeconomic crises since 1870.” *Brookings Papers on Economic Activity*, 39 (1), 255–350. [146, 148, 164]

**Ben-Ner, Avner and Louis Putterman (2001):** “Trusting and trustworthiness.” *Boston University Law Review*, 81, 523–551. [180]

**Beugelsdijk, Sjoerd, Henri L. F. De Groot, and Anton B. T. M. Van Schaik (2004):** “Trust and economic growth: a robustness analysis.” *Oxford Economic Papers*, 56 (1), 118–134. [139, 142]

**Bohnet, Iris, Fiona Greig, Benedikt Herrmann, and Richard Zeckhauser (2008):** “Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States.” *The American Economic Review*, 98 (1), 294–310. [180]

**Bohnet, Iris and Richard Zeckhauser (2004):** “Trust, risk and betrayal.” *Journal of Economic Behavior & Organization*, 55 (4), 467–484. [180]

**Bordo, Michael, Barry Eichengreen, Daniela Klingebiel, and Maria Soledad Martinez-Peria (2001):** “Is the crisis problem growing more severe?” *Economic Policy*, 16 (32), 52–82. [146]

**Cassar, Alessandra, Andrew Healy, and Carl Von Kessler (2011):** “Trust, risk, and time preferences after a natural disaster: experimental evidence from Thailand.” Mimeo. [142]

**De Luca, Giacomo and Marijke Verpoorten (2015):** “Civil war, social capital and resilience in Uganda.” *Oxford Economic Papers*, 67 (3), 661–686. [142]

**Eckel, Catherine C. and Rick K. Wilson (2004):** “Is trust a risky decision?” *Journal of Economic Behavior & Organization*, 55 (4), 447–465. [180, 181]

**Falk, Armin, Anke Becker, Thomas J. Dohmen, Benjamin Enke, and David Huffman (2017):** “Global evidence on economic preferences.” Tech. rep. National Bureau of Economic Research. [139]

**Freese, Jeremy (2004):** "Risk preferences and gender differences in religiousness: Evidence from the World Values Survey." *Review of Religious Research*, 46 (1), 88–91. [180]

**Fuchs-Schündeln, Nicola and Matthias Schündeln (2015):** "On the endogeneity of political preferences: Evidence from individual experience with democracy." *Science*, 347 (6226), 1145–1148. [143, 144, 149, 153]

**Funke, Manuel, Moritz Schularick, and Christoph Trebesch (2016):** "Going to extremes: Politics after financial crises, 1870–2014." *European Economic Review*, 88, 227–260. [160]

**Gambetta, Diego (1988):** *Trust: Making and Breaking Cooperative Relations*. Blackwell. [139]

**Giuliano, Paola and Antonio Spilimbergo (2014):** "Growing up in a recession." *The Review of Economic Studies*, 81 (2), 787–817. [141–144, 153, 155, 158, 170]

**Guiso, Luigi, Paola Sapienza, and Luigi Zingales (June 2004):** "The role of social capital in financial development." *The American Economic Review*, 94 (3), 526–556. [142]

**Hetherington, Marc J. and Jason A. Husser (2012):** "How trust matters: The changing political relevance of political trust." *The American Journal of Political Science*, 56 (2), 312–325. [142, 160]

**Jaime-Castillo, Antonio M. (2016):** *Social trust and demand for redistribution. Is there a crowding out effect?* Mimeo. [142]

**Kelemen, Deborah and Evelyn Rosset (2009):** "The human function compunction: Teleological explanation in adults." *Cognition*, 111 (1), 138–143. [159]

**Knack, Stephen and Philip Keefer (1997):** "Does social capital have an economic payoff? A cross-country investigation." *The Quarterly Journal of Economics*, 112 (4), 1251–1288. [139, 142]

**Kosfeld, Michael and Armin Falk (2006):** "The hidden costs of control." *The American Economic Review*, 96 (5), 1611–1630. [142, 159]

**Kramer, Roderick M. (1999):** "Trust and distrust in organizations: Emerging perspectives, enduring questions." *Annual Review of Psychology*, 50 (1), 569–598. [142, 159]

**Kramer, Roderick M. and Tom R. Tyler (1996):** *Trust in Organizations: Frontiers of Theory and Research*. Sage Publishing. [139]

**Kumagai, Yoshitaka, John Edwards, and Matthew S. Carroll (2006):** "Why are natural disasters not "natural" for victims?" *Environmental Impact Assessment Review*, 26 (1), 106–119. [159]

**Laeven, Luc and Fabian Valencia (2013):** "Systemic banking crises database." *IMF Economic Review*, 61 (2), 225–270. [146]

**Leiser, David, Rinat Benita, and Sacha Bourgeois-Gironde (2016):** "Differing conceptions of the causes of the economic crisis: Effects of culture, economic training, and personal impact." *Journal of Economic Psychology*, 53, 154–163. [159]

**Leiser, David, Sacha Bourgeois-Gironde, and Rinat Benita (2010):** "Human foibles or systemic failure—Lay perceptions of the 2008–2009 financial crisis." *The Journal of Socio-Economics*, 39 (2), 132–141. [159]

**Lewicki, Roy J. and Barbara B. Bunker (1995):** "Trust in relationships." *Administrative Science Quarterly*, 5, 583–601. [139]

**Lewicki, Roy J. and Barbara B. Bunker (1996):** "Developing and maintaining trust in work relationships." In. *Trust in Organizations: Frontiers of Theory and Research*. [139]

**Lewicki, Roy J. and Carolyn Wiethoff (2000):** "Trust, trust development, and trust repair." In. *The Handbook of Conflict Resolution: Theory and Practice*, 86–107. [139]

**Malmendier, Ulrike and Stefan Nagel (2011):** "Depression babies: Do macroeconomic experiences affect risk taking?" *The Quarterly Journal of Economics*, 126 (1), 373–416. [142, 144, 169, 171]

**Malmendier, Ulrike and Stefan Nagel (2015):** "Learning from inflation experiences." *The Quarterly Journal of Economics*, 131 (1), 53–87. [142]

**Malmendier, Ulrike, Stefan Nagel, and Zhen Yan (2017):** "The making of hawks and doves: Inflation experiences on the FOMC." Tech. rep. National Bureau of Economic Research. [142]

**Malul, Miki, Mosi Rosenboim, and Tal Shavit (2010):** "Costs of mistrust between ethnic majority and minorities: Evidence from Israel." *Review of Social Economy*, 68 (4), 447–464. [142, 159]

**Miller, Alan S. (2000):** "Going to hell in Asia: The relationship between risk and religion in a cross cultural setting." *Review of Religious Research*, 5–18. [180]

**Miller, Alan S. and John P. Hoffmann (1995):** "Risk and religion: An explanation of gender differences in religiosity." *Journal for the Scientific Study of Religion*, 63–75. [180]

**Naeff, M. and J. Schupp (2009):** "Measuring trust: experiments and surveys in contrast and combination." *Berlin: Deutsches Institut für Wirtschaftsforschung*. [146]

**Nannestad, Peter (2008):** "What have we learned about generalized trust, if anything?" *Annual Review Political Science*, 11, 413–436. [139]

**Porta, Rafael La, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny (1997):** "Trust in Large Organizations." *The American Economic Review*, 87 (2), 333–338. [139]

**Rainer, Helmut and Thomas Siedler (2009):** "Does democracy foster trust?" *Journal of Comparative Economics*, 37 (2), 251–269. [153, 154]

**Reinhart, Carmen M and Kenneth Rogoff (2009):** *This time is different: eight centuries of financial folly*. Princeton University Press. [146, 148, 164]

**Robinson, Robert V. and Elton F. Jackson (2001):** "Is trust in others declining in America? An age–period–cohort analysis." *Social Science Research*, 30 (1), 117–145. [139]

**Romer, Christina D. and David H. Romer (Oct. 2017):** "New Evidence on the Aftermath of Financial Crises in Advanced Countries." *American Economic Review*, 107 (10), 3072–3118. [146]

**Rosset, Evelyn (2008):** "It's no accident: Our bias for intentional explanations." *Cognition*, 108 (3), 771–780. [159]

**Sapienza, Paola, Anna Toldra-Simats, and Luigi Zingales (2013):** "Understanding trust." *The Economic Journal*, 123 (573), 1313–1332. [146, 181]

**Sapienza, Paola and Luigi Zingales (2012):** "A trust crisis." *International Review of Finance*, 12 (2), 123–131. [152]

**Schwerter, Frederik and Florian Zimmermann (2016):** *Determinants of Trust: Personal Experiences of Unrelated Social Interactions*. Mimeo. [142]

**Slovic, Paul (1993):** "Perceived risk, trust, and democracy." *Risk Analysis*, 13 (6), 675–682. [139]

**U.S. Bureau of Labor Statistics (2017):** *Consumer Price Index for All Urban Consumers: All Items [CPIAUCNS], retrieved from FRED, Federal Reserve Bank of St. Louis*. URL: https://fred.stlouisfed.org/series/CPIAUCNS (visited on 06/30/2017). [156]

**Yamamura, Eiji (2014):** "Trust in government and its effect on preferences for income redistribution and perceived tax burden." *Economics of Governance*, 15 (1), 71–100. [142, 160]

## Appendix 4.A Data

### 4.A.1 Individual-Level Data

Our cross-country level analysis in Section 4.3 uses survey data from the *World Values Survey* (WVS), available at http://www.worldvaluessurvey.org/wvs.jsp. Table 4.A.1 displays summary statistics for our individual-level controls in the WVS. We do not exclude specific countries or waves of the WVS from our analysis. The sample of available countries and survey years depends on the included variables. Our main results for banking crises are robust to excluding specific controls and including additional measures.

For the state-level analysis in the U.S., we use survey data from the *General Social Survey* (GSS), available at http://gss.norc.org/get-the-data. We additionally obtained geo-coded sensitive data from the NORC Institute that allows to link GSS observations to the respondent's state of residence at the time of the survey as well as to the state in which the respondent lived at the age of 16. We restrict our analyses to observations with geo-coded information.

### 4.A.2 Country-Level Crisis Data

Our macroeconomic crisis data comes from two sources, Reinhart and Rogoff (2009) and Barro and Ursua (2008). Figure 4.A.1 provides an overview of the distribution of different type of crises in this sample. There is substantial heterogeneity across countries and across time, providing the necessary variation in our derived measures of experience.

### 4.A.3 U.S. Bank Failures

The *Federal Deposit Insurance Corporation* (FDIC) provides Historical Statistics on Banking (HSOB) as a "reference and source document for those interested in banking history and in performing analyses on major trends in banking".[15] All data a freely available at https://www5.fdic.gov/hsob/SelectRpt.asp?EntryTyp=30. This data provide comprehensive information on failures of financial institutions insured by the FDIC since 1934, which includes both commercial banks and savings institutions. The effective date used for bank failures is defined as the "date that the failed / assisted institution ceased to exist as a privately held going concern." To create a continuous measure for the severity of a bank failure we use the Estimated Loss (previously called Estimated Cost), defined as "the difference between the amount disbursed from the Deposit Insurance Fund (DIF) to cover obligations to insured depositors and the amount

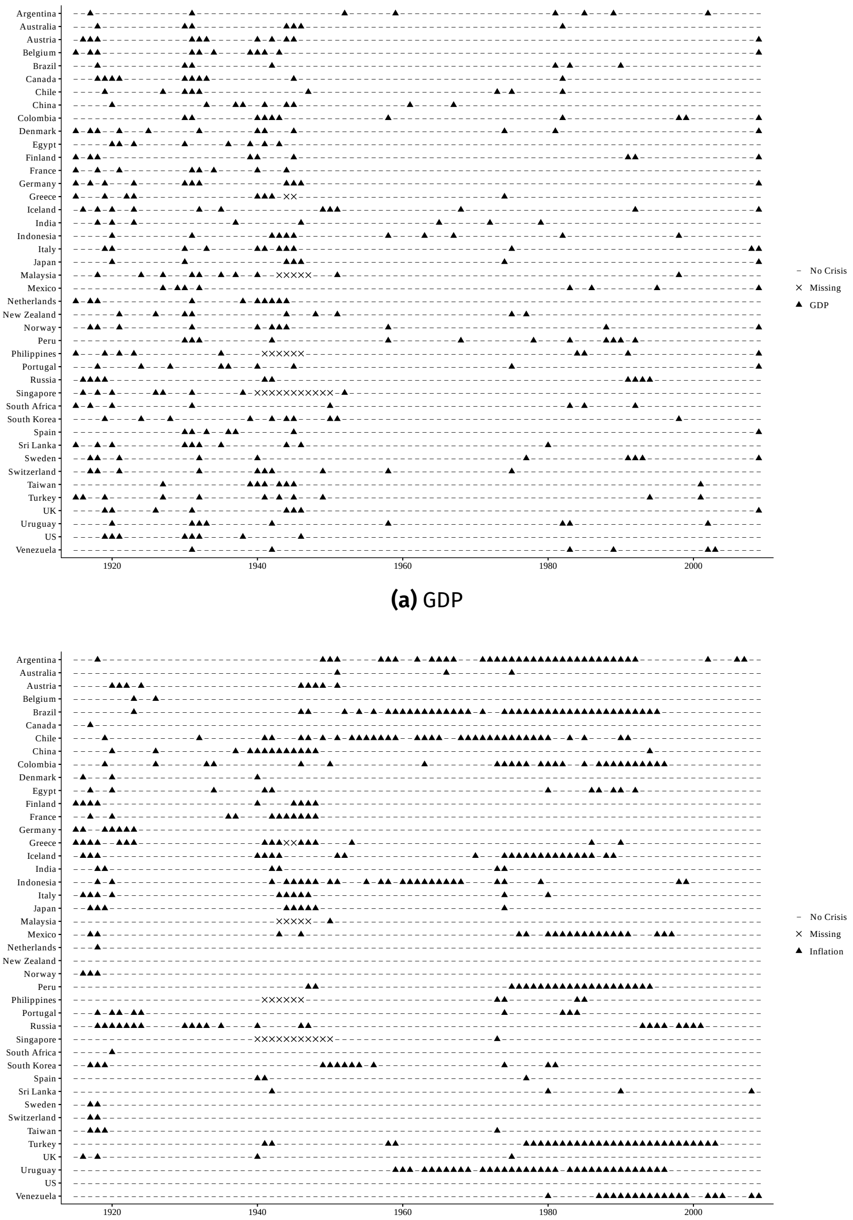
[15] See also https://www5.fdic.gov/hsob/help.asp.

**Table 4.A.1.** Summary statistics for individual-level controls

| | | | Percentiles | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std | 10th | 25th | 50th | 75th | 90th | N |
| Male | 0.49 | 0.50 | 0 | 0 | 0 | 1 | 1 | 151695 |
| Age | 40.34 | 15.61 | 21 | 27 | 38 | 51 | 64 | 154852 |
| Number of Children | 2.12 | 2.01 | 0 | 0 | 2 | 3 | 5 | 154859 |
| Religious | 0.41 | 0.49 | 0 | 0 | 0 | 1 | 1 | 144940 |
| Primary Education | 0.25 | 0.44 | 0 | 0 | 0 | 1 | 1 | 154859 |
| Secondary Education | 0.49 | 0.50 | 0 | 0 | 0 | 1 | 1 | 154859 |
| Post-secondary Education | 0.13 | 0.34 | 0 | 0 | 0 | 0 | 1 | 154859 |
| Married | 0.64 | 0.48 | 0 | 0 | 1 | 1 | 1 | 151516 |
| Separated | 0.05 | 0.22 | 0 | 0 | 0 | 0 | 0 | 151516 |
| Widowed | 0.05 | 0.22 | 0 | 0 | 0 | 0 | 0 | 151516 |
| Single | 0.25 | 0.43 | 0 | 0 | 0 | 1 | 1 | 151516 |
| Employed | 0.45 | 0.50 | 0 | 0 | 0 | 1 | 1 | 151217 |
| Self-employed | 0.12 | 0.32 | 0 | 0 | 0 | 0 | 1 | 151217 |
| Retired | 0.10 | 0.30 | 0 | 0 | 0 | 0 | 1 | 151217 |
| Student | 0.08 | 0.27 | 0 | 0 | 0 | 0 | 0 | 151217 |
| Unemployed | 0.08 | 0.27 | 0 | 0 | 0 | 0 | 0 | 151217 |
| Other | 0.17 | 0.38 | 0 | 0 | 0 | 0 | 1 | 151217 |
| Upper class | 0.02 | 0.13 | 0 | 0 | 0 | 0 | 0 | 120481 |
| Upper middle class | 0.20 | 0.40 | 0 | 0 | 0 | 0 | 1 | 120481 |
| Lower middle class | 0.37 | 0.48 | 0 | 0 | 0 | 1 | 1 | 120481 |
| Working class | 0.28 | 0.45 | 0 | 0 | 0 | 1 | 1 | 120481 |
| Lower class | 0.13 | 0.34 | 0 | 0 | 0 | 0 | 1 | 120481 |

*Notes.* This table shows summary statistics for control variables used in the study. The employment category *Other* contains respondents that characterized their employment status as “Housewife.” Social class is evaluated on a five-point scale and gives a subjective assessment. *N* denotes the number of non-missing observations.

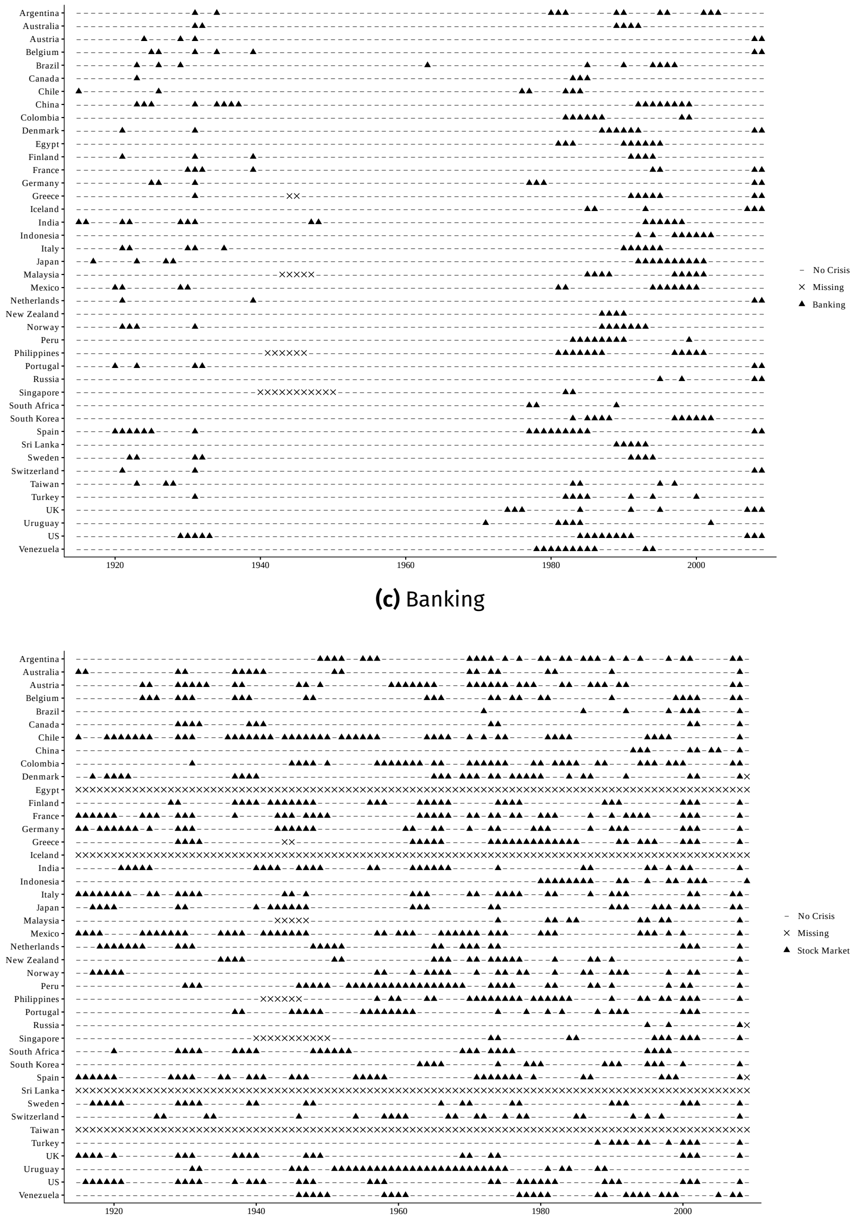
estimated to be ultimately recovered from the liquidation of the receivership estate”.



**(a)** GDP

**(b)** Inflation

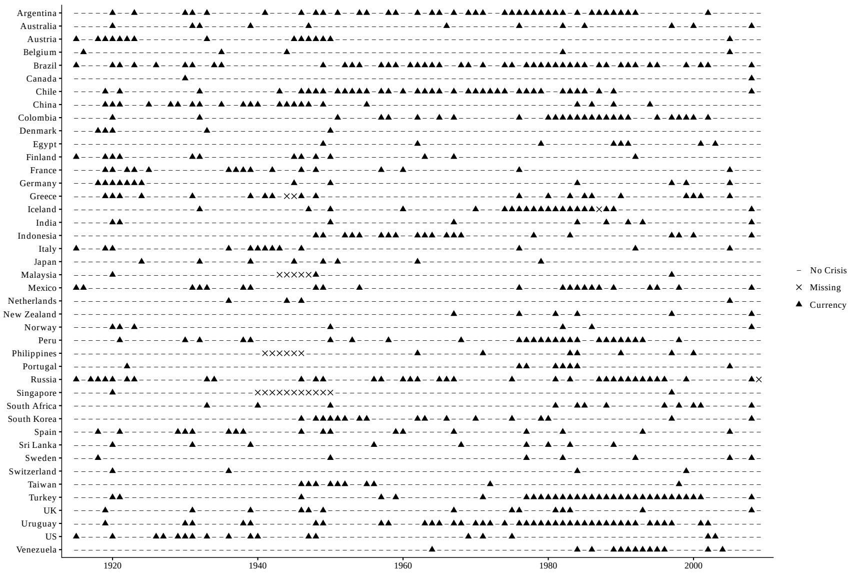
**Figure 4.A.1.** Distribution of crises over countries and years



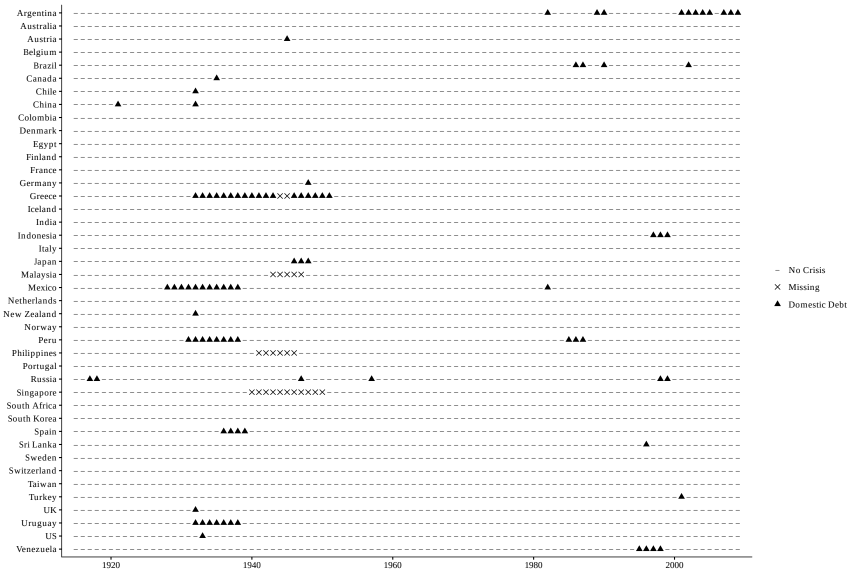
**(c)** Banking

**(d)** Stock Market

**Figure 4.A.1.** Distribution of crises over countries and years (cont.)



**(e)** Currency



**(f)** Domestic Debt

**Figure 4.A.1.** Distribution of crises over countries and years (cont.)

# Appendix 4.B Alternative Specifications and Robustness

## 4.B.1 Alternative Definitions of the Experience Stock

We investigate the robustness of our results to alternative specifications of crisis experience that have been suggested in the literature. Our first alternative, which we label *lambda experience*, is also a weighted average of years in which a crisis was experienced. We adopt the parsimonious one-parameter formulation of Malmendier and Nagel (2011), which allows weights to be constant, decline or increase over time:

$$\begin{aligned} C_{it}^{lambda}(\lambda) &= \sum_{s=1}^{age_{it}-1} w_{it}(s,\lambda)\mathbb{1}_{\{crisis_{t-s}\}}, \\ \text{where} \quad w_{it}(s,\lambda) &= \frac{(age_{it}-s)^{\lambda}}{\sum_{s=1}^{age_{it}-1}(age_{it}-s)^{\lambda}}. \end{aligned} \tag{4.B.1}$$

Here, $s$ denotes the time lag relative to $t$, and $\lambda$ is the weighting parameter. The indicator function takes value 1 if the individual experienced a crisis in year $t-s$ and 0 otherwise. The parameter $\lambda$ determines the shape of the weighting function, where $\lambda > 0$ assigns lower weights to more distant years, $\lambda < 0$ gives higher weight to years in the more remote past and $\lambda = 0$ weights all past years equally (see Figure II of Malmendier and Nagel (2011) for an illustration of the non-linearity implied by the functional form).

This specification produces an age-specific weighting function for past experience and furthermore only allows for a weakly monotonic weight distribution. The latter precludes a bimodal, e.g., a hump-shaped or U-shaped weight distribution.

Our second alternative, which we label *age range experience*, acknowledges that the effect of a crisis may differ depending on the life stage in which it is experienced. For example, events during the f*ormative years* from 18 to 25 may be particularly important for shaping preferences and beliefs (Spilimbergo and Giuliano, 2014). We define age range crisis experience as the average number of years in which a crisis was experienced during the respective age range.

### 4.B.1.1 Results for the WVS

Table 4.B.2 mirrors Table 4.3 from the main text but uses lambda experience instead of delta experience as the experience stock measure. Again, banking crisis and trust are robustly negatively related.

Figure 4.B.2 illustrates the decay effect of past experiences that we already documented in the main text. Furthermore, this specification allows for heterogeneity across individuals of different age, holding constant the time distance between the macroeconomic crisis event and today.

**Table 4.B.2.** World Values Survey: Trust and crisis experience using lambda approach

| Dependent variable: 1 if trusting | $\lambda = 1.5$ | | | | | $\lambda = 1.0$ | $\lambda = 2.0$ |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Experience stock of banking crises | -0.119** | -0.139*** | -0.146** | -0.194** | -0.217*** | -0.263*** | -0.183*** |
| | (0.047) | (0.047) | (0.059) | (0.079) | (0.071) | (0.081) | (0.064) |
| Constant | 0.139*** | 0.153*** | 0.060* | -0.034 | 0.019 | 0.028 | 0.011 |
| | (0.032) | (0.030) | (0.035) | (0.035) | (0.137) | (0.136) | (0.132) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Crisis at time of survey | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional controls | | | Yes | Yes | Yes | Yes | Yes |
| Country-age trends | | | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.121 | 0.122 | 0.127 | 0.131 | 0.146 | 0.146 | 0.146 |
| N | 120427 | 120427 | 100405 | 100405 | 76921 | 76921 | 76921 |
| # countries | 44 | 44 | 43 | 43 | 30 | 30 | 30 |
| # survey years | 17 | 17 | 16 | 16 | 13 | 13 | 13 |

*Notes.* Ordinary least squares regression. Robust standard errors clustered at the country level in parentheses. Levels of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Other macroeconomic experiences include average percentage change in GDP, as well as inflation, stock market and domestic debt crises. Baseline controls include gender, level of education and (subjective) social class. The additional controls include income decile, indicators for unemployment and marital status, subjectively assessed religiousness and number of children.

We run age range regressions to investigate explicitly whether the effects of banking crises on trust depend on the stage of life at which an individual experiences a banking crisis. Giuliano and Spilimbergo (2014) test whether macroeconomic recessions are especially relevant for political attitudes when they are experienced between the ages of 18 and 25. We follow them in their division of age ranges.

Table 4.B.3 shows that the effects of banking crises on trust are stronger when an individual is younger at the time of the crisis. In particular, we observe that the effect on trust is negative when an individual is below around 40 years old during a banking crisis. If a crisis is experienced at a later age, the effect is not statistically significant and even positive.

The *formative years* hypothesis states that social preferences are formed in the early twenties. Our result suggests that beliefs about trust also form in the earlier years of an individual's life (at least as far as trust is affected by the experience of macroeconomic crises).

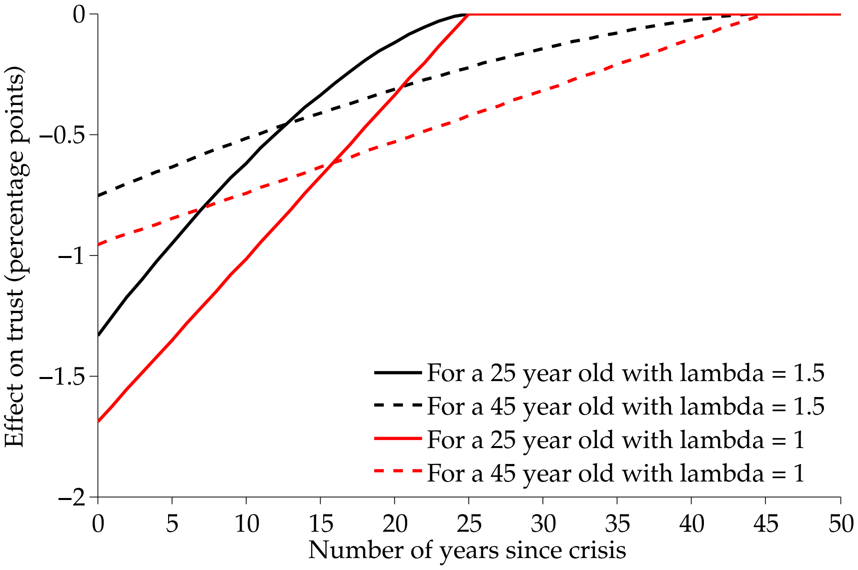


**Figure 4.B.2.** Estimated effect of experiencing a banking crisis X years ago on trust today, for different for different parameter values. Calculated effects sizes are based on our estimation results using the Malmendier and Nagel (2011) one-parameter experience specification. The red curves assume a weighting parameter $\lambda = 1$, the dark curves assume $\lambda = 1.5$.

### 4.B.1.2 Results for the GSS

Mirroring the preceding analysis of the WVS results to alternative experience stock definitions, we show that GSS results in Section 4.4 are robust across different experience stock specifications.

Table 4.B.4 shows results using lambda experience. We again find significant effects for both the binary and the continuous version of the bank default definition.

Similarly, table 4.B.5 shows age range specifications for the GSS data. Owing to the relatively small number of observations, most coefficients are imprecisely estimated. Estimates based on the binary bank default indicator suggest that bank defaults experienced in early adulthood leave a stronger footprint on interpersonal trust. The sign of coefficient estimates based on the continuous bank default variable are also consistent with that conclusion but are not statistically significant.

**Table 4.B.3.** World Values Survey: Trust and crisis experience using the age range approach

| Dependent variable: 1 if trusting | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Exp. crises 16–25 years | -0.015* | -0.015 | | | | | | |
| | (0.009) | (0.010) | | | | | | |
| Exp. crises 26–35 years | | | -0.026*** | -0.022** | | | | |
| | | | (0.009) | (0.008) | | | | |
| Exp. crises 36–45 years | | | | | -0.020* | -0.007 | | |
| | | | | | (0.012) | (0.011) | | |
| Exp. crises 46–50 years | | | | | | | 0.004 | -0.001 |
| | | | | | | | (0.007) | (0.006) |
| Constant | 0.121*** | 0.051 | 0.287 | 0.258 | -0.010 | -0.076 | -0.013 | -0.237** |
| | (0.033) | (0.041) | (0.191) | (0.230) | (0.056) | (0.091) | (0.057) | (0.091) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country-age trends | | Yes | | Yes | | Yes | | Yes |
| Other macro crisis experience | | Yes | | Yes | | Yes | | Yes |
| Additional controls | | Yes | | Yes | | Yes | | Yes |
| $R^2$ | 0.148 | 0.159 | 0.150 | 0.160 | 0.151 | 0.162 | 0.151 | 0.160 |
| N | 128217 | 111007 | 91866 | 79393 | 59593 | 51272 | 46257 | 39672 |
| # countries | 45 | 44 | 45 | 44 | 45 | 44 | 45 | 44 |
| # survey years | 19 | 18 | 19 | 18 | 19 | 18 | 19 | 18 |

*Notes.* Ordinary least squares regression. Robust standard errors clustered at the country level in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Other macroeconomic experiences include average percentage change in GDP, as well as inflation, stock market and domestic debt crises. Baseline controls include gender, level of education and (subjective) social class. The additional controls include income decile, indicators for unemployment and marital status, subjectively assessed religiousness and number of children.

**Table 4.B.4.** General Social Survey: Trust and crisis experience using the lambda approach

| Dependent variable: 1 if trusting | $\lambda = 1.5$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Experience stock of bank failures | -0.099*** | -0.129*** | -0.135*** | -0.130*** | | | | |
| | (0.025) | (0.025) | (0.025) | (0.024) | | | | |
| Estimated losses | | | | | -51.538*** | -56.553*** | -57.192*** | -55.956*** |
| | | | | | (15.197) | (17.326) | (17.691) | (17.994) |
| Constant | 1.329*** | 1.293*** | 1.272*** | 1.782*** | 1.328*** | 1.289*** | 1.267*** | 1.776*** |
| | (0.044) | (0.080) | (0.083) | (0.091) | (0.043) | (0.080) | (0.083) | (0.091) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bank failures at time of survey | Yes | Yes | Yes | Yes | | | | |
| Estimated losses in year of survey | | | | | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | | Yes | Yes | Yes | | Yes | Yes | Yes |
| Additional controls | | | Yes | Yes | | | Yes | Yes |
| State-age trends | | | | Yes | | | | Yes |
| $R^2$ | 0.136 | 0.139 | 0.141 | 0.145 | 0.136 | 0.139 | 0.141 | 0.144 |
| N | 14951 | 14950 | 14911 | 14911 | 14951 | 14950 | 14911 | 14911 |
| # states | 51 | 51 | 51 | 51 | 51 | 51 | 51 | 51 |
| # survey years | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 |

Ordinary least squares regression. Robust standard errors clustered at the country level in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Baseline controls include gender, age, income bracket and level of education. The additional controls further include unemployment and marital status, religion and number of children.

**Table 4.B.5.** General Social Survey: Trust and crisis experience using the age range approach

| Dependent variable: 1 if trusting | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Exp. failures 16–25 years | -0.069** | | | | | | | |
| | (0.029) | | | | | | | |
| Estim. losses 16–25 years | | -20.920 | | | | | | |
| | | (20.679) | | | | | | |
| Exp. failures 26–35 years | | | -0.058* | | | | | |
| | | | (0.031) | | | | | |
| Estim. losses 26–35 years | | | | -38.777 | | | | |
| | | | | (32.716) | | | | |
| Exp. failures 36–45 years | | | | | 0.057 | | | |
| | | | | | (0.062) | | | |
| Estim. lossess 36–45 years | | | | | | 49.859 | | |
| | | | | | | (50.221) | | |
| Exp. failures 46–50 years | | | | | | | -0.033 | |
| | | | | | | | (0.050) | |
| Estim. losses 46–50 years | | | | | | | | -77.953** |
| | | | | | | | | (35.290) |
| Constant | 1.968*** | 1.961*** | 4.256 | 4.263 | 4.257 | 4.315 | 3.597 | 3.648 |
| | (0.072) | (0.072) | (2.941) | (2.942) | (3.549) | (3.518) | (2.409) | (2.416) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bank failures at time of survey | Yes | | Yes | | Yes | | Yes | |
| Estimated losses at time of survey | | Yes | | Yes | | Yes | | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State-age trends | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.138 | 0.138 | 0.143 | 0.143 | 0.164 | 0.164 | 0.182 | 0.183 |
| N | 11800 | 11800 | 9080 | 9080 | 5976 | 5976 | 3725 | 3725 |
| # states | 51 | 51 | 51 | 51 | 48 | 48 | 49 | 49 |
| # survey years | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 |

*Notes.* Ordinary least squares regression. Robust standard errors clustered at the country level in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Baseline controls include gender, age, income bracket and level of education. The additional controls further include unemployment and marital status, religion and number of children.

## Appendix 4.C Nonlinear Specifications: Probit and Ordered Probit

For our main analyses we estimated linear probability models. Below we display analogous results to those in the main text using non-linear specifications estimated as probit or ordered probit models.

Table 4.C.6 corresponds to table 4.2, table 4.C.7 to table 4.3, table 4.C.8 to table 4.4, and table 4.C.9 to table 4.5. All results are qualitatively identical.

**Table 4.C.6.** World Values Survey: Experience of different types of crises and the effect on trust

| Dependent variable: 1 if trusting | $\delta = 0.98$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
| Most people can be trusted | | | | | | | | | | | | | | |
| Experience stock of output crises | 0.091 | 0.128 | | | | | | | | | | | | |
| | (0.117) | (0.171) | | | | | | | | | | | | |
| Experience stock of inflation crises | | | 0.004 | 0.006 | | | | | | | | | | |
| | | | (0.029) | (0.038) | | | | | | | | | | |
| Experience stock of banking crises | | | | | -0.066*** | -0.081*** | | | | | | | | |
| | | | | | (0.022) | (0.025) | | | | | | | | |
| Experience stock of stock market crises | | | | | | | 0.025 | 0.039 | | | | | | |
| | | | | | | | (0.030) | (0.042) | | | | | | |
| Experience stock of currency crises | | | | | | | | | -0.005 | -0.004 | | | | |
| | | | | | | | | | (0.025) | (0.042) | | | | |
| Experience stock of domestic debt crises | | | | | | | | | | | 0.045 | 0.165 | | |
| | | | | | | | | | | | (0.041) | (0.124) | | |
| Experience stock of foreign debt crises | | | | | | | | | | | | | 0.011 | 0.031 |
| | | | | | | | | | | | | | (0.045) | (0.068) |
| Constant | -1.420*** | -1.635*** | -1.161*** | -1.645*** | -1.186*** | -1.794*** | -1.316*** | -2.164*** | -1.130*** | -1.666*** | -1.147*** | -1.701*** | -1.162*** | -1.702*** |
| | (0.143) | (0.096) | (0.225) | (0.348) | (0.102) | (0.112) | (0.242) | (0.322) | (0.152) | (0.247) | (0.086) | (0.106) | (0.090) | (0.110) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Crisis at time of survey | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional controls | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes |
| Country-age trends | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes | | Yes |
| Pseudo R$^2$ | 0.113 | 0.120 | 0.106 | 0.112 | 0.106 | 0.113 | 0.110 | 0.117 | 0.106 | 0.112 | 0.106 | 0.113 | 0.106 | 0.113 |
| N | 97672 | 81700 | 120396 | 100387 | 120396 | 100387 | 108796 | 89640 | 120396 | 100387 | 120396 | 100387 | 120396 | 100387 |
| # countries | 33 | 32 | 44 | 43 | 44 | 43 | 39 | 38 | 44 | 43 | 44 | 43 | 44 | 43 |
| # survey years | 15 | 14 | 17 | 16 | 17 | 16 | 15 | 14 | 17 | 16 | 17 | 16 | 17 | 16 |

*Notes.* Probit regression. Robust standard errors clustered at the country level in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Crisis at the time of survey is a dummy for whether there was a respective type of crisis in the year in which the respondent answered the survey. Baseline controls include gender, level of education and (subjective) social class. The additional controls include income decile, indicators for unemployment and marital status, subjectively assessed religiousness and number of children.

**Table 4.C.7.** World Values Survey: Trust and banking crisis experience

| Dependent variable: 1 if trusting | $\delta = 0.98$ | | | | | $\delta = 0.95$ | $\delta = 0.90$ |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Most people can be trusted | | | | | | | |
| Experience stock of banking crises | -0.041*** | -0.042*** | -0.049*** | -0.053*** | -0.056*** | -0.064*** | -0.068*** |
| | (0.015) | (0.016) | (0.017) | (0.018) | (0.017) | (0.019) | (0.025) |
| Constant | -0.997*** | -0.935*** | -1.139*** | -1.396*** | -1.758*** | -1.624*** | -1.592*** |
| | (0.120) | (0.122) | (0.156) | (0.155) | (0.400) | (0.424) | (0.383) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Banking crisis at time of survey | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional controls | | | Yes | Yes | Yes | Yes | Yes |
| Country-age trends | | | | Yes | Yes | Yes | Yes |
| Other macro crisis experience | | | | | Yes | Yes | Yes |
| Pseudo $R^2$ | 0.105 | 0.106 | 0.110 | 0.113 | 0.113 | 0.113 | 0.114 |
| N | 120412 | 120396 | 100387 | 100387 | 100387 | 100387 | 100387 |
| # countries | 44 | 44 | 43 | 43 | 43 | 43 | 43 |
| # survey years | 17 | 17 | 16 | 16 | 16 | 16 | 16 |

*Notes.* Probit regression. Robust standard errors clustered at the country level in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Other macroeconomic experiences include average percentage change in GDP, as well as inflation, stock market and domestic debt crises. Baseline controls include gender, level of education and (subjective) social class. The additional controls include income decile, indicators for unemployment and marital status, subjectively assessed religiousness and number of children.

**Table 4.C.8.** World Values Survey: Confidence in institutions, $\delta = 0.98$

| | Confidence in institutions (1=“None at all” to 4=“A great deal”) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dependent variable: | Government | | Parliament | | Political parties | | Banks | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Experience stock of banking crises | -0.041*** | -0.074*** | -0.045** | -0.077** | -0.031** | -0.055** | 0.013 | 0.028 |
| | (0.013) | (0.022) | (0.019) | (0.030) | (0.014) | (0.022) | (0.018) | (0.032) |
| Experience stock of output crises | 0.166** | 0.229*** | 0.118*** | 0.182*** | 0.091** | 0.133*** | 0.025 | -0.548** |
| | (0.078) | (0.084) | (0.042) | (0.061) | (0.042) | (0.049) | (0.060) | (0.277) |
| Experience stock of inflation crises | 0.019** | 0.023* | 0.016*** | 0.012 | 0.012** | 0.008 | 0.021*** | 0.131*** |
| | (0.008) | (0.012) | (0.006) | (0.010) | (0.005) | (0.010) | (0.007) | (0.044) |
| Experience stock of stock market crises | -0.003 | 0.001 | -0.000 | 0.019 | 0.002 | 0.018 | -0.008 | 0.045 |
| | (0.007) | (0.014) | (0.006) | (0.013) | (0.006) | (0.012) | (0.010) | (0.092) |
| Experience stock of domestic debt crises | 0.057** | 0.055 | 0.041*** | 0.023 | 0.039*** | -0.025 | -0.040 | -0.196 |
| | (0.025) | (0.045) | (0.014) | (0.037) | (0.012) | (0.025) | (0.026) | (0.193) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional controls | | Yes | | Yes | | Yes | | |
| Any crisis at time of survey | | Yes | | Yes | | Yes | | Yes |
| Country-age trends | | Yes | | Yes | | Yes | | Yes |
| Pseudo R$^2$ | 0.051 | 0.055 | 0.068 | 0.064 | 0.063 | 0.056 | 0.075 | 0.106 |
| N | 145673 | 86754 | 147606 | 88492 | 146911 | 88815 | 37495 | 4118 |
| # countries | 44 | 36 | 44 | 37 | 44 | 37 | 27 | 3 |
| # waves | 4 | 4 | 5 | 4 | 4 | 4 | 2 | 2 |
| # survey years | 17 | 13 | 18 | 13 | 17 | 13 | 4 | 2 |

*Notes.* Ordered probit regression. Robust standard errors clustered at the country level in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The dependent variables have four levels: 1=“None at all”, 2=“Not very much”, 3=“Quite a lot”, and 4=“A great deal”. Alternative specifications with ordinal regressions are reported in appendix 4.C. Other macroeconomic experiences include average percentage change in GDP, as well as inflation, stock market and domestic debt crises. Baseline controls include gender, level of education and (subjective) social class. The additional controls include income decile, indicators for unemployment and marital status, subjectively assessed religiousness and number of children. We do not use additional controls in specification (8) as this reduces the available sample to a single wave.

**Table 4.C.9.** General Social Survey: Trust and crisis experience

| Dependent variable: 1 if trusting | $\delta = 0.98$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| can people be trusted | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Experience stock of bank failures | -0.039*** | -0.040*** | -0.043*** | -0.039*** | | | | |
| | (0.007) | (0.007) | (0.008) | (0.007) | | | | |
| Estimated losses | | | | | -21.370*** | -21.859*** | -22.814*** | -19.612*** |
| | | | | | (4.125) | (3.979) | (3.960) | (4.231) |
| Constant | 2.643*** | 2.497*** | 2.419*** | 3.872*** | 2.632*** | 2.484*** | 2.404*** | 3.857*** |
| | (0.143) | (0.260) | (0.268) | (0.292) | (0.143) | (0.257) | (0.265) | (0.290) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bank failures at time of survey | Yes | Yes | Yes | Yes | | | | |
| Estimated losses in year of survey | | | | | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | | Yes | Yes | Yes | | Yes | Yes | Yes |
| Additional controls | | | Yes | Yes | | | Yes | Yes |
| State-age trends | | | | Yes | | | | Yes |
| Pseudo $R^2$ | 0.112 | 0.115 | 0.116 | 0.120 | 0.112 | 0.115 | 0.116 | 0.120 |
| N | 14950 | 14949 | 14910 | 14910 | 14950 | 14949 | 14910 | 14910 |
| # states | 51 | 51 | 51 | 51 | 51 | 51 | 51 | 51 |
| # survey years | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 |

Probit regression. Robust standard errors clustered at the country level in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Baseline controls include gender, age, income bracket and level of education. The additional controls further include unemployment and marital status, religion and number of children.

## Appendix 4.D Additional Analyses

### 4.D.1 Trust as a Belief Versus Trust as a Preference

Our discussion leaves open whether interpersonal trust is a belief or is related to risk preferences. For instance, Ben-Ner and Putterman (2001) argue that trusting others can be risky and that trust should therefore be related to individual risk tolerance.[16] If individuals become less willing to take risks as a result of experiencing a banking crisis, this argument suggests that we should also see a reduction in interpersonal trust. In this section, we test this channel by regressing different measures of risk tolerance on the banking crisis experience stock.

The WVS does not elicit risk tolerance directly. Instead, we rely on three suggestive measures that have been used in the literature (e.g., Freese, 2004; Miller, 2000; Miller and Hoffmann, 1995). We use potential risk measures as outcome variables in otherwise unchanged versions of our regression specification in equation (4.1). Our three measures are:

- *Act boldly versus cautiously*: The WVS asks respondents whether they think it is important to act boldly to achieve major life changes or be cautious about major changes. The answer is coded on a 10-point scale with lower values corresponding to "more cautious."

- *Adventure and risk taking*: The WVS asks respondents to assess their similarity to a person for which adventure and risk taking are important. The answer is coded on a six-point scale, and we inverted the scale such that a higher value corresponds to "This person is very much like me."

- *Job security*: Respondents are shown several (16) dimensions of a job (e.g., good hours, good chance for promotion) and indicate which dimensions they find important. Job security is coded as 1 if it is mentioned by the respondent.

Table 4.D.10 shows that we do not find a consistent effect of banking crises on any of the three measures of risk. Coefficients are small and insignificant for two of the measures. The coefficient's sign in column (4) is significant and in the expected direction, but we are skeptical because the sign switches only when additional controls are included. If banking crises affect risk preferences, the effect appears to be rather weak.

[16] The empirical evidence on this hypothesis is not clear-cut: Eckel and Wilson (2004) do not find a strong correlation between trust and risk preferences in laboratory experiments. Bohnet and Zeckhauser (2004) find *betrayal* aversion in a trust game, and Bohnet et al. (2008) extend the evidence to several countries.

We interpret these results as suggestive evidence for an effect of banking crises on trust beliefs rather than via a risk preference channel. This interpretation is consistent with Sapienza et al. (2013), who demonstrate that the trust question in the WVS largely captures beliefs, and, as mentioned above, with Eckel and Wilson (2004) who do not find a strong relationship between trust and risk tolerance.

**Table 4.D.10.** World Values Survey: Risk preferences, $\delta = 0.98$

| | Self-assessment: Risk attitude | | | | | |
|---|---|---|---|---|---|---|
| Dependent variable: | Act boldly vs. cautious to achieve major life changes | | Adventure and taking risks is important | | Job security is important | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Experienced banking crises | -0.000 | -0.005 | 0.116 | -0.229** | -0.008 | 0.003 |
| | (0.030) | (0.027) | (0.110) | (0.096) | (0.007) | (0.005) |
| Experience stock of output crises | -0.153* | -0.095** | -0.081 | 0.029 | -0.001 | 0.006 |
| | (0.087) | (0.043) | (0.135) | (0.089) | (0.014) | (0.016) |
| Experience stock of inflation crises | 0.018 | -0.006 | -0.043 | 0.045* | -0.002 | 0.000 |
| | (0.013) | (0.015) | (0.028) | (0.024) | (0.002) | (0.004) |
| Experience stock of stock market crises | -0.010 | 0.006 | 0.052* | -0.018 | -0.001 | -0.005 |
| | (0.013) | (0.015) | (0.030) | (0.031) | (0.001) | (0.004) |
| Experience stock of domestic debt crises | 0.009 | 0.018 | -0.007 | 0.131** | -0.002 | -0.002 |
| | (0.032) | (0.024) | (0.021) | (0.050) | (0.002) | (0.006) |
| Constant | 3.434*** | 4.002*** | 5.235*** | 7.603*** | 0.422*** | 0.441*** |
| | (0.232) | (0.261) | (1.049) | (0.687) | (0.063) | (0.057) |
| Age FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Additional controls | | Yes | | Yes | | Yes |
| Any crisis at time of survey | | Yes | | Yes | | Yes |
| Country-age trends | | Yes | | Yes | | Yes |
| $R^2$ | 0.195 | 0.187 | 0.108 | 0.112 | 0.116 | 0.133 |
| N | 78541 | 35025 | 35919 | 28570 | 68353 | 53140 |
| # countries | 40 | 27 | 28 | 24 | 35 | 30 |
| # waves | 2 | 2 | 1 | 1 | 3 | 2 |
| # survey years | 8 | 5 | 5 | 4 | 10 | 8 |

*Notes.* Ordinary least squares regression. Robust standard errors clustered at the country level are in parentheses. Levels of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Other macroeconomic experiences include average percentage change in GDP, as well as inflation, stock market and domestic debt crises. Baseline controls include gender, level of education and (subjective) social class. The additional controls include income decile, indicators for unemployment and marital status, subjectively assessed religiousness and number of children.

# 5

# Negative Long-run Effects of Prosocial Behavior on Happiness

*Joint with Armin Falk*

## 5.1 Introduction

Happiness is a key concept and building block of modern societies. Philosophers put happiness center stage as a fundamental driving force, life goal, and even natural right of humans (Cumberland and Maxwell, 1727; Leibniz, 1988; Locke, 1988; Wollaston, 1759). Recently, subjective well-being has gained increasing acceptance as a relevant indicator of a society's welfare and various countries have incorporated national happiness levels into their economic policy objectives (Layard, 2011; Oswald and Wu, 2010). Likewise, the nature of prosocial behavior has attracted sustained interest for centuries as an identifying feature of human existence. Scholars across diverse fields such a philosophy, psychology, economics, organizational science, political science and neuroscience have studied the consequences of other-regarding as opposed to selfish behavior at the individual and societal levels (Batson and Powell, 2003; Bénabou and Tirole, 2006; Latané and Darley, 1970). A long-standing and divisive hypothesis connects happiness to prosociality, suggesting that prosocial behavior is a reliable source of happiness.[1]

Recently, scientific interest in this topic has surged anew, putting the empirical validity of the proposed relationship to the test. This empirical literature forcefully argues for a positive association between prosocial behavior and happiness (Aknin et al., 2015, 2012; Dunn et al., 2008, 2014; Lyubomirsky et al., 2005; Thoits and Hewitt, 2001), up to the point of proclaiming it a new "psycho-

[1] This hypothesis dates back at least two millennia to Aristotle's (1987) *Nicomachean Ethics*, in which he suggests a fundamental link between well-being, *eudemonia*, and moral behavior.

logical universal" (Aknin et al., 2013a). In an influential experiment, Elizabeth Dunn, Lara Aknin and Michael Norton randomly gave a small amount of money to subjects and told one group to spend it on a gift for themselves and another group to spend it on a gift to others (Dunn et al., 2008). Happiness elicited later that day was significantly greater among subjects instructed to spend the money on others. Yet, as this example illustrates, rigorous evidence on a universal effect of prosocial behavior on happiness is hard to obtain. First, real-life decisions of interest have more far-reaching consequences than the small-stakes decisions typically studied in experimental research. More substantial choices might not only affect happiness more strongly, but also in different ways. Second, all forms of giving come at a cost to the giver. The alternative of donating money, for instance, is not simply the absence of that donation, but it is what the person would otherwise buy with that money. Awareness about the self-benefiting counterfactual is a key feature of prosocial behavior. Third, small-stakes decisions are unlikely to have lasting effects on happiness. This and other experimental constraints impede investigating the temporal profile of happiness derived from prosocial as opposed to selfish behavior. Virtually all activities, however, generate distinct time-varying patterns of happiness. A tempting dessert creates immediate happiness that quickly fades, while an exhausting workout can reduce happiness in the short term but be a source of happiness with some delay. Fourth, to identify causal relationships, experiments typically "force" people into prosocial or selfish behavior so as to create random variation. This prevents observing what subjects *would have chosen* for themselves. Such choice data, however, is crucial to classify subjects into more prosocial versus more selfish "types", who plausibly react in different ways to an enforced prosocial outcome. Fifth, social activities are easily mistaken for prosocial, other-regarding behaviors. Prosocial behavior *can* imply more social interactions, which are known to increases happiness (Aknin et al., 2013b; Diener and Seligman, 2002). However, this is a mere consequence of prosocial behavior that is often absent, as for instance in monetary donations. In fact, social connections can similarly result from purely selfish behavior.

We took the existing evidence as a point of departure and put the happiness hypothesis to a new test. We designed a behavioral experiment that addresses previous limitations and the above-noted concerns, expecting to generate robust affirmative evidence for the hypothesis of a positive causal relationship from prosocial behavior to happiness.

## 5.2 Saving a Human Life Paradigm

To examine a meaningful prosocial decision that plausibly affects happiness, we designed a high-stakes study paradigm in which subjects faced a life-and-death situation. Each participant took a binary decision to either save a human life in expectation, or not to save a human life. We developed the *Saving a Life* paradigm in cooperation with the Indian non-profit organization Operation ASHA. Operation ASHA specializes in the treatment of tuberculosis, an infectious disease caused by bacteria. With an estimated global death toll of 1.7 million people in 2016, tuberculosis kills more people than HIV or malaria, making it the deadliest infectious disease of today (World Health Organization, 2018). Highly effective and low-cost treatment with antibiotics is available for drug-susceptible tuberculosis. We calculated the cost of a life saved by Operation ASHA based on public information on the charity's operations in combination with estimates from peer-reviewed epidemiological studies on tuberculosis mortality for the specific type of treatment and location considered (Kolappan et al., 2008; Straetemans et al., 2011; Tiemersma et al., 2011) Under conservative assumptions, a donation of 350 euros – roughly 400 dollars at the time – covers all costs incurred by Operation ASHA to identify, treat and cure five patients, which is equivalent to saving one human life in expectation (see Appendix for more details). To construct a strong alternative to saving a life that constitutes the individual "opportunity cost" of a moral outcome, every study participant who did not save a life received a payment of 100 euros. This payment implements a salient self-benefiting counterfactual to the prosocial outcome. Our binary setting generates an extreme personal conflict, with little room for interpretation of what is the "right" thing to do. In fact, the *Saving a Life* paradigm fundamentally presents not just a prosocial, but a moral tradeoff.[2]

This study is not a thought experiment with only hypothetical consequences. Instead, for each subject, we actually implemented a donation of 350 euros or an experimental payment of 100 euros. Moreover, our paradigm accommodates the common critique of donation experiments that subjects can take the money to donate it for other purposes, such that the seemingly selfish option is in fact altruistic. In our setting this is essentially ruled out given that foregoing 100 euros generates a substantially larger donation of 350 euros.

Identifying the causal effect of prosocial behavior on happiness requires exogenous variation in whether subjects save a life or receive the money. However, randomly allocating subjects into either condition, i.e., force them to save a life or take the money, does not allow observing which option a subject would choose for herself. This decision reflects a subject's personal intention and reveals whether she is a more altruistic or a more selfish "type". Not knowing a subject's choice, we would not be able to tell, for example, whether a subject who saved a life and is later found to be relatively unhappy is simply dissatisfied with not getting her desired outcome. To circumvent the dilemma between

[2] A general notion of morality defines immoral behavior as harming others in an unjustified and intentional way. See, e.g., B. Gert and J. Gert (2017).

observing voluntary choice and generating causal variation, we implemented a lottery procedure where subjects had to choose between two lotteries, Lottery A and Lottery B. Lottery A was the prosocial lottery. Choosing this lottery, the subject saved a life with 60% probability and received money with 40% probability, i.e., $L_A = (0.6,$ €350 donation; 0.4, €100 payment). Lottery B, the selfish lottery, featured the reverse probabilities, i.e., this lottery saved a life with only 40% probability and generated additional earnings of 100 euros with 60% probability, so that $L_B = (0.4,$ €350 donation; 0.6, €100 payment). Our procedure simultaneously provided random variation allowing for a causal identification of the effect of prosociality on happiness, as well as information on subject's prosocial intentions. Specifically, by choosing lottery A rather than lottery B, a subject made the prosocial outcome 50% more likely. The lottery procedure produced four different study groups. Outcomes were aligned with choices for those subjects who chose the altruistic lottery and ended up saving a life, and for those who chose the selfish lottery and received 100 euros. Outcomes and choices were not aligned among subjects who picked the altruistic lottery, but received money, and for subjects who chose selfishly, but nevertheless saved a life, respectively.

Our main measure of interest was subjects' self-reported happiness. Subjects completed the statement "In general, I consider myself" with responses ranging from 1 = "not a very happy person" to 7 = "a very happy person" on a Likert scale (Lyubomirsky and Lepper, 1999). To account for potential temporal patterns, we elicited happiness at three points in time. The first measurement, $H_{Baseline}$, provided a baseline level of happiness at the very beginning of the laboratory session, before the *Saving a Life* paradigm was presented. The second measurement, $H_{Short-run}$, was elicited shortly after the lottery was drawn, i.e., after subjects had learned about the respective outcome. Finally, $H_{Long-run}$ was measured four weeks after the laboratory session in an online survey. In between the second and third measurements, we sent three emails to all subjects, exactly one, three, and four weeks after the laboratory session. The purpose of these emails was to remind subjects of the content and outcome of the laboratory session.[3] The lottery outcome from the experiment was implemented exactly after two weeks, between the first and the second reminder. The corresponding transfer was said to be "in process" in the first email and "executed" in the second email. By the time the third email invited subjects to participate in the online survey after four weeks, the donation or personal bank transfer had already been carried out two

[3] Specifically, each email stated that in the study "you could make a decision about whether you rather want a human life to be saved for you, or whether you rather want to receive an additional payment of €100". Moreover, the text reminded that the subject would "receive an additional payment of €100" or that she "arranged for a donation in the amount of €350 for the treatment and cure of tuberculosis patients, such that one human life is saved in expectation" (see full text in Appendix).

weeks earlier. In addition to the happiness measures, we obtained self-reports of subjects' self-image, measured as agreement with the statement "I am a good person", and of subjects' mood. The happiness, self-image and mood questions formed part of a questionnaire including various other items, so as to obfuscate the purpose of the study and to reduce experimenter demand effects.

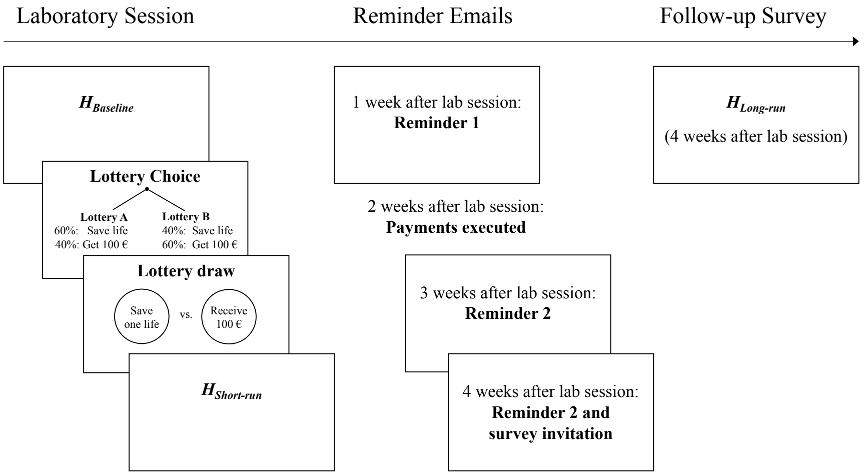


**Figure 5.1.** The lottery choice paradigm and repeated happiness measurements. In the initial laboratory session, each participant chose between two lotteries, the "prosocial" Lottery A and the "selfish" Lottery B. Based on their individual lottery draw, subjects either received 100 euros or saved one human life in expectation, in which case the experimenter transferred a donation of 350 euros to a charity that fights tuberculosis. The laboratory session was followed by an online survey four weeks later. We elicited happiness three times, at the beginning ($H_{Baseline}$) and at the end ($H_{Short\text{-}run}$) of the laboratory session, and again in the survey ($H_{Long\text{-}run}$). We sent out two personalized emails reminding each participant of their individual lottery outcome between the laboratory session and the survey.

We ran 10 laboratory sessions with a total of 325 subjects. Of those, 297 also participated in the follow-up online survey four weeks later and constitute our sample for the main analysis.[4] Several additional treatments allow for further analyses, see Section 5.5. In sum, we used data from 591 subjects.

For the short run, we made the following predictions. Our main hypothesis was a positive causal effect of saving a life on happiness. This would be in line with the existing body of evidence that analyzes the short-run effect of prosocial outcomes (Aknin et al., 2015, 2012; Dunn et al., 2008, 2014). In addition, we expected a positive (non-causal) effect of the prosocial choice itself, in as much as choosing the altruistic lottery might improve a person's self-image (Bénabou and Tirole, 2002, 2011), feel like "the right thing to do", or generate positive emotions ("warm glow", Andreoni (1990)). Moreover, we hypothesized that the alignment of choice and lottery outcome, i.e., whether people *got what they wanted*, affects mood, which might also be reflected in happiness. A person who picked the selfish lottery but saved a life might partly be less happy simply due to not getting her preferred outcome. Note that the distinction between choice, outcome and the interaction between these two is a central feature of our paradigm. Subjects initially opted for one of the two lotteries, and conditional on the lottery choice we can identify the causal effect of saving a life versus receiving money.

For the long run, the existing literature cannot provide specific hypotheses, as the causal evidence is limited to short-term effects. First, with respect to the causal effect of the prosocial outcome, one may expect that timing matters. Saving a life can provide happiness through a mental form of consumption from thoughts or memories that occurs entirely in the mind (Ariely and Norton, 2009). Such "conceptual consumption" presumably occurs in temporal proximity to the experiment but fades as time passes. Money, by contrast, generates happiness based on what a person buys with it and when. Happiness is then linked to the time when actual consumption occurs, which can be spread out over time. The long-term causal effect of the prosocial versus the selfish outcome is therefore ambiguous, i.e., even a negative effect is conceivable. Second, we further hypothesized that the effect of prosocial choice wanes over time, because past choices become less and less accessible to people's mind and thus lose relevance for generating positive self-image or feelings of warm-glow. At the same time, there is no reason to expect a reversed effect in the long run – unlike for prosocial outcomes.

To analyze the happiness data, we split the sample along two dimensions. The lottery choice provided an endogenous dichotomy between more altruistic and more selfish subjects. The outcome as determined by the individual lottery draw was fully random conditional on lottery choice and allowed for causal inference. For each subject in each of the four study groups, we calcu-

[4] Note that the relative frequency of altruistic lottery choice as well as other personality measures did not systematically differ for subjects who did not complete the follow-up survey (see Appendix).

lated two individual differences of reported happiness scores in the short and long run relative to the baseline level of happiness. That is, for each subject we obtained $\Delta_{Short-run} = H_{Short-run} - H_{Baseline}$ and $\Delta_{Long-run} = H_{Long-run} - H_{Baseline}$. These measures indicated individual-level *changes* in happiness over time. By comparing group averages of those individual changes, we can assess whether different lottery choices and lottery outcomes were associated with different changes in happiness. We complemented this group comparison with regression analyses to assess the size and significance of the main and interaction effects. We regressed the standardized level of a short-run or long-run measurement on an indicator variable that equals 1 if a subject chose the altruistic lottery and 0 otherwise, an indicator variable that is 1 if the subject's lottery draw determined that she would save a life and 0 if she received the money, an interaction term between these two variables, and the baseline level of the dependent variable. For these analyses, we standardized happiness scores at each point in time.

## 5.3 Prosocial Behavior Promoted Happiness in the Short Run

Overall, 60% of subjects ($N = 178$) chose the altruistic lottery. Based on the random lottery draws, 53% of the entire sample actually saved a human life ($N = 158$). Panel A of Figure 5.2 shows our main finding for the short run. Displayed is the mean *change* in happiness from the beginning to the end of the laboratory session, $\Delta_{Short-run}$. We made three observations on the short-run change in happiness. First, we found a positive relationship between prosocial behavior and happiness, in line with previous evidence. Irrespective of lottery choice, the outcome of saving a life was *causally* related to higher mean changes in happiness. The main effect of saving a life is an increase in happiness by 0.06 SD ($p < 0.1$, Table 5.1, column 1).[5] Moreover, we found that choosing the prosocial lottery was associated with a positive change in happiness, irrespective of the actual lottery outcome. The magnitude of the (non-causal) relationship was 0.14 SD ($p < 0.1$, Table 5.1, column 1). Note, however, that size and significance of both main effects were moderate. Considering point estimates of mean changes in happiness in Figure 5.2, we observed a gradual decline from the prosocial-choice-prosocial-outcome group to the selfish-choice-selfish-outcome group. Moreover, the insignificant interaction effect of lottery choice and lottery outcome in the regression implied that changes in happiness were not driven by

[5] The main effect of saving a life (as compared to receiving money) is the average of the effects for those who chose the altruistic lottery and those who chose the selfish lottery. This main effect is reported in the footer of Table 5.1. Note that the coefficients reported in row 2 of the regression table indicate the estimated effect of saving a life for those who chose the selfish lottery.

participants getting or not getting what they wanted, as judged by their lottery choice (Table 5.1, column 1).

Second, we documented pronounced effects of both prosocial choice and the prosocial outcome on changes in the measure of subjects' self-image. Subjects who chose the altruistic lottery reaped self-image benefits as indicated by a quantitatively large correlational effect (0.38 SD, $p < 0.01$, Table 5.1, column 2). More compellingly, the random variation in lottery outcomes identified a sizable and highly significant causal effect of saving a life conditional on preceding choice (0.25 SD, $p < 0.01$, Table 5.1, column 2). As with happiness, we found no significant interaction between choice and outcome, suggesting that the fact of (not) getting the desired outcome did not affect changes in happiness or self-image.

Third, in contrast to happiness and self-image, changes in mood strongly reflected whether subjects got what they wanted. Choosing prosocially was associated with an increase in a person's mood, but only if she indeed ended up saving the life (0.54 SD, $p < 0.01$, Table 5.1, column 3). Mood declined by a similar magnitude after opting for the prosocial lottery if that person instead got the money (-0.62 SD, $p < 0.01$, Table 5.1, column 3), such that the main effect of choice across both groups was close to zero and insignificant (-0.04 SD, $p = 0.91$, Table 5.1, column 3). Considering the effect of randomized lottery outcomes, we found that saving a life did not significantly affect mood of subjects who chose the selfish lottery (0.21 SD, $p = 0.38$, Table 5.1, column 3), but led to a strongly positive and highly significant effect on those who picked the altruistic lottery (1.36 SD, $p < 0.01$, Table 5.1, column 3).

## 5.4 Negative Effect of Prosocial Behavior on Happiness in the Long Run

For the long-run effects on happiness, we performed analyses that are analogous to the short run, this time based on group averages of changes in happiness observed four weeks later, $\Delta_{Long-run}$. The key finding is presented in panel B of Figure 5.2. We found a strongly *negative* causal effect of saving a life on happiness. Notably, this effect was independent of the lottery choice. That means not only those subjects who picked the selfish lottery, but also those who opted for the altruistic lottery reported substantially higher levels happiness if they ended up receiving the money, rather than saving a life. Moreover, the effect on happiness was quantitatively large, especially in light of the four-week delay. Regression results showed that saving a life decreased long-run happiness by 0.26 SD ($p < 0.01$, Table 5.1, column 4) relative to receiving money, controlling for individual-specific baseline levels of happiness. This finding was in marked contrast to our own and other studies' short-term finding of a prevailing positive association. It also implied a pronounced temporal pattern of happiness derived from prosocial behavior. In fact, the sign of the causally induced change in happiness switched over time. Moreover, over time the outcome got relatively more important than the choice in determining happiness: In the short run, our data revealed marginally significant positive effects of prosocial lottery choice (0.14 SD, correlational) and of saving a life (0.06 SD, causal), while long term happiness was swayed by a highly significant and sizable negative causal effect of saving a life (0.26 SD).

Interestingly, the only short-term relationship that persisted in the long run was a positive association between happiness and altruistic lottery choice among those who ended up receiving money (0.28 SD, $p < 0.05$, Table 5.1, column 4). Put differently, choosing prosocially made people happy even in the long run, but only if they did not have to bear the "cost" of that decision in the sense of foregoing their payment. These subjects credibly revealed their prosocial intentions, but were randomly selected to experience the self-benefitting outcome ex post. While this kept them from actually saving a life, they did not have to sacrifice the high payment, *and* they were able to tell themselves (and/or others) that they had done what was in their power to bring about the moral outcome. We completed our analysis of the long-term results by considering self-image and mood. The strong effects of both choice and outcome on these measures observed in the short term faded out as time passed (Table 5.1, columns 5 and 6).

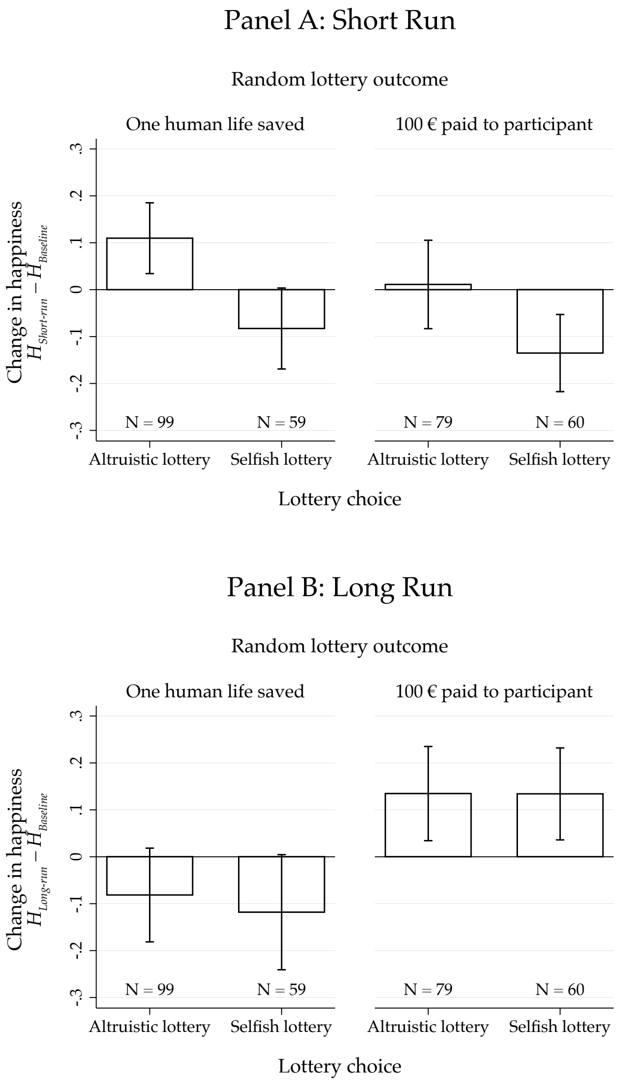


**Figure 5.2.** Prosocial behavior increased happiness in the short run but decreased it in the long run. Panel A shows the mean changes in self-reported happiness between the end and the beginning of the laboratory session for each of the four study groups. Regression analyses confirmed a casual main effect of saving a life (0.06 SD, $p < 0.1$, Table 5.1, column 1) and a (non-causal) positive main effect of choosing the prosocial lottery (0.14 SD, $p < 0.1$, Table 5.1, column 1). Panel B displays mean changes in happiness after four weeks compared to the beginning of the laboratory session. The causal effect of saving a life on happiness after four weeks was negative and quantitatively large (-0.25 SD, $p < 0.01$, Table 5.1, column 4). Happiness scores were standardized at each point in time. N = 297. Error bands indicate ± 1 SEM.

**Table 5.1.** Regression analyses of the effect of prosocial behavior on happiness, self-image and mood in the short run and the long run

| | Short run | | | Long run | | |
|---|---:|---:|---:|---:|---:|---:|
| Dependent variable (standardized): | $H_{Short-run}$ | Self-image | Mood | $H_{Long-run}$ | Self-image | Mood |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Lottery choice: 1 if altruistic, 0 if selfish | 0.11 | 0.41*** | -0.62*** | -0.05 | 0.28** | 0.17 |
| | (0.112) | (0.135) | (0.149) | (0.129) | (0.128) | (0.172) |
| Lottery outcome: 1 if life saved, 0 if money received | 0.03 | 0.28** | 0.21 | -0.29* | 0.25 | 0.14 |
| | (0.116) | (0.132) | (0.179) | (0.150) | (0.159) | (0.190) |
| Altruistic lottery choice * life saved | 0.06 | -0.05 | 1.16*** | 0.06 | -0.27 | -0.14 |
| | (0.159) | (0.166) | (0.212) | (0.190) | (0.193) | (0.235) |
| $H_{Baseline}$ | 0.77*** | | | 0.61*** | | |
| | (0.065) | | | (0.056) | | |
| Self-image at begin of session | | 0.52*** | | | 0.46*** | |
| | | (0.033) | | | (0.041) | |
| Mood at begin of session | | | 0.10** | | | 0.18*** |
| | | | (0.037) | | | (0.041) |
| Constant | -3.65*** | -4.45*** | -0.84*** | -2.64*** | -3.83*** | -1.47*** |
| | (0.316) | (0.279) | (0.302) | (0.294) | (0.344) | (0.356) |
| Main effect choice: altruistic lottery | .14* | .38*** | -.04 | -.02 | .14 | .11 |
| Main effect outcome: life saved | .06* | .25*** | .78*** | -.26*** | .12 | .07 |
| $R^2$ | .5426 | .5564 | .3127 | .3642 | .391 | .08326 |
| N | 297 | 297 | 297 | 297 | 297 | 297 |

*Notes*: Displayed are regression results that complement the findings in Figure 5.2. In each column, we regressed a standardized happiness, self-image, or mood score on an indicator variable that equals 1 if a subject chose the altruistic lottery and 0 otherwise, an indicator variable that is 1 if the subject's lottery draw determined that she would save a life and 0 if she received the money, an interaction term between these two variables, and the baseline level of the dependent measure. Ordinary least squares regression. Analogous ordered probit regressions are reported in Table S2. Robust standard errors in parentheses. Significance levels: $*$ p < 0.1, $*$ $*$ p < 0.05, $*$ $*$ $*$ p < 0.01.

## 5.5 Robustness

A number of additional analyses examined whether our findings were the artifact of a specific experimental design. One feature that distinguishes the *Saving a Life* paradigm from previous studies is stake size. We validated the credibility of the paradigm in a separate calibration treatment run with a different set of subjects. For each subject, we elicited the minimum amount of money that she preferred over saving a life by triggering a donation of 350 euros. This was done using an incentive compatible price-list method, such that each subject indeed either saved a life or received money (see Appendix). In a sample of 45 students, we found that the median valuation was 200 euros, an amount close to the average monthly disposable income of a German student. That subjects were willing to forgo substantial amounts of money indicated that the high-stakes experimental paradigm was credible. Based on the calibration we chose an amount of 100 euros for our main treatment, which provides a greater contrast to the amount of the donation. We therefore expected more than 50% of subjects to be willing to save a life in our main experiment, which at 60% was the case.

Next, we investigated the confirmatory validity of personality measures that the existing literature has linked to prosociality, i.e., we tested whether prosocial lottery choice in our data was predicted in a plausible manner by these measures. Reassuringly, we found that higher cognitive skills, higher self-control and stronger empathic concern were all positively associated with the propensity to choose the altruistic lottery (see Appendix for details). Most of all, our measure of altruism (Falk et al., 2017) strongly predicted lottery choice.

One concern about using lotteries rather than a deterministic choice to assess participants' individual preference is that some people who picked the prosocial (selfish) lottery might have wanted to choose money (to save a life) in a deterministic environment: on the one hand, the lottery gives the opportunity to perform a prosocial act without necessarily incurring the cost of foregoing the payment, which might increase prosocial choice. On the other hand, picking the prosocial lottery might be perceived as less virtuous than saving a life directly, decreasing the motivation for prosocial choice. To address this concern, we ran a treatment in which a separate set of subjects made a direct choice between saving a life and receiving 100 euros ($N = 221$). Reassuringly, 60% of subjects in our main treatment chose the altruistic lottery, and 57% of subjects in the deterministic condition chose to save a life ($p = 0.49$, two-sample test of proportions). Moreover, none of the personality measures elicited in the survey differentially predicted lottery choice and the direct choice to save a life (see Appendix). Hence, our data showed that the lottery choice closely captures what subjects would have chosen if they had been offered the direct and deterministic choice between saving a life and the payment.

## 5.6 Discussion

In the short run, our results squared with the current consensus of a positive relationship between prosocial behavior and happiness. Extending previous results, we documented that prosocial choices and prosocial outcomes were independently associated with happiness. Whether subjects received their desired outcome affected mood, but not happiness or self-image. In addition, the data hinted at the importance of image concerns in the short run, a key driver of prosocial behavior in a large class of economic models of prosociality (Bénabou and Tirole, 2002, 2011). In the long run, our main finding was that prosocial behavior causally *reduced* happiness. This changes our previous understanding of the relationship between prosociality and happiness, which – based on short-run evidence – suggested a uniformly positive effect.

We speculate that happiness derived in the short run and the long run is partly driven by different factors. In the short term, happiness may be governed by visceral factors and the "conceptual consumption" (Ariely and Norton, 2009) associated with the act of giving, such as "warm glow" (Andreoni, 1990) and a favorable self-image (Bénabou and Tirole, 2002, 2011). That these factors play a role in shaping happiness in the short run was explicitly confirmed by our data. As time passes, however, a prosocial act might become less salient or even be forgotten, and the effect of these positive emotions and thoughts fades out. Instead, however, a substantial amount of money can still be source of happiness after some delay if it is spent gradually and thus leads to consumption that is spread out over time. In fact, that giving money to people can have a positive impact on their happiness is a frequent finding in empirical research (Haushofer and Shapiro, 2016). Moreover, in our design, the payment of 100 euros deliberately constructed a "cost of saving a life" that is *known to all subjects*, which in itself might affect happiness derived from saving a life. Research on the nature of prosocial behavior has repeatedly shown that the set of available alternatives plays a key role, e.g., in determining what people deem fair (Falk et al., 2003).

Another interpretation of the main finding is that a single prosocial act can be a two-sided sword: The short-run elation takes a toll on happiness in the long run. In our experiment, for example, the survey after four weeks might act like a reminder that brings up the positive feelings experienced in the short run among those who saved a life. At the same time, however, it may remind them that since the time of saving a life, they have not performed a prosocial action anymore, or their prosocial deeds from the meantime appear minuscule relative to saving a life. In other words, prosocial behavior can raise the bar for future happiness because it establishes a standard of comparison against which people judge their own behavior. This could either have a discouraging effect or provide a motivation for repeated prosocial behavior. We did not, however, find a negative effect on self-image in the long run, which would have been in line this interpretation.

Adding to a recent debate, the results are relevant for the debate on *effective altruism*, an evidence-based philosophy that advocates finding the most efficient ways to benefit others (MacAskill, 2015; Singer, 2015). A positive correlation between prosocial behavior and happiness is a central empirical justification for the quest to donate more: Philosopher Peter Singer forcefully argues that altruism is not about self-sacrifice, but that the greatest happiness arises from helping other people (Singer, 2015). Our findings indicate that this notion is incomplete.

Finally, our evidence speaks to a puzzle pointed out in previous work. If the proposed relationship between prosociality and happiness were as strong and unambiguously positive as suggested therein, people should behave much more prosocially in practice if they wanted to maximize their happiness (Dunn et al., 2008). Previous authors point out that people might mis-predict their own happiness. In contrast, our findings are qualitatively compatible with selfish behavior in many contexts, assuming people trade off the short-term benefit of prosocial behavior against the delayed costs.

Rather than suggesting that prosocial behavior does not promote happiness, our data reconfirm that it does, but only so in the short term. However, proclaiming this relationship to be a "psychological universal" (Aknin et al., 2013a) may be premature. A comprehensive understanding on the effects of prosocial behavior on happiness requires a more nuanced view that accounts for delayed effects.

## References

**Aknin, Lara B., Christopher P. Barrington-Leigh, Elizabeth W. Dunn, John F. Helliwell, Justine Burns, Robert Biswas-Diener, Imelda Kemeza, Paul Nyende, Claire E. Ashton-James, and Michael I. Norton (2013a):** "Prosocial spending and well-being: Cross-cultural evidence for a psychological universal." *Journal of Personality and Social Psychology*, 104 (4), 635–652. [184, 196]

**Aknin, Lara B., Tanya Broesch, J. Kiley Hamlin, and Julia W. Van de Vondervoort (2015):** "Prosocial behavior leads to happiness in a small-scale rural society." *Journal of Experimental Psychology*, 144 (4), 788–795. [183, 188]

**Aknin, Lara B., Elizabeth W. Dunn, Gillian M. Sandstrom, and Michael I. Norton (2013b):** "Does social connection turn good deeds into good feelings?: On the value of putting the "social" in prosocial spending." *International Journal of Happiness and Development*, 1 (2), 155–171. [184]

**Aknin, Lara B., J. Kiley Hamlin, and Elizabeth W. Dunn (2012):** "Giving leads to happiness in young children." *PLoS One*, 7 (6), e39211. [183, 188]

**Andreoni, James (1990):** "Impure altruism and donations to public goods: A theory of warm-glow giving." *The Economic Journal*, 100 (401), 464–477. [188, 195]

**Ariely, Dan and Michael I. Norton (2009):** "Conceptual consumption." *Annual Review of Psychology*, 60, 475–499. [188, 195]

**Aristotle (1987):** *The Nicomachean Ethics (translated by J.E.C. Welldon)*. Prometheus Books. [183]

**Batson, C. Daniel and Adam A. Powell (2003):** "Altruism and prosocial behavior." In *Handbook of Psychology*. Wiley Online Library. [183]

**Bénabou, Roland and Jean Tirole (2002):** "Self-confidence and personal motivation." *The Quarterly Journal of Economics*, 117 (3), 871–915. [188, 195]

**Bénabou, Roland and Jean Tirole (2006):** "Incentives and prosocial behavior." *The American Economic Review*, 96 (5), 1652–1678. [183]

**Bénabou, Roland and Jean Tirole (2011):** "Identity, morals, and taboos: Beliefs as assets." *The Quarterly Journal of Economics*, 126 (2), 805–855. [188, 195]

**Chen, Daniel L., Martin Schonger, and Chris Wickens (2016):** "oTree—An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance*, 9, 88–97. [200]

**Cumberland, Richard and John Maxwell (1727):** *A Treatise of the Laws of Nature*. R. Phillips. [183]

**Davis, Mark H. (1983):** "Measuring individual differences in empathy: Evidence for a multidimensional approach." *Journal of Personality and Social psychology*, 44 (1), 113–126. [200]

**Diener, Ed and Martin E. P. Seligman (2002):** "Very happy people." *Psychological Science*, 13 (1), 81–84. [184]

**Dunn, Elizabeth W., Lara B. Aknin, and Michael I. Norton (2008):** "Spending money on others promotes happiness." *Science*, 319 (5870), 1687–1688. [183, 184, 188, 196]

**Dunn, Elizabeth W., Lara B. Aknin, and Michael I. Norton (2014):** "Prosocial spending and happiness using money to benefit others pays off." *Current Directions in Psychological Science*, 23 (1), 41–47. [183, 188]

**Falk, Armin, Anke Becker, Thomas J. Dohmen, Benjamin Enke, and David Huffman (2017):** "Global evidence on economic preferences." Tech. rep. National Bureau of Economic Research. [194, 200, 202]

**Falk, Armin, Ernst Fehr, and Urs Fischbacher (2003):** "On the nature of fair behavior." *Economic Inquiry*, 41 (1), 20–26. [195]

**Gert, Bernard and Joshua Gert (2017):** "The Definition of Morality." In *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2017. Metaphysics Research Lab, Stanford University. [185]

**Haushofer, Johannes and Jeremy Shapiro (2016):** "The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya." *The Quarterly Journal of Economics*, 131 (4), 1973–2042. [195]

**Kolappan, C., R. Subramani, V. Kumaraswami, T. Santha, and P. R. Narayanan (2008):** "Excess mortality and risk factors for mortality among a cohort of TB patients from rural south India." *The International Journal of Tuberculosis and Lung Disease*, 12 (1), 81–86. [185]

**Latané, Bibb and John M. Darley (1970):** *The Unresponsive Bystander: Why Doesn't He Help?* Appleton-Century-Crofts. [183]

**Layard, Richard (2011):** *Happiness: Lessons from a New Science*. Penguin UK. [183]

**Leibniz, Gottfried Wilhelm (1988):** *Leibniz: Political Writings*. Cambridge University Press. [183]

**Locke, John (1988):** *Locke: Two Treatises of Government*. Cambridge University Press. [183]

**Lyubomirsky, Sonja and Heidi S. Lepper (1999):** "A measure of subjective happiness: Preliminary reliability and construct validation." *Social Indicators Research*, 46 (2), 137–155. [186, 199]

**Lyubomirsky, Sonja, Kennon M. Sheldon, and David Schkade (2005):** "Pursuing happiness: The architecture of sustainable change." *Review of General Psychology*, 9 (2), 111. [183]

**MacAskill, William (2015):** *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference*. Guardian Faber Publishing. [196]

**Oswald, Andrew J. and Stephen Wu (2010):** "Objective confirmation of subjective measures of human well-being: Evidence from the USA." *Science*, 327 (5965), 576–579. [183]

**Rammstedt, Beatrice and Oliver P. John (2005):** "Kurzversion des big five inventory (BFI-K)." *Diagnostica*, 51 (4), 195–206. [200]

**Singer, Peter (2015):** *The Most Good You Can Do: How Effective Altruism Is Changing Ideas About Living Ethically*. Castle Lectures Series. Yale University Press. [196]

**Straetemans, Masja, Philippe Glaziou, Ana L. Bierrenbach, Charalambos Sismanidis, and Marieke J. van der Werf (2011):** "Assessing tuberculosis case fatality ratio: a meta-analysis." *PLoS One*, 6 (6), e20755. [185]

**Tangney, June P., Roy F. Baumeister, and Angie Luzio Boone (2004):** "High self-control predicts good adjustment, less pathology, better grades, and interpersonal success." *Journal of Personality*, 72 (2), 271–324. [200]

**Thoits, Peggy A. and Lyndi N. Hewitt (2001):** "Volunteer work and well-being." *Journal of Health and Social Behavior*, 115–131. [183]

**Tiemersma, Edine W., Marieke J. van der Werf, Martien W. Borgdorff, Brian G. Williams, and Nico J. D. Nagelkerke (2011):** "Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review." *PLoS One*, 6 (4), e17601. [185]

**Wollaston, William (1759):** *The Religion of Nature Delineated*. Samuel Palmer. [183]

**World Health Organization (2018):** *Tuberculosis fact sheet*. URL: http://www.who.int/mediacentre/factsheets/fs104/en/ (visited on 03/29/2018). [185]

## Appendix 5.A Sample

The data used in this paper comprise three between-subjects conditions with a total of $N = 591$ participants.

1. **Main sample: Lottery choice**
   Each subject had a choice between two lotteries, Lottery A and Lottery B. $N = 325$ participated in the laboratory sessions, $N = 297$ of those also completed the follow-up online survey four weeks later.

2. **Baseline sample: Deterministic choice**
   As in condition Lottery choice, but each subject directly chose between saving a life and receiving 100 euros. $N = 221$.

3. **Calibration sample: Price list method**
   Using a price list method, we elicited the minimal monetary amount that would make a participant indifferent to saving a life. N = 45.

## Appendix 5.B Details of the Experiment

Participants were recruited from the subject pool of the *BonnEconLab* and re a fixed payment of 10 euros transferred to their bank account for participation in the laboratory experiment. In between the laboratory session and the follow-up online survey, we sent two reminder emails to subjects, stating their individual lottery outcome in the laboratory experiment. Exact wording of the experimental instructions and email texts is reproduced in Appendix 5.D. Subjects received 15 euros transferred to their bank account for participating in the online survey. The study was approved by the Ethics Committee of the Economics Department at the University of Bonn (reference no. 2016-02).

Our focus was on measuring two concepts, subjective well-being (SWB) and self-image. As to the former, the current consensus in the literature is that SWB constitutes a multidimensional concept with several components. Rather than evaluation of life, i.e., life satisfaction, or a sense of meaning or purpose in one's life, the notion of happiness used in this study most closely relates to the emotional, or affective element of subjective well-being. We chose our main measure of happiness to fulfill two requirements: It should be suited to capture both short-term as well as long-term variation in happiness, and it should be a widely used and validated by the previous literature. Our measure is based on the Subjective Happiness Scale (Lyubomirsky and Lepper, 1999), also referred to as the General Happiness Scale. In particular, we use the first item, which is an assessment of the statement "*In general, I consider myself:*" with possible responses raging from 1 = "*not a very happy person*" to 7 = "*a very happy person*" on a 7-point Likert scale.

Our measure of self-image is an assessment of the statement "*I am a good person*" on a 10-point Likert scale ranging from 1 = "*fully disagree*" to 10 = "*fully agree*". "Good person" is a direct translation of the original German phrase "guter Mensch" used in the experiment, which may also be translated as "good man" or "good human" here. Importantly, this is a typical expression with a clear meaning in the German language featuring a strong moral connotation, with essentially the opposite meaning of being a "bad person" or "evil". Moreover, mood was elicited using the frequently used question "*What is your mood at the moment*?", and a 11-point response scale from 0 = "*very bad*" to 10 = "*very good*".

All of the above measures were asked at three points in time during the study. The first instance was at the beginning of the laboratory session, before subjects were informed about the content of the study. This measure serves as an unpolluted individual measure which we exploit as an individual-specific benchmark for comparison against later measurements. The second elicitation took place after subjects had taken their lottery choice and the lottery had been resolved, i.e., after knowing the outcome of the lottery. Note that we abstained from asking the set of questions again between the choice and lottery resolution, mainly because this would have cluttered the experimental procedure and might have been indicative of the experimenters' objectives. The third set of measures was elicited during the online survey four weeks after the laboratory session. At the end of the laboratory session, we elicited a range of further measures based on standard questionnaires. In particular, we elicited cognitive skills using a set of 10 incentivized Raven matrices, self-control (Tangney et al., 2004), the Interpersonal Reactivity Index including a measure of empathic concern (Davis, 1983), a short version of the big five personality inventory (Rammstedt and John, 2005) and a measure of altruism Falk et al. (2017).

The laboratory sessions were run in the main auditorium of the University of Bonn, Germany, in September 2016. We recruited 325 subjects for the main lottery sample, mostly students at the University of Bonn, studying in various fields. 297 subjects completed both the laboratory session and the follow-up survey four weeks later, corresponding to an attrition rate of 9.4%. Attrition was not significantly predicted by lottery choice ($p > 0.1$) or outcome ($p > 0.1$) in a linear regression of a dummy variable for participation on indicators for lottery choice and lottery outcome and their interaction. The experiment was fully computerized and conducted using the software oTree (Chen et al., 2016). Subjects sat in cubicles to allow full privacy – no other person could see their screen during the experiment. Participants could ask questions to an experimenter at all times. The average completion time was 45 minutes.

## Appendix 5.C Additional Analyses

### 5.C.1 Deterministic Choice Treatment

We ran the Deterministic Choice treatment to examine whether the lottery choice in the main treatment is informative for which outcome the participant would have chosen had he had the direct, deterministic choice between life and money. We compare lottery choice in our main sample ($N = 297$) to the direct deterministic choice in an independent baseline experiment with a different set of subjects ($N = 221$). In particular, subjects in this comparison study received identical instructions about the two outcomes, except that they could directly choose one of them.

First of all, note that the fraction of subjects choosing the prosocial option is almost exactly identical in both samples. 60% choose the prosocial lottery (58% when including subjects who did not participate in the follow-up) and 57% choose to save a life directly ($p = 0.49$). Second, we analyze whether those who choose prosocially in each sample differ systematically based on the personality measures that we elicited at the end of the laboratory session. Table 5.C.1 shows results from regressions that investigate which measures are correlated with prosocial choice in both treatments. Column 1 indicates which measures predict choice of the prosocial lottery. In line with previous evidence, we find that higher cognitive skills, higher levels of self-control, higher levels of altruism, and stronger emphatic concerns all positively predict altruistic choice. Our data shows no direct effects of agreeableness – a component of the big five personality inventory – and gender, once other factors are controlled for. The correlates of altruistic lottery choice reported in column 1 square with previous evidence.

Column 2 reports a regression run on the joint sample of the lottery treatment and the deterministic choice treatment. We again include the above mentioned personality measures as regressors as well as a full set of interaction terms of our personality measures with an indicator variable that equals 1 for observations from the baseline condition and 0 otherwise. We only display estimates of the interaction effects in the table, since the main effects are identical to those reported in column 1. We find that none of the measures differentially predicts altruistic choice in the baseline sample relative to the lottery sample ($p > 0.1$ for all interaction terms). Taken together, these results strongly suggest that our lottery choice data allows for a categorization of more altruistic versus more selfish types that is essentially identical to the categorization that we would have obtained from having subjects choose directly between saving a life and receiving 100 euros.

**Table 5.C.1.** Correlates of prosocial choice in lottery treatment and the deterministic choice treatment

| Dependent variable:<br>**Lottery choice** (1 if altruistic, 0 if selfish) | Lottery sample | Full sample:<br>Measure * **1**(baseline) |
|---|---|---|
| | (1) | (2) |
| Female | 0.046 | -0.004 |
| | (0.056) | (0.089) |
| Cognitive intelligence (Raven) | 0.027* | -0.007 |
| | (0.015) | (0.021) |
| Self-control | 0.003 | 0.001 |
| | (0.003) | (0.004) |
| Big 5 - agreeableness | 0.005 | -0.017 |
| | (0.010) | (0.015) |
| Preferences module: altruism | 0.220*** | -0.025 |
| | (0.034) | (0.049) |
| Preferences module: positive reciprocity | -0.040 | 0.000 |
| | (0.038) | (0.054) |
| Preferences module: trust | 0.026 | -0.015 |
| | (0.027) | (0.043) |
| IRI - empathic concern | 0.028** | 0.005 |
| | (0.012) | (0.017) |
| Self-image at begin of session | -0.026 | 0.022 |
| | (0.020) | (0.029) |
| $R^2$ | .2551 | .2286 |
| N | 297 | 518 |

*Notes*: Column 1 tests the predictive power of different personality measures for the choice of the prosocial instead rather than the selfish lottery. Column 2 is the same regression on a joint sample including the condition with a deterministic choice between life and money instead of lotteries. The displayed coefficients in column 2 are interactions terms with an indicator that equals 1 for observations from the deterministic choice sample. Ordinary least squares regression. Robust standard errors in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

### 5.C.2 Robustness of Results

Table 5.C.2 shows that the regressions analyses in the main text are robust to including a battery of control variables. The regression specifications are identical to those in Table 5.1 but further include the set of personality measures (all big five personality traits, the four measures of the Interpersonal Reactivity Index, all measures of the preferences module (Falk et al., 2017), the self-control score), our measure of cognitive skills, and a gender dummy. The results are similar to those in Table 5.1.

**Table 5.C.2.** Additional regression specifications with control variables

| | Short run | | | Long run | | |
|---|---|---|---|---|---|---|
| Dependent variable (standardized): | $H_{Short-run}$ | Self-image | Mood | $H_{Long-run}$ | Self-image | Mood |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Lottery choice: 1 if altruistic, 0 if selfish | -0.001 | 0.444*** | -0.694*** | -0.116 | 0.244* | 0.139 |
| | (0.118) | (0.142) | (0.171) | (0.129) | (0.140) | (0.187) |
| Lottery outcome: 1 if life saved, 0 if money received | 0.027 | 0.248* | 0.161 | -0.277* | 0.259* | 0.067 |
| | (0.104) | (0.143) | (0.180) | (0.143) | (0.154) | (0.184) |
| Altruistic lottery choice * life saved | 0.087 | -0.018 | 1.243*** | 0.064 | -0.266 | -0.084 |
| | (0.147) | (0.176) | (0.218) | (0.181) | (0.192) | (0.228) |
| Constant | -4.389*** | -4.282*** | 0.736 | -2.740*** | -4.116*** | -0.460 |
| | (0.732) | (0.713) | (0.818) | (0.751) | (0.730) | (1.043) |
| $H_{Baseline}$ | Yes | | | Yes | | |
| Self-image at begin of session | | Yes | | | Yes | |
| Mood at begin of session | | | Yes | | | Yes |
| Big 5 | Yes | Yes | Yes | Yes | Yes | Yes |
| Preferences module | Yes | Yes | Yes | Yes | Yes | Yes |
| IRI | Yes | Yes | Yes | Yes | Yes | Yes |
| Self-control | Yes | Yes | Yes | Yes | Yes | Yes |
| Cognitive intelligence | Yes | Yes | Yes | Yes | Yes | Yes |
| Gender | Yes | Yes | Yes | Yes | Yes | Yes |
| Main effect choice: altruistic lottery | .04 | .43*** | -.07 | -.08 | .11 | .1 |
| Main effect outcome: life saved | .07 | .24*** | .78*** | -.24*** | .13 | .03 |
| $R^2$ | .6068 | .5746 | .3629 | .4295 | .457 | .2047 |
| N | 297 | 297 | 297 | 297 | 297 | 297 |

*Notes*: The regressions results shown in this table replicate the results obtained in Table 5.1 of the main text but include a set of control variables as a test of robustness. Ordinary least squares regression. Robust standard errors in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

In addition, we recognize that a least squares regression implicitly interprets the measurements of self-reported happiness, self-image and mood scores as if they were interval data. Table 5.C.3 shows estimates from ordered probit regressions, which accommodate the fact that these data are better characterized as having an ordinal scale instead. We show estimates for ordered response model specifications that are equivalent to the least squares specifications in the main text (Table 5.1). The qualitative results are similar to those in the least squares analysis.

**Table 5.C.3.** Alternative probit specification for main analyses

| | Short run | | | Long run | | |
|---|---|---|---|---|---|---|
| Dependent variable (standardized): | $H_{Short-run}$ | Self-image | Mood | $H_{Long-run}$ | Self-image | Mood |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| main | | | | | | |
| Lottery choice: 1 if altruistic, 0 if selfish | 0.223 | 0.549** | -0.728*** | -0.117 | 0.339** | 0.163 |
| | (0.191) | (0.214) | (0.172) | (0.182) | (0.170) | (0.179) |
| Lottery outcome: 1 if life saved, 0 if money received | 0.098 | 0.357* | 0.221 | -0.432** | 0.376* | 0.145 |
| | (0.190) | (0.206) | (0.215) | (0.200) | (0.201) | (0.195) |
| Altruistic lottery choice * life saved | 0.091 | 0.026 | 1.399*** | 0.144 | -0.379 | -0.132 |
| | (0.266) | (0.261) | (0.247) | (0.255) | (0.250) | (0.245) |
| $H_{Baseline}$ | 1.188*** | | | 0.829*** | | |
| | (0.157) | | | (0.099) | | |
| Self-image at begin of session | | 0.822*** | | | 0.618*** | |
| | | (0.084) | | | (0.065) | |
| Mood at begin of session | | | 0.126** | | | 0.196*** |
| | | | (0.050) | | | (0.043) |
| Main effect choice: altruistic lottery | .269** | .562*** | -.028 | -.045 | .15 | .098 |
| Main effect outcome: life saved | .143** | .369*** | .921*** | -.359*** | .187 | .08 |
| N | 297 | 297 | 297 | 297 | 297 | 297 |

*Notes*: The displayed regression results replicate the main results from Table 1 using ordered probit estimation, which recognizes that the response data is ordinal rather than interval. The table shows regression coefficients, not partial effects. Coefficients for cut points not displayed. Robust standard errors in parentheses. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

## Appendix 5.D Instructions

Instructions used in the laboratory experiment, the reminder emails and the follow-up online survey were translated from German into English. Please contact the authors for the German instructions.

### 5.D.1 Laboratory Session

Welcome and thank you for your interest in this study!
For your participation you will receive a fixed payment of 10.00 €, which will be paid to you by bank transfer after the study. In this study you will take decisions on the computer. Depending on how you decide you can earn additional money.
During the entire study it is not allowed to talk to other participants. Please turn off your mobile phone now, so that other participants will not be disturbed. Please only use the designated functions on the computer and make your entries using the keyboard and the mouse. If you have questions, please make a hand signal. Your question will be answered at your seat.
On the next screens you will see detailed information concerning the participation in this study. After reading this information you can confirm or refuse your participation.
To proceed click "Next".

[end of screen]

**Information on Participation in this Study of the BonnEconLab**
The following information have been sent to you via email together with the confirmation of your registration for this study. You receive this information again now. Once you have read the subsequent declaration of consent you can confirm your participation by clicking on “I agree”.
[ followed by mandated exclusion restrictions for participation in this study ]

[end of screen]

**Information**
In the following you will see important information, which are relevant for your subsequent decisions. They are about the disease tuberculosis and its possible treatment. Please read all information carefully.

[end of screen]

### Information about Tuberculosis

*What is tuberculosis?*

Tuberculosis – also called consumptiveness or White Death – is an infectious disease, which is caused by bacteria. Roughly one third of all humans are infected with the pathogen of tuberculosis. Active tuberculosis breaks out among 5 to 10 % of all those infected. Tuberculosis is primarily airborne. This is also why a quick treatment is necessary.

*What are the symptoms of tuberculosis?*

Tuberculosis patients often suffer from very inspecific symptoms like fatigue, feeling of weakness, lack of appetite and weight loss. At an advanced stage of lung tuberculosis, the patient coughs up blood, leading to the so-called rush of blood. **Without treatment a person with tuberculosis dies with a probability of 43 %**.

*How prevalent is tuberculosis?*

In the year 2014, 6 million people have been recorded as falling ill with active tuberculosis. **Almost 1.5 million people die of tuberculosis each year.** This means more deaths due to tuberculosis than due to HIV, malaria or any other infectious disease.

*Is tuberculosis curable?*

Today tuberculosis is curable. Treatment is administered by giving antibiotics several times each week over a period of 6 months. It is important that there is no interruption of treatment. In the years 2000 to 2014 approximately 43 million human lives could be saved due to an effective diagnosis and treatment of tuberculosis. **The success rate of treatment for a new infection is often above 85 %.**

The preceding numbers and information were provided by the World Health Organization (WHO), the United Nations' institution for the international public health, and are freely available. You can check this information on the web page of the WHO after this study.

[end of screen]

### Description of the Decision

In the course of this study there is an Option A and an Option B. Option A and Option B have different consequences. One of these two options will be implemented for you. That means, this option will be implemented with all its consequences exactly as described.

Next the consequences of Option A and Option B will be explained to you in de-

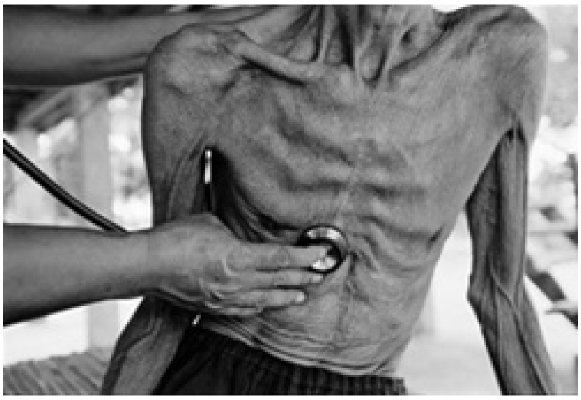


**Figure 5.D.1.** Picture shown to subjects in instructions. Typical symptoms of a tuberculosis patient. Source: http://www.opasha.org.

tail. After that you will see a decision situation, in which you will have to make a choice. By means of your choice in this decision situation you can influence *which* of the two options – Option A or Option B – will be implemented for you.

**Option A**

If Option A is implemented for you, you will be paid an **additional monetary amount of 100.00 €** by bank transfer after the study.

**Option B**

If Option B is implemented for you, you will not receive an additional payment. This option has another consequence: **You save one human life** if Option B is implemented.

After it has emerged which option will be implemented for you, it will be carried out exactly as described.

On the next tab you will receive more information about the implementation of Option B.

[end of screen]

**Information about Option B**

*How will the human life be saved?*

Only if Option B is carried out for you, you will save a human life.

If this option is implemented, a donation of 350.00 € will be arranged on your behalf to an organization which identifies and treats people suffering from tuberculosis. This donation will be executed for you by the BonnEconLab after the study. The entire donation amount will be used by the organization for the direct treatment of tuberculosis.

*What does it mean to “save a life”?*

To save a human life here means the successful cure from tuberculosis for one person, who otherwise would have died *due to* his tuberculosis. That means in particular: The donation amount is sufficient to identify and cure as many sick persons such that there is at least one person among these, who would otherwise have died from tuberculosis in expectation. The calculation of the amount accommodates the fact that there are other ways (e.g., the national health care system) through which people can be cured.

That means: **The amount of 350.00 € was calculated in such a way that the organization can save at least one additional human from death.**

On the next tab you will receive additional information about the possible saving of a human life and details about the organization that treats tuberculosis patients.

[end of screen]

***Operation ASHA***

In case of Option B being implemented you will save one human life. For this an amount of 350.00 € will be transferred to the organization *Operation ASHA* after the study.

**Figure 5.D.2.** Picture shown to subjects in instructions: Logo of organization *Operation ASHA*. Source: http://www.opasha.org.

*Operation ASHA* is a charity organization specialized since 2005 on treating tuberculosis in disadvantaged communities. The work of *Operation ASHA* is based on the insight that the biggest obstacle for the treatment of tuberculosis is the interruption of the necessary 6-month-long regular intake of medication. For a successful treatment the patient has to come to a medical facility twice a week – more than 60 times in total – to take the medication. An interruption or termination of the treatment is fatal, because this strongly enhances the development of a drug-resistant form of tuberculosis. This form of tuberculosis is much more difficult to treat and almost always leads to death.

To overcome this problem, *Operation ASHA* developed a concept that guarantees the regular treatment through immediate spatial proximity to the patient. A possible non-adherence is additionally prevented by visiting the patient at home.

By now *Operation ASHA* runs more than 360 treatment centers, almost all of which are located in the poorer regions of India. More than 60,000 sick persons have been identified and treated that way.

*Operation ASHA* is an internationally recognized organization, and their successes haven been covered by the New York Times, BBC and Deutsche Welle, for example. The MIT and University College London have already conducted research projects about the fight against tuberculosis in cooperation with *Operation ASHA*. The treatment method employed by *Operation ASHA* is described by the World Health Organization (WHO) as “highly efficient and cost-effective”.

[end of screen]

**Figure 5.D.3.** icture shown to subjects in instructions: An employee of *Operation ASHA* provides medication to a tuberculosis patient. Source: http://www.opasha.org.

**What determines the donation amount for saving a human life?**

the donation amount makes sure that at least one human life is saved in

expectation.
The information used for the calculation of the donation amount exclusively consists of public statements by the World Health Organization (WHO), peer-reviewed research studies, statistical releases from the Indian government as well as published figures from Operation ASHA. In the calculation all information was interpreted in a conservative way and more pessimistic estimates were used in case of doubt, such that the donation amount of 350.00 € is, if anything, higher than the actual costs associated with saving a human life. Moreover, the calculation was based on the treatment success rate of *Operation ASHA*, the mortality rate of an alternative treatment by the national tuberculosis program in India, and different detection rates for new cases of tuberculosis have been accounted for.

Based on a very high number of cases, one can illustrate the contribution of your donation as follows:

**With your donation *Operation ASHA* can treat 5 additional tuberculosis patients.**
If these 5 sick persons would not be treated by *Operation ASHA*, **one** patient would die in expectation.
If 5 persons are treated by means of your donation, **no** patient dies in expectation.

Based on these expected values this means that one human life will be saved with your donation. This relationship is depicted in the following diagram.

**Without** treatment by *Operation ASHA*, **one** of 5 persons sick of tuberculosis will die in expectation.

**With** the donation 5 persons sick of tuberculosis can be treated by *Operation ASHA* and **none** of these persons will die in expectation.

An agreement with *Operation ASHA* for the purpose of this study ensures that 100 % of the donation amount will exclusively be used for the diagnosis and treatment of tuberculosis patients. That means that every euro of the donation amount will directly go into saving human lives and no other costs will be covered with it.

[end of screen]

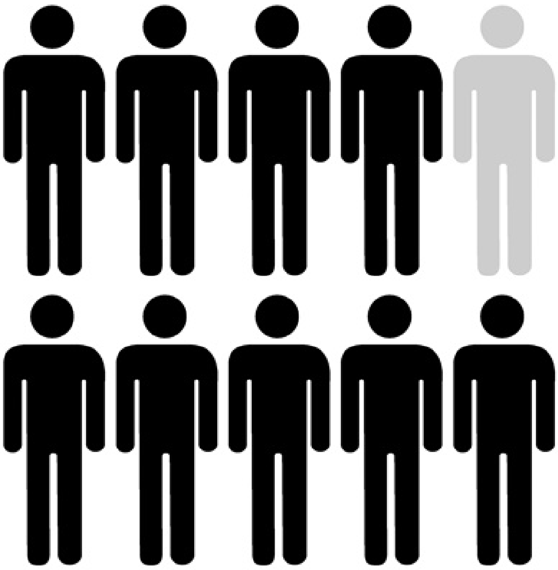


**Figure 5.D.4.** Picture shown to subjects in instructions. Top: Illustration of Option A. Without treatment by Operation ASHA, one of 5 persons sick of tuberculosis will die in expectation. Bottom: Illustration of Option B. With the donation 5 persons sick of tuberculosis can be treated by Operation ASHA and none of these persons will die in expectation.

**Summary**

*Tuberculosis*

The success rate of medical treatment for a new infection is very high. Despite this 1.5 million people die from tuberculosis each year. The biggest obstacle for the cure of tuberculosis is a possible termination of the regular treatment with antibiotics. The concept of *Operation ASHA* is therefore based on the direct spatial proximity to their patients and on the control and recording of the regular intake of medication.

*Option A, Option B and your decision*

Exactly one of the two options will be carried out for you after the study. The options have different consequences:

- In case of an implementation of Option A you will receive an additional amount of money.
- In case of an implementation of Option B you will save a human life. Concretely, for Option B a donation of 350.00 € will be paid on your behalf,

which is sufficient not only to cure one person, but to actually save that person from death by tuberculosis.

In the following decision situation you will take a choice through which you can influence *which* of the two options – Option A or Option B – will be implemented for you.

*How is the human life saved?*
The donation amount already accounts for the fact that a sick person could also have survived without treatment by *Operation ASHA*; or that he could instead have been treated by the national health care system. This is why the amount is sufficient for the diagnosis and complete treatment of **several** affected persons.

**Please note:**
**This is not a hypothetical game.** The option to be implemented for you will actually be carried out – exactly as described – on behalf of the BonnEconLab. As a proof you will receive the money in case of Option A; in case of an implementation of Option B we will allow inspection of the confirmed bank transfer to the organization *Operation ASHA* on request.

If you have individual questions, you can also direct these by email after the study to nachbesprechung@uni-bonn.de. You find this email address on the back of your seating card. You can take it home with you.

Click on "Next", if you have carefully read the information on this page.
*Please note*: You can only click on the button "Next" once you have spent at least 5 minutes on the seven tabs of this page.

[end of screen]

**Your decision**
On the next screen you can choose between two lotteries, **Lottery 1** and **Lottery 2**.
**Lottery 1**
With **60 % probability**, Option A is implemented for you.
With **40 % probability**, Option B is implemented for you.
**Lottery 2**
With **40 % probability**, Option A is implemented for you.

With **60 % probability**, Option B is implemented for you.

This means: With your choice of a lottery you can determine whether rather Option A or rather Option B shall be implemented for you.
The lottery is played as follows: After you have chosen one of the two lotteries, the computer will draw a random number. The drawn random number is one of the numbers from 1 to 10.
If you have opted for **Lottery 1**, Option A will be implemented only if the drawn random number is a 1, 2, 3, 4, 5, or 6. Option B will be implemented if the drawn random number is a 7, 8, 9 or 10.
If you have opted for **Lottery 2**, Option A will be implemented only if the drawn random number is a 1, 2, 3 or 4. Option B will be implemented if the drawn random number is a 5, 6, 7, 8, 9 or 10.

**Remember**:

- In case of an implementation of Option A you will receive an additional amount of money of 100.00 €.
- In case of an implementation of Option B you will save a human life. Concretely, for Option B a donation of 350.00 € will be paid on your behalf, which is sufficient not only to cure one person, but to actually save that person from death by tuberculosis.

**Please note**:

1. All statements in these instructions are true. In particular, all consequences that are described in the instructions will be implemented exactly as described. This holds generally for all studies of the Bonn lab for research in experimental economics, and also for this study.
2. Anonymity: No other participant of this study can see your decision. The subsequent analysis of all data is performed in an anonymized way, such that your decisions cannot be linked to your person anymore.

[end of screen]

**Decision**

**I choose Lottery 1**
With **60 %**: I receive 100.00 €.
With **40 %**: I save one human life.

**I choose Lottery 2**
With **40 %**: I receive 100.00 €.
With **60 %**: I save one human life.

[end of screen]

**Result**

You chose [**Lottery 1 / Lottery 2**].

The random number drawn for you is a [**1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10**].

Bank transfer to you: [0.00 € / 100.00 €] Bank transfer to *Operation ASHA*: [0.00 € / 350.00 €]

### 5.D.2 Reminder Emails and Survey

**Email 1 (after 1 week)**

Dear [ first name last name ]!

Thank you very much for your participation in our study from [ date of study, time of study ]. In the context of this study, you could make a decision about whether you rather want a human life to be saved for you, or whether you rather want to receive an additional payment of 100 €.

[ Either: ] At the end of the study you were informed that you receive an additional payment of 100 €.
[ Or: ] At the end of the study you were informed that you arranged for a donation in the amount of 350 € for the treatment and cure of tuberculosis patients, such that one human life is saved in expectation.

The corresponding bank transfer is currently in process.

We will notify you again as soon as the bank transfers are entered.

Yours sincerely,
BonnEconLab

**Email 2 (after 3 weeks)**

Dear [ first name last name ]

Thank you very much for your participation in our study from [ date of study, time of study ].

[Either:] In the context of this study you received an additional payment of 100 €!

The corresponding bank transfer has been executed in the meantime and is credited to your bank account.
[Or:] In the context of this study you have arranged for a donation in the amount of 350 € for the treatment and cure of tuberculosis patients. This way you saved one human life in expectation!
The corresponding bank transfer has been executed in the meantime and is credited to the bank account of Operation ASHA.

In the next days you will receive a further email including the link to the announced online survey.

Yours sincerely,
BonnEconLab

**Survey (after 4 weeks)**

To remind you: The study that you participated in was about either saving a human life or receiving an amount of money.

[Either:] Due to your participation you have received an additional payoff of 100 €for yourself! This considerable amount of money has been transferred to you in the meantime.
With your amount of money you saved no human life in expectation!

[Or:] Due to your participation you have arranged for a donation in the amount of 350 € for the treatment and cure of tuberculosis patients. This considerable amount of money has been transferred in the meantime and will now benefit people in great distress.
With your amount of money you saved one human life in expectation!