On the Formation and Economic Implications of Subjective Beliefs and Individual Preferences

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

Benjamin Enke

aus Bonn-Bad Godesberg

2016

Dekan:	Prof. Dr. Rainer Hüttemann
Erstreferent:	Prof. Dr. Armin Falk
Zweitreferent:	Prof. Dr. Thomas Dohmen
Tag der mündlichen Prüfung:	1. Juli 2016

Acknowledgments

In Spring 2011 – I had just finished my first semester as a graduate student – my understanding of economics was pretty much restricted to taking derivatives of various sorts. It was then that I had my first meeting with Armin Falk. "So?", he asked after what would soon become a regular chit-chat about soccer, "what are you interested in?" In the previous months, I had read cross-cultural work in sociology and psychology as well as the motivated cognition papers by Roland Benabou, so I recall myself exclaiming "beliefs and culture!" Back then, it was beyond my wildest dreams that these opening lines would form the starting point of the intellectually most exciting period of my life, and that, five years later, my dissertation would indeed consist of three papers on beliefs and three on the cross-country (cultural?) variation in preferences. However, one way or another, Armin made it happen. I am deeply grateful to him for providing the environment, guidance, freedom, and resources that I needed to conduct my research. In retrospect, I think I have tremendously benefited not just from his enormous creativity and beliefs about what are important economic questions, but perhaps even more from his insistence that "this is not good enough".

I have been very fortunate to have had as further advisors and coauthors a considerable number of people with quite heterogeneous views and skills. I am especially grateful to Thomas Dohmen for many stimulating discussions about preference heterogeneity, to Hans-Martin von Gaudecker for inspiring thoughts on structural work and causality, to Uwe Sunde for introducing me to the world of comparative development, and to Florian Zimmermann for probably hundreds of fascinating (and fun!) discussions about beliefs, bounded rationality, the scope of experimental economics, and economics in general. I also benefited from a great many discussions and software lessons with Tilman Drerup and Holger Gerhardt. Apart from my coauthors and advisors, my fellow graduate students Ralph Lütticke, Andreas Kleiner, Sinem Hacioglu, and Frederik Schwerter provided a superb network of economists and friends. I benefited from the great infrastructure and financial support of the Bonn Graduate School of Economics.

Nothing of what is contained in this thesis would have been possible without the lifelong (but mostly "early-career") love and support of my parents. I am deeply grateful to have had the tremendous luck of being born into this family. Finally, my biggest thanks goes to my wife, Anke. We met during a time when I was more frustrated about graduate school and economics than ever – introspection suggests to me that she was the turnaround not just for my private life, but also for my dissertation. She cannot possibly imagine how much of her inspiration, imagination, and love is contained in all of the below.

Contents

List of Figures

List of Tables

1

Introduction

The conceptual framework of neoclassical economics posits that individual decision-making processes can be represented as maximization of some objective function. In this framework, people's goals and desires are expressed through the means of preferences over outcomes; in addition, in choosing according to these objectives, people employ subjective beliefs about the likelihood of unknown states of the world. For instance, in the subjective expected utility paradigm, people linearly combine their probabilistic beliefs and preferences over outcomes to form an expected utility function.

Much of the parsimony and power of theoretical economic analysis stems from the striking generality and simplicity of this framework. At the same time, the crucial importance of preferences and beliefs in our conceptual apparatus in combination with the heterogeneity in choice behavior that is observed across many economic contexts raises a number of empirical questions. For example, how much heterogeneity do we observe in core preference or belief dimensions that are relevant for a broad range of economic behaviors? If such preferences and beliefs exhibit heterogeneity, then what are the origins of this heterogeneity? How do beliefs and preferences form to begin with? And how does variation in beliefs and preferences translate into economically important heterogeneity in choice behavior?

This thesis is organized around these broad questions and hence seeks to contribute to the goal of providing an improved empirical understanding of the foundations and economic implications of individual decision-making processes. The content of this work reflects the deep belief that understanding and conceptualizing decision-making requires economists to embrace ideas from a broad range of fields. Accordingly, this thesis draws insights and techniques from the literatures on behavioral and experimental economics, cultural economics, household finance, comparative development, cognitive psychology, and anthropology.

Chapters 2 through 4 combine methods from experimental economics, household finance, and cognitive psychology to investigate the effects of bounded rationality on the formation and explanatory power of subjective beliefs. Chapters 5 through 7 use tools from cultural economics, anthropology, and comparative development to study the cross-country variation in economic preferences as well as its origins and implications.

The formation of beliefs about payoff-relevant states of the world crucially hinges on an adequate processing of incoming information. However, oftentimes, the information people receive is rather complex in nature. Chapters 2 and 3 investigate how boundedly rational people form beliefs when their information is subject to sampling biases, i.e., when the information pieces people receive are either not mutually independent or systematically selected.

Chapter 2 is motivated by Akerlof and Shiller's 2009 popular narrative that from time to time some individuals or even entire markets undergo excessive belief swings, which refers to the idea that sometimes people are overly optimistic and sometimes overly pessimistic over, say, the future development of the stock market. In particular, Akerlof and Shiller (2009) argue that such "exuberance" or excessive pessimism might be driven by the pervasive "telling and re-telling of stories". In fact, many real information structures such as the news media generate correlated rather than mutually independent signals, and hence give rise to severe double-counting problems. However, clean evidence on how people form beliefs in correlated information environments is missing. Chapter 2, which is joint work with Florian Zimmermann, provides clean experimental evidence that many people neglect such double-counting problems in the updating process, so that beliefs are excessively sensitive to well-connected information sources and follow an overshooting pattern. In addition, in an experimental asset market, correlation neglect not only drives overoptimism and overpessimism at the individual level, but also gives rise to a predictable pattern of over- and underpricing. Finally, investigating the mechanisms underlying the strong heterogeneity in the presence of the bias, a series of treatment manipulations reveals that many people struggle with identifying double-counting problems in the first place, so that exogenous shifts in subjects' focus have large effects on beliefs.

Chapter 3 takes as starting point the big public debate about increased political polarization in the United States, which refers to the fact that political beliefs tend to drift apart over time across social and political groups. Popular narratives by, e.g., Sunstein (2009), Bishop (2009), and Pariser (2011) posit that such polarization is driven by people selecting into environments in which they are predominantly exposed to information that confirms their prior beliefs. This pattern introduces a selection problem into the belief formation process, which may result in polarization if people failed to take the non-representativeness among their signals into account. However, again, we do not have meaningful evidence on how people actually form beliefs in such "homophilous" environments. Thus, Chapter 3 shows experimentally that many people do not take into account how their own prior decisions shape their informational environment, but rather largely base their views on their local information sample. In consequence, beliefs excessively depend on people's priors and tend to be too extreme, akin to the concerns about "echo chambers" driving irrational belief polarization across social groups. Strikingly, the distribution of individuals' naïveté follows a pronounced bimodal structure – people either fully account for the selection problem or do not adjust for it at all. Allowing for interaction between these heterogeneous updating types induces little learning: neither the endogenous acquisition of advice nor exogenously induced dissent lead to a convergence of beliefs across types, suggesting that the belief heterogeneity induced by selected information may persist over time. Finally, the paper provides evidence that selection neglect is conceptually closely related to correlation neglect in that both cognitive biases appear to be driven by selective attentional patterns.

Taken together, chapters 2 and 3 show that many people struggle with processing information that is subject to sampling issues. What is more, the chapters also show that these biases might share common cognitive foundations, hence providing hope for a unified attention-based theory of boundedly rational belief formation.

While laboratory experimental techniques are a great tool to study the formation of beliefs, they cannot shed light on the relationship between beliefs and economically important choices. In essentially all economic models, beliefs mechanically map into choice behavior. However, it is not evident that people's beliefs play the same role in generating observed behavior across heterogeneous individuals: while some people's decision process might be well-approximated by the belief and preference-driven choice rules envisioned by economic models, other people might use, e.g., simple rules of thumb instead, implying that their beliefs should be largely irrelevant for their choices. That is, bounded rationality might not only affect the formation of beliefs, but also the mapping from beliefs to choices. In Chapter 4, Tilman Drerup, Hans-Martin von Gaudecker, and I take up this conjecture in the context of measurement error problems in household finance: while subjective expectations are important primitives in models of portfolio choice, their direct measurement often yields imprecise and inconsistent measures, which is typically treated as a pure measurement error problem. In contrast to this perspective, we argue that individual-level variation in the precision of subjective expectations measures can actually be productively exploited to gain insights into whether economic models of portfolio choice provide an adequate representation of individual decision processes. Using a novel dataset on experimentally measured subjective stock market expectations and real stock market decisions collected from a large probability sample of the Dutch population, we estimate a semiparametric double index model to explore this conjecture. Our results show that investment decisions exhibit little variation in economic model primitives when individuals provide error-ridden belief statements. In contrast, they predict strong variation in investment decisions for individuals who report precise expectation measures. These findings indicate that the degree of precision in expectations data provides useful information to uncover heterogeneity in choice behavior, and that boundedly rational beliefs need not necessarily map into irrational choices.

In the standard neoclassical framework, people's beliefs only serve the purpose of achieving a given set of goals. In many applications of economic interest, these goals are well-characterized by a small set of preferences, i.e., risk aversion, patience, and social preferences. Prior research has shown that these preferences vary systematically in the population, and that they are broadly predictive of those behaviors economic theory supposes them to. At the same time, this empirical evidence stems from often fairly special samples in a given country, hence precluding an analysis of how general the variation and predictive power in preferences is across cultural, economic, and institutional backgrounds. In addition, it is conceivable that preferences vary not just at an individual level, but also across entire populations – if so, what are the deep historical or cultural origins of this variation, and what are its (aggregate) economic implications? Chapters 5 through 7 take up these questions by presenting and analyzing the Global Preference Survey (GPS), a novel globally representative dataset on risk and time preferences, positive and negative reciprocity, altruism, and trust for 80,000 individuals, drawn as representative samples from 76 countries around the world, representing 90 percent of both the world's population and global income.

In joint work with Armin Falk, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde, Chapter 5 presents the GPS data and shows that the global distribution of preferences exhibits substantial variation across countries, which is partly systematic: certain preferences appear in combination, and follow distinct economic, institutional, and geographic patterns. The heterogeneity in preferences across individuals is even more pronounced and varies systematically with age, gender, and cognitive ability. Around the world, the preference measures are predictive of a wide range of individual-level behaviors including savings and schooling decisions, labor market and health choices, prosocial behaviors, and family structure. We also shed light on the cultural origins of preference variation around the globe using data on language structure.

The magnitude of the cross-country variation in preferences is striking and raises the immediate question of what brought it about. Chapter 6 presents joint work with Anke Becker and Armin Falk in which we use the GPS to show that the migratory movements of our early ancestors thousands of years ago have left a footprint in the contemporary cross-country distributions of preferences over risk and social interactions. Across a wide range of regression specifications, differences in preferences between populations are significantly increasing in the length of time elapsed since the respective groups shared common ancestors. This result obtains for risk aversion, altruism, positive reciprocity, and trust, and holds for various proxies for the structure and timing of historical population breakups, including genetic and linguistic data or predicted measures of migratory distance. In addition, country-level preference endowments are non-linearly associated with migratory distance from East Africa, i.e., genetic diversity.

In combination with the relationships between language structure and preferences established in Chapter 5, these results point to the importance of very long-run events for understanding the global distribution of some of the key economic traits. Given these findings on the very deep roots of the cross-country variation in preferences, an interesting – and conceptually different – question is whether such country-level preference profiles might have systematic aggregate economic implications. Indeed, according to standard dynamic choice theories, patience is a key driving factor behind the accumulation of productive resources and hence ultimately of income not just at an individual, but also at a macroeconomic level. Using the GPS data on patience, Chapter 7 (joint work with Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde) investigates the empirical relevance of this hypothesis in the context of a micro-founded development framework. Around the world, patient people invest more into human and physical capital and have higher incomes. At the macroeconomic level, we establish a significant reduced-form relationship between patience and contemporary income as well as medium- and long-run growth rates, with patience explaining a substantial fraction of development differences across countries and subnational regions. In line with a conceptual framework in which patience drives income through the accumulation of productive resources, average patience also strongly correlates with aggregate human and physical capital accumulation as well as investments into productivity.

Taken together, this thesis has a number of unifying themes and insights. First, consistent with the vast heterogeneity in observed choices, people exhibit a large amount of variation in beliefs and preferences, and in how they combine these into choice rules. Second, at least part of this heterogeneity is systematic and has identifyable sources: preferences over risk, time, and social interactions appear to have very deep historical or cultural origins, but also systematically vary with individual characteristics; belief heterogeneity, on the other hand, is partly driven by bounded rationality and its systematic, predictable effects on information-processing. Third, and finally, this heterogeneity in beliefs and preferences is likely to have real economic implications: across cultural and institutional backgrounds, preferences correlate with the types of behaviors that economic models envision them to, not just across individuals, but also at the macroeconomic level; subjective beliefs are predictive of behavior, too, albeit with the twist that certain subgroups of the population do not appear to entertain stable belief distributions to begin with. In sum, (I believe that) much insight is to be gained from further exploring these fascinating topics.

References

Akerlof, George A. and Robert J. Shiller (2009): *Animal spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism*. Princeton University Press. [2]

Bishop, Bill (2009): *The Big Sort: Why the Clustering of Like-Minded America is Tearing us Apart*. Houghton Mifflin Harcourt. [2]

Pariser, Eli (2011): *The Filter Bubble: What the Internet is Hiding From You*. Penguin UK. [2]

Sunstein, Cass R. (2009): *Republic.com 2.0*. Princeton University Press. [2]

2

Correlation Neglect in Belief Formation*

2.1 Introduction

A pervasive feature of information structures is that decision makers are exposed to correlated signals. For example, various news media frequently share common information sources such as press agencies, so that the contents of different news reports (newspaper articles, television shows, online print) tend to be correlated. Similarly, in social networks, the opinions of different network members are often partly based on information from a mutually shared third party, so that, in communicating with these people, one is confronted with correlated information. A common feature of these information structures is that similar "stories" are getting told and retold multiple times, implying the presence of informational redundancies, i.e., potential double-counting problems.

Taking this observation as point of departure, we employ a series of laboratory experiments to make three key contributions. First, we provide clean evidence that even in transparent settings people neglect redundancies in information sources when forming beliefs, albeit with a strong heterogeneity at the individual level.[1] As a consequence, just like recent models of boundedly rational social learning predict, people's beliefs are excessively sensitive to well-connected information sources and hence follow an overshooting pattern. In a second step, we examine whether the bias persists in markets.

* For valuable comments and discussions we are grateful to Steffen Altmann, Sandro Ambuehl, Anke Becker, Doug Bernheim, Stefano DellaVigna, Thomas Dohmen, Erik Eyster, Armin Falk, Hans-Martin von Gaudecker, Paul Heidhues, Holger Herz, Alex Imas, David Laibson, Rosemarie Nagel, Muriel Niederle, Yagiz Özdemir, Matthew Rabin, Frederik Schwerter, Andrei Shleifer, Joel Sobel, Charlie Sprenger, Georg Weizsäcker, and Matthias Wibral. Helpful comments were also received from seminar audiences at Amsterdam, WZB Berlin, Bonn, Carnegie Mellon, Frankfurt, Harvard, HBS NOM, INSEAD, Kellogg MEDS, Maastricht, Pompeu Fabra, UCLA Anderson, UCSD, Wharton, and Zurich, as well as from conference participants at NYU BRIC 2015, Russell Sage Foundation ECBE 2015, CESifo Munich 2013, M-BEES Maastricht 2013, SITE Stanford 2013, TIBER Tilburg 2013, and the North-American ESA conference 2012. Financial support from the Russell Sage Foundation (Small Grants Program in Behavioral Economics), the International Foundation for Research in Experimental Economics (Small Grants Program), the Bonn Graduate School of Economics and the Center for Economics and Neuroscience Bonn is gratefully acknowledged.

[1] Throughout the paper, a correlation is implicitly understood as being conditional on a state realization. Also, we only refer to positive correlations.

Recently, Shiller (2000) and Akerlof and Shiller (2009) have argued that "exuberant" public opinions or "panics", driven by the multiple occurrence of similar stories, may be a driver of aggregate distortions. In this spirit, we establish that, in an experimental asset market, the incidence of correlated (and hence partially recurring) news leads to pronounced and predictable price distortions. Finally, we examine the mechanisms underlying the cognitive mistake. A series of treatment variations suggests that people possess the mathematical skills that are necessary to solve the updating task, but do not identify the double-counting problem in the first place, so that exogenous shifts in focus debias the majority of subjects.

In the baseline experiment, subjects need to estimate an ex ante unknown state of the world and are paid for accuracy. The key idea of our experimental design is to construct two sets of information (one with and one without a known and simple correlation) that are identical in terms of informational content, and should thus result in the same belief. In a between-subjects design, one group of subjects receives correlated, the other uncorrelated information. All pieces of information are generated by computers to ensure that subjects know the precise process generating the data. Specifically, four unbiased iid signals about the state of the world are generated by four computers (A through D). In the uncorrelated condition, subjects observe these four independent signals. In the correlated condition, participants obtain the signal of computer A as well as the average of the signals of A and B, of A and C, as well as of A and D. Thus, just as in the motivating examples, the signal of the common source A is partially recurring in the averages, implying the presence of informational redundancies. In this setting, the correlation structure has a particularly simple form because the signal of computer A is known, so that subjects only need to invert averages to back out the underlying independent signals. If subjects correctly took the redundancies into account, beliefs should be identical across treatments. However, despite extensive instructions and control questions, our results indicate that a considerable fraction of subjects treats all incoming information as approximately independent and hence double-counts the signal of the common source A. Thus, while beliefs remain statistically unbiased ex ante, they are highly sensitive to well-connected information sources and exhibit excessive swings: whenever the relatively low (high) signal of the common source repeatedly emerges through other messages, people on average become overpessimistic (overoptimistic) relative to the control condition, an effect that is sizable, significant, and causes lower payoffs. In light of the strong *average* tendency to neglect correlations, we proceed by specifying the precise and possibly heterogeneous updating rules subjects employ. We find that beliefs follow a bimodal distribution: most people are either fully sophisticated or very naïve, emphasizing the presence of different belief formation types. In particular, those subjects that do not succesfully process correlations form beliefs by following a particular simple heuristic of averaging the correlated messages. These results are robust to a number of variations in the experimental design such as the precise information structure, the experimental frame, or the incentive scheme.

An immediate question is whether these biased, but heterogeneous, beliefs persist in competitive markets and have systematic implications beyond the individual level, or whether market interaction induces naïve subjects to learn (see, e.g., Camerer, 1987;

Gneezy et al., 2003, for other studies of biases in market settings). To approach this issue, we embed our individual belief elicitation design into a standard continuous double-auction environment in which subjects trade financial assets of ex ante unknown value. To keep the market environment as simple as possible, subjects are allowed to either buy or sell assets, but not both. Before each trading round, all subjects receive the same sets of information about the true state as in the individual treatments. Again, we form treatment (control) groups by providing correlated (uncorrelated) signals about the fundamental value of the assets. The results show that our experimental market interaction does not induce naïfs to learn: market prices differ between treatments in the direction one would expect if subjects disregard correlations. In periods in which correlation neglect leads to overly optimistic beliefs (because the signal of the common source A is relatively high), market prices in the correlated treatment are too high relative to both the control treatment and the fundamental level. Likewise, when neglecting redundancies implies overpessimism, market prices are too low. Thus, correlation neglect causes a predictable pattern of over- and underpricing. In addition, within the correlated market treatment, subjects' propensity to ignore correlations predicts both individual trading behavior and the degree of price distortions in a market. These findings are reminiscent of the narratives provided by Akerlof and Shiller (2009) who emphasize how the excessive confidence swings that may be generated by the "telling and re-telling" of stories could drive aggregate distortions. While other theories can be invoked to explain either (collective) overoptimism or -pessimism, correlation neglect provides a unified view on these phenomena and relates them to the informational network structure.

Next, we investigate whether the updating error we observe is driven by a simple "face value" heuristic. This hypothesis posits that people *never* think through the process generating their information and instead treat each number as if it were an unmanipulated independent signal realization, *regardless of whether the signals are correlated or distorted in other ways*. We design two treatments to evaluate the empirical validity of such an extreme heuristic. The results reject a face value heuristic, and correlation neglect persists even when face value bias makes opposite predictions, suggesting that subjects indeed struggle with correlations as such.

Based on this set of findings, we implement further treatment variations to delve into the cognitive mechanisms underlying correlation neglect. Understanding the cognitive underpinnings of belief biases provides crucial inputs into formalizing these errors. Corresponding insights may also facilitate predictions about where the bias is likely (not) to occur in applied work, or how to debias people. For instance, are people less likely to neglect informational redundancies when the financial stakes are high, or when the double-counting problem is very salient? A key innovation of this paper is to move beyond the identification of a particular bias and to develop an experimental technology that allows an investigation of the underlying cognitive mechanisms. We start our corresponding quest by establishing the crucial role of complexity: just like other behavioral biases, correlation neglect only arises if the informational environment is sufficiently complex, but not if only two computers generate signals (also see Charness and Levin, 2009). In addition, we show that subjects' propensity to double-count signals is significantly related to both past academic achievement and an IQ test score. To better understand why and how low cognitive skills produce correlation neglect, we conceptualize belief formation in our more complex information setup as a simple two-step process: first, people need to identify the double-counting problem inherent in our experimental environment; second, they ought to execute the mathematical computations that are necessary to solve the double-counting problem and develop unbiased beliefs. Which of these two steps do subjects struggle most with, and why?

To address this question, we first show through an additional treatment that once we solve the first step for subjects by explicitly instructing them to back out the underlying independent signals from the correlated messages, the vast majority of our participants is both willing to and mathematically capable of performing the necessary calculations. Thus, a key challenge in successfully processing correlations appears to be to identify the double-counting problem in the first place. Even though our experimental procedures ensure that subjects understand the information structure in an abstract sense, it seems that they do not detect the informational redundancy when approaching a specific belief formation task. According to this logic, the first step of our simple belief formation process may act as a threshold to developing unbiased beliefs. We bolster this interpretation empirically: if people struggle with identifying double-counting problems, then nudging their focus towards the mechanics that generate the correlation may attenuate the bias. We find that two different treatment variations along these lines indeed debias the large majority of subjects, hence suggesting that many people are in principle capable of dealing with the informational redundancies in our experimental task, but only so when their focus is directed to the problematic aspect of the updating environment. In a final step, we explore whether these results reflect "rational ignorance", i.e., subjects trading off the benefits of more precise beliefs against lower cognitive effort costs as resulting from not even thinking about the problem in detail (Caplin et al., 2011; Caplin and Dean, forthcoming). We find that an increase in the stake size significantly affects subjects' effort levels, but not their beliefs, which again exhibit a bimodal pattern. These findings are consistent with the idea that – if left to their own devices – subjects attempt to identify the critical aspect of the informational environment, and do so harder when the stakes are higher. However, if they do not succeed in passing the threshold of identifying the double-counting problem, they make use of a specific heuristic.

This paper contributes to the literature on boundedly rational belief formation by identifying a novel error in statistical reasoning that is associated with a pervasive feature of real information structures such as the news media Charness et al. (see, e.g., 2010), Andreoni and Mylovanov (2012), and Esponda and Vespa (2014, for other recent contributions). In addition, our paper moves beyond existing work on belief formation by studying in great detail the cognitive mechanisms underlying an updating error. Our finding that variation in focus might affect the formation of beliefs dovetails with recent empirical work that highlights the effectiveness of nudging people into paying attention to certain features of the informational environment (Hanna et al., 2014). Gennaioli and Shleifer (2010), Bordalo et al. (2015b), and Schwartzstein (2014) provide related theoretical models.[2]

[2] Brocas et al. (2014) highlight the relevance of attention in strategic settings.

Our individual belief elicitation treatments admit a natural interpretation in terms of learning in networks. Eyster and Rabin (2014) develop a model to show that, in many network structures other than the canonical sequential herding example, rationality requires people to anti-imitate predecessors because of the need to subtract off sources of correlations. In consequence, these authors argue, empirical tests are needed to separate whether people follow others for rational reasons or due to correlation neglect. Our experimental design provides the first assessment of this issue by making explicit use of the advantages of laboratory experiments in studying statistical inference: our static experimental environment with exogenous signals and a *known* data-generating process allows for a clean identification of people's tendency to ignore redundancies in information sources that does not require ancillary assumptions on people's models of others' decision rules in the presence of no common knowledge of rationality.[3] In consequence, our findings support the assumptions underlying recent theories of inferential naïveté in social interactions (e.g., DeMarzo et al., 2003; Golub and Jackson, 2010; Eyster and Rabin, 2010; Bohren, 2013).[4] Levy and Razin (forthcoming) and Ortoleva and Snowberg (2015) investigate the implications of correlation neglect in political economy settings.[5]

Finally, in a broader sense, our paper also relates to work on financial decision-making in the presence of correlated asset returns (Eyster and Weizsäcker, 2011; Kallir and Sonsino, 2010). Here, apart from the different context (porfolio choice versus belief formation), the term "correlation neglect" also has a conceptually different meaning than in our paper. For instance, portfolio choice problems do not feature the double-counting problem that is at the heart of our analysis. Also, unlike in the case of informational redundancies, dealing with correlated asset returns requires contingent reasoning (state by state). None of the papers in this literature studies correlation neglect in information

[3] While our paper is concerned with updating under a *known* data-generating process, a literature in cognitive psychology explores how people aggregate potentially correlated opinions in settings in which the structure generating the information is left ambiguous to subjects (Budescu and Rantilla, 2000; Budescu and Yu, 2007). These papers focus on non-incentivized confidence ratings. Kahneman and Tversky (1973) note that correlated information sources tend to produce consistent signals and may hence lead to an "illusion of validity" (also see Maines, 1990, 1996).

[4] The findings from our individual belief elicitation task contribute to an active empirical literature that tests key predictions of naïve social learning models and finds mixed results. In the context of sequential herding experiments, Kübler and Weizsäcker (2005) argue that many people fail to recognize herding behavior of others. At the same time, such experiments consistently yield the result that people vastly overweight their private signals, which is the antithesis of correlation neglect in such environments: people overwhelmingly herd *less* than the rational model predicts (Weizsäcker, 2010), while correlation neglect predicts that they herd *more* (Eyster and Rabin, 2010). Similarly, in dynamic social network experiments, some studies find belief patterns that are broadly consistent with naïve updating (Brandts et al., 2014; Chandrasekhar et al., 2015). At the same time, Corazzini et al. (2012) find that exogenously increasing the number of outgoing links of an agent does not affect his social influence; Grimm and Mengel (2014) find heavy overweighting of private signals, again at odds with correlation neglect, while Möbius et al. (2013) cannot reject Bayesian rationality. While these experiments are insightful, the mixed results need to be interpreted with care because these studies focus on social interactions, implying that signals consist of the actions of other players; thus, when there is no common knowledge of rationality, such designs potentially conflate erroneous updating with people's models of other's decision rules in attempting to identify updating mistakes. In consequence, the mixed results may or may not reflect a combination of correlation neglect and people theorizing that fellow subjects follow certain decision rules.

[5] Spiegler (2015) uses Bayesian networks to model boundedly rational belief formation.

sources, corresponding implications (such as overshooting beliefs and market behavior), or the underlying mechanisms.

The remainder of the paper is organized as follows. In the next section, we present our baseline experiments including the market treatments. Sections 2.3 and 2.4 investigate the validity of face value bias and the mechanisms underlying correlation neglect, respectively. Section 2.5 concludes.

2.2 Correlation Neglect and its Implications

We developed a simple experimental design which allows for both the clean identification of correlation neglect and an investigation of its implications in market settings in a unified and coherent framework. We first describe the basic belief elicitation design and then explain how these treatments were extended into market treatments. After stating our predictions, we present the results.

2.2.1 Experimental Design

2.2.1.1 Individual Belief Formation Treatments

An environment in which updating from correlated sources can be studied requires (i) control over signal precision and correlation, (ii) subjects' knowledge of the data-generating process, (iii) a control condition that serves as benchmark for updating in the absence of correlated information, and (iv) incentivized belief elicitation.

Our design accommodates all these features. Subjects were asked to estimate an ex ante unknown continuous state of the world μ and were paid for accuracy. The task was framed as guessing how many items are contained in an imaginary container. In order to keep the experiment as simple as possible, we refrained from inducing prior beliefs.[6] The only information provided to participants consisted of unbiased computer-generated signals about the true state. The key idea of the between-subjects design was to construct two sets of signals (one with and one without a known and simple correlation), which are identical in terms of their objective informational content. As depicted in Figure 2.1, subjects in the *Correlated* treatment received correlated and subjects in the *Uncorrelated* condition uncorrelated information about μ.

The computers A-D generated four unbiased iid signals about μ, which were identical across treatments. Technically, this was implemented by random draws from a truncated discretized normal distribution with mean μ and standard deviation $\sigma = \mu/2$.[7] In the *Uncorrelated* treatment (left panel), the intermediaries 1 to 3, who are fictitious computers themselves, observed the signals of computers B through D, respectively, and simply transmitted these signals to the subject. Thus, subjects received information from computer A as well as from the three intermediaries. For example, in one belief formation task, the signals of computers A through D were given by 12, 9, 10, and 0, respectively. We will refer to all numbers that are communicated to subjects as "messages".

[6] Section 2.2.4.1 shows that inducing prior beliefs does not affect our findings.
[7] Truncation was at $\mu \pm 2\sigma = \mu \pm \mu$ in order to avoid negative signals.

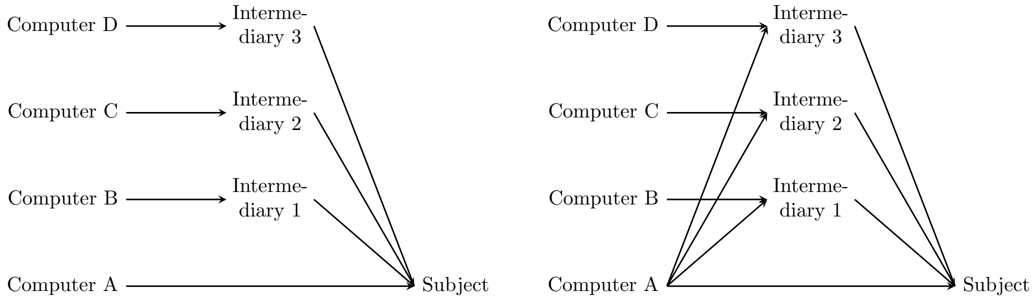


Figure 2.1. Uncorrelated (left panel) and correlated (right panel) information structure

In the *Correlated* treatment (right panel), the intermediaries 1 to 3 observed both the signal of computer A and of computers B to D, respectively, and then reported the average of these two signals. Again, subjects were provided with information from computer A as well as from the three intermediaries. Throughout the paper, we will also refer to computer A's signal as common source signal. Since subjects knew this signal, they could extract the other independent signals from the intermediaries' reports. Continuing the example from above, each of the three intermediaries took the average of 12 and the corresponding signal of the other computer it communicated with. Thus, computer A reported 12, intermediary 1 reported 10.5, intermediary 2 reported 11, and intermediary 3 reported 6. In the terminology of Eyster and Rabin (2014), this information structure constitutes a "shield". Here, people need to "anti-imitate" because they predominantly see messages larger than 9, while the majority of signals and the rational belief are smaller than 9. In particular, given that the common source signal of computer A is known, being rational requires subjects to back out the underlying independent signals from the messages of the intermediaries, i.e., to invert averages.

Notice that our identification strategy relies solely on the identical informational content of the two sets of signals. Differences in beliefs between the *Correlated* and *Uncorrelated* condition can only be attributed to variations in the information structure since all other factors are held constant. Thus, comparing beliefs between the *Correlated* treatment and the *Uncorrelated* benchmark allows us to identify subjects' potential naïveté when updating from correlated information.[8] Crucially, using computers as opposed to human subjects in the signal-generating process ensures that subjects have complete knowledge of how their data are being generated, leaving no room for, e.g., beliefs about the rationality of the intermediaries. Also note that the correlated information structure mirrors the examples provided in the introduction. For example, one could think of computer A as a press agency which sells information to various newspapers, which in turn each have an additional independent information source. Alternatively, in a social learning context, the intermediaries could be viewed as network members who each received an independent piece of information, yet have all also talked to a common acquaintance before communicating their opinion.

[8] This holds provided that the treatment did not affect prior beliefs. As we show in Section 2.2.4.1, our results are robust to explicitly inducing equal priors across treatments.

Table 2.1. Overview of the belief formation tasks

True State	Computer A	Intermed. 1 uncorr.	Intermed. 2 uncorr.	Intermed. 3 uncorr.	Intermed. 1 corr.	Intermed. 2 corr.	Intermed. 3 corr.	Rational Belief	Correlation Neglect Belief
10	12	9	10	0	10.5	11	6	7.75	9.88
88	122	90	68	5	106	95	64	71.25	96.63
250	179	295	288	277	237	234	228	259.75	219.38
732	565	847	650	1,351	706	608	958	853.25	709.13
1,000	1,100	1,060	629	1,100	1,085	870	1,105	974.75	1,042.38
4,698	1,608	7,240	4,866	5,526	4,424	3,237	3,567	4,810.00	3,209.00
7,338	9,950	1,203	11,322	11,943	5,577	10,636	10,947	8,604.50	9,277.25
10,000	2,543	10,780	6,898	8,708	6,662	4,721	5,626	7,232.25	4,887.63
23,112	15,160	21,806	20,607	47,751	18,483	17,884	31,456	26,331.00	20,745.50
46,422	12,340	32,168	49,841	61,293	22,254	31,091	36,817	38,910.50	25,625.25

The reports of intermediaries 1 through 3 in the *Uncorrelated* condition directly reflect the draws of computers B-D. The rational belief is computed by taking the average of the signals of computers A-D. The correlation neglect belief is given by the average of the signal of computer A and the reports of intermediaries 2-4 in the *Correlated* condition. Note that subjects faced the ten rounds in randomized order, which was identical across treatments. Given that we did not induce priors, we could select the true states ourselves. This was done in a fashion so as to be able to investigate the effects of computational complexity, i.e., we implemented true states of different magnitude.

Upon receiving the information pieces, a subject had five minutes to state a belief. Subjects completed a total of ten independent belief formation tasks without feedback between tasks. We used three different randomized orders of tasks, see Appendix 2.B. At the end of the experiment, subjects were paid according to the precision of their belief in one randomly selected task using a quadratic scoring rule (Selten, 1998).[9] Table 1 provides an overview over the ten tasks. In order to provide an indication of both the direction and the extent of a potential bias, we also provide the benchmarks of rational beliefs and "full correlation neglect", which we define to be the average of the four signals subjects receive in the *Correlated* treatment (see Section 2.2.2 for details). Throughout, we employ the term "belief" to denote the mean of the belief distribution.

Subjects received extensive written instructions which explained the details of the task and the incentive structure.[10] In particular, the signals of the four computers, how these signals mapmed into the reports of the intermediaries, and the fact that the four computers are of identical quality, were explained in great detail. For instance, the instructions included the applicable panel from Figure 2.1. The instructions also contained an example consisting of four computer signals as well as the respective messages of the three intermediaries, given a certain state of the world. Subjects were provided with a visual representation of an exemplary distribution function and the concept of unbiasedness was elaborated upon in intuitive terms. A summary of the instructions was read out aloud. In addition, subjects completed a set of control questions with a particular focus on the information structure. For example, in both treatments, subjects had to compute the reports of intermediaries 1 and 2 given exemplary signals of the four computers in order to make sure that subjects understood the (un)correlated nature of the messages. Subjects could only participate in the experiment once they had answered all control questions correctly.[11]

[9] Variable earnings in euros were given by $\pi = max\{0, 10 - 160 \times (\text{Belief / True state} - 1)^2\}$.

[10] See Appendix H for a translation of the instructions and control questions for all treatments. The instructions can also be accessed at https://sites.google.com/site/benjaminenke/.

[11] We can rule out that subjects solved the control questions by trial-and-error. The quiz was implemented on two consecutive computer screens that contained three and four questions, respectively. If at least one question was answered incorrectly, an error message appeared, but subjects were not notified

At the end of the experiment, we conducted a questionnaire in which we collected information on sociodemographics. To capture dimensions of cognitive ability, we asked subjects for their high school GPA (German "Abitur") and had them solve ten rather difficult IQ test Raven matrices.

2.2.1.2 Market Treatments

In the market treatments, the belief formation task was embedded into a standard double-auction setting with uncertainty over the value of the assets. In each trading round, an asset's value corresponded to the true state of the world from the individual belief formation treatments. Before each round, all traders received the same sets of signals about the state as participants in the baseline design (see Table 1). In the *Correlated market* treatment, all market participants received correlated, in the *Uncorrelated market* treatment they received uncorrelated information. Before each trading round, subjects were given five minutes to think about an asset's value and to provide a non-incentivized belief. Afterwards, subjects traded the assets.

In order to keep the experiment as simple as possible and to retain subjects' focus on the information structure, participants were assigned to be in the role of a buyer or a seller, so that each subject could either buy or sell assets, but not both. A market group consisted of four buyers and four sellers. Subjects were randomly assigned to be in either role and kept their roles throughout the experiment; they also remained in the same market groups. Before each of the ten rounds, each seller was endowed with four assets. Also, at the beginning of each round, each buyer received a monetary endowment that was sufficient to purchase between three and six assets at fundamental values.[12]

In a standard double-auction format, buyers could post buying prices and accept selling offers from the sellers. Sellers could post selling prices and accept buying offers from the buyers. Buying and selling offers were induced to converge by the standard procedure, i.e., a new buying (selling) offer had to be higher (lower) than all previous offers. An accepted offer implied a trade and erased all previous offers. Trading lasted for four minutes. Profits per trading period for both buyers and sellers corresponded to the value of the assets owned plus the amount of money held at the end of the respective trading round minus some known fixed costs.

---

which question(s) they had gotten wrong. For instance, the computer screen which contained two questions that asked subjects to compute the reports of the intermediaries given exemplary signal draws (which arguably constitute the key control questions) had a total of 13 response options across four questions (i.e., $2 \times 3 \times 4 \times 4 = 96$ combinations of responses), making trial-and-error *extremely* cumbersome. In addition, the BonnEconLab has a control room in which the decision screens of all subjects can be monitored. From this monitoring, no attempts to solve the control questions by random guessing were detectable. Furthermore, whenever a subject appeared to have trouble solving the control questions, an experimenter approached the subject, clarified open questions, and (very rarely) excluded the subject if they did not show an adequate understanding of the task.

[12] Throughout the experiment, profits, prices etc. were described in points rather than euros. Since the true state differed in magnitude from round to round, we had to adjust the point / euro exchange rate across rounds. This was made clear in the instructions. In principle, the exchange rate as well as the budget was informative of the true state. However, the relationship between these variables was chosen to be non-constant across rounds, so that the informational content was weak (see Appendix 2.E.7 for details). In any case, since budgets and exchange rates were identical across treatments, this procedure cannot explain potential treatment differences.

We used two different randomized orders of rounds. After each round, subjects received feedback about the true state of the world and the resulting profits in that round. At the end of the experiment, one of the ten rounds was randomly selected and implemented, i.e., payoff-relevant for the subjects. The written instructions included the same information on the information structure as in the individual belief formation treatments. A summary of the instructions was read out aloud. In addition to the control questions about the information structure, we asked several questions related to the trading activities. After the control questions, we implemented a test round after which participants again had the opportunity to ask questions.

2.2.2 Hypotheses

In the information structure described above, the computers generated four iid signals of the form $s_h \sim \mathcal{N}(\mu, (\mu/2)^2)$ (truncated at $(0, 2\mu)$) for $h \in \{1, \dots, 4\}$. In the *Correlated* condition, subjects observed messages s_1 and $\tilde{s}_h = (s_1 + s_h)/2$ for $h \in \{2, 3, 4\}$. When prompted to estimate μ, a rational decision maker would extract the underlying independent signals from the messages $\tilde{s}_h$ and compute the mean rational belief as $b_B = \sum_{h=1}^{4} s_h/4$, which by design also equals the rational belief in the *Uncorrelated* condition.[13]

However, now suppose that the decision maker suffers from correlation neglect, i.e., he does not fully take into account the extent to which $\tilde{s}_h$ reflects s_1, but rather treats $\tilde{s}_h$ (to some extent) as independent. Call such a decision maker naïve and let his degree of naïveté be parameterized by $\chi \in [0, 1]$ such that $\chi = 1$ implies full correlation neglect. A naïve agent extracts s_h from $\tilde{s}_h$ according to the rule

$$\hat{s}_h \ = \ \chi\tilde{s}_h + (1 - \chi)s_h \ = \ s_h + \frac{1}{2}\chi(s_1 - s_h) \tag{2.1}$$

where $\hat{s}_h$ for $h \in \{2, 3, 4\}$ denotes the agent's (possibly biased) inference of s_h. He thus forms mean beliefs according to

$$b_{CN} \ = \ \frac{s_1 + \sum_{h=1}^{3} \hat{s}_h}{4} \ = \ \bar{s} + \frac{3}{8}\chi(s_1 - \bar{s}_{-1}) \tag{2.2}$$

where $\bar{s} = (\sum_{h=1}^{4} s_h)/4$ and $\bar{s}_{-1} = (\sum_{h=2}^{4} s_h)/3$. Thus, a (perhaps partially) naïve belief is given by the rational belief plus a belief bias component which depends on the degree of naïveté and the magnitude of the common source signal relative to the other signals.

[13] For simplicity, when computing the rational belief, we ignore the truncation in the signal distribution and assume that subjects hold vague priors. Note that the quantitative errors resulting from this are likely to be very small in magnitude. Given the information provided to subjects, potential priors are very likely to be weak. Also, the tails outside the truncation are fairly thin. Moreover, our definition of the rational belief conforms with observed behavior in the *Uncorrelated* treatment, where subjects tended to merely take the average of the four signals. Finally, and most importantly, this definition of the rational benchmark has no effect on the qualitative predictions of our treatment comparison. Regardless of the precise definition, beliefs should be identical across treatments.

Hypothesis 1. *Assuming that $\chi > 0$, beliefs in the Correlated treatment exhibit an overshooting pattern. Specifically, given a high common source signal, i.e., $s_1 > \bar{s}_{-1}$, beliefs in the Correlated treatment are biased upward compared to the Uncorrelated treatment. Conversely, if $s_1 < \bar{s}_{-1}$, beliefs in the Correlated condition are biased downward. The degree of the belief bias increases the relative magnitude of the common source signal.*

Intuitively, by partially neglecting the redundancies among the signals, the decision maker double-counts the first signal, so that beliefs are biased in the corresponding direction. Throughout the paper, we will call a belief above (below) the rational benchmark overoptimistic (overpessimistic). Note that the beliefs of a naïve agent remain statistically unbiased. Since the first signal is unbiased, any double-counting leads to a zero expected error. The upshot of this is that naïve agents are correct on average, yet exhibit excessive swings in their beliefs.

In the market treatments, the standard theoretical prediction is that the competitive market equilibrium price is given by the rational belief.[14] Since it is well-established that experimental double-auctions tend to converge to the theoretical competitive equilibrium, this is also the standard experimental prediction. However, this prediction changes in the presence of naïve traders. If, for instance, all traders are homogenous in their degree of naïveté, the equilibrium price level is given by the corresponding level of distorted beliefs. More generally, as we detail in Appendix 2.E.1, under heterogeneity the magnitude of a potential price distortion will depend on the naïveté of the marginal traders.[15]

Hypothesis 2. *Assuming that $\chi > 0$, the excessive belief swings induced by correlation neglect translate into over- and underpricing. If $s_1 > \bar{s}_{-1}$, market prices in the Correlated market treatment are too high relative to the Uncorrelated treatment, and if $s_1 < \bar{s}_{-1}$ they are too low.*

On the other hand, it has been argued that the influence of cognitive biases on aggregate variables is limited. In the market we implement, two channels in particular may attenuate such effects. First, competitive forces and market incentives could induce subjects to think harder and thus cause a reduction of correlation neglect. Second, markets provide ample opportunities for traders to learn. For instance, traders may learn from realized profits in each trading round. In this respect, we gave rather extensive feedback between rounds, providing subjects with realized profits as well as the true asset value. Perhaps more importantly, markets also allow participants to learn from the actions of more rational traders. For instance, an overly optimistic market participant who observes others trading at relatively low prices may become inclined to rethink his valuation of the assets. While all these channels could mitigate the effect of individual biases on market

[14] Since every subjects got the same signals about the value of the assets, under homogenous risk preferences there should be no trade, unless market participants trade at the rational belief.

[15] For instance, intuitively, suppose that a fraction α fully ignores correlations and a fraction $1-\alpha$ holds rational beliefs. Further suppose that each seller owns four assets and each buyer has a budget sufficient to buy four assets at fundamental values. Then, assuming that subjects do not learn from others' trading behavior and are risk-neutral, the supply and demand curves will be step functions which overlap at the correlation neglect belief if $\alpha \to 1$. Similar arguments apply if a fraction α exhibits only partial (or heterogeneous degrees of) correlation neglect.

outcomes, the learning arguments in particular would suggest that correlation neglect (and its consequences) is reduced in the last trading rounds.[16]

2.2.3 Procedural Details

The experiments were conducted at the BonnEconLab of the University of Bonn. Subjects were mostly students from the University of Bonn and were recruited using the online recruitment system by Greiner (2004). No subject participated in more than one session. The experiment was run using the experimental software z-Tree (Fischbacher, 2007). A total of 94 subjects participated in the individual belief formation treatments, which were randomized within session. Sessions lasted about 1.5 hours and average earnings equalled 11.60 euros (≈ USD 15 at the time). 288 subjects participated in the market treatments. These sessions lasted about 2.5 hours and subjects earned 19.40 euros (≈ USD 25) on average. In all treatments, payments included a 6 euros show-up fee.

2.2.4 Results

Our analysis proceeds in two steps. First, we provide evidence for correlation neglect across the ten belief formation tasks. Second, we investigate how the neglect of informational redundancies plays out in markets.

2.2.4.1 Clean Evidence for Correlation Neglect

Beliefs Across Treatments

Result 1. *In all but one belief formation task, beliefs differ significantly between treatments in the direction predicted by correlation neglect.*

Figure 2.2 visualizes the pattern of beliefs across tasks. Recall the key implication of the hypotheses developed above that subjects' beliefs should be too high (low) relative to the rational benchmark if the signal of the common source A is relatively high (low) compared to the other signals. Thus, for each of the ten tasks and both treatments, the figure plots the difference between the respective median belief and the rational benchmark against the relative magnitude of the signal of the common source (i.e., the difference between the signal of computer A and the average signal of the other computers). By construction of the figure, the rational prediction is a flat line at zero (no belief bias), while full correlation neglect predicts an upward-sloping relationship. Beliefs in the *Uncorrelated* condition follow the rational prediction very closely. In contrast, median beliefs in the *Correlated* condition always lie between the rational benchmark and the full correlation neglect prediction, and the magnitude of the belief bias exhibits a clear relationship with the relative magnitude of the common source signal, as predicted in Section 2.2.2.

Table 2.2 provides summary statistics for all tasks and reveals that in nine out of ten cases do beliefs in *Correlated* significantly differ from those in the *Uncorrelated* treat-

[16] Camerer (1987) provides a more extensive discussion of these feedback and learning effects. Similar to our approach, he uses experimental markets to test if other updating mistakes (e.g., base-rate neglect) matter for market outcomes. See also Ganguly et al. (2000) and Kluger and Wyatt (2004) for similar studies.

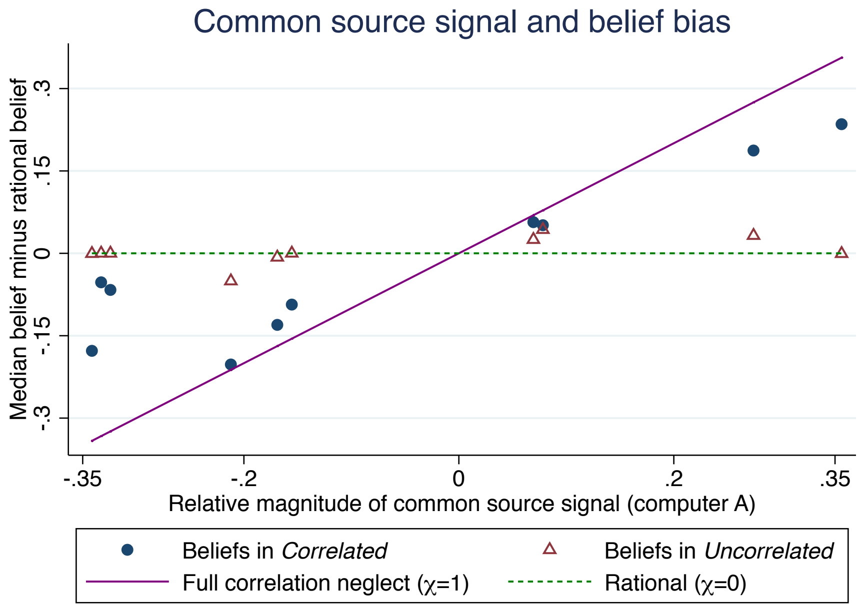


Figure 2.2. Beliefs in the *Correlated* and *Uncorrelated* treatments plotted against the relative magnitude of the signal of computer A. The logic of the figure is that if the signal of computer A is relatively high (low) compared to the other signals, correlation neglect predicts that beliefs should be above (below) the rational benchmark. Accordingly, the x-axis measures the signal of computer A minus the average signal of the other computers, while the y-axis represents the median belief for the given signal realizations minus the corresponding rational belief. Both differences are then rescaled across tasks by dividing them through the Bayesian belief. That is, in terms of the notation introduced in Section 2.2.2, the variable on the x-axis is computed as $3 \cdot (s_1^j - \bar{s}_{-1}^j)/(8\bar{s}^j)$ and the variable on the y-axis as $(b^j - \bar{s}^j)/\bar{s}^j$. The dashed line represents the rational prediction, while the solid line denotes the full correlation neglect benchmark across the ten different signal realizations (tasks).

ment.[17] The bias is very stable across tasks and does not seem to depend on the magnitude of the true state.[18] Also note that we do not find order effects, i.e., subjects do not seem to learn to deal with correlations over time (see Appendix 2.C.4).

Because beliefs in the *Correlated* treatment are consistently further away from the rational belief than beliefs in the *Uncorrelated* condition, these subjects earned roughly 2.70 euros less than those in the *Uncorrelated* group, which amounts to almost 50 % of subjects' average variable earnings. The earnings difference is significant (p-value = 0.0025, Wilcoxon ranksum test).

[17] The non-significant true state is also the only one in which beliefs and prices did not differ in the market treatments to be presented below. Notice, however, that subjects' beliefs indeed reflected correlation neglect, but beliefs in the *Uncorrelated* condition were also tinted into that direction. A potential reason for this is that, in the *Uncorrelated* condition, subjects received three signals in the ballpark of 10,000 and one which equalled 1,203. It is conceivable that subjects viewed the latter signal as implausible and formed beliefs based on the other signals, coincidentally leading to a belief which is biased towards the correlation neglect prediction.

[18] Appendix 2.C.1 illustrates the robustness of this first main result by excluding outliers from the analysis and by providing kernel density estimates for each of the ten belief formation tasks.

Distribution of Naïveté

Thus far, we have established a significant amount of correlation neglect *on average*. However, these average patterns may mask a substantial amount of heterogeneity. To investigate this, we develop a measure of an individual's belief type. To this end, we aggregate the data across tasks into a one-dimensional measure per individual. Specifically, our experimental design in combination with the simple model of belief formation introduced in Section 2.2.2 allows us to derive a simple estimator for the individuals' naïveté χ. As a first step, we normalize beliefs across tasks such that they equal the naïveté parameter $\chi \in [0,1]$ in eq. 2.2, i.e., we express the normalized belief $\tilde{b}_i^j$ of individual i in round j as function of his stated belief b_i^j and the realized signals s^j.[19] We then compute the median normalized belief of each individual for further analysis, yielding the following estimator for the naïveté parameter:

$$\hat{\chi}_i \equiv med(\tilde{b}_i^j) = med\left(\frac{8(b_i^j - \bar{s}^j)}{3(s_1^j - \bar{s}_{-1}^j)}\right) \tag{2.3}$$

The left panel of Figure 2.3 provides kernel density estimates of the distribution of these naïveté parameters for both the *Correlated* and the *Uncorrelated* treatment.[20] The plots reveal that in the *Uncorrelated* treatment the vast majority of subjects approximately behaves rational, as indicated by the spike around zero. In the *Correlated* treatment, on the other hand, we observe two peaks around the rational benchmark and the full correlation neglect parameters, respectively, which suggests the presence of different types of subjects. In particular, those subjects that do not succesfully process correlations form beliefs by following a particular simple heuristic that is essentially fully naïve. As

Table 2.2. Correlation neglect by belief formation task

True State	Rational Belief	Correlation Neglect Belief	Median Belief *Uncorr.* Treatment	Median Belief *Correlated* Treatment	Ranksum Test (p-value)
10	7.75	9.88	8	9.2	0.0048
88	71.25	96.63	71.2	88	0.0005
250	259.75	219.38	259.75	235.5	0.0067
732	853.15	709.13	847	742	0.0044
1,000	974.75	1,042.38	999	1,030	0.0484
4,698	4,810	3,209	4,810	4,556	0.0082
7,338	8,604.5	9,277.25	8,975	9,044.5	0.8657
10,000	7,232.25	4,887.63	7,232	6,750	0.0087
23,112	26,331	20,745.5	25,000	21,000	0.0001
46,422	38,910.5	25,625	38,885.5	32,000	0.0527

See Table 2.1 for details of the computation of the rational and the correlation neglect benchmarks. Note that subjects faced the ten tasks in randomized order.

[19] This normalization procedure takes into account that the (percentage) difference between rational and correlation neglect belief differs across tasks. Note that, naturally, in the actual data, not all χ map into [0,1]. For example, a subject who fully neglects redundancies may in addition make a computational mistake to end up with a χ higher than one. Likewise, a subject who aims at computing the rational belief may make a small error, so that their χ may be below zero.

[20] In what follows, we use the terms normalized belief and naïveté parameter interchangeably.

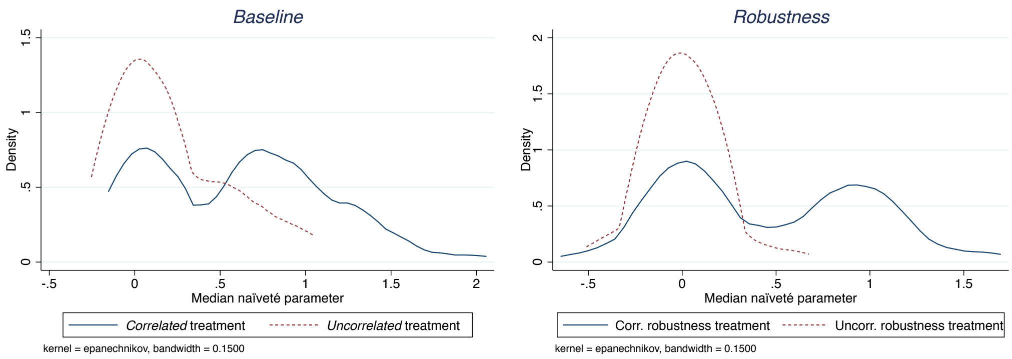


Figure 2.3. Kernel density estimates of median naïveté parameters. The left panel depicts the distribution of naïveté in the baseline treatments, and the right panel in the robustness treatments.

visual inspection suggests, comparing median normalized beliefs across treatments also reveals a pronounced treatment difference (p-value < 0.0001, Wilcoxon ranksum test). Appendix 2.C.2 confirms that the bimodal structure of the belief distribution in *Correlated* is not an artifact of our particular aggregation procedure, but is also clearly visible in the disaggregated data.[21]

Our procedure of computing an individual's belief type only makes use of the first moment of the distribution of each subject's set of beliefs (the median), and hence ignores the variability in beliefs. In Appendix 2.C.5, we pursue a different approach by structurally estimating the belief formation rule proposed in Section 2.2.2 through a finite mixture model, which allows for heterogeneity in both the mean and the error rate of subjects' belief formation type. The picture resulting from these estimations is very similar to what can be inferred from Figure 2.3. For example, the estimations also identify a group of rationals as well as group of fully naïve subjects.

Robustness

Our belief elicitation design made a number of design choices, whose overarching goal was to create a relatively simple updating environment. To illustrate that none of our design features was critical in generating the results, we now investigate the robustness of our treatment comparison. To this end, we conducted a robustness treatment (both *Correlated* and *Uncorrelated*) which was identical to the baseline treatments, with the exception of variations along four design dimensions.

First, the data-generating process was altered slightly. We induced a prior belief by informing subjects that μ would be drawn from $\mathcal{N}(0; 250,000)$, while the signal distribution was given by $s_h \sim \mathcal{N}(\mu; 250,000)$. As a consequence, negative true states were possible and we eliminated the truncation of the signal distribution. Both prior and signal distributions were explained to subjects in great detail, and the instructions included the corresponding formulas. Control questions ensured that subjects understood the key

[21] Appendix 2.C.3 analyzes the stability of the individual-level naïveté parameters across tasks.

features of the prior distribution as well as the equal variance of the prior and signal distributions.

Second, we introduced a fourth intermediary which, in both the *Uncorrelated* and the *Correlated* condition, simply transmitted the signal of computer A to the subject. Thus, subjects only communicated with intermediaries.

Third, subjects' payment was determined by the binarized scoring rule, which is incentive-compatible regardless of subjects' risk attitudes (Hossain and Okui, 2013).[22]

Fourth, instead of framing the experimental task as guessing how many items are contained in an imaginary container, we explicitly told subjects that they would have to estimate a hypothetical true state, which would be drawn by the computer.

96 subjects participated in these treatments and earned 11.10 euros on average. Appendix 2.D presents details on all ten belief formation tasks as well as the corresponding results. To summarize, the results of these robustness treatments are very similar to those in the baseline treatments. The right panel of Figure 2.3 illustrates this by plotting median naïveté parameters for both conditions.[23] As in the baseline treatments, the type distribution in the *Correlated* condition exhibits a bimodal structure, according to which some fraction of subjects fully neglects informational redundancies, while others state the same beliefs as subjects in the *Uncorrelated* condition. Accordingly, the belief distributions in the *Correlated* and *Uncorrelated* treatments significantly differ from each other ($p<0.0001$, Wilcoxon ranksum test). This is also reflected by lower earnings of subjects in the *Correlated* condition (earnings difference=2.30 euros, p-value=0.0255, Wilcoxon ranksum test).

2.2.4.2 Market Treatments – Over- and Underpricing

Price Levels Across Treatments

In both market treatments, we have observations from 18 market groups that trade in ten trading rounds each. For each market group and trading round, we define the price of the last concluded trade to be the market price.[24] We first consider the effect of our treatment variation on price levels.

Result 2. *Market prices differ between treatments as predicted by correlation neglect. In the Correlated market treatment, we observe frequent over- or underpricing, depending on the relative magnitude of the common source signal. Neither prices nor subjects' beliefs reflect learning over time.*

Table 2.3 provides summary statistics for all ten trading rounds. We present two price predictions (consisting of the rational benchmark and the full correlation neglect belief,

---

[22] Specifically, we computed a penalty term by squaring the distance between a subject's belief and the true state. The subject then received 10 euros if the penalty was smaller than a randomly drawn number $k \sim U[0; 100,000]$, and nothing otherwise.

[23] Given that we induced a prior in these treatments, computing individual-level naïveté towards correlations requires an assumption on potential base rate neglect. We base this assumption on behavior in the *Uncorrelated* robustness condition, where subjects uniformly essentially fully neglect the base rate. Accordingly, we assume full base rate neglect, i.e., normalized beliefs are computed using equation (2.3), also see Appendix 2.D. This assumption has no bearing on our treatment comparison, but only serves to illustrate the population distribution of naïveté.

[24] All results are robust to other definitions of the market price, see Appendices 2.E.2 and 2.E.3.

Table 2.3. Market prices by trading round

True State	Rational Belief	Correlation Neglect Belief	Median Market Price *Uncorr.* Treatment	Median Market Price *Correlated* Treatment	Ranksum Test (p-value)	Beliefs Differ?
10	7.75	9.88	8.35	9.05	0.0093	Yes
88	71.25	96.63	86.5	93.45	0.0338	Yes
250	259.75	219.38	275	260	0.0113	Yes
732	853.15	709.13	820	737	0.0001	Yes
1,000	974.75	1,042.38	1,000	1,039	0.0723	Yes
4,698	4,810	3,209	5,200	4,470.5	0.0085	Yes
7,338	8,604.5	9,277.25	9,124	8,999	0.6087	No
10,000	7,232.25	4,887.63	7,575	6,250	0.0534	Yes
23,112	26,331	20,745.5	24,100	21,300	0.0007	Yes
46,422	38,910.5	25,625	41,000	35,000	0.0015	Yes

Median market prices are defined as the median of all market prices over the 18 markets in the respective round. Beliefs are said to differ between treatments in a particular round if and only if p-value < 0.05, Wilcoxon ranksum test. Note that subjects faced the ten rounds in randomized order.

respectively), actual price levels, as well as an indicator for whether subjects' beliefs (as stated prior to trading) differ significantly across treatments. In all rounds but one, prices significantly differ between treatments in the direction one would expect from a correlation neglect perspective. While market prices in the *Uncorrelated* treatment follow the rational prediction rather closely, we observe frequent instances of over- and underpricing in the *Correlated market* treatment. Thus, the magnitude of the common source signal relative to the other signals consistently predicts whether assets sell above or below the values from the *Uncorrelated market* treatment.

In Appendices 2.E.2 and 2.E.3, we establish the robustness of the treatment difference in price levels by excluding outliers from the analysis and by providing density estimates of the price kernel, both at an aggregated level across periods and separately for each period. Strikingly, the (aggregated) price kernel is centered around $\chi \approx 0.5$, suggesting that rational and naïve types negotiate prices between the two extreme predictions. We also show that the treatment difference in prices is entirely driven by subjects' beliefs: In an OLS regression of all prices from all market groups on a treatment dummy, the latter vanishes after accounting for elicited beliefs. Thus, the overshooting beliefs that are implied by neglecting informational redundancies indeed cause overshooting price levels.

Next, we provide a visual representation of the temporal pattern of the market price volatility induced by correlation neglect. To this end, we first normalize market prices to make them comparable across rounds. This is done using a procedure akin to the belief normalization in the individual belief formation treatments (see eq. (2.3)), so that, for each market group and trading period, we essentially compute the naïveté inherent in the market price (which, in principle, should be between zero and one). However, by construction, this normalization does not allow us to distinguish the occurrence of over- from that of underpricing. Thus, we slightly reformulate this normalization: In trading rounds in which correlation neglect predicts overoptimism, the normalization remains the same, so that a normalized price of one (zero) indicates full correlation neglect (rational price levels). On the other hand, in periods in which neglecting correlations leads to overpessimism, we normalize prices such that full correlation neglect is indicated by

(−1) and the rational benchmark by zero, respectively.[25] For each trading round, we then compute the difference between the median market price in the *Correlated market* treatment and the median market price in the *Uncorrelated* condition, which gives us an indication of the price distortion in the *Correlated market* treatment *relative* to its appropriate benchmark.

The two panels in Figure 2.4 plot this difference in market prices against the theoretical predictions across our ten trading rounds (we used two different orderings of rounds). First note that, by construction, the rational prediction is always given by zero; if correlation neglect did not impact aggregate outcomes, prices would not differ across conditions. The full correlation neglect prediction, on the other hand, alternates between one and (-1) depending on whether correlation neglect implies overoptimism or -pessimism. The plots show that in almost all periods the price difference follows the correlation neglect prediction, so that prices frequently overshoot. As a result, the excessive belief swings implied by correlation neglect directly translate into volatile price levels. In addition, as visual inspection suggests, this pattern does not attenuate over time. Appendix 2.E.4 formally confirms that the bias reflected in market prices does not become smaller over the course of the ten trading periods. Appendix 2.E.5 analyzes the time trend of the beliefs subjects stated prior to trading started. Again, the results provide no indication that subjects learn to deal with correlated signals over time. Appendix 2.E.6 discusses potential reasons why the market does not debias subjects.

Beliefs, Prices, and Individual Trading Behavior

So far, we have shown that correlated information structures have predictable consequences for experimental market outcomes, i.e., price levels. Next, we demonstrate that individual-level heterogeneity in the capability to process informational redundancies predicts both the magnitude of price distortions across markets and individual trading behavior.

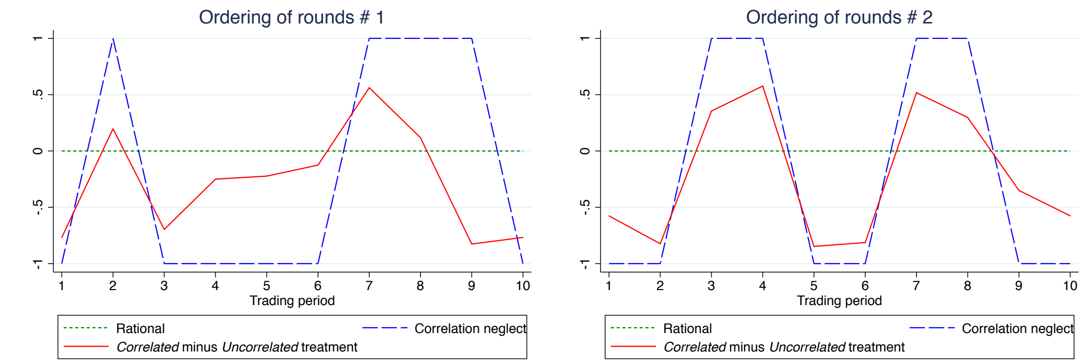


Figure 2.4. Difference between median normalized market prices in the *Correlated* and *Uncorrelated* treatments across trading rounds for the two randomized orders of rounds

[25] Formally, the new set of normalized prices p_i^j is given by $p_i^j = \chi_i^j \times (2 \times 1_{s_1^j > \bar{s}_{-1}^j} - 1)$.

Result 3. *In the Correlated market treatment, the pervasiveness of the belief bias within a market group predicts the degree of price distortions. Additionally, correlation neglect is reflected in individual trading behavior. When ignoring correlations predicts an upward (downward) biased belief, subjects with a higher propensity to overlook correlations hold significantly more (less) assets. Consequently, these subjects earn lower profits.*

The higher the degree of naïveté of the *marginal* traders in a market group, the more pronounced should be the resulting price distortion (see Appendix 2.E.1). Thus, if it is indeed correlation neglect which causes the alternating pattern of over- and underpricing, then market groups in which people are more capable of dealing with correlations should exhibit smaller price distortions. To investigate this issue, we normalize all market prices in the *Correlated market* treatment according to equation (2.3) such that they capture the size of the price distortion and then, for each trading round, relate these price levels to the naïveté which is implicit in the beliefs that subjects stated before trading started. Specifically, we employ as explanatory variable the (average) naïveté of the marginal traders, for each market group and trading round.[26] Columns (1) and (2) of Table 2.4 provide corresponding OLS estimates, with standard errors clustered at the market group level. The results show that, within the *Correlated market* treatment, a higher propensity to commit correlation neglect is indeed associated with more biased price levels.

Thus, individual-level heterogeneity in belief updating has implications for price levels. However, correlation neglect also makes clear predictions about who should hold the assets and make losses. In trading rounds in which correlation neglect leads to an overvaluation of assets, subjects who ignore correlations should own most of the assets. Likewise, when correlation neglect implies an undervaluation of assets, subjects who correctly process the correlation should hold the majority of the assets. To examine these predictions, we relate asset holdings to individual beliefs. For each individual, we employ the median naïveté parameter as explanatory variable. The OLS regressions in columns (3) through (6) establish that the magnitude of the belief bias predicts asset holdings. Columns (3) and (4) show that in trading rounds in which correlation neglect leads to an overly pessimistic belief, those subjects with a higher propensity to ignore correlations hold significantly less assets. Likewise, when the bias implies overoptimism, those subjects whose stated beliefs reveal a higher degree of correlation neglect hold more assets (columns (5) and (6)). Thus, naïve subjects buy when prices are too high and sell when they are too low. In consequence, these participants earn lower profits (columns (7) and (8)).

[26] To this end, as we detail in Appendix 2.E.1, we construct supply and demand curves from the beliefs subjects stated ex ante. We then approximate the theoretical competitive equilibrium price by identifying the buyer and seller who marginally give rise to trade and compute the average naïveté of these two traders. The results are robust to employing the simple median naïveté across all traders in a given market group and trading round as independent variable.

Table 2.4. Determinants of prices, asset holdings, and profits in the *Correlated market* treatment

	Dependent variable:							
	Normalized price distortion		Median asset holdings if underpricing		Median asset holdings if overpricing		Median profit	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Naïveté of marginal traders (χ)	0.72*** (0.12)	0.65*** (0.14)						
Individual median naïveté (χ)			-1.53*** (0.17)	-1.30*** (0.19)	0.64*** (0.12)	0.26* (0.14)	-0.12** (0.05)	-0.11** (0.05)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	152	152	143	143	143	143	143	143
R^2	0.28	0.41	0.31	0.42	0.20	0.43	0.04	0.13

OLS estimates, standard errors clustered at the market group level. In columns (1) and (2), observations include all (normalized) prices from *Correlated* excluding outliers for which the (absolute) normalized price or the naïveté of the marginal trader are larger than three. The results are robust to including these outliers when employing median regressions. See Appendix 2.E.1 for a definition of the marginal traders. Additional controls in (1)-(2) include fixed effects for each true state and the average age, average monthly disposable income, and average final high school grade as well as the proportion of females in a given market group. In columns (3) - (8), observations include median asset holdings / profits of all subjects in the *Correlated* treatment. Overpricing (underpricing) is defined as rounds in which correlation neglect predicts overoptimism (-pessimism). Median profits are computed as median normalized profit across all rounds, where for each trader and for each round a normalized profit is defined as $\pi = 10 \times \frac{\text{Money holdings + value of assets held}}{\text{Monetary value of endowment}}$, where for sellers (buyers) the value of the endowment consists of the value of the initially owned assets (the budget). The individual-level median correlation neglect parameter in (3) and (4) [(5) and (6)] is computed as median χ of the rounds in which correlation neglect predicts overpessimism [overoptimism]. In (7) and (8), the median correlation neglect parameter equals the median χ across all rounds. Additional controls in (3) - (8) include a buyer dummy, age, gender, monthly disposable income, marital status dummies, and high school GPA. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.3 A General “Face Value” Heuristic?

We have shown that many subjects employ a simplifying heuristic and often fully neglect the informational redundancies present in our environment. A possible, though perhaps extreme, conjecture is that these subjects never think through the process generating their information. Instead, they may take the visible and salient messages at “face value”, meaning that they treat each number as if it were an unmanipulated independent signal realization, *regardless of whether the signals are correlated or distorted in other ways* (see, e.g., the recent literature on the “sampling approach” towards judgment biases in cognitive psychology or the “system neglect” hypotheses articulated by Fiedler and Juslin, 2006; Massey and Wu, 2005). If true, this would imply that the updating error documented in Section 2.2 is inherently unrelated to correlations as such, but rather a special case of a rather simplistic heuristic. Based on these considerations, we now investigate the limits of such neglect patterns, i.e., we seek to understand whether people neglect signal distortions of *any* kind.[27]

If a general face value bias was at work in our experimental environment, people should also make mistakes in all other settings in which they receive distorted signals. We hence investigate the empirical validity of the face value explanation by introducing two further treatment variations, in which the source of the distortion is not (just) a correlation. Key idea behind both designs is to introduce a simple *external* distortion of the signals, i.e., a distortion which does not arise from the interplay of various signals, but rather from the intervention of some external source. According to a simple face value

[27] Evidently, the goal of this exercise is not to claim that people *only* fall prey to correlation neglect.

heuristic, these environments should also produce a particular pattern of biased beliefs. First, we designed treatment *Multiply*, which was identical to the baseline *Uncorrelated* condition, except that each of the three intermediaries obtained one of the true signals, and multiplied it by 1.5. Thus, subjects received messages $(s_1, s_2 \times 1.5, s_3 \times 1.5, s_4 \times 1.5)$. Note that, across tasks, the signal of computer A is well within the range of the distorted messages, just like in the *Correlated* treatment. If subjects take all information they see at face value, this treatment should produce biased beliefs, hence allowing for a first assessment of the empirical validity of face value bias. We implemented the same true states, signals, and procedures as in the baseline conditions. 46 subjects participated in this treatment and earned an average of 11.70 euros.

In a second treatment variation (*Face value*), we created an information environment in which (i) the rational benchmark belief coincides with that in the *Uncorrelated* treatment, (ii) correlation neglect predicts the same beliefs as in the *Correlated* condition, and (iii) the correlation neglect and face value predictions do not coincide. Specifically, as depicted in Figure 2.5, we amended the baseline *Correlated* treatment by introducing three further "machines" which communicated with subjects. Computers A through D generated four unbiased iid signals, and the intermediaries 1-3 again took the average of the respective signals of the computers. The machines M1 through M3 each observed one of these averages, and added a *known* constant X ("noise"). Thus, subjects' decision screens contained the signal of computer A as well as the messages of the three machines. In addition, the written instructions included a table in which X was provided, separately for each task. In the instructions, the machines were described in a manner that was comparable to how we introduced the intermediaries, and we made it clear that the value of X was unrelated to the solution of the task. In this treatment, both the rational and the full correlation neglect predictions are identical to those in the baseline conditions. By tailoring X, the face value prediction can be constructed to take on any desired value. In five of the tasks, we chose X such that the face value prediction is *equal to the rational belief*, i.e., the average of the independent signals. Thus, in these tasks, behaving "rationally" is computationally very simple and can be achieved by either taking messages at

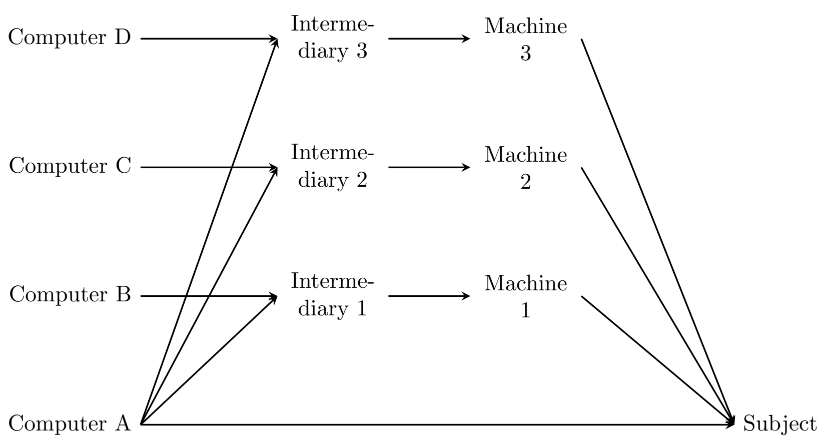


Figure 2.5. Treatment *Face value*. The machines add X to the reports of the intermediaries.

face value or going through the full debiasing process. On the other hand, neglecting correlations alone requires subjects to subtract X from the messages of the machines and then stop in further debiasing the messages. In the other five tasks, we chose X such that – after normalizing beliefs – the face value prediction was exactly opposite to the correlation neglect prediction, relative to the rational benchmark. For example, if the signal of computer A was relatively high, so that correlation neglect predicts an inflated belief, X assumed a negative value such that face value predicts a normalized belief of (-1). We implemented the same true states, signals, and procedures as in the baseline conditions, so that this treatment allows for a sharp separation between correlation neglect and a face value heuristic. 45 subjects participated in *Face value* and earned 8.10 euros on average.

Result 4. *Across contexts, face value bias explains a negligible fraction of beliefs.*

The results from both treatments indicate that subjects do not take all information at face value without reflecting upon the data-generating process. As we discuss in detail in Appendix 2.F.5, virtually all subjects behave fully rational in treatment *Multiply*, suggesting that subjects attend to and are capable of correcting for the biased messages.

A similar picture emerges for treatment *Face value*, see Appendix 2.F.6. Here, the distribution of beliefs is very similar to the baseline *Correlated* condition, suggesting that subjects again fall prey to correlation neglect, but not to face value bias. For instance, we cannot reject the hypothesis that beliefs in *Face Value* do not differ from those in the *Correlated* condition ($p = 0.3670$). In addition, beliefs in *Face value* clearly differ from both beliefs in the *Uncorrelated* treatment ($p = 0.0086$, Wilcoxon ranksum test) and the respective “face value” predictions. This implies that subjects again detect and correct for the external distortion introduced through the machines, but then stop in further debiasing the (still correlated) messages. Thus, we identify evidence for correlation neglect even when it makes a prediction different from face value bias.

In sum, we have shown that - unlike a simplistic face value bias would prescribe - people struggle considerably more with distortions that arise from the interdependence of multiple signals than with externally biased messages. Of course, these findings do not imply that correlations are the *only* type of complexity that induce people to make systematic errors. However, they show that rather simple distortions of signals such as adding or multiplying a constant do not suffice to lead people astray. One possible interpretation of these results is that correlations are more complex and less intuitively wrong than more simple signal distortions.[28]

[28] In Appendix 2.F.7, we further investigate the relevance of face value bias in our setup from a different angle, using two additional treatment variations. These treatments build on the idea underlying face value bias, namely the notion that people do not attend to the process generating the data and instead excessively focus on the visible messages. According to this logic, exogenous measures to steer attention towards the underlying process should mitigate the bias. We implemented two treatments in which we attempt to shift subjects’ focus on the information structure (but not on the correlation as such) using two nudges. Our findings reveal that both nudges were rather ineffective in mitigating correlation neglect. This provides further suggestive evidence that face value bias is an unlikely driver of correlation neglect.

2.4 The Mechanisms Underlying Correlation Neglect

This section investigates the mechanisms underlying correlation neglect. This is important for at least two reasons. First, regarding theory, studying cognitive underpinnings may prove valuable in supporting efforts to formalize the bias. Second, for applied work, one may wish to understand how the neglect of informational redundancies depends on features of the environment such as the stake size, the degree of mathematical complexity, or how salient the existence of the double-counting problem is, in order to derive predictions in which type of environments correlation neglect is (less) likely to occur and which type of interventions are likely to mitigate the bias. Likewise, the strong heterogeneity in subjects' tendency to neglect correlations may be systematically related to individual characteristics, hence allowing predictions which sub-groups of the population are more likely to suffer from the consequences of boundedly rational belief formation.

2.4.1 The Role of Complexity and Cognitive Skills

A common theme in the literature is that the degree of complexity of the problem exerts a substantial effect on the existence and magnitude of cognitive biases (e.g., Charness and Levin, 2009). To investigate the role of complexity in our setup, we implemented a new set of treatments in which we manipulated the overall complexity of the information structure while keeping the nature of the correlation constant. In our reduced complexity treatments, only two computers (A and B) generated unbiased iid signals, see Figure 2.6. In the *Uncorrelated* treatment, the only intermediary directly transmitted the signal of computer B. In the *Correlated* treatment, the intermediary reported the average of the signals of computers A and B. Thus, the type of correlation is identical to the baseline condition and requires the same conceptual understanding of double-counting, yet the complexity of the environment is severely reduced. We implemented the same ten belief formation tasks as in the baseline treatments using the same incentive structure, instructions and procedures. In total, 94 subjects participated in these treatments, which lasted 80 minutes on average and yielded average earnings of 11.60 euros.

Result 5. *An extreme reduction in the environment's complexity mitigates the bias.*

Consistent with previous documentations of the role of complexity in different contexts, we find that correlation neglect is severely reduced in our low complexity treatments. In none of the ten tasks do we find statistically significant evidence for double-counting.[29] This finding is noteworthy because it suggests that in (admittedly extremely)

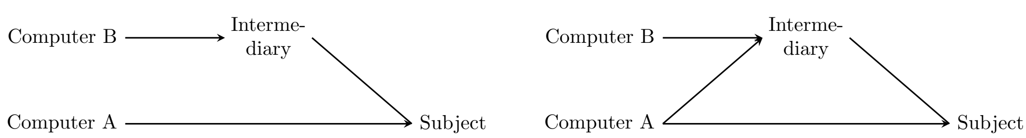


Figure 2.6. Simple uncorrelated (left panel) and correlated (right panel) information structure

[29] Appendix 2.F.1 provides a full analysis of these reduced complexity treatments.

simple informational environments subjects do grasp the implications of correlated information structures.[30]

We proceed by establishing the importance of (low) cognitive ability for correlation neglect. Table 2.5 presents the results of OLS regressions of each subject's median naïveté parameter from the baseline *Correlated* treatment from Section 2 on two proxies for cognitive ability, scholastic achievement in high school and the test score on a Raven matrices IQ test. Results show that falling prey to double-counting is significantly related to low cognitive skills.

In sum, it appears as if low cognitive skills in combination with a sufficient degree of complexity are crucial inputs into generating the updating bias. In the remainder of this section, we seek to develop a more specific understanding of how the combination of high complexity and low cognitive skills produces correlation neglect. To address this issue in a systematic manner, we conceptualize the process of belief formation in a simplified way. Intuitively, solving our more complex experimental task requires people to complete two sequential steps of reasoning, each of which potentially pertains to a conceptually distinct aspect of how cognitive skills matter in our environment:

1. Subjects need to *identify* and think through the problematic feature of our updating environment. That is, they need to notice that the workings of the intermediaries introduce a double-counting problem that they need to take care of. After all, it may not be a priori clear to participants which part of the problem they need to focus on and think through in detail.

2. Subjects need to actually *solve* the problem mathematically, i.e., conditional on noticing and understanding the problem, they ought to execute the computations that are necessary to debias the messages of the intermediaries.

While such a procedural view of the belief formation process is obviously stylized, it will nevertheless prove useful in further developing and empirically assessing several

Table 2.5. Correlation neglect and cognitive skills

	Dependent variable: Median naiveté χ			
	(1)	(2)	(3)	(4)
High school grade point average	-0.24**		-0.25**	-0.29**
	(0.10)		(0.10)	(0.11)
Raven test score		-0.10***	-0.11**	-0.11**
		(0.04)	(0.04)	(0.04)
Additional controls	No	No	No	Yes
Observations	47	47	47	46
R^2	0.10	0.10	0.22	0.27

OLS estimates, robust standard errors in parantheses. Observations include all subjects from the baseline *Correlated* treatment. Additional controls include age, gender, monthly disposable income, and marital status dummies. High school GPA 1 (worst) - 5 (best). Raven test score 0-10. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[30] Note, however, that this context is very simplistic: Since we did not induce priors, the report of the intermediary in the correlated treatment equals the rational belief, rendering actual computations by the subjects unnecessary.

competing explanations for correlation neglect that can account for the important role of cognitive ability and complexity.

2.4.2 Solving the Problem Mathematically

Suppose for now that people do not struggle with the first step, i.e., they think through the mechanics that generate the correlation and detect the resulting double-counting problem. Then, subjects still need to execute the computations that are necessary to develop rational beliefs. However, two issues may prevent them from actually doing so and hence drive the observed neglect of correlations. First, subjects may lack the mathematical skills needed to invert the averages computed by the intermediaries. Second, even if participants could in principle solve the problem, they may incur thinking costs in doing the necessary calculations.[31] Both of these potential channels would account for the importance of cognitive ability and complexity in a straightforward way; for instance, the level of mathematical skills or the effort cost function likely depend on cognitive ability. Likewise, higher complexity requires higher levels of mathematical skills.[32]

Note that both of these channels rest on the presumption that subjects know and understand that they need to compute the average of the four signals of computers A-D to develop rational beliefs. To evaluate the empirical validity of this hypothesis, we introduced treatment *Math*. In this treatment variation, we altered the instructions relative to the *Correlated* treatment by explicitly advising subjects to back out the underlying independent signals from the correlated messages.[33] In essence, this treatment solves the first step of the belief formation process outlined above. Thus, any remaining systematic mistake can be attributed to either cognitive effort costs or mathematical problems in executing the calculations. 47 subjects took part in this treatment and earned an average of 11.40 euros.

Result 6. *Provided that subjects know how to solve the problem, a large majority are both willing to and capable of executing the necessary calculations.*

Appendix 2.F.2 provides a detailed analysis of treatment *Math*. To summarize, the vast majority of subjects states rational beliefs once they know how to solve the problem. For instance, the (median) naïveté parameter of the median individual in this treatment is $\chi = 0.00$, down from $\chi = 0.68$ in *Correlated*. Formally, the distribution of median naïveté parameters in this treatment is significantly different from that in the *Correlated* treatment ($p = 0.0003$) and does not significantly differ from that in the *Uncorrelated* condition ($p = 0.7593$, Wilcoxon ranksum tests). Thus, while a small fraction of our subjects appear to struggle with the mere task of computing the average signal of the computers and state fully naïve beliefs, low mathematical skills or prohibitively high

[31] The idea that the processing of information is associated with thinking costs can be traced back to Simon (1956) and has been formalized in different models (see, e.g., Caplin et al., 2011; Caplin and Dean, forthcoming; Gabaix et al., 2011).

[32] In fact, as discussed above, in the low complexity treatment, actual computations are unnecessary since the intermediary directly reports the rational belief.

[33] For instance, the instructions stated: "Important hint: . . . You should attempt to determine the average of the signals of the computers." We also introduced a corresponding control question, see Appendix H for details.

effort costs in executing the necessary calculations are unlikely drivers of correlation neglect for the majority of subjects.

2.4.3 Identifying and Thinking Through the Problem

The previous results suggest that many subjects struggle more with identifying and thinking through the critical aspect of our updating problem than with its mathematical solution per se (i.e., with the first step of the two-step belief formation process outlined above). In particular, identifying the double-counting feature may work as a threshold which subjects do or do not pass, giving rise to a bimodal type distribution. After all, if subjects do not identify the double-counting issue in the first place, they cannot solve this problem mathematically. Based on this logic, we proceed by investigating whether people become better at processing correlated signals once they are (exogenously) induced to focus on the double-counting problem. Key idea behind the corresponding treatment variations – relative to treatment *Math* – is to directly increase subjects' focus on the correlated nature of the signals (i.e., the workings and implications of the intermediaries) without providing any additional information on the double-counting problem or its mathematical solution. That is, we explicitly alert subjects *what* to think about, but not *how*.

To this end, we introduced two variations of the baseline *Correlated* treatment. The first treatment (*Intermediaries*) was inspired by the evidence in Hanna et al. (2014) who show that people become better at optimizing behavior once they are induced to focus on previously overlooked dimensions of a decision problem. To shift subjects' focus while forming beliefs, we conducted a treatment variation that is identical to the baseline *Correlated* condition except for one additional short paragraph which was provided both at the end of the instructions and on subjects' decision screens along with the graphical representation of the information structure (see Figure 2.1):

> ***Hint for solving the task:*** *Again consider the figure which depicts the information you will receive. Think carefully about what the intermediaries do! What does that imply for the estimates of the intermediaries?*

Note that this constitutes a rather strong intervention in the sense that we explicitly told subjects what to focus on when approaching the task. However, the paragraph did not provide any additional information on how to solve the problem and compute rational beliefs. Subjects completed the same ten belief formation tasks as in the baseline *Correlated* condition.

In a second treatment (*Alternating*), we nudged our participants by varying the nature of the information structure (correlated or uncorrelated) within subjects between tasks. This allowed us to alert subjects to the workings and implications of the intermediaries in a more indirect manner. The instructions for this treatment introduced both the correlated and the uncorrelated information structure from our baseline design, which were framed as “Scenario I” and “Scenario II”, respectively. Subjects were told that in some tasks they would receive information according to Scenario I and in some tasks according to Scenario II and that, in each task, they would be informed of the scenario before seeing the messages of computer A and of the intermediaries. Consequently, subjects solved five tasks with correlated and five with uncorrelated information. In the instructions, we emphasized to subjects that they would have to pay special attention to the prevailing scenario and the corresponding change in the intermediaries' behavior. In addition, the control questions in this treatment required subjects to compute the messages of intermediaries 1 and 2 for exemplary computer signals for both the correlated and the uncorrelated scenario, which presumably further increased the salience of the intermediaries. 46 (47) subjects took part in the *Intermediaries* (*Alternating*) treatment and earned 12.70 (13.10) euros on average.

Result 7. *Exogenously increasing subjects' focus on the correlation reduces the bias.*

To illustrate, Figure 2.7 visualizes the distribution of median beliefs across the ten tasks in the *Intermediaries* treatment, again plotted against the relative magnitude of the common source signal. As visual inspection suggests, median beliefs are very close or often identical to those in the *Uncorrelated* condition, and clearly differ from those in *Correlated*. As Appendix 2.F.3 visualizes, very similar results obtain in *Alternating*. Appendix 2.F.3 provides a complete analysis of these treatments and shows that in 50 % of all tasks, beliefs in the nudge treatments significantly differ from those in *Correlated* at the 5 % level (Wilcoxon ranksum tests).

A different way to grasp this pattern is to consider the previously identified type heterogeneity at the individual level, i.e., to aggregate the data across tasks at the individual level, rather than across individuals for each task. To this end, Figure 2.8 plots kernel density estimates of the median naïveté parameters for both additional treatments. The median subject in these two treatments exhibits a naïveté of only $\chi = 0.09$ (*Intermedi-*

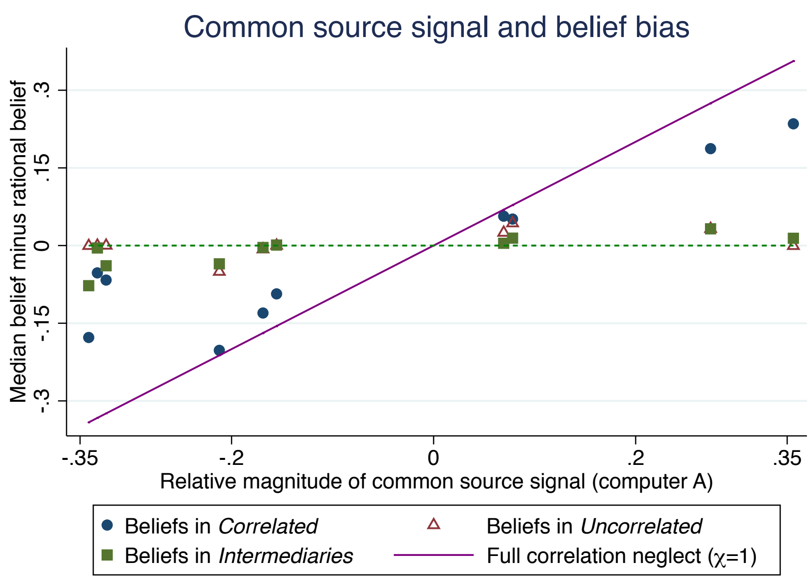


Figure 2.7. Beliefs in the *Correlated*, *Uncorrelated* and *Intermediaries* treatments plotted against the relative magnitude of the signal of computer A. See the notes of Figure 2.2 for details on the construction of this figure.

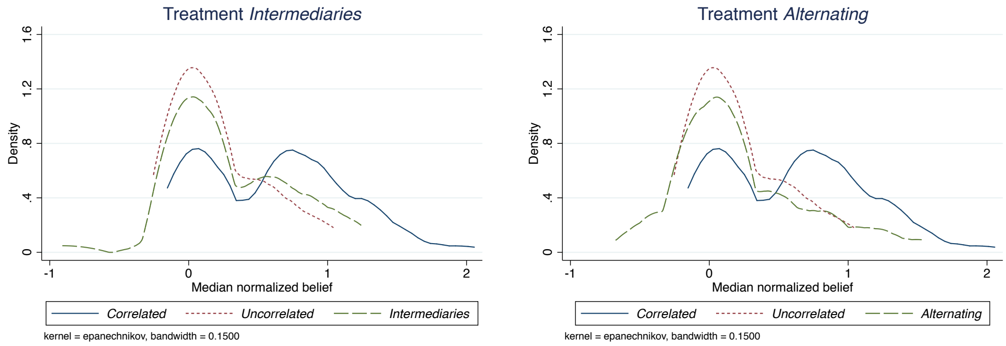


Figure 2.8. Kernel density estimates of median normalized beliefs in the *Intermediaries* treatment (median of ten tasks) and the *Alternating* treatment (median of five tasks), each compared with median beliefs in the baseline *Correlated* and *Uncorrelated* treatments

aries) and $\chi = 0.03$ (*Alternating*), respectively. The parameter distributions are centered significantly closer to the rational level and are clearly distinguishable from the *Correlated* condition (p = 0.0023 for *Intermediaries* and p = 0.0230 for *Alternating*, Wilcoxon ranksum tests). In addition, beliefs do not statistically differ from those in the *Uncorrelated* condition (p = 0.2906 for *Intermediaries* and p = 0.1361 for *Alternating*).

In sum, if subjects are nudged to focus on the critical feature of the informational environment, the bias is substantially reduced. Notably, most subjects do not adjust partially, but rather develop fully unbiased beliefs. These findings are consistent with our results from treatment *Math*: once subjects focus on thinking about the double-counting problem, they possess the mathematical skills to solve our experimental belief formation task. In combination, these results lend support to the idea that the first step of our simple two-step belief formation process may act as a threshold towards developing rational beliefs, and hence give rise to a bimodal type distribution. In addition, these results are also consistent with the relationship between correlation neglect and complexity as well as cognitive skills. After all, subjects may have more problems in identifying the problematic feature of the updating environment when the problem is more complex; likewise, subjects with high cognitive skills may find it easier to focus on and think through the double-counting problem.

A possible conjecture is that the results on the relationship between beliefs and nudges reflect cognitive effort costs: reflecting upon the information structure and identifying the double-counting problem may be cognitively costly. While the above results show that effort costs do not prevent participants from executing the necessary calculations, they may induce subjects to refrain from even thinking about what the correct solution may be, implying that subjects remain “rationally” inattentive towards the double-counting problem, akin to rational inattention behavior established in, e.g., Caplin et al. (2011) and Caplin and Dean (forthcoming).

We evaluate the explanatory power of this rational ignorance hypothesis by making use of its straightforward and testable implication that an increase in the marginal financial incentives to hold correct beliefs should increase cognitive effort and hence reduce the amount of correlation neglect. Accordingly, we triple both the absolute and the marginal level of the financial incentives in the *Correlated* and *Uncorrelated* treatments.[34] Apart from the increase in stake size, these treatments were identical to the baseline *Correlated* and *Uncorrelated* treatments, respectively. 94 subjects participated in these experiments, which lasted 90 minutes on average and yielded average earnings of 21.90 euros.

Result 8. *In our experiments, a moderate increase in financial incentives affects cognitive effort, but not subjects' tendency to disregard correlations.*

Support for this claim is provided by Table 2.6. Columns (1)-(3) show the results of a difference-in-difference OLS estimation of each subject's median naïveté parameter on (i) a treatment dummy, (ii) a stake size dummy, and (iii) an interaction term equal to one if subjects were in the high-stakes *Correlated* treatment. If the increase of the stake size by 200% lead to more accurate beliefs, then this interaction term should have a negative coefficient. However, the point estimate is actually slightly positive, and despite the relatively large sample size, the only sizable and significant effect is the treatment difference, which is robust to increasing the (marginal) financial incentives. To further illustrate this result, Appendix 2.F.4 shows that the distribution of naïveté again exhibits a roughly bimodal structure according to which many subjects essentially fully neglect correlations.[35]

While the higher stake size does not induce more belief accuracy, it does affect cognitive effort, as proxied for by response times. Relative to the baseline conditions, subjects take on average more than 20% longer to solve each task, which indicates that they indeed provide higher effort when being confronted with higher stakes (columns (5) and (6)). However, this higher effort level does not translate into more accurate beliefs. This is noteworthy as column (4) shows that, within the *Correlated* treatments, higher effort (as proxied by higher response times) is indeed associated with higher belief accuracy. Columns (7) and (8) of Table 2.6 contrast these findings with the response time patterns in treatments *Intermediaries* and *Alternating*. Results show that these nudge treatments have a large positive effect on response times, which increase by almost one minute on average.

2.4.4 Discussion

People do not always neglect correlations, but only when the updating problem is sufficiently complex. Starting with the strong relationship between correlation neglect and cognitive skills in such complex environments, we have decomposed the cognitive bias using a stylized two-step process of belief formation. We have seen that – at least in the context considered in this paper – people are both willing to and mathematically capable of executing the calculations that are necessary to debias correlated messages. What is more, treatments *Intermediaries* and *Alternating* have highlighted that people do not

[34] In these high-stakes conditions, variable earnings in euros were given by $\pi = max\{0, 30 - 480 \times (\text{Belief / True state} - 1)^2\}$.

[35] Unreported regressions confirm that all results on the relationship between stake size, response times, and beliefs hold if we do not consider the median normalized belief of each subject, but instead all beliefs, i.e., ten observations per subject.

Table 2.6. Correlation neglect, stake size, and response times

	Dependent variable:							
	Median χ				Median response time			
	Corr. + Uncorr.			Correlated	Corr. + Uncorr.		Corr. + nudge	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 if correlated	0.41***	0.39***	0.40***		0.49***	0.45***		
	(0.07)	(0.09)	(0.09)		(0.13)	(0.14)		
1 if high stakes		-0.046	-0.029	0.074	0.25*	0.29**		
		(0.07)	(0.07)	(0.10)	(0.13)	(0.13)		
1 if correlated high stakes		0.029	0.012					
		(0.13)	(0.13)					
Median response time				-0.21***				
				(0.07)				
1 if *Intermediaries*, 0 if *Baseline corr.*							0.94***	
							(0.22)	
1 if *Alternating*, 0 if *Baseline corr.*								0.95***
								(0.22)
Constant	0.20***	0.23***	-0.0060	0.33	0.94***	1.12**	1.52	2.53***
	(0.04)	(0.05)	(0.24)	(0.50)	(0.09)	(0.54)	(1.10)	(0.82)
Additional controls	No	No	Yes	Yes	No	Yes	Yes	Yes
Observations	188	188	186	92	188	186	92	93
R^2	0.17	0.17	0.19	0.21	0.08	0.13	0.22	0.23

OLS estimates, robust standard errors in parentheses. In columns (1)-(3), the dependent variable consists of median naïveté parameters from all subjects in the baseline and the high stakes treatments (both *Correlated* and *Uncorrelated*). In column (4), the sample is restricted to subjects in the *Correlated* conditions, both high stakes and baseline. In columns (5)-(6), the dependent variable is the median response time of all subjects in the baseline and high stakes conditions. In columns (7) and (8), the sample consists of subjects in the baseline correlated condition and the respective nudge treatment. Response time in minutes. Additional controls include age, gender, monthly disposable income, and marital status fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

even need to be told *how* to solve the problem. Rather, exogenously inducing them to think about the problematic aspect of the updating environment already has large effects on beliefs.

Bounded rationality has often been considered as a continuous concept. In contrast, in our context, identifying and thinking through the double-counting problem appears to constitute a threshold which people do or do not pass, resulting in a somewhat (discrete) bimodal distribution of types (see Gabaix, 2014, for a model in which limited attention acts as a binary threshold). An interesting question is whether the (non-) passing of this threshold results from costs of thinking ("rational ignorance") or whether subjects attempt to, but do not succeed in, devising an appropriate problem-solving strategy. While one could argue that studying belief biases in a controlled laboratory context comes at the cost of relatively small financial incentives, the response time patterns nevertheless provide suggestive evidence that exogenously increasing effort through moderate increases in incentives may change behavior along the intensive, but not along the extensive margin: in our experiments, higher stakes induce subjects to invest higher effort, yet people appear to not alter their problem-solving strategy as such. After all, in the high stakes treatments, the distribution of beliefs also has a mass point at full naïveté. In contrast, shifting subjects' focus on the correlation has large effects on beliefs. A plausible interpretation of these findings is that – if left to their own devices – subjects attempt to identify the critical aspect of the informational environment (i.e., to solve the first step of the simple two-step belief formation process outlined above), and do so harder when the stakes are higher. However, if they do not succeed in passing this threshold, they make use of a specific simple heuristic. On the other hand, once people are told which aspect of the problem they need to consider in detail, they pass the threshold of step 1 and subsequently take considerably longer to solve each task, because they need to go through the additional mathematical steps of debiasing the correlated messages.

Our findings on the effects of nudges in debiasing subjects lend themselves to a natural interpretation in terms of limited and selective attention: if decision-makers face limits in the level of attention they can allocate to the different features of the task, they may lack focus on important aspects of the data-generating process. In our context, subjects could in principle focus on a variety of features of the environment, such as the nature of the distribution generating the signals, the relative signal precisions, the payment scheme etc. Consequently, attention may be lacking on the precise workings of the intermediaries and corresponding implications, so that drawing subjects' attention towards the mechanics which generate the correlation should attenuate correlation neglect. This selective attention interpretation bears a natural relationship with a small recent theoretical literature which models the idea that, in forming beliefs, people may naturally attend to some aspects of the problem, but not to others (Gennaioli and Shleifer, 2010; Bordalo et al., 2015b; Schwartzstein, 2014). However, as they are, these theories do not posit specific attentional frames in processing correlations.[36]

2.5 Concluding Remarks

Using experiments with more than 1,000 subjects, this paper provides clean evidence for people's tendency to neglect correlations in information sources when forming beliefs and the corresponding cognitive mechanisms. While we deliberately designed a tightly controlled and abstract information structure to obtain a clean view on the cognitive bias and corresponding remedies, an interesting question is whether correlation neglect persists in more natural informational environments. While studying belief formation using naturalistic information naturally comes at the loss of some internal validity, in Appendix 2.G, we explore one possible avenue by investigating subjects' behavior when they are confronted with real newspaper reports covering correlated information. To this end, we make use of a naturally occurring informational redundancy in professional GDP forecasts that arose because a German research institute contributed to a joint forecast, but also issued a separate (different) forecast at the same time. Again, the (incentivized) beliefs subjects state when they are confronted with these correlated forecasts are consistent with the neglect of informational redundancies, hence suggesting that the bias we identify in this paper also plays out in more naturalistic environments.

Economists have recently increased their efforts to explicitly model erroneous probability judgments (see, e.g., the discussion in Rabin, 2013). While most of the literature has focused on formalizing specific biases and drawing out corresponding economic implications (Rabin and Schrag, 1999; Rabin, 2002; Rabin and Vayanos, 2010; Benjamin et al., forthcoming), more recently economists have started to model the mental process of belief formation (Gennaioli and Shleifer, 2010; Bordalo et al., 2015b; Schwartzstein,

[36] Also see, e.g., Bordalo et al. (2013), Bordalo et al. (2015a), Taubinsky (2014), Kőszegi and Szeidl (2013), and Gabaix (2014) for the application of limited attention to consumer choice.

2014). While none of these theories are designed to apply in the settings we considered, our empirical results are broadly supportive of this type of models in that we emphasize the interplay of complexity and focus in generating correlation neglect. An interesting question is which other prevalent and economically important features of real information structures induce the neglect patterns we document in this paper, and how the resulting biases are conceptually linked to correlation neglect. As our "face value" treatments have shown, the tendency to naïvely process distorted signals is not universal across contexts.

References

Akerlof, George A. and Robert J. Shiller (2009): *Animal spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism*. Princeton University Press. [8, 9]

Andreoni, James and Tymofiy Mylovanov (2012): "Diverging Opinions." *American Economic Journal: Microeconomics*, 4 (1), 209–232. [10]

Benjamin, Dan, Collin Raymond, and Matthew Rabin (forthcoming): "A Model of Non-Belief in the Law of Large Numbers." *Journal of the European Economic Association*. [37]

Bohren, Aislinn (2013): "Informational Herding with Model Misspecification." *Working Paper*. [11]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2013): "Salience and Consumer Choice." *Journal of Political Economy*, 121 (5), 803–843. [37]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2015a): "Memory, Attention and Choice." *Working Paper*. [37]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2015b): "Stereotypes." *Working Paper*. [10, 37]

Brandts, Jordi, Aya Ebru Giritligil, and Roberto A. Weber (2014): "An Experimental Study of Persuasion Bias and Social Influence in Networks." *Working Paper*. [11]

Brocas, Isabelle, Juan D. Carrillo, Stephanie W. Wang, and Colin F. Camerer (2014): "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games." *Review of Economic Studies*, 81 (3), 944–970. [10]

Budescu, David V. and Adrian K. Rantilla (2000): "Confidence in Aggregation of Expert Opinion." *Acta Psychologica*, 104, 371–398. [11]

Budescu, David V. and Hsiu-Ting Yu (2007): "Aggregation of Opinion Based on Correlated Cues and Advisors." *Journal of Behavioral Decision Making*, 20, 153–177. [11]

Camerer, Colin F. (1987): "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence." *American Economic Review*, 77 (5), 981–997. [8, 18]

Caplin, Andrew and Mark Dean (forthcoming): "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review*. [10, 31, 34]

Caplin, Andrew, Mark Dean, and Daniel Martin (2011): "Search and Satisficing." *American Economic Review*, 101 (7), 2899–2922. [10, 31, 34]

Chandrasekhar, Arun G., Horacio Larreguy, and Juan P. Xandri (2015): "Testing Models of Social Learning on Networks: Evidence from a Framed Field Experiment." *Working Paper*. [11]

Charness, Gary, Edi Karni, and Dan Levin (2010): "On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda." *Games and Economic Behavior*, 68 (2), 551–556. [10]

Charness, Gary and Dan Levin (2009): "The Origin of the Winner's Curse: A Laboratory Study." *American Economic Journal: Microeconomics*, 207–236. [9, 29]

Corazzini, Luca, Filippo Pavesi, Beatrice Petrovich, and Luca Stanca (2012): "Influential Listeners: An Experiment on Persuasion Bias in Social Networks." *European Economic Review*, 56, 1276–1288. [11]

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel (2003): "Persuasion Bias, Social Influence, and Unidimensional Opinions." *Quarterly Journal of Economics*, 118 (3), 909–968. [11]

Esponda, Ignacio and Emanuel Vespa (2014): "Hypothetical Thinking and Information Extraction in the Laboratory." *American Economic Journal: Microeconomics*, 6 (4), 180–202. [10]

Eyster, Erik and Matthew Rabin (2005): "Cursed Equilibrium." *Econometrica*, 73 (5), 1623–1672. [58]

Eyster, Erik and Matthew Rabin (2010): "Naïve Herding in Rich-Information Settings." *American Economic Journal: Microeconomics*, 2 (4), 221–243. [11]

Eyster, Erik and Matthew Rabin (2014): "Extensive Imitation is Irrational and Harmful." *Quarterly Journal of Economics*, 129 (4), 1861–1898. [11, 13]

Eyster, Erik, Matthew Rabin, and Dimitri Vayanos (2013): "Financial Markets Where Traders Neglect the Informational Content of Prices." *Working Paper*. [58]

Eyster, Erik and Georg Weizsäcker (2011): "Correlation Neglect in Financial Decision-Making." *Working Paper*. [11]

Fiedler, Klaus and Peter Juslin (2006): *Information Sampling and Adaptive Cognition*. Cambridge University Press. [26]

Fischbacher, Urs (2007): "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10 (2), 171–178. [18]

Gabaix, Xavier (2014): "A Sparsity-Based Model of Bounded Rationality." *Quarterly Journal of Economics*, 129 (4), 1661–1710. [36, 37]

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2011): "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." *American Economic Review*, 96 (4), 2899–2922. [31]

Ganguly, Ananda R., John H. Kagel, and Donald V. Moser (2000): "Do Asset Market Prices Reflect Traders' Judgment Biases?" *Journal of Risk and Uncertainty*, 20 (3), 219–245. [18]

Gennaioli, Nicola and Andrei Shleifer (2010): "What Comes to Mind." *Quarterly Journal of Economics*, 125 (4), 1399–1433. [10, 37]

Gneezy, Uri, Arie Kapteyn, and Jan Potters (2003): "Evaluation Periods and Asset Prices in a Market Experiment." *Journal of Finance*, 58 (2), 821–838. [8]

Golub, Benjamin and Matthew O. Jackson (2010): "Naïve Learning in Social Networks and the Wisdom of Crowds." *American Economic Journal: Microeconomics*, 2 (1), 112–149. [11]

Greiner, Ben (2004): "The Online Recruitment System ORSEE 2.0 – A Guide for the Organization of Experiments in Economics." *University of Cologne, Working paper series in economics*, 10 (23), 63–104. [18]

Grimm, Veronika and Friederike Mengel (2014): "An Experiment on Belief Formation in Networks." *Working Paper*. [11]

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein (2014): "Learning through Noticing: Theory and Experimental Evidence in Farming." *Quarterly Journal of Economics*, 129 (3), 1311–1353. [10, 32]

Hossain, Tanjim and Ryo Okui (2013): "The Binarized Scoring Rule." *Review of Economic Studies*, 80 (3), 984–1001. [22]

Kahneman, Daniel and Amos Tversky (1973): "On the Psychology of Prediction." *Psychological Review*, 80 (4), 237–251. [11]

Kallir, Ido and Doron Sonsino (2010): "The Neglect of Correlation in Allocation Decisions." *Southern Economic Journal*, 75 (4), 1045–1066. [11]

Kluger, Brian D. and Steve B. Wyatt (2004): "Are Judgment Errors Reflected in Market Prices and Allocations? Experimental Evidence Based on the Monty Hall Problem." *Journal of Finance*, 59 (3), 969–998. [18]

Kőszegi, Botond and Adam Szeidl (2013): "A Model of Focusing in Economic Choice." *Quarterly Journal of Economics*, 128 (1), 53–104. [37]

Kübler, Dorothea and Georg Weizsäcker (2005): "Are Longer Cascades More Stable?" *Journal of the European Economic Association*, 3 (2-3), 330–339. [11]

Levy, Gilat and Ronny Razin (forthcoming): "Correlation Neglect, Voting Behaviour and Polarization." *American Economic Review*. [11]

Maines, Laureen (1990): "The Effect of Forecast Redundancy on Judgments of a Consensus Forecast's Expected Accuracy." *Journal of Accounting Research*, 28, 29–47. [11]

Maines, Laureen (1996): "An Experimental Examination of Subjective Forecast Combination." *International Journal of Forecasting*, 12, 223–233. [11]

Massey, Cade and George Wu (2005): "Detecting Regime Shifts: The Causes of Under- and Overreaction." *Management Science*, 51 (6), 932–947. [26]

Möbius, Markus M., Thuan Phan, and Adam Szeidl (2013): "Treasure Hunt: Social Learning in the Field." *Working Paper*. [11]

Ortoleva, Pietro and Erik Snowberg (2015): "Overconfidence in Political Behavior." *American Economic Review*, 105 (2), 504–535. [11]

Rabin, Matthew (2002): "Inference by Believers in the Law of Small Numbers." *Quarterly Journal of Economics*, 117 (3), 775–816. [37]

Rabin, Matthew (2013): "Incorporating Limited Rationality into Economics." *Journal of Economic Literature*, 51 (2), 528–43. [37]

Rabin, Matthew and Joel L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics*, 114 (1), 37–82. [37]

Rabin, Matthew and Dimitri Vayanos (2010): "The Gambler's and Hot-Hand Fallacies: Theory and Applications." *Review of Economic Studies*, 77 (2), 730–778. [37]

Schwartzstein, Joshua (2014): "Selective Attention and Learning." *Journal of the European Economic Association*, 12 (6), 1423–1452. [10, 37]

Selten, Reinhard (1998): "Axiomatic Characterization of the Quadratic Scoring Rule." *Experimental Economics*, 1 (1), 43–62. [14]

Shiller, Robert J. (2000): *Irrational Exuberance*. Princeton University Press. [8]

Simon, Herbert A. (1956): "Rational Choice and the Structure of the Environment." *Psychological Review*, 63, 129–138. [31]

Spiegler, Ran (2015): "Bayesian Networks and Boundedly Rational Expectations." *Working Paper*. [11]

Taubinsky, Dmitry (2014): "From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice." *Working Paper*. [37]

Weizsäcker, Georg (2010): "Do we Follow Others When we Should? A Simple Test of Rational Expectations." *American Economic Review*, 100 (5), 2340–2360. [11]

Appendix 2.A Overview of Treatments

Table 2.7. Treatment overview

Treatment	# of subjects	Session length (mins)	Ave earnings (euros)
Baseline correlated	47	90	10.25
Baseline uncorrelated	47	90	12.92
Robustness correlated	48	80	9.96
Robustness uncorrelated	48	80	12.25
Market correlated	144	150	19.40
Market uncorrelated	144	150	19.33
Reduced complexity correlated	47	80	12.52
Reduced complexity uncorrelated	47	80	11.60
Math	47	90	11.40
High stakes correlated	47	90	19.17
High stakes uncorrelated	47	90	24.58
Intermediaries	46	90	12.70
Alternating	47	90	13.13
Multiply	46	90	11.70
Face value	45	90	8.10
Structure	47	90	10.58
Messages	47	90	12.86

Appendix 2.B Order of Belief Formation Tasks in Main Treatments / Trading Rounds

In all individual belief elicitation treatments we implemented three different randomized orders of rounds. These orders (by true state) are as follows:

1. 10’000, 88, 46’422, 4’698, 250, 23’112, 1’000, 10, 7’338, 732
2. 732, 23’112, 88, 1’000, 250, 4’698, 10, 7’338, 10’000, 46’422
3. 250, 7’338, 10’000, 10, 4’698, 88, 46’422, 732, 1’000, 23’112

In the market treatments, we implemented the first two of these randomizations. Neither in the individual nor in the market treatments do we find any evidence that the order of rounds matters.

Appendix 2.C Additional Analyses for Individual Baseline Treatments

2.C.1 Robustness of Results in Individual Decision Making Treatments

This section demonstrates the robustness of our results in the baseline individual treatments. First, Table 2.8 provides the p-values of ranksum tests for each of the ten belief formation tasks if we exclude all “outliers”, i.e., all observations which are not within [50 %, 150 %] of the rational belief. Figures 2.9 and 2.10 provide kernel density estimates of the beliefs in each of the ten tasks to provide a visual representation of the robustness of our results. As the ranksum tests above, these densities exclude beliefs which are not within [50 %, 150 %] of the rational belief (on average, this resulted in the exclusion of 4 out of 94 beliefs per true state).

Table 2.8. P-values of ranksum tests in the individual treatments excluding outliers

True state	10	88	250	732	1’000	4’698	7’338	10’000	23’112	46’422
p-value	0.0109	0.0038	0.0067	0.0099	0.0940	0.0096	0.9968	0.0122	0.0002	0.0261

Observations include all beliefs in the low-stakes treatment within a 50 % range around the rational belief. The p-values refer to a Wilcoxon ranksum test.

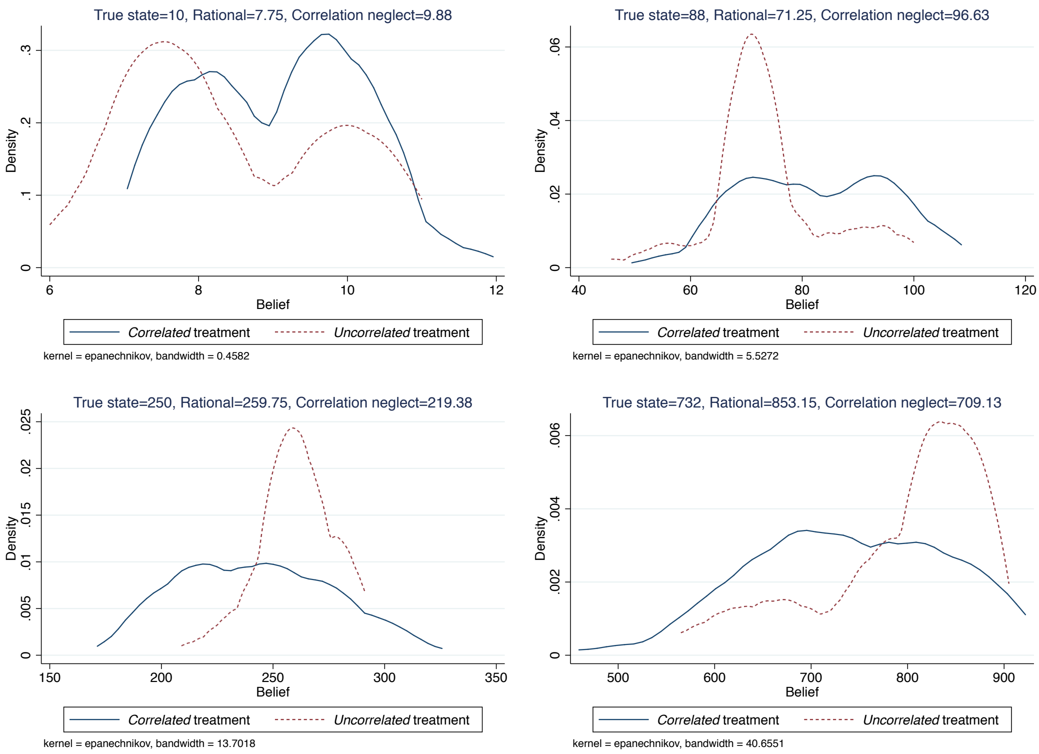


Figure 2.9. Kernel density estimates of beliefs in individual belief formation treatments (1/2)

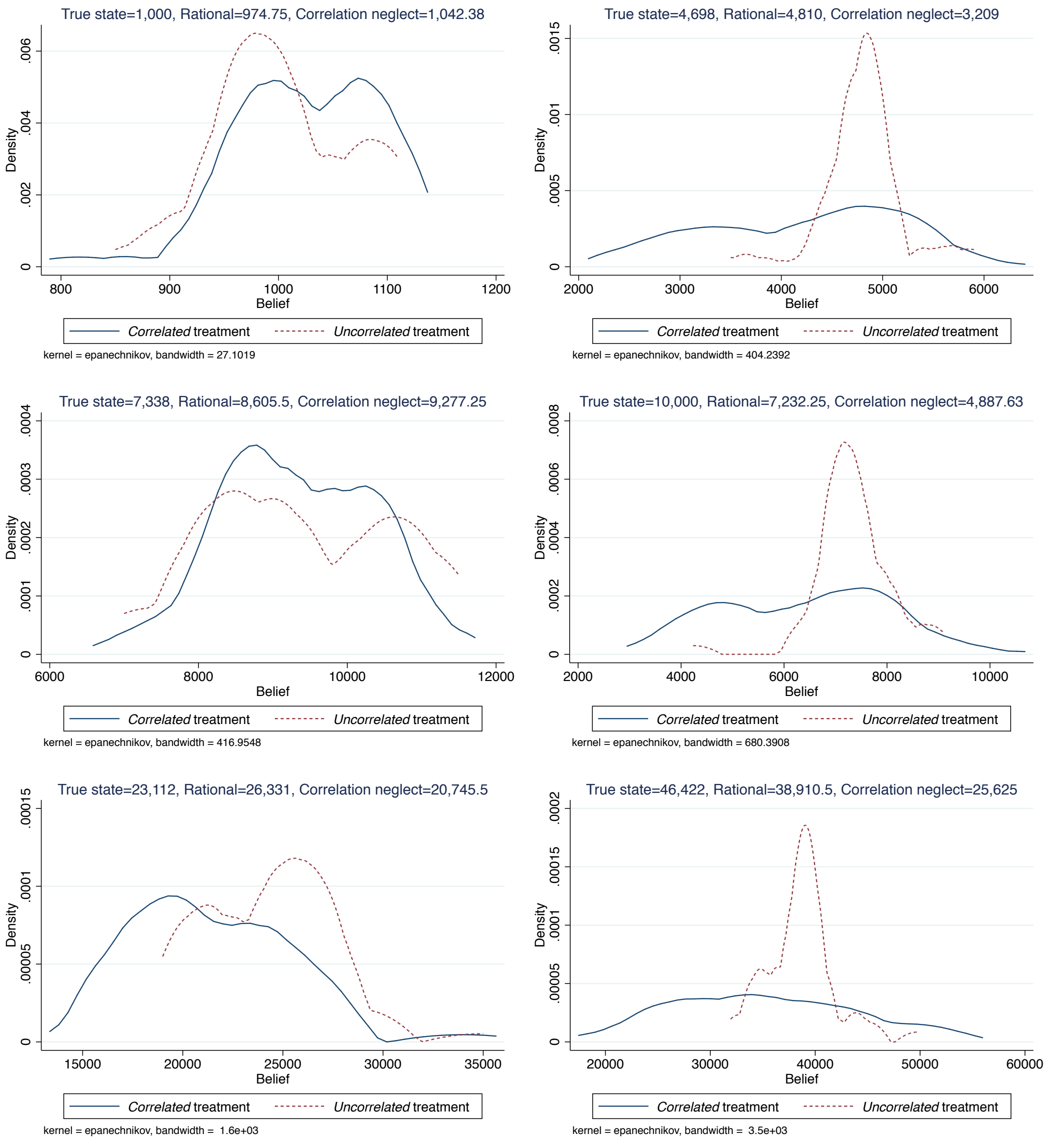


Figure 2.10. Kernel density estimates of beliefs in individual belief formation treatments (2/2)

2.C.2 Individual Treatments: Treatment Comparison and the Role of Cognitive Abilities

This section establishes that the baseline treatment difference between the *Correlated* and the *Uncorrelated* individual decision making treatments is robust to pooling beliefs across all ten tasks. To this end, Figure 2.11 plots kernel density estimates of all normalized individual beliefs in the baseline *Correlated* and *Uncorrelated* treatments, excluding 4 (out of 462) observations with $|b_i^j| > 10$. As the plots show, the disaggregated data confirm the visual impression arising from plotting the median naïveté parameters of each individual. Specifically, in the *Uncorrelated* treatment, the vast majority of beliefs is approximately rational, while those in the *Correlated* treatment tend to be either rational (normalized belief = 0) or almost fully naïve (normalized belief = 1).

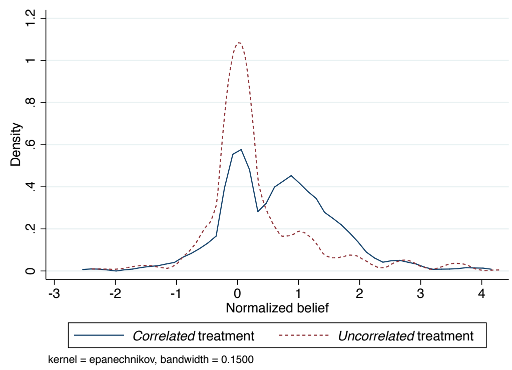


Figure 2.11. Kernel density estimates of all normalized individual beliefs

To statistically confirm this visual impression, columns (1) and (2) of Table 2.9 present the results of an OLS regression of all normalized beliefs in the *Correlated* and *Uncorrelated* baseline conditions on a treatment dummy and thereby establishes a quantitatively large amount of correlation neglect. As columns (3) and (4) indicate, however, such correlation neglect is not uniform, but significantly stronger for subjects with low cognitive skills, as proxied for by subjects' high school grades and their score on a ten-item Raven matrices IQ test.

Table 2.9. Correlation neglect and cognitive ability

	Dependent variable: Normalized belief			
	Full sample		Correlated treatment	
	(1)	(2)	(3)	(4)
1 if correlated	0.36***	0.37***		
	(0.09)	(0.09)		
High school grade point average			-0.29***	-0.33***
			(0.08)	(0.09)
Raven score			-0.100**	-0.087**
			(0.04)	(0.03)
Constant	0.32***	0.49	2.32***	2.68***
	(0.05)	(0.30)	(0.40)	(0.56)
Additional controls	No	Yes	No	Yes
Observations	924	914	458	448
R^2	0.04	0.04	0.07	0.09

OLS regressions, standard errors (clustered at individual) in parentheses. Observations in column (1) and (2) include all normalized beliefs from all rounds in the baseline treatments excluding extreme outliers with normalized belief $|b_i^j| > 10$. In columns (3) and (4), observations include all normalized beliefs from all rounds in the baseline *Correlated* treatment excluding extreme outliers with normalized belief $|b_i^j| > 10$. All results are robust to including these observations when employing median regressions. Additional controls include gender, age, marital status fixed effects, and monthly disposable income. † Scale: 1 (worst) - 5 (best). In the German system, the high school GPA ("Abitur") is a summary statistic of grades in the final years of secondary education and serves as primary university entrance criterion. ‡ Scale: 0 (worst) - 10 (best). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.C.3 Stability of (Median) Naïveté Parameters

To provide an illustration of the stability of the naïveté parameters, we conduct the following empirical exercise. For each subject, we set the belief to missing whose implied naïveté parameter is closest to that subject's median naïveté parameter. Then, we recompute the median naïveté parameters on the remaining (nine) beliefs and calculate the difference between the original and the "modified" naïveté parameter. If this difference is small, this indicates that the median naïveté parameter is stable. For instance, in the example above, if a median naïveté parameter was 0.5 because the respective subject switched between implied naïveté parameters of 0 and 1 across the ten belief formation tasks, throwing out one belief should move the naïveté parameter by 0.5.

The left panel of Figure 2.12 plots a histogram of the difference between the naïveté parameters if we exclude one belief. The right-hand panel displays the difference between the original naïveté parameter and a modified naïveté parameter if we exclude those two beliefs that are closest to that subject's median naïveté parameter. The results show that the vast majority of naïveté parameters is very stable, as indicated by the mass points around zero.

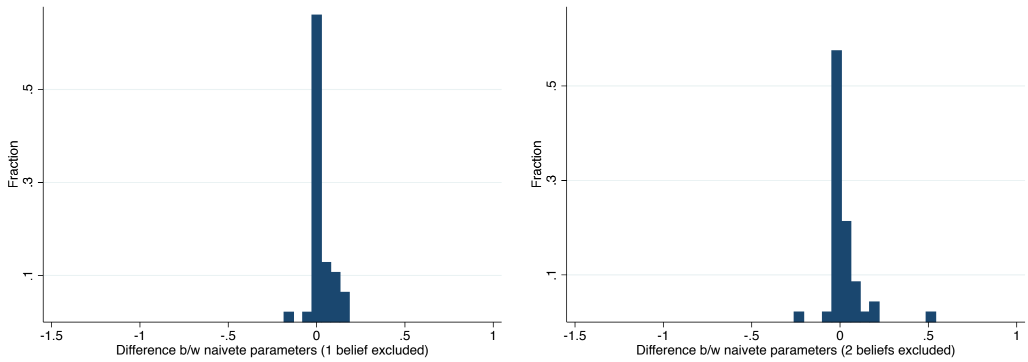


Figure 2.12. Histograms of the difference between original naïveté parameters and modified naïveté parameters when excluding the one or two beliefs that are closest to the original (implied) naïveté parameter

2.C.4 Individual Treatments: (No) Learning Over Time

Column (1) of Table 2.10 provides the results of an OLS regression of all normalized beliefs in the individual *Correlated* treatment on a time trend. This estimation shows that normalized beliefs do not become smaller over time, i.e., they do not converge to the rational belief of zero. In column (2), we show that beliefs do not converge to their counterparts in the *Uncorrelated* treatment, either. To this end, we take all normalized beliefs from the *Correlated* treatment, subtract the median normalized belief in the respective belief formation task in the *Uncorrelated* treatment and then regress this modified belief on a time trend (in essence, this accounts for potential fixed effects of specific belief formation tasks). The results show that the difference between the *Correlated* and the *Uncorrelated* treatment does not become smaller over time.

Table 2.10. Time trend of beliefs in the *Correlated* treatment

	Dependent variable:			
	Normalized belief		Normalized belief minus median in uncorrelated	
	(1)	(2)	(3)	(4)
# of round	-0.0067	0.024	-0.024	-0.0065
	(0.02)	(0.03)	(0.02)	(0.03)
Constant	0.72***	0.22	0.69***	0.31
	(0.12)	(0.55)	(0.12)	(0.56)
Additional controls	No	Yes	No	Yes
Observations	458	448	458	448
R^2	0.00	0.08	0.01	0.09

OLS regressions, standard errors (clustered at individual) in parentheses. Observations include all normalized beliefs from all rounds in the baseline correlated treatment excluding extreme outliers with normalized belief $|b_i^j| > 10$. The results are robust to including these outliers. Additional controls include age, gender, final high school grade, monthly disposable income, marital status fixed effects, and fixed effects for each true state. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.C.5 Finite Mixture Model

For the purpose of the finite mixture model, we assume that every individual belongs to a discrete set of two-dimensional types $\theta_k = (\chi_k, \sigma_k)$ with $k \in \{1, \ldots, K\}$, where the population weights w_k are estimated along with θ_k. Following equation (2.3), the normalized belief of subject i in round j, who is of type k, can be expressed as $\tilde{b}_i^j = \chi_k + u_i^j$, where $u_i^j \sim \mathcal{N}(0, \sigma_k)$ can be thought of as individual- and task-specific random computational error. In allowing for heterogeneity both in χ and σ, we will employ standard maximum likelihood procedures to analyze the prevalence of particular types. The likelihood contribution of individual i is given by

$$L_i(\chi, \sigma, w) \;=\; \sum_{k=1}^{K} w_k \prod_{j=1}^{10} P(\tilde{b}_i^j | \chi_k, \sigma_k) \tag{2.4}$$

where the interior product term computes the likelihood of observing the collection of (normalized) beliefs given a certain type $\theta_k = (\chi_k, \sigma_k)$. This term is then weighted by the respective population share w_k. The grand likelihood is obtained by summing the logs of the individual likelihood contributions, which is then maximized by simultaneously choosing $(\chi_k, \sigma_k, w_k)\ \forall\ k$.

Table 2.11 presents the key results from these estimations. The table reports the estimated parameters of our belief formation model for three different specifications, which differ in the number of types we impose. The results show that if we restrict the model to only one updating rule, the maximum likelihood procedure estimates a substantial degree of naïveté along with a rather high error rate (variance). However, this model masks a considerable degree of heterogeneity: If we allow for the existence of two types of subjects, the model fit increases substantially. In particular, the model indicates that the data are explained as a mixture of two clearly distinguishable groups of subjects. For the first group, the estimation generates a naïveté parameter very close to the rational level of $\chi = 0$. The second group, on the other hand, is characterized by a large degree of correlation neglect with little adjustment from full naïveté. The high variance estimated for the second type motivates us to allow for the presence of further sub-groups in the data. Accordingly, if we allow for three classes of updating rules, the model fit further improves, but not dramatically so. While the parameter estimates for the first (rational) group remain intact, the model now distinguishes between a fully naïve type of subjects (estimated with a rather small error rate) and an intermediate group which is characterized by a rather high degree of naïveté.[37] In sum, our individual-level analysis has shown that the strong *average* tendency to ignore informational redundancies masks a considerable heterogeneity.

[37] Further extending the estimations to allow for four types of subjects does not lead to noteworthy changes of the spirit of our results. These estimations break the rational type up into a fully and almost fully naïve type.

Table 2.11. Results of finite mixture model

Model	Type	*Model parameters* χ	σ	w (%)	*Goodness of fit* LL	AIC	BIC
$K = 1$	$k = 1$	0.68	0.91	100	−607	1219	1222
		(0.07)	(0.05)				
$K = 2$	$k = 1$	0.05	0.26	19.1	−531	1073	1082
		(0.02)	(0.03)	(5.8)			
	$k = 2$	0.83	0.95	80.9			
		(0.06)	(0.06)	(5.8)			
$K = 3$	$k = 1$	0.05	0.26	19.1	−512	1041	1056
		(0.02)	(0.03)	(5.8)			
	$k = 2$	0.74	1.08	55.7			
		(0.19)	(0.07)	(15.8)			
	$k = 3$	1.02	0.52	25.2			
		(0.27)	(0.18)	(15.4)			

47 subjects, standard errors (clustered at the subject level) in parentheses. All estimations exclude a few extreme outliers, which are likely due to typing mistakes: For each task and individual, an observation is set to missing if the implicit normalized belief satisfies $|\tilde{b}_i^j| > 10$ (see eq. (2.3)). This resulted in the exclusion of 4 (out of 462) observations.

Appendix 2.D Details for Individual Robustness Treatments

2.D.1 Design

The design of the robustness treatments closely followed the one in the baseline treatments, with the exceptions discussed in the main text. Table 2.12 provides details on all ten belief formation tasks, including true states, signal draws, and reports of the intermediaries. In addition, we again provide the benchmarks of full correlation neglect and rational beliefs. Note that these theoretical benachmarks are computed assuming full base rate neglect.

Table 2.12. Overview of the belief formation tasks in the robustness treatment

True State	Intermed. 1 uncorr.	Intermed. 2 uncorr.	Intermed. 3 uncorr.	Intermed. 4 uncorr.	Intermed. 2 corr.	Intermed. 3 corr.	Intermed. 4 corr.	Rational Belief	Correlation Neglect Belief
-563	-446	-1,374	-1,377	-1,475	-910	-911.5	-960.5	-1,168	-807
-279	44	90	-388	137	67	-172	90.5	-29.25	7.38
-241	249	-699	-139	70	-225	55	159.5	-129.75	59.63
-33	170	21	225	-128	95.5	197.5	21	72	121
-28	248	83	-110	-364	165.5	69	-58	-35.75	106.13
-23	810	-822	-99	409	-6	355.5	609.5	74.5	442.25
38	442	173	58	233	307.5	250	337.5	226.5	334.25
154	314	206	-229	711	260	42.5	512.5	250.5	282.25
548	-73	-559	181	910	-316	54	418.5	114.75	20.88
1,128	1,989	781	440	2,285	1,385	1,214.5	2,137	1,373.75	1,681.38

The reports of intermediaries 1 through 4 in the *Uncorrelated* condition directly reflect the draws of computers A-D. The report of intermediary 1 in the *Correlated* condition equals the report of intermediary 1 in the *Uncorrelated* treatment. The rational benchmark is computed by taking the average of the signals of computers A-D, i.e., assuming full base rate neglect. The correlation neglect benchmark is given by the average of the reports of intermediaries 1-4 in the *Correlated* condition, i.e., also assuming full base rate neglect. Note that defining the rational belief assuming base rate neglect has no consequences for our treatment comparison. Also note that subjects faced the ten rounds in randomized order, which was identical across treatments.

Table 2.13. Correlation neglect by belief formation task, robustness treatments

True State	Rational Belief	Correlation Neglect Belief	Median Belief *Uncorr.* Treatment	Median Belief *Correlated* Treatment	Ranksum Test (p-value)
-563	-1,168	-807	-1,168	-912.5	0.0189
-279	-29.25	7.38	-29.25	20	0.0031
-241	-129.75	59.63	-126.25	13	0.0052
-33	72	121	72.25	78.5	0.8456
-28	-35.75	106.13	-35.35	36.25	0.0006
-23	74.5	442.25	75	208.5	0.0009
38	226.5	334.25	224.5	226.5	0.0202
154	250.5	282.25	250.5	262.5	0.2133
548	114.75	20.88	115	100	0.1074
1,128	1,373.75	1,681.38	1,373.35	1,412.1	0.0227

See Table 2.12 for details of the computation of the rational and the correlation neglect benchmarks. Note that subjects faced the ten rounds in randomized order.

2.D.2 Results

Table 2.13 reports the results for all ten belief formation tasks. As can be inferred by comparing columns (2) and (4), median beliefs in the *Uncorrelated* condition closely follow our definition of the "rational" belief, suggesting that subjects indeed fail to take into account base rates. Median beliefs in the *Correlated* condition, however, are always biased away in the direction of the full correlation neglect prediction. For seven out of ten tasks, beliefs differ significantly at the 5% level (Wilcoxon ranksum test).

Appendix 2.E Details, Hypotheses, and Robustness Checks for Market Treatments

2.E.1 Derivation of Market Hypotheses

This section derives predictions for our market experiments. In particular, we will highlight the role of the marginal trader in setting the price in the experimental double-auction.

2.E.1.1 Basic Set-Up

A market is populated by 4 buyers and 4 sellers. Sellers own 4 assets that they can sell. Buyers have a monetary endowment that roughly allows them to buy up to 4 goods at fundamental value, see Appendix 2.E.7.[38] The true value of the goods is identical for all traders, and all traders obtain the same signals about the true value. We denote individual beliefs about the value of the assets by b_{si}, $i \in \{1, 2, 3, 4\}$ for the sellers and b_{bj}, $j \in \{1, 2, 3, 4\}$ for the buyers. Likewise, χ_{si}, $i \in \{1, 2, 3, 4\}$ denotes individual-level naïveté for the sellers, and χ_{bj}, $j \in \{1, 2, 3, 4\}$ for the buyers respectively. Without loss of

[38] For ease of exposition, in what follows, we will assume that buyers can buy up to four goods at *any* price. None of the theoretical predictions hinge on this assumption.

generality we assume

$$\chi_{s1} \leq \chi_{s2} \leq \chi_{s3} \leq \chi_{s4} \text{ and } \chi_{b1} \leq \chi_{b2} \leq \chi_{b3} \leq \chi_{b4}.$$

Traders are assumed to be risk-neutral, to behave as price-takers and to not learn from others' trading behavior. Thus, supply of seller i given a market price p is denoted by $xs_i(p)$, where

$$xs_i(p) = \begin{cases} 4 & \text{if } p > b_{si} \\ \{0,1,2,3,4\} & \text{if } p = b_{si} \\ 0 & \text{if } p < b_{si} \end{cases}$$

Likewise, demand of buyer j given price p is denoted by $xd_j(p)$, where

$$xd_j(p) = \begin{cases} 4 & \text{if } p < b_{bj} \\ \{0,1,2,3,4\} & \text{if } p = b_{bj} \\ 0 & \text{if } p > b_{bj} \end{cases}$$

It is well-established that experimental double-auctions converge to the theoretical perfectly competitive equilibrium. Accordingly, we base our market predictions on the notion of competitive equilibrium.

Definition 1. *A price p, market supply $xs = \sum_i xs_i$ and market demand $xd = \sum_j xd_j$ constitute a perfectly competitive equilibrium if $xs = xd$, $xs_i \in xs_i(p), \forall i$ and $xd_j \in xd_j(p), \forall j$.*

2.E.1.2 Homogenous Beliefs

If all traders hold identical beliefs b about the value of the asset, then there are no gains from trade. In the competitive equilibrium, there will be $p = b$, and since all traders will be indifferent between trading and not trading, all possible numbers of trades can be sustained in equilibrium. Thus, for example, if all traders are rational ($\chi = 0$), the prediction would be that $p = b_B = \bar{s}$. If on the other hand all traders are fully naïve ($\chi = 1$), then the prediction is that $p = b_{CN} = \bar{s} + \frac{3}{8}(s_1 - \bar{s}_{-1})$. Thus, prices will be distorted in the direction of the first signal. For intermediate degrees of naïveté, the price is predicted to be $p = b_{CN} = \bar{s} + \frac{3}{8}\chi(s_1 - \bar{s}_{-1})$. Trivially, the higher the degree of naïveté in the market, the more pronounced the resulting price distortion.

2.E.1.3 Heterogeneous Beliefs

The more interesting and also empirically more relevant case are heterogenous beliefs, i.e., different degrees of naïveté in the market. The key question is for what compositions of rational and naïve types equilibrium prices will be distorted and under which conditions rational traders drive prices to the rational level. We focus on signal realizations where $s_1 > \bar{s}_{-1}$, such that correlation neglect distorts beliefs upwards. It is straightforward to show that results are symmetric for the opposite case ($s_1 < \bar{s}_{-1}$). It will be useful to define the following:

- $\#_{rs}$ = number of rational sellers ($\chi = 0$) in a market
- $\#_{nb}$ = number of of naïve buyers ($\chi > 0$) in a market

We enumerate three different cases:

1. $\#_{rs} < \#_{nb}$
Suppose $p = \bar{s}$ (rational level). We would have that

$$\begin{aligned} xs(p = \bar{s}) &\in \{0, ..., 4 \cdot \#_{rs}\} \text{ and} \\ xd(p = \bar{s}) &\in \{4 \cdot \#_{nb}, ..., 16\} \end{aligned}$$

Thus, markets do not clear at $p = \bar{s}$ because $xs(p = \bar{s}) < xd(p = \bar{s})$. In order to equilibrate supply and demand, the price must increase such that either naïve buyers reduce their demand, naïve sellers increase their supply, or both. The equilibrium price level will depend on the degree of naïveté of the marginal traders. Thus, prices will overshoot in the direction predicted by correlation neglect, and

$$\bar{s} < p \leq \bar{s} + \frac{3}{8}(s_1 - \bar{s}_{-1})$$

2. $\#_{rs} = \#_{nb}$
For $p = \bar{s}$, again

$$\begin{aligned} xs(p = \bar{s}) &\in \{0, ..., 4 \cdot \#_{rs}\} \text{ and} \\ xd(p = \bar{s}) &\in \{4 \cdot \#_{nb}, ..., 16\} \end{aligned}$$

Since $\#_{rs} = \#_{nb}$ there exists a market equilibrium at $p = \bar{s}$. However, if the price increases, the market stays in equilibrium, until either the first naïve seller has incentives to sell or the first naïve buyer no longer has incentives to buy. Thus, there exists a range of prices (including the rational price) for which the market is in equilibrium. Importantly, the range is such that, if prices overshoot, they overshoot in the direction of correlation neglect, and the maximum degree of overshooting depends on the naïveté of the marginal traders. Specifically,

$$\bar{s} \leq p \leq min\{\bar{s} + \frac{3}{8}\chi_{s(\#_{rs}+1)}(s_1 - \bar{s}_{-1}), \bar{s} + \frac{3}{8}\chi_{b(4-\#_{nb}+1)}(s_1 - \bar{s}_{-1})\}.$$

3. $\#_{rs} > \#_{nb}$
Again, we start with $p = \bar{s}$ where

$$\begin{aligned} xs(p = \bar{s}) &\in \{0, ..., 4 \cdot \#_{rs}\} \text{ and} \\ xd(p = \bar{s}) &\in \{4 \cdot \#_{nb}, ..., 16\} \end{aligned}$$

Since $\#_{rs} > \#_{nb}$, $p = \bar{s}$ constitutes a market equilibrium. If we (marginally) increase the price, all rational sellers will want to sell all their assets ($xs \geq 4 \cdot \#_{rs}$)

while only naïve buyers will want to buy ($xd \leq 4 \cdot \#_{nb}$), such that supply exceeds demand. Therefore, the only equilibrium is $p = \bar{s}$.

2.E.1.4 Summary

In sum, with homogenous beliefs, higher naïveté implies more distorted price levels. With heterogeneity, the effect of naïveté on prices depends on the composition and overall number of naïve traders. While under certain conditions market prices will remain at the rational level even if some traders are naïve, we have identified different empirically relevant cases where market prices will overshoot in the direction predicted by correlation neglect. Regardless of the particular case discussed above, the magnitude of a potential price distortion depends on the degree of naïveté of the marginal traders.

2.E.1.5 Empirical Identification of Marginal Traders

To compute the naïveté of the marginal traders for a given market group and trading round, we proceed as follows. First, we construct supply and demand curves from the beliefs subjects stated before trading started by sorting the beliefs of buyers in ascending and those of sellers in descending order, which gives rise to four pairs of beliefs. We then identify the lowest belief of a buyer which is still above the belief of the corresponding seller, i.e., we identify the buyer who is located on the demand curve right above the supply curve. We then compute the average naïveté of this buyer and the seller who is located beneath him on the supply curve, to approximate the competitive equilibrium price, and use it for further analysis as detailed in the main text.

2.E.2 Robustness of Treatment Difference in Market Prices

This section provides a robustness check for our main treatment effect in the market treatments. To this end, as in the individual treatments, we first provide a visual representation of our results by plotting kernel density estimates of the market prices in each of the ten trading periods. As above, for this purpose, we restrict the sample to market prices which lie within [50 %, 150 %] of the rational belief (on average, this resulted in the exclusion of one market price per trading period).

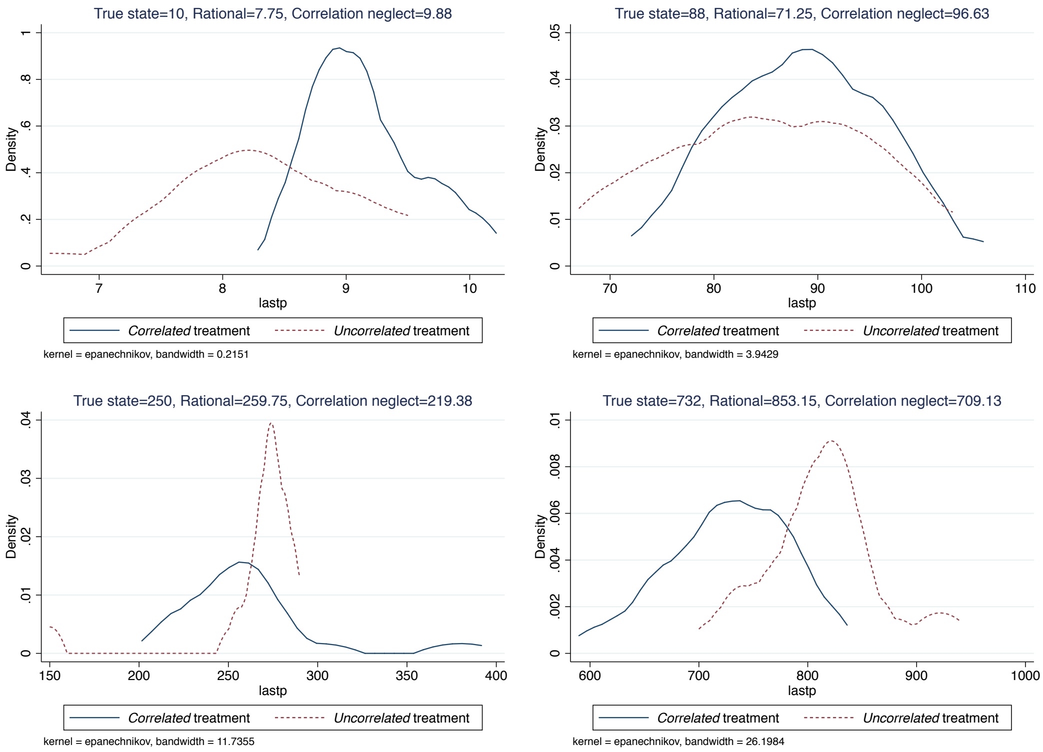


Figure 2.13. Kernel density estimates of market prices (1/2)

Next, we show that the strong treatment difference in price levels is not driven by our definition of the market price. Table 2.14 provides p-values of Wilcoxon ranksum tests for the equality of market prices across treatments for two alternative definitions of the market price. The exposition is akin to Table 2.3 from the main text, but now additionally defines the market price to be either the median or mean trading price (rather than the price of the last concluded trade).

2.E.3 Additional Illustrations of Treatment Difference in Prices

This section provides alternative ways to describe the treatment difference in the market treatments. For this purpose, analogously to the belief normalization, we first normalize the market price of each round and market group such that it equals the naïveté parameter χ, see equation (2.3). We then pool the normalized market prices from all market groups, trading rounds, and both treatments and regress these prices on a treatment dummy. Column (1) of Table 2.15 shows that this treatment difference is highly significant and large in magnitude. As columns (2) and (3) demonstrate, this treatment effect operates entirely through beliefs. After conditioning on the beliefs participants stated before trading started, the treatment effect collapses to zero and becomes insignificant. These results show that it is indeed subjects' beliefs which cause the treatment difference in market prices.

In order to get a visualization of the aggregate treatment difference, we next aggregate the normalized market prices across rounds akin to our procedure in the individual

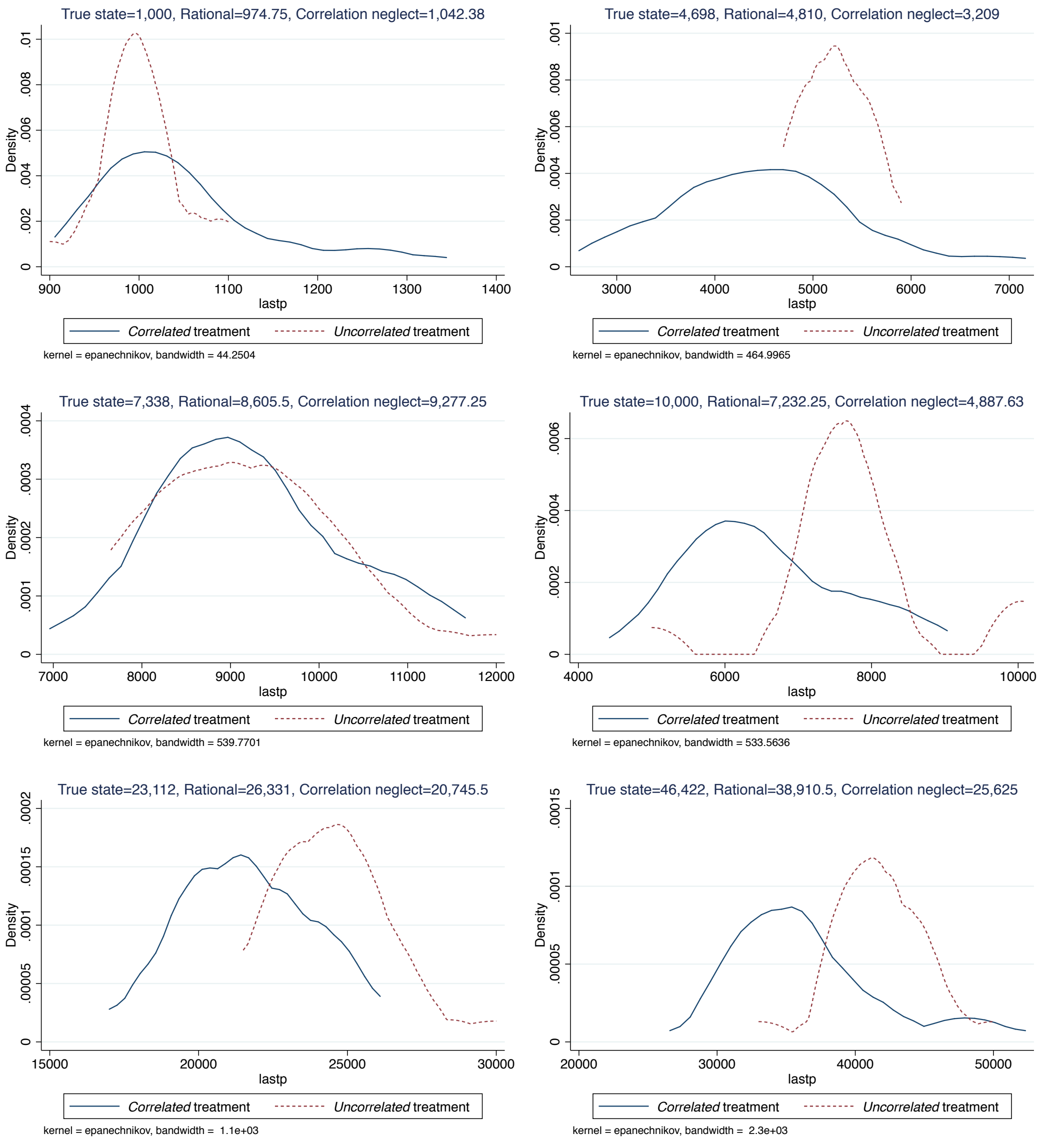


Figure 2.14. Kernel density estimates of market prices (2/2)

Table 2.14. P-values for equality of market prices by trading round for alternative price definitions

True State	*Market Price ≡* Last trading price	Median trading price	Average trading price
10	0.0093	0.0053	0.0075
88	0.0338	0.0200	0.0665
250	0.0113	0.0107	0.0138
732	0.0001	0.0000	0.0000
1,000	0.0723	0.1108	0.1681
4,698	0.0085	0.0025	0.0050
7,338	0.6087	0.7042	0.5092
10,000	0.0534	0.0045	0.0014
23,112	0.0007	0.0061	0.0515
46,422	0.0015	0.0003	0.0095

This table provides p-values of Wilcoxon ranksum tests of the equality of market prices across treatments. For this purpose, for each market group and trading round, the market price is defined as (i) last trading price, (ii) median price, or (iii) average price.

Table 2.15. Beliefs drive treatment difference in market prices

	Dependent variable: Normalized market price (1)	(2)	(3)
1 if correlated	0.32^{***}	-0.052	-0.051
	(0.08)	(0.08)	(0.10)
Group-level median belief (χ)		0.75^{***}	0.70^{***}
		(0.08)	(0.12)
Constant	0.19^{***}	0.040	0.75
	(0.04)	(0.04)	(0.63)
Additional controls	No	No	Yes
Observations	330	330	330
R^2	0.05	0.33	0.39

OLS estimates, standard errors clustered at market group. Observations include all normalized prices from both market treatments excluding four extreme outliers for which the normalized price satisfies $|p_i^j| > 10$. All results are robust to including these observations when employing median regressions. Additional controls include fixed effects for each true state, average age, average monthly disposable income, average final high school grade, and the proportion of females within a given group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

decision making treatments. Specifically, for each market group we use the median normalized market price over the ten rounds to plot the distribution of market prices across treatments.

Figure 2.15 provides kernel density estimates of these aggregated data. It reveals a pronounced and statistically significant difference between the two treatment groups (p-value < 0.0001, Wilcoxon ranksum test). Normalized prices in the *Uncorrelated* treatment are centered close to zero, confirming the standard result that double-auctions tend to produce price levels close to fundamentals. Prices in the *Correlated* treatment, however, are centered around 0.6, i.e., prices systematically overshoot in the direction predicted by correlation neglect.

Again, this treatment difference hinges neither on our aggregation procedure nor on the definition of the market price. Using three definitions of market prices and two different aggregation procedures (for aggregating the market prices of ten trading rounds into a single price per market group), Table 2.16 presents the p-value of ranksum tests for the equality of the aggregated market price between treatments.

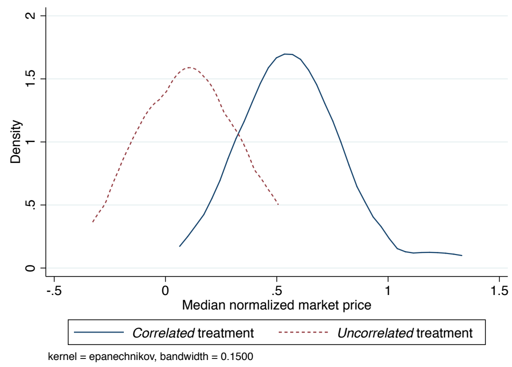


Figure 2.15. Kernel density estimates of median market prices

Table 2.16. P-values of Wilcoxon ranksum tests for equality of aggregated market price between treatments

	Definition of market price:		
Aggregation mechanism	Median price	Average price	Last trading price
Median market price	0.0000	0.0000	0.0000
Average market price	0.0001	0.0002	0.0054

2.E.4 Time Trend of Market Prices

In our market setup, subjects could learn by observing others as well as through the feedback provided at the end of each trading round. If learning played an important role, then the price distortion should be reduced towards the end of the experiment. However, we find no evidence for such an effect – neither beliefs nor prices in the *Correlated market* treatment show any sign of converging to their counterparts in the *Uncorrelated market* treatment. For instance, if we take the last round from all market groups and normalize the market price (to make it comparable between different orderings of rounds), we still find a significant treatment difference (p-value = 0.0290, Wilcoxon ranksum test). Similarly, Table 2.17 gives an overview of the time trend of market prices. In columns (1) and (2), we report the results of an OLS regression of all normalized market prices in the *Correlated market* treatment on a time trend, which indicate that market prices do not converge to rational levels.[39] We also show that prices do not converge to their counterparts in the *Uncorrelated market* treatment (columns (3)-(4)). To this end, we take all normalized market prices and then subtract the normalized market price of the median market group in that round in the *Uncorrelated market* treatment. Again, there is no sign of convergence to the levels in the *Uncorrelated* treatment. In sum, these results show that there is no learning across rounds.

Table 2.17. Time trend of market prices in the *Correlated market* treatment

	Dependent variable:			
	Normalized market price		Normalized market price minus median price in uncorrelated	
	(1)	(2)	(3)	(4)
# of trading period	-0.018 (0.03)	-0.0091 (0.02)	-0.024 (0.03)	-0.0069 (0.02)
True state FE	No	Yes	No	Yes
Observations	167	167	167	167
R^2	0.00	0.18	0.01	0.05

OLS regressions, standard errors (clustered at market group level) in parentheses. Observations include the market prices from all trading rounds in the correlated market treatment excluding market prices which satisfy $|p_i^j| > 10$.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.E.5 Time Trend of Beliefs in Market Experiments

Table 2.E.5 presents the results of OLS regressions of subjects' (normalized) beliefs in the *Correlated market* treatment on a linear time trend. If the market interaction induces naïve subjects to learn, we should observe a negative coefficient. However, we do not find any significant effects, regardless of the specification we employ. In column (1), we include beliefs which satisfy $|b_i^j| \leq 10$, i.e., we only exclude very extreme outliers. In columns (2)-(5), we use beliefs which satisfy $b_i^j > -1$ and $b_i^j < 2$, i.e., we focus on beliefs in a reasonable range, which likely don't reflect typing errors. Regardless of the sample, the coefficient on the time trend is small and insignificant, both with and without fixed effects for a particular market group, individual subjects, and particular true states.

2.E.6 Why Does the Market not Reduce the Bias?

This section discusses potential reasons, why our double-auction market environment did not eliminate correlation neglect. In short, three reasons in particular could play a role. First, given that we implemented a common value environment with identical information across subjects (but potentially heterogeneous processing thereof), a feature of our market is that it allows subjects to learn from the behavior of (potentially more rational) others. For instance, suppose a seller in the correlated environment neglects the correlation and arrives at a belief that the value of the asset is, say, 10. If this seller observes

[39] Similar results obtain if we run the corresponding regressions using subjects' beliefs as dependent variable.

Table 2.18. Time trend of normalized beliefs in the *Correlated market* treatment

	Dependent variable: Naïveté χ				
	(1)	(2)	(3)	(4)	(5)
# of trading period	0.015	-0.0087	-0.0088	-0.0094	-0.0016
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Market FE	No	No	Yes	No	No
Subject FE	No	No	No	Yes	Yes
True state FE	No	No	No	No	Yes
Observations	1404	1241	1241	1241	1241
R^2	0.00	0.00	0.04	0.27	0.35

OLS regressions, standard errors (clustered at market group level) in parentheses. Observations include the market prices from all trading rounds in the correlated market treatment. In column (1), we only excluce beliefs which satisfy $|b_i^j| > 10$. In columns (2)-(5), we use beliefs which satisfy $b_i^j > -1$ and $b_i^j < 2$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

all buyers offering to buy the asset at, say, 20, this could induce him to reconsider his valuation of the asset. For instance, that seller might conjecture that he misinterpreted his signals. In this sense, the existence of even one rational type in a given market group could in principle debias all other subjects. Furthermore, even if observing others' trading behavior does not debias subjects, it might at least reduce their confidence in their valuation of the good. Both of these channels should attenuate the impact of correlation neglect on market outcomes. The fact that we do not find evidence for this is consistent with the idea that people might neglect that the trading behavior of others carries informational content, perhaps akin to the idea of "cursedness" (Eyster and Rabin, 2005; Eyster et al., 2013) with the twist that there is no heterogeneous private information in our setup, but rather heterogeneous processing of the same signals.[40]

Second, the rational types might not be able to bring prices to fundamental values due to institutional features of our trading environment. In particular, our setup did not allow the same subject to both buy and sell. Each subject's influence on the market price was hence restricted to selling four assets as a seller, and buying a small number of assets as a buyer. In the data, an average of 3.8 subjects (out of 8) per market group had a median naïveté parameter of $\chi \in [-.25; .25]$, implying that these rational subjects would have needed to trade excessively to bring prices to fundamentals by themselves.

However, third, even if some subjects hold correct beliefs and could in principle bring prices to fundamentals, they might not be willing to do so. For instance, if the rational types are slightly risk averse and have some subjective uncertainty over the true state (as they should), they could attempt to diversify, i.e., hold a mix of both assets and cash. Indeed, in the data, we see strong evidence of this. For instance, in trading periods in which correlation neglect predicts underpricing, those subjects with a (median) naïveté parameter of $\chi \in [-.25; .25]$ only held a total of 7.7 (out of a total of 16) assets on average, i.e., the rational subjects do not buy all assets when prices are too low, i.e., when assets are a bargain. The fact that rational agents seemed to limit their trading activity

[40] Alternatively, our empirical pattern is consistent with the idea that people are overconfident about their ability to process correlations.

suggests that these types were cautious in fully exploiting their superior knowledge about the true value of the asset.

2.E.7 Endowments and Exchange Rates in Market Treatments

Table 2.19. Overview of the ten trading rounds

True state	Budget buyer (points)	Exchange rate points / euros	Fixed costs buyer
10	40	2.67	4
88	450	30	45
250	1,500	100	150
732	3,000	200	300
1,000	5,000	333.33	500
4,698	25,000	1,666.67	2,500
7,338	25,000	1666.67	2,500
10,000	50,000	3,333.33	5,000
23,112	90,000	6,000	9,000
46,422	200,000	13,333.33	20,000

Sellers did not incur any fixed costs. Buyers' fixed costs amounted to 10 % of the respetive budget. The relationship between budget and true state was non-constant across rounds. The exchange rate is computed as budget / 15.

Appendix 2.F Treatments to Investigate the Mechanisms Underlying the Bias

2.F.1 Reduced Complexity

In the reduced complexity treatments, we implemented the same basic structure as in the baseline design, yet there were only 2 independent computer signals and one intermediary. Both the true states and the signals of computer A were identical to the baseline conditions, while the signal of computer B in the reduced complexity treatments always equalled the signal of computer C in the baseline condition.[41]

Table 2.20 provides an overview of each of the ten belief formation tasks, including median beliefs in the *Correlated* and the *Uncorrelated* condition as well as the p-value of a Wilcoxon ranksum test. In none of the ten tasks is the treatment difference significant at the 5 % level.

We can again normalize each belief (i.e., compute the naïveté parameter implicit in a belief) to make it comparable across belief formation tasks.[42] Figure 2.16 provides kernel density estimates of the distributions of the median naïveté parameters in the *Correlated* and the *Uncorrelated* treatment. As visual inspection suggests, beliefs in the two treatment are statistically indistinguishable from each other (Wilcoxon ranksum test, p-value = 0.1505). Table 2.21 confirms this result using OLS regressions and also shows that – unlike in the baseline treatments – there is no difference in response time between the *Correlated* and the *Uncorrelated* treatments. Interestingly, there is also no relationship between response times and beliefs within the *Correlated* treatment. While in the baseline *Correlated* treatment higher response times are associated with better beliefs, this association breaks down in the low complexity case, suggesting that at least a considerable fraction of subjects understood that the report of intermediary 2 already reflected the rational belief.

[41] this was determined randomly.

[42] Formally, a normalized belief of individual i in task j of the low-complexity conditions is given by

$$\tilde{b}_i^j = \chi_i^j = \frac{s_1^j + s_2^j}{2} + \frac{1}{4}(s_1^j - s_2^j)$$

Table 2.20. Overview of belief formation tasks in the reduced complexity treatments

True state	Computer A	Interm. uncorr.	Interm. corr. = Rational belief	Correlation neglect belief	Median belief *Uncorrelated*	Median belief *Correlated*	Ranksum test (p-value)
10	12	10	11	11.5	11	11	0.9808
88	122	68	95	108.5	95	95	0.7141
250	179	288	233.5	206.25	233.5	234	0.2752
732	565	650	607.5	586.25	607	600	0.9184
1,000	1,100	629	869.5	989.75	869.5	870	0.0967
4,698	1,608	4,866	3,237	2,422.5	3237	3237	0.1686
7,338	9,950	11,322	10,636	10,293	10,500	10,636	0.1154
10,000	2,543	6,898	4,720.5	3,631.75	4,720	4,721	0.5180
23,112	15,160	20,607	17,883.5	16,521.8	17,883	17,884	0.3479
46,422	12,340	49,841	31,090.5	21,715.3	31,090.5	31,090	0.7534

The reports of the intermediary in the *Uncorrelated* condition directly reflect the draw of computer B. The rational belief is computed by taking the average of the signals of computers A and B. The correlation neglect belief is computed assuming $\chi = 1$, i.e., full correlation neglect. Thus, this benchmark is given by the average of the signal of computer A and the message of the intermediary in the *Correlated* condition. Note that subjects faced the ten rounds in randomized order.

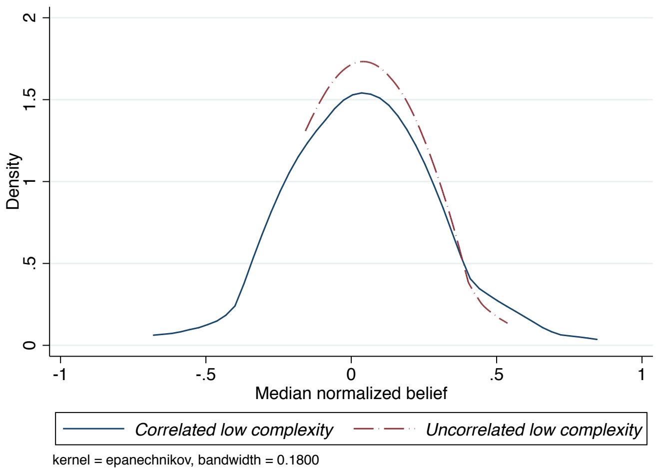


Figure 2.16. Kernel density estimates of beliefs in the reduced complexity treatments

Table 2.21. Reduced complexity treatments

	Dependent variable:			
	Median χ		Median response time	
	(1)	(2)	(3)	(4)
1 if correlated	-0.013	-0.013	0.022	0.020
	(0.03)	(0.03)	(0.12)	(0.13)
Constant	0.051^{***}	-0.050	0.65^{***}	0.19
	(0.02)	(0.15)	(0.07)	(0.44)
Additional controls	No	Yes	No	Yes
Observations	94	93	94	93
R^2	0.00	0.07	0.00	0.03

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.F.2 Treatment *Math*

Figure 5.11 provides kernel density plots of the median naïveté parameters in treatment *Math* as well as the two baseline conditions. As can be inferred, while a minority of subjects remains fully naïve, a large fraction now states rational beliefs. Table 2.22 provides an overview of each separate belief formation task and shows that in six out of ten tasks do beliefs statistically significantly differ between *Math* and the baseline *Correlated* condition. Notably, in all ten tasks is the median belief closer to the median belief in the *Uncorrelated* condition than the median belief in the *Correlated* treatment, also see Figure 2.18.

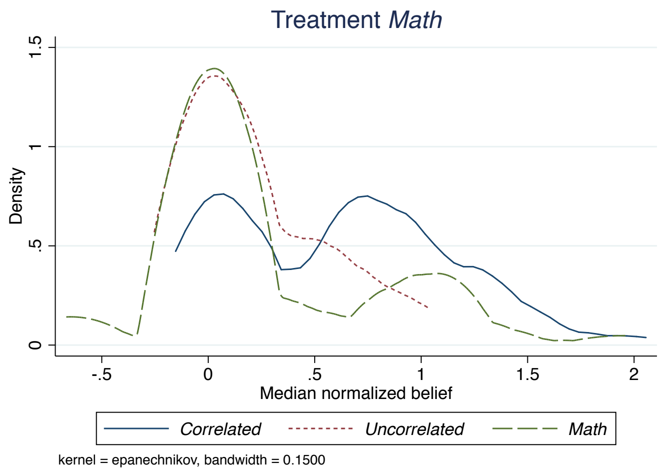


Figure 2.17. Kernel density estimates of beliefs in the *Math* treatment

Table 2.22. Correlation neglect by belief formation task in the *Math* treatment

True State	Median Estimate *Uncorr.*	Median Belief *Correlated*	Median Belief *Math*	Ranksum Tests (p-value) *Correlated*	*Uncorrelated*
10	8	9.2	7.75	0.0005	0.8376
88	71.2	88	72	0.1647	0.0132
250	259.75	235.5	260	0.0952	0.2431
732	847	742	853.5	0.0013	0.1470
1,000	999	1,030	975	0.0026	0.4827
4,698	4,810	4,556	4,792.5	0.4880	0.0100
7,338	8,975	9,044.5	8,605	0.3588	0.6475
10,000	7,232	6,750	7,100	0.7424	0.0095
23,112	25,000	21,000	26,215.5	0.0001	0.1732
46,422	38,885.5	32,000	38,500	0.7063	0.3385

See Table 1 for details of the computation of the rational and the correlation neglect benchmarks. Note that subjects faced the ten rounds in randomized order. The ranksum tests refer to a comparison between the *Math* treatment and the baseline *Correlated* / *Uncorrelated* treatment, respectively.

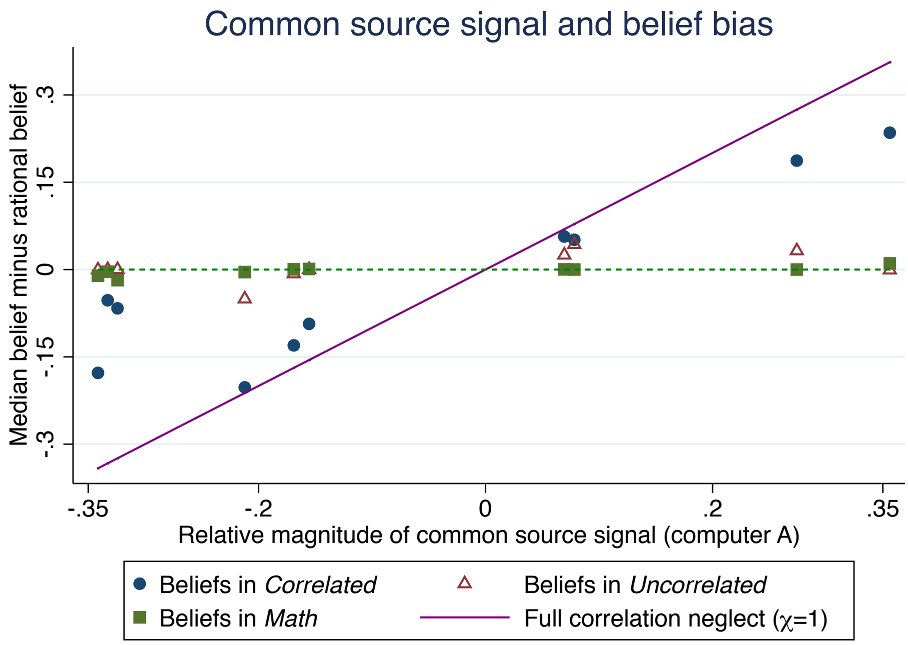


Figure 2.18. Beliefs in the *Correlated*, *Uncorrelated* and *Math* treatments plotted against the relative magnitude of the signal of computer A. See the notes of Figure 2.2 for details on the construction of this figure.

2.F.3 Treatments *Intermediaries* and *Alternating*

In the *Intermediaries* treatment, subjects went through the same ten belief formation tasks as in the *Correlated* treatment, but subjects' attention was steered towards the correlation by including (i) the paragraph provided in the main text and (ii) by repeating the visual representation of the information structure both at the end of the instructions and on subjects' decision screens. In the *Alternating* treatment, attention was shifted in a more indirect way, by varying the information structure (correlated versus uncorrelated) between rounds. This was again made rather salient to subjects since they were asked to pay special attention to the prevailing scenario and to consider the corresponding implications. In the main text, we presented aggregated results from these treatments; now, we detail the results from each of the separate tasks by comparing the corresponding beliefs with those in the baseline *Correlated* condition.

Table 2.23 summarizes the results from the different belief formation tasks for both treatments. The table provides rational and full correlation neglect beliefs for all ten tasks, as well as median beliefs from the *Correlated* treatment, the *Intermediaries* treatment and the *Alternating* treatment. In addition, p-values of Wilcoxon ranksum tests, testing for differences between *Intermediaries* and the *Correlated* treatment, as well as between *Alternating* and the *Correlated* treatment, are provided. First note that, in all ten rounds, beliefs in the *Intermediaries* treatment are closer to the rational belief compared to the *Correlated* treatment. However, in five rounds, beliefs do not differ from each other statistically at the 5 % level. Likewise, in all five rounds of the *Alternating* treatment in which correlated information was provided, beliefs are closer to the rational belief relative to the *Correlated* treatment. However, again, this difference is only significant in two out of five belief formation tasks.

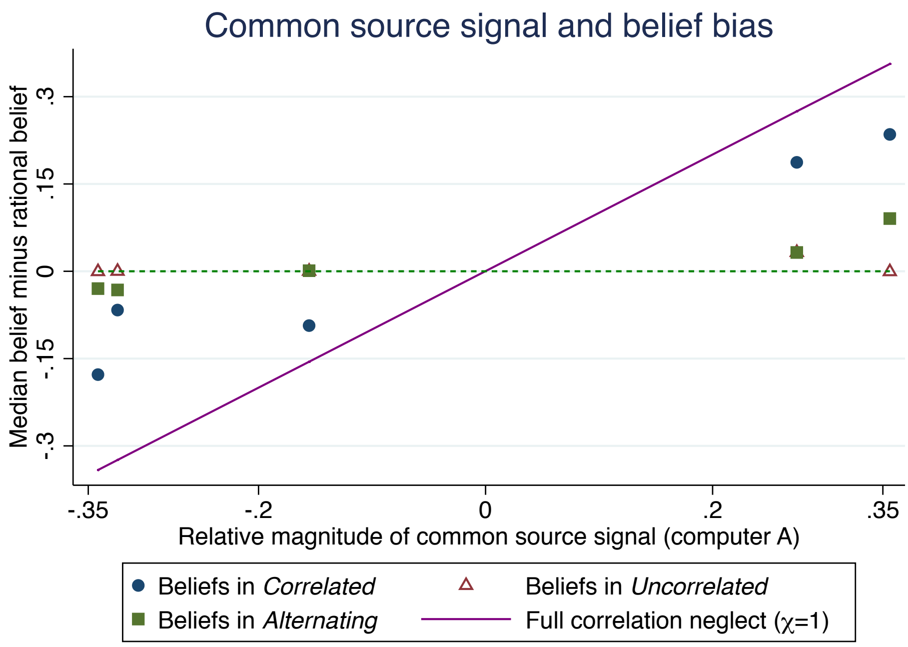


Figure 2.19. Beliefs in the *Correlated*, *Uncorrelated* and *Alternating* treatments plotted against the relative magnitude of the signal of computer A. See the notes of Figure 2.2 for details on the construction of this figure.

Table 2.23. Correlation neglect by belief formation task: *Intermediaries* and *Alternating* treatments

				Intermediaries Treatment			*Alternating* Treatment		
True State	Rational Belief	Correlation Neglect Belief	Median Belief *Correlated* Treatment	Median Belief *Intermediaries* Treatment	Ranksum Tests (p-value) *Correlated*	Ranksum Tests (p-value) *Uncorrelated*	Median Belief *Alternating* Treatment	Ranksum Tests (p-value) *Correlated*	Ranksum Tests (p-value) *Uncorrelated*
10	7.75	9.88	9.2	8	0.0367	0.2782	8	0.0224	0.4304
88	71.25	96.63	88	72.25	0.0051	0.3173	77.7	0.1182	0.0064
250	259.75	219.38	235.5	260	0.0751	0.5258	260	0.0193	0.9386
732	853.15	709.13	742	850	0.0030	0.5815			
1,000	974.75	1,042.38	1,030	979	0.0039	0.4959			
4,698	4,810	3,209	4,556	4,787.5	0.2980	0.0774			
7,338	8,604.5	9,277.25	9,044.5	8,727.5	0.2433	0.2558			
10,000	7,232.25	4,887.63	6,750	6,950	0.8716	0.0027	7,000	0.2128	0.1040
23,112	26,331	20,745.5	21,000	25,399.7	0.0001	0.5951			
46,422	38,910.5	25,625	32,000	35,894	0.4624	0.0920	37,750	0.4055	0.3011

See Table 1 for details of the computation of the rational and the correlation neglect benchmarks. Note that subjects faced the ten rounds in randomized order. The ranksum tests refer to a comparison between the baseline *Correlated* (*Uncorrelated*) treatment and the *Intermediaries / Alternating* treatments, respectively.

2.F.4 High Stakes Conditions

In the high-stakes conditions, we implemented the same procedure as in the baseline conditions using a different incentive scheme. For all ten belief formation tasks, the results in these treatments are virtually identical to those in the baseline conditions. Figure 2.20 provides kernel density estimates of the median naïveté parameters (see equation (2.3)) in the baseline and high-stakes conditions, which suggest that beliefs in these treatments are almost indistinguishable from each other. As Figure 2.21 shows, median beliefs in each task are sometimes marginally closer to the rational benchmark than in the baseline treatment, and sometimes further away. Detailed results for each belief formation task are available upon request.

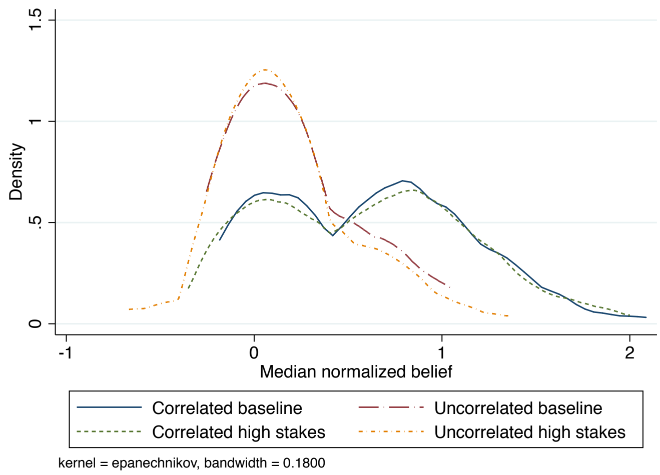


Figure 2.20. Kernel density estimates of beliefs in the baseline and high stakes conditions

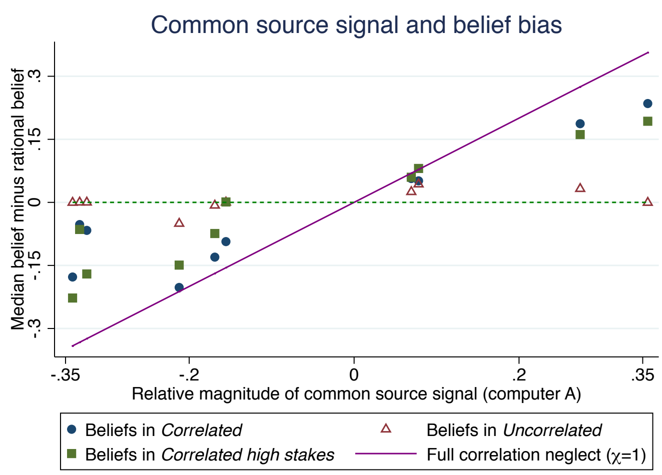


Figure 2.21. Beliefs in the baseline *Correlated* and *Uncorrelated* as well as the *Correlated high stakes* treatments plotted against the relative magnitude of the signal of computer A. See the notes of Figure 2.2 for details on the construction of this figure.

2.F.5 *Multiply* Treatment

In treatment *Multiply*, the intermediaries 1-3 each received one signal and multiplied it by *1.5*. Table 2.24 presents an overview over all ten belief formation tasks. For reference, the table provides the rational as well as the face value prediction. As can be inferred, in all ten tasks are median beliefs in *Multiply* very close to those from the baseline *Uncorrelated* condition. In consequence, none of the tasks exhibits a significant treatment difference compared to the benchmark treatment. Figure 2.22 visualizes this result by plotting median normalized beliefs (median naïveté parameters) for *Multiply*. The large spike around zero indicates that virtually all subjects behave approximately rational in this context.

The OLS regressions presented in columns (5) and (6) of Table 2.27 show that beliefs in *Multiply* are indeed significantly less biased compared to those in the *Correlated* treatment (a comparison between the two treatments can be facilitated by computing naïveté parameters). In addition, subjects in *Multiply* took substantially longer to solve the tasks. Notice that this pattern is consistent with the idea that, once people notice the "bias" in the information structure, they successfully correct for it and hence need more time to do the necessary calculations.

Table 2.24. Overview of belief formation tasks in the *Multiply* treatment

True state	Rational belief	Face value belief	Median belief *Uncorrelated*	Median belief *Multiply*	Ranksum test (p-value)
10	7.75	10.125	8	8.3	0.3755
88	71.25	91.625	71.2	71.25	0.8233
250	259.75	367.25	259.75	260	0.8085
732	853.25	1209.25	847	805	0.8747
1,000	974.75	1,323.375	999	1,000	0.3054
4,698	4,810	7,014	4,810	4,818	0.8474
7,338	8,604.5	11,663	8,975	8,750	0.3097
10,000	7,232.25	10,530.5	7,232	7,100	0.3959
23,112	26,331	37,601.5	25,000	23,000	0.2270
46,422	38,910.5	56,823.25	38,885.5	38,573.75	0.9525

The rational belief is computed by taking the average of the signals of computers A through D. The face value belief is given by $(s_A + 1.5s_B + 1.5s_C + 1.5s_D)/4$. Note that subjects faced the ten rounds in randomized order. The ranksum tests refer to a comparison between the baseline *Uncorrelated* and the *Multiply* treatments.

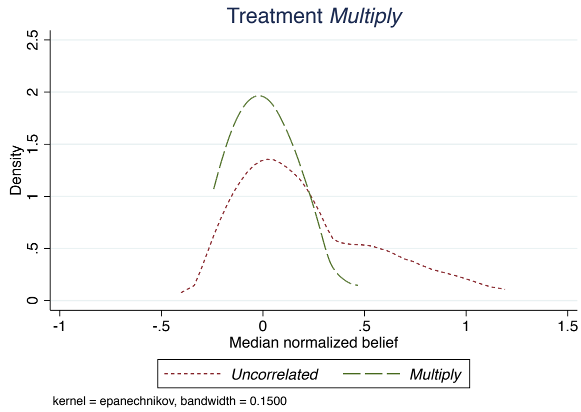


Figure 2.22. Kernel density estimates of median naïveté parameters in *Multiply* and the *Uncorrelated* treatment. Naïveté parameters are computed akin to the procedure in eq. (2.3)

2.F.6 *Face Value* Treatment

In treatment *Face Value*, the computers A-D generated the same sets of signals as in the baseline conditions, while the intermediaries 1-3 computed the same averages as in the baseline *Correlated* treatment. Table 2.25 presents an overview of the value of X in each task and the resulting messages of machines M1 through M3. Further notice that this treatment allows the separate computation of rational, correlation neglect, and face value benchmarks.

To illustrate the results, Figure 2.23 compares kernel density estimates of the belief distributions between the *Face value* treatment and the two baseline treatments. The left panel depicts median normalized beliefs (median naïveté parameters) for tasks in which face value bias coincides with the rational prediction of zero. The right panel displays median normalized beliefs for tasks in which face value bias and correlation neglect make opposite predictions, i.e., after normalization the face value prediction is (-1) and the correlation neglect prediction is 1. In both panels, the belief distribution in the *Face value* treatment is closest to the belief distribution in the baseline *Correlated* treatment and clearly differs both from beliefs in the *Uncorrelated* treatment as well as from the face value predictions.[43] A Wilcoxon ranksum test confirms that beliefs in *Face value* significantly differ from those in the *Uncorrelated* condition ($p = 0.0086$), but not from those in the baseline *Correlated* treatment ($p = 0.3670$).[44] Thus, even in a treatment in which face value bias makes a prediction different from correlation neglect, we identify significant evidence for people's neglect of correlations.

Table 2.26 presents an overview of the corresponding results for all separate belief formation tasks. Beliefs in *Face value* typically closely follow beliefs in the baseline *Corre-*

[43] If anything, beliefs are slightly less rational in *Face value*. It is conceivable that some subjects immediately noticed that the messages of the machines are biased due to X and, once they understood this, stopped to reflect upon potential further problems in the data-generating process.

[44] Beliefs in *Face value* do not significantly differ between tasks in which face value predicts zero or (-1), providing further evidence for the low explanatory power of a simple face value bias.

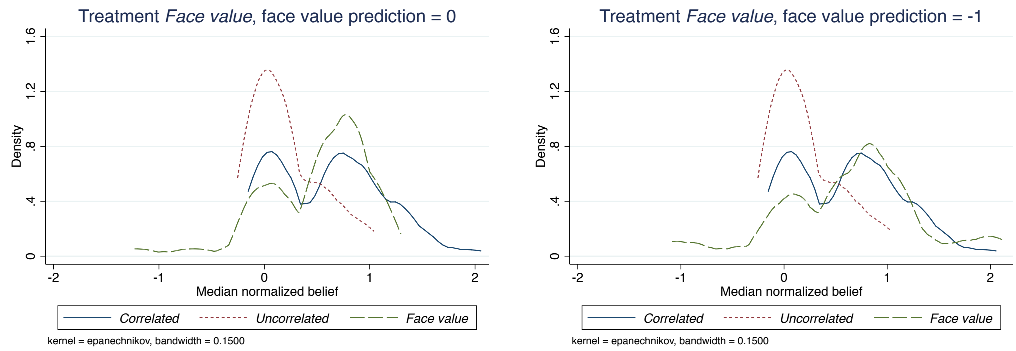


Figure 2.23. Kernel density estimates of median normalized beliefs in the *Face value* treatment, compared with those from the baseline *Correlated* and *Uncorrelated* conditions. The left panel illustrates the five tasks in which the face value belief equals the rational belief, while the right panel depicts the five tasks in which the face value belief makes the opposite prediction compared to correlation neglect (relative to the rational belief). To ease readability, the densities exclude (4 / 3, respectively) subjects with median normalized belief of less than (-2). All statistical tests include these outliers.

lated condition, suggesting that subjects do not fall prey to a simple face value heuristic, but instead extract X from the reports of the machines. In consequence, in the vast majority of tasks, beliefs significantly differ between the *Uncorrelated* and the *Face value* treatments in the direction predicted by correlation neglect, while the comparison between *Face value* and the baseline *Correlated* treatment is usually far from significant.

Columns (1) and (2) of Table 2.27 confirms this finding using OLS regressions. In addition, columns (3) and (4) show that response times are substantially higher in *Face value* compared to the baseline *Correlated* condition. This reflects the fact that in this treatment virtually all subjects engage in some computations to debias the messages (almost everybody corrects for X), while in the baseline *Correlated* treatment some part of subjects does not debias messages in any way before computing averages.

Table 2.25. Overview of the belief formation tasks, *Face Value* treatment

True State	X	Machine M1	Machine M2	Machine M3	Rational Belief	Correlation Neglect Belief	Face Value Belief
10	-6	4.5	5	0	7.75	9.88	5.38
88	-34	72	61	30	71.25	96.63	71.13
250	54	291	288	282	259.75	219.38	259.88
732	192	898	800	1,150	853.25	709.13	853.13
1,000	-90	995	780	1,015	974.75	1,042.38	974.88
4,698	4,269	8,693	7,506	7,836	4,810.00	3,209.00	6410.75
7,338	-1,794	3,783	8,842	9,153	8,604.50	9,277.25	7,931.75
10,000	3,126	9,788	7,847	8,752	7,232.25	4,887.63	7,232.13
23,112	14,895	33,378	32,779	46,351	26,331.00	20,745.50	31,916.75
46,422	35,427	57,681	66,518	72,244	38,910.50	25,625.25	52,195.50

The rational benchmark is computed by taking the average of the signals of computers A-D. The correlation neglect benchmark is given by the average of the reports of computer A and intermediaries 1-3, i.e., by extracting X from the reports of the machines. The face value belief is given by the average of the messages of computer A and machines M1-M3. Note that subjects faced the ten rounds in randomized order.

Table 2.26. Correlation neglect by belief formation task, *Face value* treatment

True State	Rational Belief	Correlation Neglect Belief	Face Value Belief	Median Belief *Face Value*	Median Belief *Correlated*	Ranksum Tests (p-value) *Correlated*	*Uncorrelated*
10	7.75	9.88	5.38	9	9.2	0.6455	0.0840
88	71.25	96.63	71.13	85	88	0.2197	0.0341
250	259.75	219.38	259.88	240	235.5	0.5761	0.0184
732	853.15	709.13	853.13	757.3	742	0.0978	0.2098
1,000	974.75	1,042.38	974.88	1,020	1,030	0.5013	0.1839
4,698	4,810	3,209	6410.75	3,742.7	4,556	0.5341	0.0001
7,338	8,604.5	9,277.25	7,931.75	8,800	9,044.5	0.0646	0.0473
10,000	7,232.25	4,887.63	7,232.13	5,669	6,750	0.5459	0.0001
23,112	26,331	20,745.5	31,916.75	21,229	21,000	0.3034	0.0937
46,422	38,910.5	25,625	52,195.50	29,574	32,000	0.3210	0.0012

See Table 2.25 for details of the computation of the rational, correlation neglect, and face value benchmarks. The ranksum tests refer to a comparison between the face value treatment and the *Correlated* (*Uncorrelated*) treatment, respectively. Note that subjects faced the ten rounds in randomized order.

Table 2.27. Beliefs and response times in *Multiply* and *Face value*

	Face value treatment				*Multiply* treatment			
	Dependent variable:							
	Median χ		Median response time		Median χ		Median response time	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 if *Face value* treatment	-0.14	-0.20	1.06***	1.19***				
	(0.13)	(0.14)	(0.22)	(0.21)				
1 if *Multiply* treatment					-0.61***	-0.62***	0.51**	0.65***
					(0.08)	(0.08)	(0.23)	(0.23)
Constant	0.62***	0.69	1.38***	0.27	0.62***	0.73**	1.38***	1.17
	(0.07)	(0.42)	(0.15)	(0.97)	(0.07)	(0.32)	(0.15)	(1.03)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	90	88	90	88	93	91	93	91
R^2	0.01	0.05	0.20	0.31	0.40	0.40	0.05	0.12

OLS estimates, robust standard errors in parentheses. Observations include all subjects from the baseline *Correlated* and the *Face value* treatments (columns (1)-(2)), and from the baseline *Correlated* and the *Multiply* treatments (columns (3)-(4)). In columns (1)-(2), the dependent variable is median normalized beliefs (naïveté parameters), while in columns (3)-(4) it is median response time. Additional controls include age, gender, monthly disposable income, and marital status fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.F.7 Treatments *Structure* and *Messages*

Following the literature, we define face value bias as excessive focus on the salient messages relative to the underlying data-generating process. Thus, as a second test of the idea of face value bias, we implement two treatments in which we direct subjects' attention towards the data-generating process. If subjects indeed take all messages at face value because they do not attend to the information structure, these treatments should be effective in debiasing subjects. The corresponding treatments *Structure* and *Messages* were identical to the baseline *Correlated* condition, except that we provided a hint both at the end of the instructions and on subjects' decision screens.

> Treatment *Structure:* ***Hint for solving the task:*** *You can only solve this problem correctly if you have understood the structure which generates your information.*
>
> Treatment *Messages:* ***Hint for solving the task:*** *The intermediaries do not generate estimates themselves.*

Arguably, these hints steer subjects' attention towards the underlying data-generating process relative to the visible messages. However, the nudges do not tell subjects on which specific features they ought to focus. 96 subjects participated in these treatments (47 each) and earned 10.60 / 12.90 euros on average, respectively.

Result 9. *Exogenous shifts in subjects' attention towards the data-generating process as a whole do not debias subjects.*

Figure 2.24 depicts the distributions of median normalized beliefs. Both nudges had a rather small and overall statistically insignificant effect on subjects' behavior. While both belief distributions appear to undergo a small shift, a Wilcoxon ranksum test indicates that median beliefs still exhibit correlation neglect compared to the *Uncorrelated* treatment ($p = 0.0134$ for *Structure* and $p = 0.0039$ for *Messages*). In addition, beliefs do not statistically differ from those in the baseline correlated condition ($p = 0.1618$ for *Structure* and $p = 0.3783$ for *Messages*). Table 2.28, we present the corresponding analyses for all ten separate belief formation tasks. We again present the rational and correlation neglect benchmarks and contrast beliefs from the nudge treatments with those in the *Uncorrelated* and *Correlated* baseline conditions. The results show that, in both salience treatments, in the large majority of tasks do beliefs significantly differ from those in the *Uncorrelated* condition, while only in at most two tasks do beliefs become more rational compared to the baseline *Correlated* condition. Thus, while it appears that these treatments might have had a small positive effect on behavior, they were not nearly sufficient to debias the majority of subjects. Unreported results also show that these treatments produce beliefs which are statistically significantly more biased than beliefs in *Intermediaries* and *Alternating*.

In sum, in contrast to what face value bias would predict, alerting subjects to the data-generating process as a whole (relative to the messages) is not sufficient to debias them.

Table 2.28. Correlation neglect by belief formation task: *Structure* and *Messages* treatments

				Structure Treatment			*Messages* Treatment		
True State	Rational Belief	Correlation Neglect Belief	Median Belief *Correlated* Treatment	Median Belief *Structure* Treatment	Ranksum Tests (p-value) *Correlated*	Ranksum Tests (p-value) *Uncorrelated*	Median Belief *Messages* Treatment	Ranksum Tests (p-value) *Correlated*	Ranksum Tests (p-value) *Uncorrelated*
10	7.75	9.88	9.2	9	0.3057	0.0627	9	0.0788	0.2179
88	71.25	96.63	88	80	0.4202	0.0016	75	0.1197	0.0618
250	259.75	219.38	235.5	260	0.0486	0.6772	250	0.2761	0.0950
732	853.15	709.13	742	785	0.0336	0.5772	800	0.0265	0.3160
1,000	974.75	1,042.38	1,030	1,020	0.7908	0.0380	1,020	0.6603	0.1242
4,698	4,810	3,209	4,556	4,750	0.4790	0.0225	4,454.22	0.9751	0.0025
7,338	8,604.5	9,277.25	9,044.5	9,251.25	0.8360	0.9357	9,284.25	0.4196	0.4912
10,000	7,232.25	4,887.63	6,750				5,555	0.2892	0.0001
23,112	26,331	20,745.5	21,000	20,133	0.2862	0.0003	21,600	0.5462	0.0055
46,422	38,910.5	25,625	32,000	38,000	0.2561	0.5898	33,158	0.8087	0.0213

See Table 1 for details of the computation of the rational and the correlation neglect benchmarks. Note that subjects faced the ten rounds in randomized order. The ranksum tests refer to a comparison between the baseline *Uncorrelated* (*Correlated*) treatment and the *Structure / Messages* treatments. Note that, in the *Structure* treatment, we lost all observations for the true state of 10′000 due to a programming error.

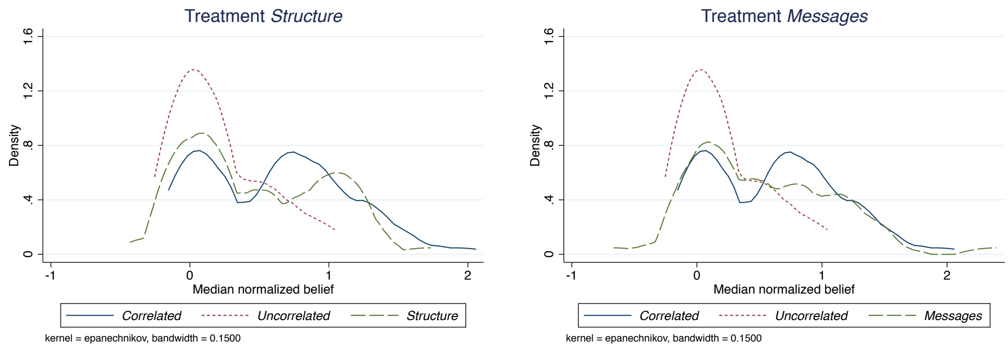


Figure 2.24. Kernel density estimates of median normalized beliefs in the *Structure* and the *Messages* treatments, each compared with median normalized beliefs in the baseline *Correlated* and *Uncorrelated* treatments

Appendix 2.G Correlation Neglect in Newspaper Articles

2.G.1 Overview

In our main experiments, we deliberately designed an abstract decision environment which allowed tight control over (subjects' knowledge of) the data-generating process. To show the robustness of our findings, we now make use of a naturally occurring correlation in an informational context with which many subjects are familiar, i.e., extracting information from newspaper articles.

In the experiment, a new set of subjects had to estimate the growth of the German economy in 2012. For this purpose, subjects were provided with (shortened) real newspaper articles discussing and summarizing growth forecasts and were asked to give an incentivized estimate. Employing the same identification strategy as in our main experiment, we again study two main treatments, one in which information is correlated and one in which it is not. In the correlated treatment, subjects received two articles. The first article discussed a joint forecast from April 2012, which is determined in a cooperation of several German research institutes, thus aggregating information from the participating institutions. It predicted that the German economy would grow at a rate of 0.9 % in 2012. The other article discussed a forecast of one particular institute from March 2012 that predicted a growth rate of 1.3 %. Importantly for our purposes, this institute also participated in the joint forecast. Consequently, the information from that institute is already incorporated in the joint forecast, implying that the two articles are correlated. This correlation was in principle known (or easy to detect), since the article reporting the joint forecast clearly stated all participating institutes. In the control condition, we merely supplied the joint forecast. Since the individual forecast is incorporated in the joint one, the joint forecast is a sufficient statistic of mean beliefs, implying that this treatment removes the correlation, yet keeps the informational content identical.

The results show that even in this rather naturalistic setting subjects exhibit a substantial degree of correlation neglect. In the control condition, the median estimate was 0.82 %, while it was 0.28 percentage points higher in the correlated treatment (p-value < 0.0001, Wilcoxon ranksum test). This finding emphasizes the robustness of correlation neglect with respect to the familiarity of the belief formation task and suggests that people exhibit the bias even in natural informational environments - while subjects may not frequently be required to predict GDP growth as such, the type of information provided in these experiments is typical for everyday information processing.

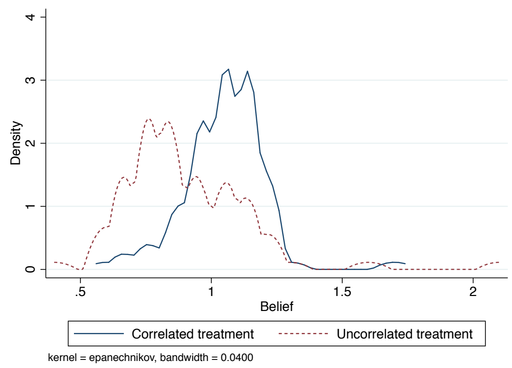


Figure 2.25. Kernel density estimates of beliefs in the two main newspaper treatments

2.G.2 Procedural Details

Overall, 151 subjects participated in the baseline experiments described above. 59 subjects took part in additional treatments (see below). Sessions were conducted using paper and pencil in the BonnEconLab at the end of different and unrelated experiments. Treatments were randomized within session. In the conditions involving two articles, the order of the articles was randomized. The study took five minutes on average. At the end of each session, one subject was randomly selected for payment. He was asked to write his address on an envelope and was reminded that his earnings will be sent to him as soon as the official growth figures are available. Earnings were 10 euros if the estimate turned out to be correct. For every 0.1 percentage point deviation, 1 euro was deducted. Negative earnings were not possible. The randomly selected subjects earned 7.30 euros on average.

2.G.3 Potential Concerns and Additional Treatments

There are five potential concerns with respect to our design. First, one could argue that the difference between the joint forecast of 0.9 % and the forecast of 1.3 % is informative because it indicates a high variance of forecasts. This variance in turn might allow inference about the signal precision of the participating institutes. Consequently, subjects in the correlated condition could put lower weight on the forecasts (relative to their own prior) when determining their estimate. Notice, however, that even if subjects actually went through this kind of inference, this would not explain our treatment difference. The estimates in our control condition reveal that subjects' priors were on average actually slightly below the joint forecast of 0.9 %. Thus, lower weight on the joint forecast in the updating process would not lead to estimates that are closer to 1.3 %.

A further potential concern might be that information from the second article is informative if subjects think that the forecast of the institute that is discussed in this article is not appropriately incorporated in the joint forecast. This does not seem plausible. However, to further address this issue, we asked a subset of subjects ($N = 56$) at the end of the experiment if they had the suspicion that this is actually the case. Only seven subjects (12.5 %) indicated such a concern. Our findings remain unchanged if we only consider those 23 subjects which explicitly stated that this was not a concern (p-value = 0.0209, Wilcoxon ranksum test).[45]

Third, subjects could interpret the mere presentation of the article discussing the forecast of 1.3 % as an indication that the article has to be of informational value. We addressed this concern by introducing an additional treatment ($N = 59$), which is identical to the correlated treatment except that it contains a second incentivized question which relates to labor market information provided in the article discussing the 1.3 % forecast.[46] Thus, there was a natural reason for the presence of the second article, which was unrelated to the question about GDP growth. Results suggest that this type of effect does not drive our results. Estimates in this treatment are almost identical to those in the standard correlated condition and significantly different from those in the control condition (p-value < 0.0001, Wilcoxon ranksum test).

Fourth, the two forecasts were published one month apart from each other. This is unproblematic, however, since the joint forecast was released at the later date. Thus, the timing as such provided no reason for subjects to place any weight on the 1.3 % forecast.

Fifth, it is possible that many subjects are not used to extracting information from newspapers, thus contradicting the purpose of our study as reflecting a more natural belief formation context. In order to ensure that this is not the case, we asked subjects at the end of the experiment whether they regularly read the newspaper, and whether they are interested in economics or economic questions. 57 percent of subjects stated that they "regularly" or "very regularly" read the newspaper. Also, 53 percent stated that they were "interested" or "very interested" in economic questions. Our treatment difference remains unchanged when we only consider subjects who regularly read the newspaper and who are interested in economic topics ($N = 74$), p-value < 0.0001, Wilcoxon ranksum test.

2.G.4 Newspaper Articles and Instructions

2.G.4.1 Paper-Based Instructions

Please read the following newspaper article(s). Please then think about how much the German economy will grow in 2012. Below you can indicate your estimate. Your payment will depend on how close your estimate is to the actual growth of the German economy.

---

[45] The precise wording of the question is: "Do you think that one of the research institutes (e.g. the IWH) was not adequately taken into account in the preparation of the joint forecast? Yes / No / Don't know"

[46] The precise wording of this second incentivized question is: "Please also think about whether the Institute for Economic Research Halle (IWH) predicts a positive development of the labor market. Below you can indicate your answer by ticking "Yes" or "No". You get 7 euros for a correct answer and 0 euros otherwise."

Maximum earnings are 10 euros - for every 0.1 percentage deviation, 1 euro will be deducted (negative earnings are not possible).

Your estimate: The growth of the German economy in 2012 will be (in percent): ...

2.G.4.2 Newspaper Articles (translated into English)

Manager-Magazin, 14.03.2012

IWH increases growth forecast

The German economy seems to be gaining speed. According to the Institute for Economic Research Halle, the short period of economic weakness is over. Thus, the researchers increase their growth forecast for Germany significantly.

On Wednesday, the institute in Halle announced that it expects the German economy to grow by 1.3 % this year. According to the IWH experts, the risks relating to the debt and trust crisis in Europe have been slightly reduced. Both the world economy and the German economy are said to have started significantly better into 2012 than was projected in autumn 2011. According to the IWH, the positive economic development will also affect the labor market.

Welt Online, 19.04.2012

Leading economic research institutes say German economy is in upswing

According to leading economic research institutes, the German economy is in upswing. In their joint “Spring 2012” forecast, published on Thursday, the institutes forecast a growth of the German economy of 0.9 %.

According to the researchers, the biggest “down-side risk” for the future remains to be the debt and trust crisis in the Euro area. While the remarkable measures of the European Central Bank relieved stress in the banking system, they are not more than a gain of time.

The forecast is prepared by the Ifo Institute in Munich, the ETH Zurich, the ZEW Mannheim, the Institute for Economic Research Halle, Kiel Economics, IHS Vienna, and the RWI Institute in Essen.

3

What You See Is All There Is*

3.1 Introduction

In forming economic, social, and political beliefs, people frequently need to process information that is selected in favor of their prior views: whenever an individual's information induces them to enter some environment, other people in this context are likely to be selected based on similar information. Thus, people's local information sample tends to reinforce their initial beliefs. For instance, if people believe a certain profession (or graduate program!) to be particularly meaningful, they enter a corresponding work environment and update beliefs by communicating with co-workers; if adolescents believe the returns to tertiary education to be high, at universities they find themselves surrounded by people with similarly rosy beliefs; if voters perceive the Republican candidate to be promising for economic prospects, they predominantly attend Republican election gatherings to support their favorite candidate. In all of these contexts, other people are likely to be selected based on the same respective mechanism, implying that people tend to be disproportionately connected to those with similar views, a phenomenon commonly referred to as (belief-based) homophily (McPherson et al., 2001). While in the above cases selection problems emerge indirectly because people with similar information tend to enter the same environments, in other contexts people may even intentionally opt for selected information sources. For example, Gentzkow and Shapiro (2006) develop a model to show that consumers rationally prefer to obtain signals from information sources that are biased towards their own beliefs because accordance with one's priors indicates higher quality of the source. Consistent with this, the empirical evidence suggests that people indeed consume political news and sort into social circles based on their ex ante views (Gentzkow and Shapiro, 2011; Pew Research Center, 2014).

Regardless of whether assortative matching on information is intentional or coincidental, a key implication of this phenomenon for belief updating is that *A* needs to infer *B*'s information from the fact that *B* is not part of *A*'s immediate informational environment, i.e., to *draw an inference from something A does not see*. For instance, if *B* does not

* For helpful discussions and comments I thank Doug Bernheim, Stefano DellaVigna, Thomas Dohmen, Tilman Drerup, Armin Falk, Ulrike Malmendier, Muriel Niederle, Frederik Schwerter, Andrei Shleifer, David Yang, and Florian Zimmermann. Financial support through the Center for Economics and Neuroscience Bonn is gratefully acknowledged.

attend university, then perhaps he received a private signal indicating that the return to education is low. If people neglect such homophily-driven selection effects, their beliefs will be biased in the direction of their immediate social environment, giving rise to popular concerns about "echo chambers" driving belief polarization across social groups (see, e.g., Sunstein, 2009; Bishop, 2009; Pariser, 2011, for narratives along these lines).[1]

However, despite the abundance of information-based selection effects in economic and social life, little is known empirically about the implications of homophily or selected information for the evolution of beliefs. In this paper, I provide the first systematic investigation of these issues by focusing on people's cognitive limitations in processing the associated selection problem. The inquiry is based on the following set of questions: at the most fundamental level, do people appreciate that their local communication network might form a non-representative sample of the available information, so that they should not base their beliefs on "what they see"? If people do not adequately process this problem, what are the precise and possibly heterogeneous updating heuristics they employ? If people differ in their capacity to deal with the selection problem, then how does the interaction of different updating types shape individual and group-level beliefs? And finally, what are the cognitive roots of selection neglect and how does the bias relate to other errors in statistical reasoning that people exhibit in social learning environments?

To address these questions, this paper provides experimental evidence for five key facts. First, on average, people neglect the selection problem induced through homophily. Second, subjects exhibit systematic heterogeneity in updating rules: people either fully account for homophily or do not adjust for it at all. Third, when subjects with heterogeneous updating rules interact, little learning takes place, hence generating persistent disagreement among people with different belief formation rules. Fourth, limited attention is a key driver of the bias. Fifth, selection neglect is strongly correlated with another error people exhibit in social learning environments, i.e., the neglect of double-counting problems.

The empirical investigation starts by providing clean evidence for people's neglect of selection effects in information sources. Identifying such a bias is challenging in that it requires not only exogenous variation in sample selection mechanisms, but also that people know the data-generating process, i.e., that they can understand the properties of the signals they do not have access to due to the selection problem. To achieve this goal, this paper proposes a novel individual decision-making experimental design in which subjects have to estimate an unknown continuous state of the world and are paid for accuracy. In the beginning, a participant as well as five computer players each obtain a private signal over the state, and then select into one out of two groups based on whether the signal is relatively high or low. Thus, the two groups exhibit strong information-based homophily. Subjects are then provided with the option to update their beliefs by communicating with a subset of the computer players. This communication stage follows a simple and known selection rule: whenever subjects do not communicate with a given computer player, that player must have entered the opposite group. Thus, in this treatment, subjects predominantly talk to those with similar signal realizations and have to infer the expected signal

[1] In yet other cases, people might wish to select certain information sources for hedonic reasons. This is not the focus of this paper. For work on motivated reasoning see, e.g., Brunnermeier and Parker (2005), Bénabou (2013), Eil and Rao (2011), and Möbius et al. (2014).

of those players they "do not see" from the fact that these players entered the other group. Using computer players with known decision and communication rules ensures that subjects know the process generating the data. To cleanly identify selection neglect, in a between-subjects design, I also implement a control treatment in which subjects obtain signals of the same objective informational content as those in the main treatment condition, yet without the presence of a selection effect. I find that beliefs significantly differ across the two treatments because a substantial fraction of people in the treatment group act as if the signals they see constitute all available information. Thus, average beliefs in the selected treatment are systematically biased in favor of people's prior belief (their private signal), implying that these subjects earn lower financial rewards. These results illustrate how selection neglect may generate irrational path-dependence in beliefs: whenever a given belief induces people to select into some environment, the resulting local information sample is likely to be biased and – due to selection neglect – reinforces the belief upon which the selection decision was made. In consequence, beliefs within a given group tend to be too extreme, akin to common notions of belief polarization across groups.

In a second step, I characterize the precise and potentially heterogeneous updating rules people employ. To this end, I estimate an individual-level naïveté parameter which pins down subjects' updating rules in relation to Bayesian rationality. The distribution of updating types follows a pronounced bimodal pattern: subjects either fully account for the selection effect or do not adjust for it at all. The underlying individual-level data are strongly clustered around these two extreme belief formation types, emphasizing that different subjects employ fundamentally different belief formation rules. This bimodal type distribution contrasts with conceptualizations of bounded rationality as a continuous process which gives rise to many different levels of naïveté. In particular, in processing selection, a considerable fraction of people appears to follow a particularly simple heuristic of full neglect.[2]

The striking bimodal type distribution raises the immediate question of how people revise their beliefs when they meet others who hold different beliefs *despite symmetric information*. Do naïfs learn by observing the beliefs of their more rational counterparts? Alternatively, do the rational types revise their beliefs after learning that the majority holds different views? Or does neither type adapt their beliefs, implying persistent belief heterogeneity? While the present paper studies these questions in the context of selection neglect, the corresponding insights may well be relevant for bounded rationality-driven disagreement more generally. After all, to date, there is no systematic evidence on how people with heterogeneous updating rules interact, learn from each other, and theorize about others' errors, even independent of the particular bias under study. Rather, previous work has typically focused on individual belief formation in isolation. Accordingly, I investigate how people revise their beliefs when they interact with different updating rules, under common knowledge of symmetric information. This question comes in two complementary variants. First, if people can choose whom to communicate with (e.g., to seek advice), do they prefer those that share their own updating rule? For example,

[2] These findings are conceptually distinct from, yet potentially related to the pronounced type heterogeneity established in experimental analyses of strategic sophistication (e.g., C. F. Camerer et al., 2004; Costa-Gomes and Crawford, 2006; Crawford et al., 2013; Fragiadakis et al., 2013).

if naïfs preferred to listen to other naïfs, that would imply a form of entrenchment in irrational beliefs. Second, how do naïfs and rationals respond if they are (exogenously) confronted with belief heterogeneity in spite of symmetric information?

I begin the corresponding analysis by showing through two additional treatment variations that when given the explicit choice to obtain access to the beliefs of one out two candidate advisors, subjects overwhelmingly pick the advisor with whom they share the same updating rule, so that little belief convergence takes place.[3] In a second step, I exogenously confront subjects with disagreement. To this end, subjects first state a belief in one of the belief formation tasks from the baseline selected treatment. Then, they are provided with the beliefs of two other randomly selected subjects who had access to exactly the same information. This procedure generates groups of subjects whose beliefs reflect disagreement. Finally, beliefs are elicited again. I find that both rationals and naïfs have a strong tendency to trust their own assessment of the available information, rather than that of their peers. Only once the opposing evidence becomes unanimous, do the naïve types start adjusting towards the rational benchmark; in contrast, rational subjects very rarely revise their beliefs in the naïve direction. Taken together, both under endogenous and exogenous selection of communication partners does communication fail to induce meaningful convergence towards a consensus, suggesting that the boundedly rational processing of information may generate persistent disagreement.[4] While these results hold in the context of selected information, the corresponding insights may well apply more broadly in contexts in which people with heterogeneous updating rules interact. In this respect, the findings have a natural relationship to models of information aggregation. For example, network analysis often focuses on the aggregation of dispersed private information, e.g., the existence and properties of a consensus belief or the speed of convergence (e.g., Golub and Jackson, 2010, 2012; Acemoglu et al., 2010, 2011; Müller-Frank, 2013). The findings from this paper suggest that an additional problem of reaching a consensus in society might be the heterogeneous *processing* of signals, rather than private information per se.

In a final step, I explore the cognitive mechanisms underlying the striking bimodality in selection neglect. Understanding these mechanisms provides insights into potential ways to formalize the bias, and also allows insights about the types of environments that are more likely to give rise to selection neglect in applied settings. I begin by establishing a strong correlation between subjects' naïveté and their cognitive skills as proxied for by academic achievement in high school. To better understand why and how low cognitive skills produce selection neglect, I conceptualize belief formation in this context as a stylized two-step process (also see Enke and Zimmermann, 2015). First, people need to *notice* the systematic holes in their information sample; second, they need to mathematically back out the expected signals they do not see from the fact that they induced the respective computer players to enter a particular group. Which of these two steps do subjects struggle with, and why? I provide evidence that limited attention plays a key role in generating the inferential naïveté. First, through an incentivized follow-up question, I verify that subjects are in principle capable of computing simple conditional expectations

[3] See Schotter (2003) and Çelen et al. (2010) for studies on decision making under advice.

[4] These findings could be related to studies of overconfidence, see, e.g., C. Camerer and Lovallo (1999) or Burks et al. (2013).

when explicitly asked to do so, hence suggesting that for a majority of people the bias is not rooted in mathematical problems, but rather in subjects' excessive focus on "what they see". To bolster this interpretation, I conduct an additional treatment variation, in which subjects' attention is nudged towards the computer players they do not communicate with. This treatment greatly reduces the fraction of naïfs, again showing that many participants are capable of drawing correct inferences once they focus on thinking about the selection problem. These findings on the importance of subjects' focus also explain the strong bimodality in subjects' naïveté types. In particular, identifying the systematic holes in one's local information sample appears to introduce a binary threshold into the belief formation process: subjects either attend to the holes in their data and (fully) adjust for them, or they do not.

As emphasized by prior work, belief formation in networks is not only complicated by selection effects, but also by informational redundancies, i.e., potential double-counting problems (DeMarzo et al., 2003; Eyster and Rabin, 2010, 2014; Eyster et al., 2015). Enke and Zimmermann (2015) show experimentally that – similarly to the present paper – the individual-level distribution of naïveté with respect to double-counting problems is roughly bimodal, and attentional nudges to focus on the mechanics generating the redundancy have large effects on beliefs. Given these similarities between selection and correlation neglect, I re-invite subjects in the selected condition and ask them to solve a belief formation task with partially redundant signals. The resulting distribution of individual-level naïveté is highly correlated with individual's propensity to neglect selection effects, conditional on proxies for cognitive ability. Thus, neglecting double-counting and selection problems is both correlated within subjects and can be attenuated using the same treatment variation, hence suggesting that these biases in how people form beliefs in social networks might share common cognitive foundations. As argued by, e.g., Fudenberg (2006), studying the micro-foundations and inter-relationships of biases is important to provide conceptual inputs into models which seek to unify erroneous updating processes.

This paper ties into several literatures. The results on people's neglect of selection effects as induced by homophily provide direct evidence for the naïveté assumption underlying Golub and Jackson's (2012) investigation of belief dynamics in homophilous networks.[5] While a small set of contributions experimentally studies the evolution of beliefs in dynamic interactive network setups (e.g., Möbius et al., 2013; Chandrasekhar et al., 2015), these papers do not aim at identifying people's neglect of selection effects, the precise heuristics people employ, the underlying cognitive mechanisms, and how heterogeneous updating types interact.[6] In consequence, in two recent surveys, Möbius et al. (2014) and Choi et al. (2015) explicitly call for more systematic experimental investigations of the precise updating rules underlying people's behavior in networks. The findings in this paper contribute to the networks literature by uncovering a novel belief bias as it arguably applies to a broad class of network problems, by providing the first systematic investigation of how people update their beliefs when they are confronted

[5] Also see Currarini et al. (2009) and Glaeser and Sunstein (2009).

[6] Grimm and Mengel (2014) show that higher homophily leads to slower convergence in experimental networks. However, in these experiments, subjects have no information about potential selection effects, so that the paper cannot address how people attempt to deal with homophily.

with belief heterogeneity that is not driven by private information, and by showing that two conceptually distinct errors in statistical reasoning in networks may be fruitfully modeled as representing the same underlying cognitive process.

The focus on identifying individual-level updating rules also makes this paper part of the experimental literature on boundedly rational belief updating and learning. Recent empirical contributions include Charness et al. (2010), Andreoni and Mylovanov (2012), and Brocas et al. (2014). Theoretical work on the relationship between attention and belief formation includes models of rational inattention (Caplin et al., 2011; Caplin and Dean, forthcoming) as well as models of heuristics (Gennaioli and Shleifer, 2010; Bordalo et al., 2015; Schwartzstein, 2014).[7] Rabin and Schrag (1999) provide a model of confirmation bias, which is an error distinct from selection neglect.

The remainder of the paper proceeds as follows. Section 3.2 studies the individual-level processing of selected information. Section 3.3 analyzes the evolution of beliefs in interactive environments. Sections 3.4 and 3.5 investigate the mechanisms underlying selection neglect and its relationship to correlation neglect. Section 3.6 concludes.

3.2 Selection Neglect in Information Sources

3.2.1 Experimental Design

Studying belief updating in homophilous environments requires (i) an abstract task that allows to flesh out people's cognitive limitations and rules out affective reasons for holding certain beliefs, (ii) full control over the data-generating process, (iii) exogenous manipulation of the degree of homophily, (iv) a control condition that serves as benchmark for updating in the absence of selected information, and (v) incentive-compatible belief elicitation. Most importantly, however, a clean identification requires subjects' full knowledge of the data-generating process, i.e., a setup in which we know that subjects can in principle understand the statistical properties of those signals they do not see due to the selection mechanism. The present between-subjects design accommodates all these features.

The key idea of the design is to construct two sets of signals which result in the same Bayesian posterior, where one information structure introduces a selection effect through information-based homophily. Subjects were asked to estimate an ex ante unknown state of the world μ and were paid for accuracy. First, the computer generated μ; to this end, the computer drew 15 times with replacement from the set $X = \{50, 70, 90, 110, 130, 150\}$. The average of these 15 draws then constitued the true state. Second, the computer generated six signals about the state. Let Y denote the set of 15 numbers that determine the state. The computer generated six signals $s_{1,\ldots,6}$ by randomly drawing from Y, without replacement. Thus, ex ante, each signal is indepen-

[7] In a paper with a different focus than the present one, Esponda and Vespa (2015) study belief formation under endogenous sample selection when the process generating the data is deliberately left unknown to subjects. Similarly, Brenner et al. (1996), Schkade et al. (2007), and Koehler and Mercer (2009) investigate updating under selected information using non-incentivized qualitative questionnaires on hypothetical scenarios, in which again the data-generating process is unknown.

dently and uniformly distributed over the set X. Note that there is residual uncertainty over μ even conditional on having access to all signals.

In the course of the experiment, a subject interacted with five computer players (called players I-V). The experimental task consisted of two stages. First, a subject as well as each of the five computer players privately observed one of the six signals and selected into a group based on the respective signal, which introduces homophily. Second, subjects communicated with the computer players and stated a belief over μ.

Specifically, in the first stage, subjects had to decide upon their group membership (blue or red group) based on their signal. As detailed below, the payoff structure was such that subjects earned higher profits as member of the blue group if $\mu < 100$ and of the red group provided that $\mu > 100$. Given this payoff structure, it was rather obvious for subjects which group to enter, and I show below that subjects indeed almost always entered the red group if their private signal was larger than 100 and the blue group otherwise. The five computer players similarly decided on their group membership using a *known* decision rule, i.e., these players opted for the blue (red) group if their private signal was smaller (higher) than 100. Notice that after this first stage, the two groups exhibit strong information-based homophily.

In the second stage, subjects communicated with (some of) the computer players to gather additional information about the state, i.e., subjects obtained the private signals of some computer players. The only difference between the *Selected* and the *Control* treatment consisted of the information subjects received from the computer players. In the *Selected* treatment, subjects talked to all computer players in their own group, but at least with three computers. Thus, for instance, if a subject's group contained only one computer player, they obtained the signal of that player and of two randomly chosen players from the other group. If a subject's group contained four players, a subject communicated (only) with these four. It was made clear to subjects that whenever they did not talk to a particular player, it would have to be that this player entered the opposite group. Thus, subjects could easily infer the number of players in each group. Note that given the simplified discretized uniform distribution over the signal space, it was rather straightforward for subjects to infer which types of signals they were missing. This provides a crucial input into a design attempting to identify selection neglect, because it ensures that subjects can in principle understand the statistical properties of the signals they do not have access to – after all, if people cannot possibly know which signals they are missing, it is difficult to speak of "selection neglect". In particular, in this setup, whenever a subject was in the red (blue) group, a missing signal was a 70 (130), in expectation.

In the *Control* condition, participants received the same signals as subjects in the *Selected* treatment, but additionally obtained a coarse version of the signals of the computer players that subjects in the selected condition did not communicate with. Specifically, if the signal of these additional computer players was in $\{50, 70, 90\}$, the respective player communicated 70 to the subject, while if the signal was in $\{110, 130, 150\}$, the computer communicated 130. Given that these coarse messages equal the expected signal condi-

Table 3.1. Overview of the experimental tasks

True State	Private signal	Observed Signal A	Observed Signal B	Observed Signal C	Observed Signal D	Unobserved Signal E	Unobserved Signal F	Rational Belief	Naïve Belief
92.66	130	110	90	70	–	50	90	90.00	100.00
106.00	130	130	150	110	–	90	50	110.00	130.00
112.67	50	70	150	130	–	110	110	110.00	100.00
85.93	110	130	110	70	–	70	90	93.33	105.00
98.00	90	70	70	90	90	130	–	90.00	82.00
95.33	130	90	150	90	–	50	70	100.00	115.00
107.33	70	90	90	110	–	110	150	103.33	90.00

Notes. Overview of the belief formation tasks in order of appearance. The categorization into observed and unobserved messages applies to the case in which subjects follow their private signal, i.e., opt for the red group if their signal was larger than 100, and for the blue group otherwise. Subjects in the *Selected* treatment observed only their own signal as well as the "observed" messages. Subjects in the *Control* condition additionally had access to a coarse version of the "unobserved" messages, i.e., if the corresponding signal was less than 100, they saw 70, and if the signal was larger than 100, they saw 130. See Section 3.2.2 for a derivation of the rational and naïve belief benchmark.

tional on group membership, the informational content of the *Selected* and the *Control* treatments is identical.[8]

Subjects completed seven independent tasks without feedback in between. All subjects solved the same tasks, summarized in Table 3.1. For instance, in the first task, subjects' private signal was 130, so that the optimal choice in the first decision was to opt for the red group. Here, subjects in the *Selected* condition would meet three computer players that obtained signals 110, 90, and 70, i.e., subjects communicated with one player from their own red group and two from the blue group. The remaining two computer players received private signals of 50 and 90, respectively. While subjects in the *Selected* condition did not communicate with these players, those in the *Control* condition received coarse versions of these signals, i.e., 70 and 70.

Four features of this experimental environment are worth noting. First, the procedure induced homophily, i.e., a selected information sample akin to the examples discussed in the introduction. In particular, a subject's initial belief (induced through the private signal) and the subsequent group entry decision determined their communication structure in the sense that the sample of communication partners consisted predominantly of computers that obtained similar signals. Second, subjects' knowledge that they would talk to every computer player in their own group allowed participants to infer which types of observations they were missing. For example, if a subject was in the blue group and one computer did not talk to them, they knew that this computer had opted for the red group. Third, drawing signals from a simplified discretized uniform distribution ensures that computing the conditional expectation of the missing signals is rather straightforward and can be done, e.g., by choosing the middle option conditional on being above or below 100. Finally, the full data-generating process was exogenous and known to subjects. Given that the other players were simulated by computers, subjects knew how to interpret the computers' messages and group entrance decisions.

[8] The control condition not only reminds people of the selection problem, but also computes the conditional expectations of the missing signals. An alternative design choice would have been to just tell people that, e.g., "computer player XY entered the red group". I chose to tell subjects the conditional expectation because this eliminates the entire selection problem, also in terms of the underlying mathematics. In any case, I verify below that the large majority of subjects are indeed themselves capable of computing this conditional expectation.

A comprehensive set of control questions ensured that subjects understood the key aspects of the process generating their data. Most importantly, subjects were asked what they knew about a computer player's private signal if they were in the red group, but did not communicate with that computer player, i.e., that this computer player must have obtained a private signal of less than 100 and hence opted for the blue group. Only once subjects had correctly solved all post-instruction questionnaire items could they proceed to the main tasks.[9] In the belief formation stage, all beliefs were restricted to be in $[0, 200]$ by the computer program. Appendix F contains the experimental instructions and control questions.[10]

The experiments were conducted at the BonnEconLab of the University of Bonn and computerized using zTree (Fischbacher, 2007). 78 student subjects participated in these two treatments (48 in *Selected* and 30 in *Control*) and earned an average of 11.60 euros including a 4 euros show-up fee. After the written instructions were distributed, subjects had 15 minutes to accommodate themselves with the task. Upon completion of the control questions, subjects entered the first task. Each task consisted of two computer screens. On the first one, subjects were informed of their private signal and decided which group to enter. On the second screen, participants received the computer players' signals and stated a point belief. Both decisions were financially incentivized, in expectation: in total, subjects took 14 decisions (seven on which group to enter and seven belief statements), one of which was selected for payment, which constitutes the best incentive mechanism in such a setup (Azrieli et al., 2015). The probability that a belief was randomly selected for payment was 80%, while a group membership was chosen with probability 20%. Beliefs were incentivized using a quadratic scoring rule with maximum variable earnings of 18 euros, i.e., variable earnings in a given task j equalled $\pi^j = max\{0; 18 - 2 \times (b^j - t^j)^2\}$, where b denotes the belief and t the state. Across the seven tasks, the average financial incentives to hold rational (relative to fully naïve) beliefs were roughly 12 euros, i.e., the marginal incentives to be rational were large. Payments for the group entrance decision were 12 euros if the subject opted for the red (blue) group when $\mu > 100$ ($\mu < 100$), and 2 euros otherwise, i.e., subjects had incentives to opt for the red (blue) group if their signal was high (low).

3.2.2 Hypotheses

Given true state $\mu = \frac{\sum_{k=1}^{15} m_k}{15}$, for $m_k \in \{50, 70, 90, 110, 130, 150\}$ with probability $1/6$ each, the signals $s_i = m_k$ for some k and $i \in \{1, \ldots, 6\}$ are unbiased. In what follows, I will distinguish between signals and messages. Given a set of six signals, the messages are given by $r_i = s_i$ if a subject communicates with the computer player who obtained the respective signal and $r_i = \emptyset$ otherwise. Let N denote the number of messages a subject actually sees, i.e., the number of communi-

[9] The control questions followed a multiple choice format, with 3-4 questions per screen. Thus, trial-and-error was very cumbersome. Moreover, the BonnEconLab has a control room in which the experimenter can monitor the decision screens of all experimental subjects. Thus, whenever a subject appeared to have problems in answering the control questions, an experimenter approached the subject, clarified open questions (if any) and excluded the subject from the experiment if they did not appear to understand the instructions.

[10] The instructions can also be accessed at https://sites.google.com/site/benjaminenke/.

cation partners. In the present setup, $E(s_i|\text{computer player in red group}) = 130$ and $E(s_i|\text{computer player in blue group}) = 70$. Given the messages, a Bayesian agent would compute the mean posterior belief b_B as

$$\begin{aligned} b_B &= E[\mu] \\ &= \frac{\sum_{i=1}^{N} s_i + \sum_{l=N+1}^{6} E[s_l|\text{computer player in blue or red group}] + E[m] \times 9}{15} \end{aligned} \tag{3.1}$$

where s_i denotes an observed signal and s_l an unobserved one. The second term in the numerator denotes the expectation of a signal conditional on the signal recipient entering a certain group. The third term in the numerator reflects the base rate $E[m] = 100$. However, starting with Grether (1980), a long stream of research has shown that people tend to neglect the base rate, especially in continuous-signal setups like the present one (Enke and Zimmermann, 2015). I thus define an alternative "rational" benchmark (in the sense of absence of selection neglect) b_R as

$$b_R = \frac{\sum_{i=1}^{N} s_i + \sum_{l=N+1}^{6} E[s_l|\text{computer player in blue or red group}]}{6} \tag{3.2}$$

That is, in the language of this paper, the rational benchmark ignores the base rate, but takes into account the systematic holes among the messages. None of the results in this paper will depend on this normalization; it only serves to illustrate the precise distribution of individual-level naïveté in processing selected information. Without assuming base rate neglect, any estimator for the naïveté parameter would be severely biased if people actually neglect the base rate (as they do, see below).

Now imagine that people neglect the selection problem, so that they merely base their beliefs on "what they see". Let $\chi \in [0, 1]$ parameterize the degree of naïveté with respect to selection such that $\chi = 1$ implies full selection neglect. I then define a selection neglect posterior b_{SN} as weighted average between the rational belief b_R and a fully naïve belief b_N, which consists of averaging all N visible signals:

$$\begin{aligned} b_{SN} &= (1-\chi)b_R + \chi b_N = (1-\chi)b_R + \chi \frac{\sum_{i=1}^{N} s_i}{N} \\ &= b_R + \chi \frac{6-N}{6}(\bar{s} - E[s_l|\text{computer player in blue or red group}]) \end{aligned} \tag{3.3}$$

where $\bar{s} \equiv 1/N \sum_{i=1}^{N} s_i$ denotes the average visible signal. That is, this selection neglect belief consists of a linear combination of the rational and the fully naïve belief. It is very flexible in that it allows for an arbitrary amount of naïveté χ, rather than just the two extreme benchmarks of rational beliefs or full neglect. In the results to be developed below, I will place special emphasis on identifying the distribution of this naïveté parameter.

Table 3.2. Overview of beliefs across tasks

True State	Private Signal	Rational Belief	Naïve Belief	Median Belief *Control* Treatment	Median Belief *Selected* Treatment	Median belief bias	p-value (Ranksum test)
92.66	High	90.00	100.00	90.00	100.00	10.00	0.0091
106.00	High	110.00	130.00	110.00	128.00	18.00	0.0001
112.67	Low	110.00	100.00	110.00	108.00	-2	0.0333
85.93	High	93.33	105.00	93.15	105.00	11.85	0.0001
98.00	Low	90.00	82.00	90.00	85.00	-5	0.0409
95.33	High	100.00	115.00	100.00	107.50	7.50	0.0001
107.33	Low	103.33	90.00	103.00	91.50	-11.50	0.0178

Notes. Overview of the estimation tasks in order of appearance. See Table 3.1 for details on the signals in each task as well as the computation of the rational and the naïve benchmarks. High (low) private signals are defined as signals above (below) 100.

Hypothesis 3. *Assuming that $\chi > 0$ (and $N < 6$), subjects' beliefs in the Selected condition are too high relative to the Control condition if the average of the visible signals is higher than the expected signal of the non-visible signals, and vice versa.*

This hypothesis says that the beliefs of naïve subjects will be upward biased if their private signal is higher than 100, so that subjects opt for the red group, implying that they will also see all other high signals ($s > 100$), but not all low signals.

3.2.3 Results

Result 10. *Beliefs significantly differ across treatments in the direction predicted by selection neglect. Consequently, beliefs in the Selected condition exhibit irrational path-dependence and are too extreme relative to the rational benchmark.*

Table 3.2 presents an overview of the results in each of the seven independent belief formation tasks. For ease of comparison, I provide the benchmarks of full selection neglect and rational beliefs, respectively. First note that, across tasks and treatments, virtually all subjects always enter the group that corresponds to their private signal realization.[11] Regarding subjects' beliefs, the results show that, reassuringly, (median) beliefs in the *Control* condition follow the rational prediction very closely, suggesting that the experimental setup was not systematically misconstrued by subjects: in the absence of homophily, people state rational beliefs. In the *Selected* treatment, however, median beliefs are always distorted away from the rational benchmark towards the full selection neglect belief. In all seven tasks do beliefs significantly differ across treatments at the 5% level (Wilcoxon ranskum test).[12]

To grasp the most basic implication of this belief bias, compare the second and seventh column of Table 3.2: whenever subjects' private signal is high ($s > 100$), the belief bias is positive. Conversely, when the initial private signal is low, the belief bias turns out negative. Thus, in essence, neglecting homophily-driven selection effects implies a form of irrational path-dependence: given a high prior belief (private signal), people select

[11] In total, in only 15 out of 546 group choice decisions did a subject enter the "wrong" group. In what follows, I exclude the beliefs from these particular subject-task combinations. All results are robust to including these observations or to entirely excluding subjects that entered the wrong group at least once.

[12] Appendix 3.B.1 visualizes the full distribution of beliefs in each task.

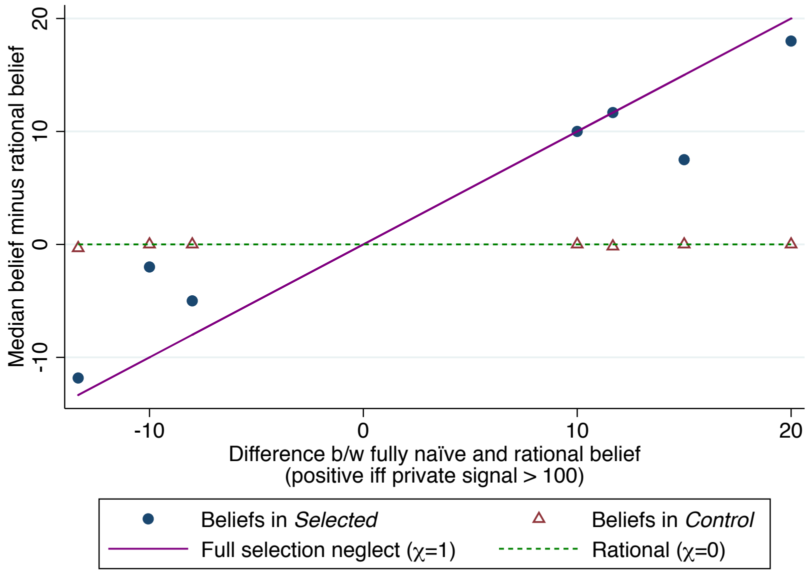


Figure 3.1. Relationship between actual and predicted belief bias. The x-axis represents the difference between the fully naïve and the rational belief as defined in Section 3.2.2. The y-axis represents the difference between the median belief in a given task and the rational prediction. Each dot represents a treatment-task combination. The figure provides evidence for path-dependence in treatment *Selected*: provided a high prior belief (private signal), the belief bias is positive and conversely.

into an environment which on average "reinforces" their prior views, if selection is not appropriately taken into account. Thus, beliefs in the red and blue group end up being too extreme (on average), akin to common notions of belief polarization across groups. Figure 3.1 visualizes this pattern. To construct this figure, for each true state, I compute the difference between the median belief in a given treatment and the rational prediction, and plot the resulting belief bias against the difference of the full selection neglect and the rational belief. Each observation represents one treatment-task combination. By construction of the figure, the rational prediction is a flat line at zero (no belief bias) and the full selection neglect prediction has a slope of one. Note that the x-axis assumes a positive value whenever subjects obtained a high signal and hence entered the red group: in this case, they mostly talked to computer players with high signals, resulting in a positive predicted belief bias. Conversely, if subjects received a signal below 100, the x-axis assumes a negative value. Figure 3.1 then shows that beliefs in *Control* follow the rational prediction very closely, but those in *Selected* are substantially upward (downward) biased depending on subjects' initial signal.

The large bias in statistical reasoning implies significantly lower earnings of subjects in the *Selected* condition. The expected profit from all seven belief formation tasks (i.e., the average hypothetical profit from each belief) is 5.00 euros in the *Selected* condition and 10.50 euros in the *Control* treatment, a statistically significant difference ($p < 0.0001$). Actual profits, which are partly based on subjects' group membership and include the show-up fee, are also significantly different from each other (13.70 vs. 10.10 euros, $p = 0.0628$, Wilcoxon ranksum test).

While these results show that subjects in the *Selected* condition do not adjust for the homophilous communication structure *on average*, such aggregate analyses reveal neither the precise quantitative degree to which subjects neglect selection nor the corresponding distribution of types. For instance, it is conceivable that all subjects intuitively adjust for the selection problem, but do not go "far enough" in debiasing their sample. On the other hand, the data might exhibit strong heterogeneity in the extent to which people can solve the problem. To investigate this issue, I proceed by estimating the individual-level naïveté parameter $\chi \in [0, 1]$ in equation (3.3) i.e., I quantify the extent to which a subject's beliefs reflect rational ($\chi = 0$), fully naïve ($\chi = 1$), or intermediate values. As a simple and transparent approach, for each subject and belief formation task, I compute the naïveté inherent in a belief and then employ the median of these seven naïveté values for further analysis.

Result 11. *In the Selected treatment, subjects' naïveté regarding the selection problem exhibits a strongly bimodal type distribution: people either fully take selection into account or do not adjust for it at all.*

The left panel of Figure 3.2 plots the distribution of median naïveté parameters for both the *Selected* and the *Control* condition. While beliefs in the *Control* condition are on average rational, as indicated by the large mass around zero, beliefs in the *Selected* condition exhibit a strongly bimodal distribution. While roughly 40% of participants are approximately rational ($\chi = 0$), the majority fully neglects the selection problem.[13] To show that the strong bimodality of types is not an artifact of the aggregation procedure of the seven beliefs per subject into one naïveté parameter, the right panel of Figure 3.2 depicts the distribution of the implied naïveté in all individual-level beliefs, i.e., seven beliefs per subject. Again, the data exhibit two large spikes at zero and one, respectively. Thus, subjects do not frequently partially adjust for selection (in particular when decision noise is taken into account). These results are noteworthy because they suggest that subjects' beliefs do not just reflect recklessness; rather, a considerable fraction of beliefs reveals that subjects approached the updating problem in a fundamentally mistaken way and *exactly* computed the fully naïve solution of $\chi = 1$. I will return to this issue in Section 3.4 when investigating the cognitive mechanisms underlying the bias.

[13] Appendix 3.B.2 shows that subjects' beliefs exhibit substantial consistency across tasks.

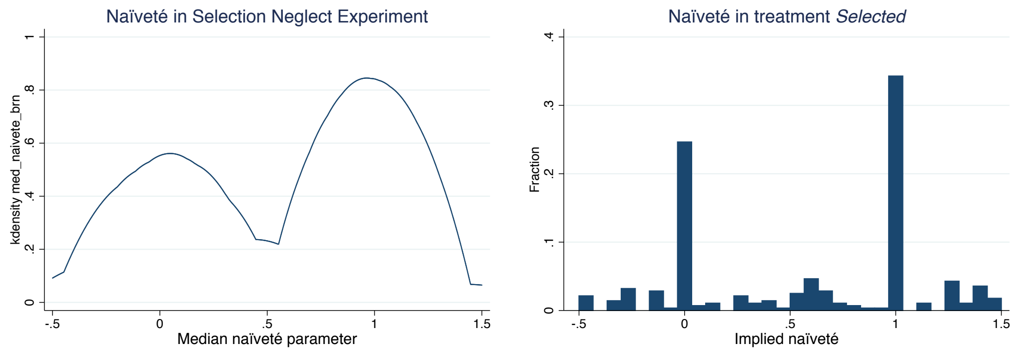


Figure 3.2. Distribution of naïveté in the *Selected* and the *Control* treatment. The left panel plots kernel density estimates of the median naïveté of each individual in both treatments (48 and 30 obs., respectively), while the right panel illustrates the distribution of naïveté implied in all individual-level beliefs in the *Selected* treatment (336 obs.).

3.3 Learning Through Interaction

The previous results highlight a pronounced heterogeneity in beliefs as arising from the presence of different updating types. However, the question of how different belief formation rules interact in shaping individual and societal beliefs has neither been adressed in theoretical nor in empirical research thus far, regardless of whether the context is selection neglect or another bias. Do the naïve types learn by observing the beliefs of their more rational counterparts? Alternatively, do the rational types revise their beliefs after observing that the majority holds different views? Or does neither type adapt their beliefs, implying persistent belief heterogeneity? To keep the setup as simple as possible, I ask how people revise their beliefs when they interact with potentially different updating rules, provided that everybody knows that everyone got the same signals. Studying this question comes in two complementary variants. First, if people can choose whom to communicate with (e.g., to seek advice), do they prefer those that share the same updating rule? Second, how do naïfs and rationals respond if they are (exogenously) confronted with belief heterogeneity in spite of symmetric information?

3.3.1 Seeking Advice – Endogenous Communication

In a first step, I examine where people tend to seek advice when given the choice. To investigate this issue, I implemented treatment *Advice*, which constitutes a simple variation of the *Selected* treatment. Subjects first completed three of the tasks in *Selected* so as to enable me to determine their naïveté χ. Then, they were unexpectedly interrupted by a computer screen which informed them that in the subsequent four tasks they would get access to the decisions of subject Y or Z (“advisors”) prior to making their own decisions.

More precisely, subjects would see the group entrance decision of the respective advisor (red or blue), then enter a group themselves, and finally observe the belief of the chosen advisor on the belief formation decision screen, along with the signals of the computer players.[14] The instructions clarified that the two potential advisors completed the same seven tasks a few weeks earlier while having access to the same information. In other words, participants knew that the two candidate advisors obtained the same private signal and communicated with the same computer players as subjects in *Advice*, implying symmetric information between advisors and advisees. In order to be able to make an informed choice between Y and Z, subjects were provided with a computer screen which contained all decisions from the first three tasks of themselves as well as of Y and Z.[15] Thus, subjects could evaluate how their own decisions compared to those of the potential advisors. Y and Z were selected such that one of them was fully naïve in all seven tasks and the other one rational in all tasks.[16] All subjects in *Advice* had access to the same two candidate advisors. After subjects had chosen their preferred advisor, they completed an additional four tasks. 59 subjects took part in this treatment, which lasted 50 minutes and yielded average earnings of 10.80 euros including a 6 euros show-up fee.[17]

In this treatment, subjects usually faced the beliefs of an advisor who stated similar beliefs to themselves and of another advisor who reported different beliefs. A perhaps natural conjecture is that the rational types understand that they are rational and hence choose the rational subject as advisor so as to reduce cognitive effort in the remaining four tasks, or to double-ckeck their own calculations against random computational errors. On the other hand, rational subjects may be uncertain about whether they pursued the correct problem-solving approach. Even more so, two competing hypotheses come to mind regarding the naïve subjects. First, just like the rational types, the naïfs may believe that they are rational and hence opt for the naïve advisor for the reasons discussed above. Second, however, the naïfs may have an intuitive feeling that their problem-solving strategy is somehow incorrect even though they weren't able to work out the correct solution themselves. Then, seeing someone state different beliefs may lead subjects to assume that (for whatever reason) this must be the correct solution.

Result 12. *Subjects overwhelmingly choose advisors whose decisions reflect their own belief formation rule. Thus, under endogenous communication, beliefs between the rational and naïve types do not converge in a meaningful way.*

Columns (1) and (2) of Table 3.3 show marginal effects at means in probit estimations of subjects' choice of advisor on their (median) naïveté parameter. Both with and without additional controls, higher naïveté is significantly associated with a higher

---

[14] Subjects saw the advisor's belief provided that the subject opted for the same group as the advisor. This restriction was put in place so as to ensure that subjects had no strategic incentives to opt for a group that contradicted their private signal.

[15] On this screen, the labeling of Y and Z and their location on the screen (left / right) were randomized across sessions. To investigate differences between fast and slow reasoning, I implemented two conditions in which subjects could not make a decision until 30 or 90 seconds after they had entered the "advisor choice" decision screen. The corresponding results are very similar, so I pool the data in what follows.

[16] No deception was used in the experimental instructions. In particular, the instructions informed subjects that Y and Z were two participants in a previous session, but it was never indicated that they were drawn at random.

[17] The show-up fee in all "interaction" treatments was 6 euros because they took slightly longer than the treatments reported above.

Table 3.3. Naïveté and choice of advisor

	Dependent variable: 1 if chose naïve advisor			
	Treatment *Advice*		Treatment *Advice only*	
	(1)	(2)	(3)	(4)
Median naïveté in first three tasks	0.33***	0.33***	0.68***	0.71***
	(0.11)	(0.11)	(0.14)	(0.16)
Additional controls	No	Yes	No	Yes
Observations	59	58	60	58

Probit estimates, robust standard errors in parentheses. The table reports partial effects at means. The base rate for the choice of the naïve advisor is 52.5% in *Advice* and 58.3% in *Advice only*. Additional controls include age, gender, log monthly income, marital status fixed effects, and high school grades. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

probability of choosing the naïve advisor. The left panel of Figure 3.3 illustrates these results by splitting the sample into subjects with $\chi \leq 0.5$ ("rationals") and $\chi > 0.5$ ("naïfs"). Here, the raw difference in the fraction who chose the naïve advisor is 39 percentage points, a statistically highly significant difference ($p = 0.0030$, two-sample test of proportions).[18]

Robustness and Extension

In treatment *Advice*, subjects have the option of choosing an advisor who states the same beliefs as them. It is conceivable that the naïve subjects seek advice in an assortative manner not because they believe that the naïve advisor is right, but rather because they may feel good if someone confirms their own assessments. Likewise, it is possible that the naïve types feel that their problem-solving strategy is incorrect, yet that they have no way of assessing how much better the strategy of the rational subject is. After all, if subjects only understand that they got it wrong, but do not know how to adequately solve the problem, they may not know whether the advisor who states different beliefs is actually superior.

In order to address these issues, I implemented treatment *Advice only*. This condition was identical to *Advice* except for three variations. First, in the advisor selection phase, subjects were not provided with the advisors' decisions from the first three tasks. Rather, subjects in *Advice only* were presented with the advisors' decisions in two belief formation tasks from the baseline *Selected* condition which subjects in *Advice only* did not complete themselves. That is, out of the seven tasks in *Selected*, subjects in *Advice only* completed three tasks without advice, two with advice, and two not at all. Accordingly, when making their decision among the advisors, they could not compare their own beliefs to the ones of the advisors (where again one advisor was essentially fully rational and one fully

[18] Subjects' propensity to choose an advisor of the their own type may depend on their confidence. To investigate this issue, I make use of a qualitative question that was asked after the first three tasks, i.e., before the choice of the advisor was introduced: "On a scale from 1 (not certain at all) to 10 (very certain), how certain are you that your previous estimates (and the underlying strategy) were correct?". In Appendix 3.E, I discuss this variable and its relationship to subjects' decisions in detail. I find that there is a moderate, statistically significant, correlation between subjects' naiïveté and their confidence ($\rho = -0.16$, $p = 0.0227$). However, the relationship between subjects' confidence and their choice of advisor is weak at best, for both rationals and naïfs.

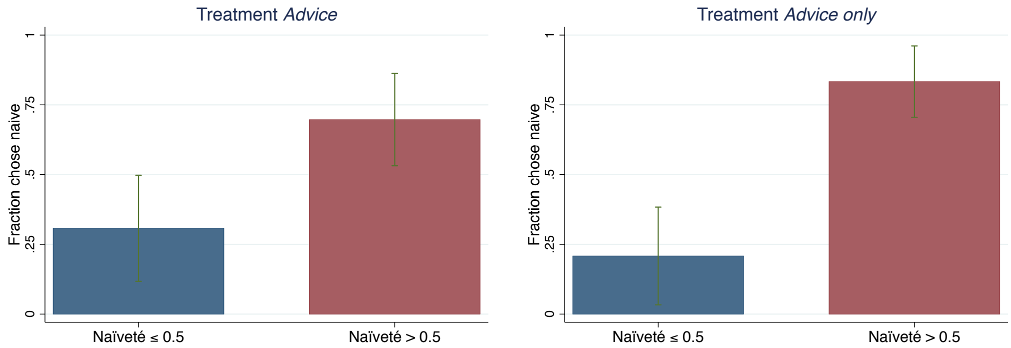


Figure 3.3. Fraction of naïve and rational subjects that chose the naïve advisor. The left panel illustrates the results from treatment *Advice* and the right panel those in treatment *Advice only*. In both panels, subjects are split according to whether their median naïveté χ in the first three tasks was larger or smaller than 0.5.

naïve). Rather, they had to work through the two tasks which they did not complete themselves in order to be able to assess the respective belief statements of the advisors. Second, and relatedly, subjects were told that the advisors' decisions would be the *only piece of information* in the subsequent tasks. In other words, in the remaining two tasks, subjects neither saw their own private signal nor did they communicate with any of the computer players. Thus, once they had chosen an advisor, subjects were essentially left only with following the group entrance decisions and belief statements of the chosen advisor. Importantly, note how these two changes to the design ensure that subjects indeed choose the advisor whom they assess to be superior, rather than someone whom they suspect to confirm their own assessments; after all, subjects did not state any beliefs when the advisors did, so that such affective reasons could play no role. Third, and finally, the instructions explicitly stated that one of the advisors solved all problems correctly. Thus, if the naïve types conjectured that their own strategy was wrong, they should immediately pick the rational advisor, even if they did not understand how the rational advisor developed their beliefs. 60 subjects participated in this treatment and earned 11.40 euros on average.

Columns (3) and (4) of Table 3.3 present the results, while the right panel of Figure 3.3 provides a graphical illustration. In short, the results are even stronger than those in *Advice*. Again, there is a strong and significant relationship between subjects' naïveté and their propensity to choose the naïve advisor. For instance, when I again split the sample into subjects with $\chi \leq 0.5$ and $\chi > 0.5$, the difference in the fraction who chose the naïve advisor is 62 percentage points ($p < 0.0001$, two-sample test of proportions).

In sum, across both treatments, subjects choose advisors in an assortative manner. I proceed by visualizing the resulting belief patterns, pooled across treatments *Advice* and *Advice only*. Figure 3.4 depicts the distribution of naïveté implied in all beliefs subjects stated when they had access to an advisor, partitioned by subjects' inherent naïveté type (as determined by the first three tasks without advice). The figure reveals that little belief convergence took place through the introduction of the advisors: the majority of rationals still states approximately rational beliefs, while the majority of naïfs remains naïve.[19] This suggests that, if people can choose whom to communicate with, pronounced belief heterogeneity may persist over time. Notably, this pattern obtains in the absence of hedonic motives, but rather because people talk to those they believe to have the correct problem-solving approach. At the same time, the relationship between subjects' beliefs with and without advice ought to be interpreted with care because subjects chose these advisors themselves. The following section discusses how subjects respond when they are confronted with the beliefs of others whom they did not select themselves.

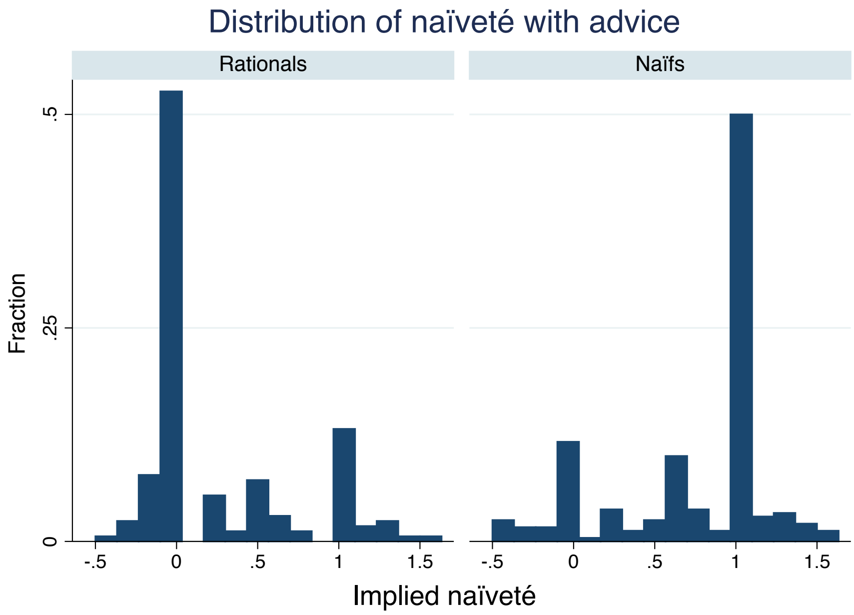


Figure 3.4. Distribution of decisions in the last four tasks (i.e., with advice), conditional on subjects' naïveté type in the first three tasks. Rationals are defined as $\chi \leq 0.5$ and naïfs as $\chi > 0.5$. The histograms exclude observations outside [-1,1.5].

3.3.2 Exogenously Induced Disagreement

In a second step, I investigate how the different types respond in their updating behavior once they are *forced* to listen to those with different beliefs. To analyze this issue, I implemented treatment *Exogenous*. Here, subjects again solved the seven tasks from the *Selected* treatment. The treatment consisted of two steps. First, subjects solved three tasks by themselves, which again allowed me to compute an out-of-sample measure of individual's naïveté to predict their subsequent behavior. In the remaining four tasks, subjects first again stated a belief.[20] Then, they were shown the beliefs of two other randomly drawn subjects ("senders") from the same experimental session. This random

[19] Appendix 3.C further analyzes the relationship between subjects' inherent naïveté, their choice of advisor, and the subsequent belief patterns.

[20] In these four tasks, subjects did not decide on their group membership. Rather, the computer decided for them that whenever their private signal was higher (lower) than 100, they entered the red (blue) group. This was done to ensure that subjects indeed had symmetric information.

matching was not constant across tasks. Rather, in each task, subjects saw the beliefs of two new (and potentially different) randomly drawn subjects. Importantly, all subjects not only solved the same tasks, they also had access to the same information, and the presence of symmetric information was made clear to participants. Subjects were then asked to state a second belief. To ensure that laziness does not affect the findings, subjects had to explicitly type in this second belief, rather than, e.g., confirm their first guess. 96 subjects took part in this condition and earned 11.60 euros on average.

I again normalize the data across tasks by computing the naïveté χ that is implied in each belief. The analysis begins by investigating the raw correlation between the naïveté implied in subjects' first and second beliefs in each of the four tasks, i.e., the beliefs before and after they saw the beliefs of the two senders. I focus on cases in which the first belief of the receiver differs from the beliefs of at least one sender in a meaningful way; after all, studying how people revise their beliefs necessitates the presence of at least partial disagreement. I define disagreement as a binary variable which equals one iff the receiver's belief differs from the belief of at least one sender in the sense that the implied naïveté of the receiver is $\chi \leq 0.5$ and that of at least one sender $\chi > 0.5$, or vice versa. Despite this disagreement, Figure 3.5 shows that pre- and post-communication beliefs exhibit a strong raw correlation ($\rho = 0.86$), providing a first piece of evidence that subjects' final belief was largely based on their own assessment of the available information, rather than the senders' beliefs. If subjects had predominantly revised their beliefs, the sophisticated types should have adjusted upwards in Figure 3.5, while the naïve types should have adjusted downwards. While the figure shows that the majority of adjustments indeed go in the expected direction, the large majority of people rarely revises their beliefs.

To provide a different perspective on this issue, I proceed by investigating how subjects revised their beliefs as a function (i) of the number of senders who state opposing views, and (ii) of the receiver's type; after all, rational and naïve types may differ in how they respond to others' solutions. Figure 3.6 presents histograms of subjects' belief revisions as a consequence of the senders' reports. To construct a measure of belief revision, I compute by how much closer the receiver's post-communication beliefs are to the average beliefs of the two senders, expressed as percentage of the pre-communication disagreement (measured as simple difference between the receiver's pre-communication belief and the two senders' average pre-communication belief). Thus, the belief revision measure describes by how much receivers altered their belief in response to the senders' beliefs, relative to how much they could have changed their beliefs given the senders' reports.

The figure provides an overview of belief revisions conditional on the receivers' updating type as well as on the number of senders whose beliefs significantly depart from the receiver's belief. To this end, I again use a coarser version of the naïveté parameter χ by calling receivers rational if both their (out-of-sample) median naïveté parameter from the first three tasks and the naïveté implied in the first belief of the respective tasks satisfy $\chi \leq 0.5$.[21] I define naïfs analogously with $\chi > 0.5$. For instance, the top left panel

[21] I use both the out-of-sample measure and the first belief to classify subjects to ensure that I do not falsely classify them as, e.g., rational merely because they (perhaps due to random errors) stated a rational belief in the respective task. Appendix 3.D.3 reports robustness checks.

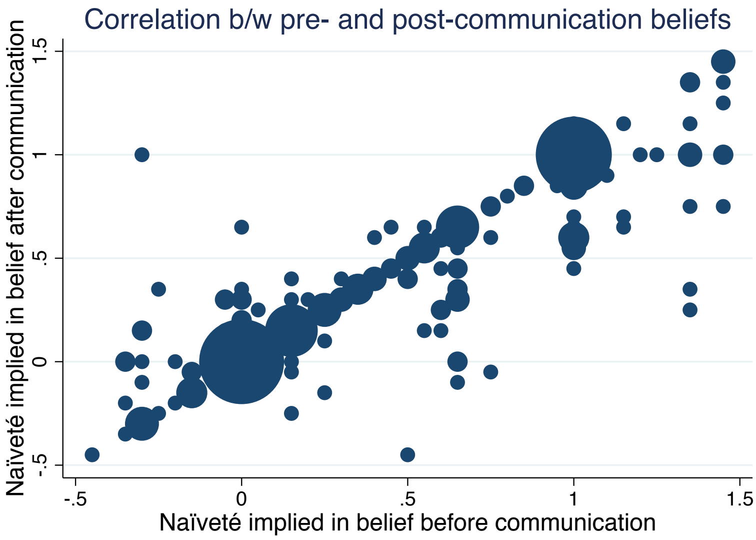


Figure 3.5. Raw correlation between the naïveté χ implied in first and second beliefs ($\rho = 0.86$). To construct this figure, subjects' pre- and post-communication naïveté is rounded to multiples of 0.05. The ball size then represents the number of observations in the respective bin. The scatter only includes observations for which there was at least partial disagreement, see the main text for details. Appendix 3.D.2 illustrates the raw correlation including the cases in which there was agreement. To ease readability, the scatter excludes 21 (out of 271) observations for which the implied naïveté of at least one belief is outside [-.5,1.5].

shows the belief adjustment of rational subjects who were confronted with one rational and one naïve sender.

The figure reveals that, consistent with the pattern reported above, subjects overwhelmingly abstain from adjusting their beliefs in response to the senders' assessments. When the senders report mixed beliefs (one rational and one naïve), the vast majority of both rationals and naïfs sticks with their own assessment, as indicated by the large spikes at belief revisions of 0%.[22] Thus, for instance, seeing one deviating response does not induce naïfs to reconsider their solution strategy. On the other hand, when subjects see two consistent beliefs that contradict their own estimate, the updating behavior differs markedly across types. While a large majority of rationals does not adjust their beliefs at all (see the top right panel), most naïfs start moving towards the rational senders (bottom right panel). This suggests that the rationals know that they are right, while at least some naïfs exhibit doubts once the evidence becomes sufficiently strong.

To analyze the preceeding patterns more rigorously, in column (1) of Table 3.4, I regress the naïveté χ implied in subjects' second belief (i.e., the belief subjects stated after they saw the beliefs of the senders) on the naïveté implied in subjects' first belief, for each subject and task. Column (2) regresses subjects' second belief on the average naïveté of the two senders.[23] Results show that, on average, subjects react only very

[22] These results may be related to studies of overconfidence (e.g., C. Camerer and Lovallo, 1999; Burks et al., 2013).

[23] I employ the average naïveté of the two senders for expositional convenience only. All results are robust to using the two measures separately.

weakly to the beliefs of their peers. While their own assessment of the available evidence explains 77.0% of the variation in the second beliefs, the beliefs of the peers only ex-

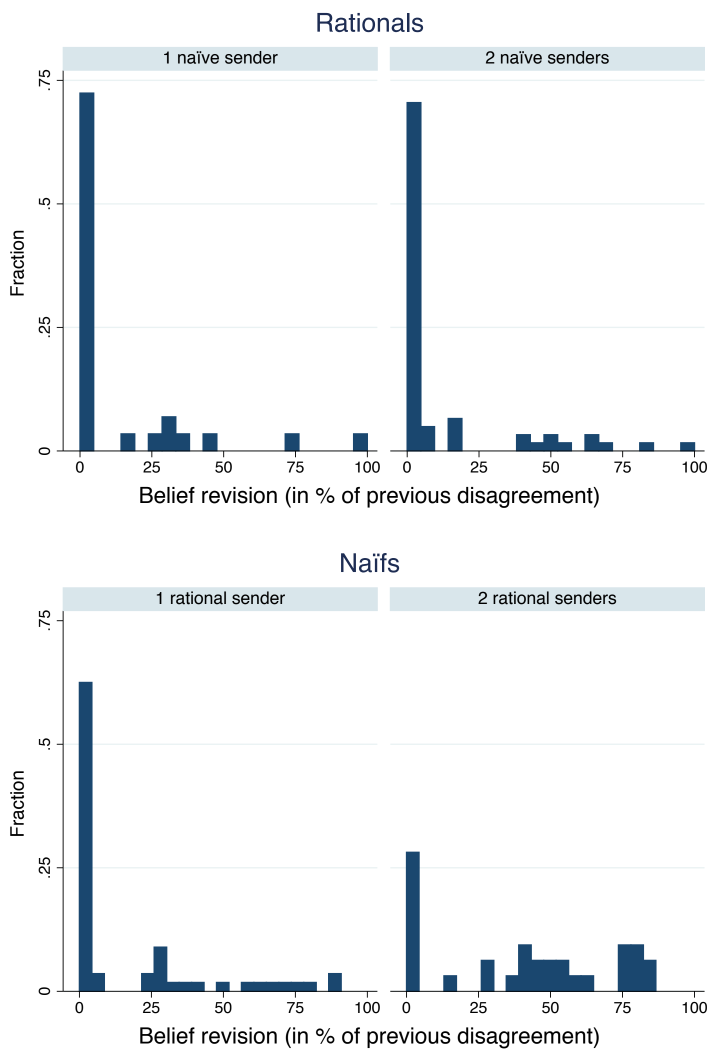


Figure 3.6. Magnitude of belief revisions. Each histogram denotes the belief revision between the first and second belief (expressed as percent of the difference between the first belief and the average belief of the two senders) conditional on the type of the subject (top / bottom panel) and on the composition of the senders. The top left panel shows the adjustment of rational subjects who face the beliefs of one naïve and one rational sender, while the top right panel illustrates the rational types' belief revision if they faced two naïfs. The bottom left panel depicts the adjustment behavior of naïfs when they faced one rational and one naïve belief, while the right panel illustrates adjustment in case of two rational senders. For a given subject and task, a subject ("receiver") is classified as rational if both the out-of-sample median naïveté parameter from the first three tasks and the first belief statement in the respective task are "rational" (i.e., $\chi \leq 0.5$), and analogously for naïfs ($\chi > 0.5$). Very similar results obtain when I define rationals and naïfs exclusively based on the out-of-sample naïveté measure or exclusively based on the first belief in the respective task, see Appendix 3.D.3. Adjustments $> 100\%$ and $< 0\%$ are excluded to ease readability (7 out of 185 obs.).

plain a miniscule 4.8%. Column (4) investigates whether the weight subjects put on other people's beliefs depends on the degree of agreement among the senders. To this end, I regress subjects' naïveté on the previously discussed variables as well as on (i) the degree of disagreement among the senders, and (ii) an interaction of the degree of disagreement with the average naïveté of the senders. Disagreement is defined as the absolute difference between the naïveté implied in the beliefs of the two senders. Results show that, consistent with intuition, subjects indeed place higher weight on the beliefs of their peers if disagreement is smaller: the negative and statistically significant interaction coefficient says that higher disagreement leads to a lower weight on the senders.

Columns (5)-(8) and (9)-(12) break these patterns down between rationals and naïfs. Notably, as suggested by Figure 3.6, the rational subjects' post-communication beliefs are not significantly correlated with the average naïveté of the senders, see columns (6)-(8). In addition, consistent with the visual evidence presented above, rationals never respond to the beliefs of their peers, regardless of whether they exhibit agreement or not. In contrast, naïfs partly respond to others' beliefs, albeit to a rather small extent: the variance in subjects' beliefs that can be explained by the beliefs of their peers is only 17.6% (column (10)), compared to 42.3% for their own pre-communication beliefs (column (9)).

Appendix 3.D.5 investigates learning over time. In particular, it is conceivable that those naïve subjects who revised their beliefs according to the beliefs of the senders state more rational beliefs in subsequent tasks. However, this is not the case, perhaps suggesting that while some subjects intuit that their strategy is incorrect, they are incapable of developing a better strategy themselves.

Result 13. *People have a strong propensity to trust their own assessment of the available evidence, rather than that of their peers. In consequence, hearing other people's beliefs does not induce meaningful convergence to a consensus.*

3.4 Cognitive Mechanisms

The strongly bimodal distribution of naïveté appears puzzling at first. What exactly is it about the belief formation task in the *Selected* treatment that some subjects fully misconstrue, while others fully take it into account? What are the deep origins of the bias, given that explicit disagreement rarely induces naïfs to reconsider their beliefs? Understanding the mechanisms underlying the neglect of selection problems is important for at least two reasons. First, from the perspective of theory, understanding mechanisms supports efforts to formalize the bias, or to provide unifying theoretical accounts of different updating errors. Second, the presence or quantitative importance of selection neglect might well depend on environmental features, so that utilizing the idea of selection neglect in applied work rests on an understanding of what exactly it is that people fail to take into account when processing selected information.

The Role of Cognitive Skills

As a starting point for the investigation of the cognitive mechanisms underlying selec-

Table 3.4. Influence of others

	Dependent variable: Naïveté implied in second belief											
	Full sample				Rationals				Naïfs			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Naïveté in first belief	0.75***		0.73***	0.73***	0.74***		0.76***	0.77***	0.63***		0.58***	0.58***
	(0.03)		(0.04)	(0.04)	(0.09)		(0.10)	(0.11)	(0.13)		(0.11)	(0.11)
Avg. naïveté of senders		0.24***	0.15***	0.22***		0.078*	0.084**	0.15**		0.32***	0.24***	0.36***
		(0.05)	(0.03)	(0.04)		(0.04)	(0.03)	(0.06)		(0.08)	(0.06)	(0.07)
Disagreement among senders				0.060*				-0.023				0.17**
				(0.03)				(0.03)				(0.07)
Avg. naïveté × disagreement of senders				-0.066***				-0.054				-0.12***
				(0.02)				(0.03)				(0.04)
Constant	0.072***	0.29***	0.040	-0.013	0.057***	-0.0024	0.11*	0.12*	0.23*	0.76***	-0.11	-0.17
	(0.02)	(0.04)	(0.04)	(0.04)	(0.02)	(0.03)	(0.06)	(0.07)	(0.13)	(0.06)	(0.12)	(0.13)
Task FE	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Observations	382	382	382	382	134	134	134	134	121	121	121	121
R^2	0.770	0.048	0.796	0.801	0.543	0.015	0.589	0.608	0.423	0.176	0.595	0.642

OLS estimates, robust standard errors (clustered at subject level) in parentheses. For a given subject and task, a subject ("receiver") is classified as rational if both the out-of-sample median naïveté parameter from the first three tasks and the first belief statement in the respective task are "rational" (i.e., $\chi \leq 0.5$), and analogously for naïfs ($\chi > 0.5$). Since some subjects' type switched between the out-of-sample naïveté parameter and the first belief statement in the respective task, the sum of rationals and naïfs does not equal the total number of subjects. Very similar results obtain when I define rationals and naïfs exclusively based on the out-of-sample naïveté measure or exclusively based on the first belief in the respective task, see Appendix 3.D.4. All regressions exclude extreme outliers with $|\chi| > 3$; the results are robust to including these outliers when employing median regressions. Disagreement among senders is defined as the absolute difference between the naïveté implied in the senders' beliefs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5. Selection neglect and cognitive ability

	Dependent variable: Median χ				
	(1)	(2)	(3)	(4)	(5)
High school grades	-0.31***	-0.32***		-0.25***	-0.28***
	(0.07)	(0.05)		(0.07)	(0.06)
1 if conditional exp. correct			-0.50***	-0.33**	-0.22
			(0.14)	(0.13)	(0.14)
Age		0.038*			0.036
		(0.02)			(0.02)
1 if female		-0.20			-0.17
		(0.14)			(0.14)
Log [Monthly income]		-0.074			-0.063
		(0.09)			(0.08)
Constant	0.65***	0.41	0.94***	0.87***	0.51
	(0.07)	(0.63)	(0.10)	(0.07)	(0.62)
Marital status FE	No	Yes	No	No	Yes
Observations	48	48	48	48	48
R^2	0.291	0.416	0.195	0.363	0.446

OLS estimates, robust standard errors in parentheses. High school grades are the z-score of the unweighted average of the z-scores of subjects' overall high school GPA and their final math grade. The conditional expectation item is coded as 1 (0) if a subject answered 130 (anything else) on the follow-up question. Response time in minutes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

tion neglect, columns (1) and (2) of Table 3.5 present the results of OLS regressions of each subject's median naïveté parameter on their cognitive skills, as proxied for by achievement in high school. This score is constructed as first factor of a subject's overall high school GPA (which in Germany serves as primary university entrance criterion) and their final math grade in high school. The results show that high cognitive ability participants are significantly less likely to commit selection neglect, conditional on other sociodemographics.

While this result illustrates that the bimodal distribution of selection neglect is driven by cognitive skills, it leaves open the question of which dimension of cognitive skills the naïve types are missing. Relative to the control treatment, the selected treatment requires subjects to engage in two steps of reasoning:

1. *Noticing the systematic holes:* Subjects need to notice that they are missing a systematic subset of the available information.

2. *Computing conditional expectations:* Conditional on noticing the selection problem, subjects need to correctly back out the missing signals, i.e., to compute the expected signal conditional on group membership of the respective computer player.

Computing Conditional Expectations

To investigate subjects' capability of computing conditional expectations in this context, the experiment contained an incentivized follow-up question that was asked of every subject in the *Selected* treatment after they had finished the seven belief formation tasks. This question reads as follows:

> *In the course of this experiment, in total, you did not communicate with five computer players because you were part of the blue group, while these computer players opted for the red group. Based on this information, please estimate which signals these players in the red group have gotten, on average. You will receive an additional 2 euros if your guess is exactly right, 50 cents if your estimate is off by at most five, and nothing otherwise.*

In essence, the question asks subjects to compute the conditional expectation of a signal. The left panel of Figure 3.7 plots the distribution of participants' estimates. About two thirds of all subjects correctly computed the correct conditional expectation of 130. Columns (3)-(4) of Table 3.5 show that these subjects perform significantly better in solving the belief formation task relative to those who provided a different response: according to the OLS estimates, correctly answering this follow-up question is associated with a reduction of estimated naïveté of 0.5.

At the same time, even those subjects that did not answer 130 did understand that the signals of the computer players must have been larger than 100, on average: only 2 out of 48 subjects provided a response below 100, suggesting that subjects understood the experimental setup and were capable of making qualitatively appropriate inferences from the behavior of the computer players. In particular, if subjects had attributed no informational content to the computer players' group choice, they should have guessed 100 (the prior). In addition, as the right panel of Figure 3.7 shows, there is substantial heterogeneity in subjects' naïveté in the belief formation tasks even conditional on correctly answering the conditional expectation question. The left subpanel shows that the vast majority of subjects who provided a response larger than 100, but did not answer 130, exhibit full selection neglect ($\chi = 1$). This is remarkable in that if subjects understand the direction (even if not the magnitude) of the signals they do not see, they should at least partially adjust from full selection neglect. However, they don't.

Even more puzzling, those subjects that provided *exactly* the correct response of 130 (depicted in the right subpanel), also exhibit strong heterogeneity in their naïveté. While the fraction of rational subjects is higher in this subgroup, many people still fully neglect the selection problem in the belief formation tasks. These findings emphasize that being able to compute the correct conditional expectation is not sufficient to develop unbiased beliefs. Thus, the roots of selection neglect seem to be more than purely mathematical, but rather rooted in how subjects approach the problem in the first place.

Identifying the Problem: Subjects' Focus

The results presented so far show that the majority of subjects is capable of drawing (at least qualitatively) correct inferences from the computer players' group entrance decisions once they are explicitly prodded to do so, and to back out the signals they do not see. At the same time, these same subjects often exhibit (full) naïveté in the face of the full belief formation problem. This suggests that when people face complex updating problems, they fail to even think about the existence and properties of the information sources they do not directly interact with, akin to a "what you see is all there is" heuristic (Kahneman, 2011). Indeed, the follow-up question might well be interpreted as steering subjects' focus towards the computer players in the other group. Do subjects correctly

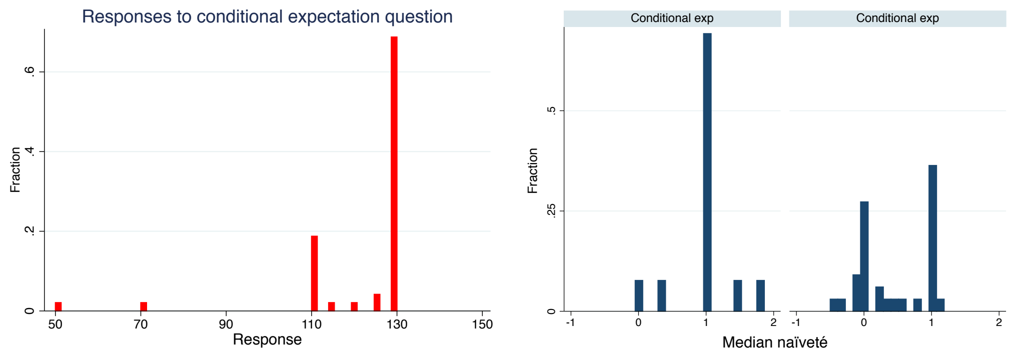


Figure 3.7. The left panel plots the distribution of responses to the follow-up question in the *Selected* treatment. The right panel illustrates the distribution of naïveté conditional on providing a response of larger than 100, but different from 130 (left subpanel), and conditional on answering exactly 130 (right subpanel).

solve the entire belief formation problem if they are nudged to focus on what they don't see?

To address this question, I implemented treatment variation *Salience*. This condition exogenously shifted subjects' focus towards the holes in their information samples, albeit without instructing them what to do about these holes. To this end, the treatment provided a hint both at the end of the instructions and on subjects' decision screens:[24]

> *HINT about the solution: Also think about the computer players you do not communicate with!*

Arguably, this hint alerts subjects to reflect upon the missing computer players and the information they have gotten. At the same time, the treatment does not manipulate subjects' mathematical skills or their motivation to solve the problem. 48 subjects participated in this treatment and earned 11.60 euros on average.

The left panel of Figure 3.8 provides kernel density plots of subjects' median naïveté in this *Salience* treatment compared to the two baseline treatments, while the right panel plots the distribution of naïveté implied in all individual-level beliefs. As visual inspection suggests, this treatment had a large effect on subjects' beliefs relative to the *Selected* condition ($p = 0.0009$, Wilcoxon ranksum test), and reduced the fraction of naïfs by 60%.[25] Notably, in this condition, most subjects did not adjust partially from full to partial naïveté; rather, they develop beliefs which exactly reflect $\chi = 0$. This is in line with the findings from the conditional expectation follow-up item: once people are prodded to actively think about the computer players they do not see, most are capable of drawing

[24] The quote provided in the main text applies to subjects' decision screen. To avoid confusion on the part of participants, the hint at the end of the instructions read as: "HINT about the solution: When you estimate the number X, always also think about the computer players you do not communicate with!"

[25] Appendix 3.E.1 provides a detailed analysis of the seven separate belief formation tasks, which confirms the findings from the aggregate analysis: in six out of seven tasks do beliefs significantly differ between *Salience* and *Selected*.

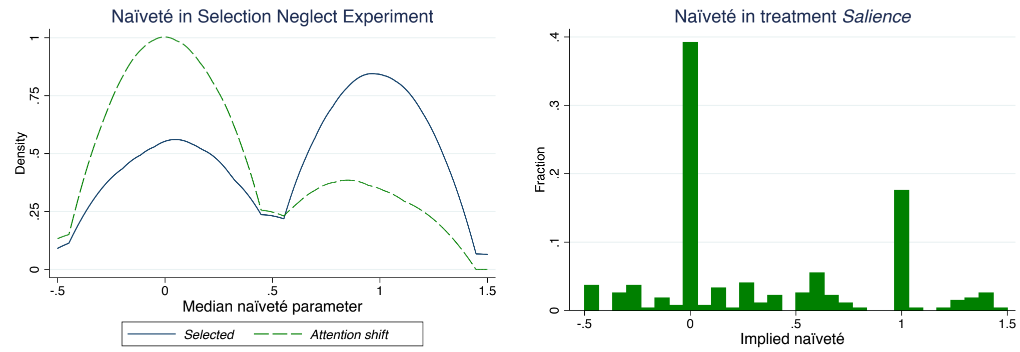


Figure 3.8. Distribution of naïveté in the *Selected*, *Control*, and *Salience* treatments. The left panel plots kernel density estimates of the median naïveté of each individual in all three treatments (48, 48, and 30 obs., respectively), while the right panel illustrates the distribution of naïveté implied in all individual-level beliefs in the *Salience* treatment (336 obs.).

appropriate inferences and to recognize the selection mechanism upon which the initial group entrance decision was based.[26]

Taken together, the findings on the cognitive step of paying attention to and identifying the systematic holes in one's information sample provide an intuitive explanation for the striking bimodality in subjects' types. In other contexts, bounded rationality (in the sense of costs of thinking) is often considered as a continuous concept in the sense that different people might exhibit any value of naïveté between zero and one. In contrast, the type of belief formation problem discussed in this paper appears to have a strong threshold logic, that is driven by subjects' focus: people either attend to the holes in their data and (fully) adjust for them, or they do not.

Result 14. *Selection neglect is significantly associated with cognitive ability. While the majority of subjects appear to possess the mathematical skills necessary to adjust for the selection problem at least in a qualitatively correct manner, many people do not pay attention to the systematic holes in their information sample: Exogenously shifting subjects' focus towards the information sources they do not directly interact with debiases a large fraction of participants.*

3.5 Relationship to Correlation Neglect

Forming beliefs in network environments is frequently complicated not only by selection problems, but also by double-counting problems that arise through correlated information sources (DeMarzo et al., 2003; Eyster and Rabin, 2010, 2014). For instance, if A talks to B and C and both have previously communicated with D, A runs the risk of double-counting D's information. As Enke and Zimmermann (2015) have shown experimentally,

[26] Indeed, this treatment did not affect subjects' ability to compute conditional expectations per se: in the follow-up question, the distribution of guesses is statistically indistinguishable from that in the *Selected* treatment ($p = 0.3810$, Wilcoxon ranksum test), suggesting that the positive effect of this treatment variation can indeed be attributed to a shift in attention rather than increased mathematical skills.

many people neglect these redundancies and hence fall prey to double-counting the signals of well-connected information sources.

Are the neglect of informational redundancies and selection effects conceptually related?[27] Answering this question is important because identifying common cognitive underpinnings of belief biases might allow theorists to develop unifying models of boundedly rational belief formation in network environments, based on primitives such as limited attention (see, e.g, the discussion in Fudenberg, 2006). Indeed, there are two strong ex ante reasons to study the relationship between correlation and selection neglect in particular, rather than between selection neglect and other updating errors. First, both biases arguably have some of their most powerful implications in social network or other social learning environments. Second, the results in Enke and Zimmermann (2015) provide a first indication for a possible relationship between the two errors: using a treatment intervention akin to the *Salience* treatment, they established that prodding subjects to actively think about the mechanics that generate the correlation in their setup, debiased a large fraction of subjects.

To further delve into the relationship between the two mistakes, I re-invited all subjects from the *Selected* treatment to take part in a follow-up study, and 32 out of 48 agreed to participate at least two weeks after the first experiment and earned 12.00 euros on average. In this follow-up experiment, subjects solved five of the tasks used by Enke and Zimmermann (2015) to establish the neglect of redundancies in information sources. While Appendix 3.F contains details on these experiments, the basic idea can be grasped from Figure 3.9. Subjects again had to estimate a hypothetical true state; computers A-D generated unbiased iid signals about the state and transmitted these to the subject and the intermediaries as depicted in Figure 3.9. The intermediaries, which were simulated by computers, computed the average of the signals they have access to, and transmitted that average to the subject. Thus, subjects were at the risk of double-counting the signal of computer A because all messages contained that signal. On the other hand, given that subjects knew the signal of computer A, being rational only required them to back out the underlying independent signals from the intermediaries' messages.

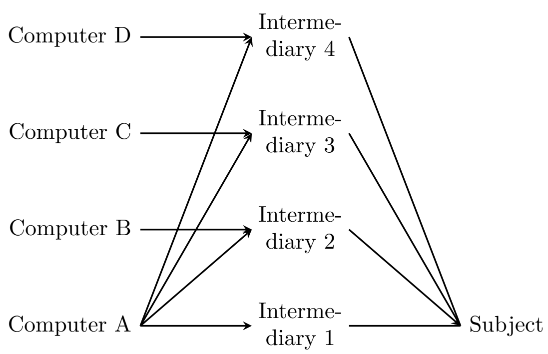


Figure 3.9. Correlation neglect information structure

[27] Formally, selection and correlation neglect are related because both homophily and informational redundancies can be formalized as introducing correlated error terms into the signals people have access to (Glaeser and Sunstein, 2009). Nevertheless, the biases apply to two distinct problems, one of non-representative samples and one of double-counting certain signals.

Table 3.6. Selection neglect and correlation neglect

	Dependent variable: Median χ				
	(1)	(2)	(3)	(4)	(5)
Median correlation neglect naïveté	0.35***	0.29**	0.31**	0.27**	0.27**
	(0.11)	(0.12)	(0.12)	(0.12)	(0.13)
High school grades		-0.097		-0.064	-0.17*
		(0.08)		(0.08)	(0.09)
1 if conditional exp. correct			-0.31**	-0.29*	-0.20
			(0.15)	(0.15)	(0.16)
Constant	0.40***	0.45***	0.65***	0.66***	1.48*
	(0.12)	(0.13)	(0.15)	(0.14)	(0.86)
Additional controls	No	No	No	No	Yes
Observations	32	32	32	32	32
R^2	0.202	0.227	0.279	0.290	0.432

OLS estimates, robust standard errors in parentheses. High school grades are the z-score of the unweighted average of the z-score of subjects' overall high school GPA and their final math grade. The conditional expectation item is coded as 1 (0) if a subject answered 130 (anything else) on the follow-up question. Additional controls include age, gender, log monthly income, and marital status fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Result 15. *Subjects' naïveté in updating from selected signals is significantly correlated with their propensity to neglect redundancies in information sources.*

The five belief formation tasks again allow the derivation of an individual-level naïveté parameter, i.e., subjects' propensity to adjust for the double-counting problem. To investigate the relationship between the two belief biases, Table 3.6 reports the results of OLS estimations of subjects' (median) selection neglect naïveté parameter on their (median) correlation neglect naïveté. Results show that the two types of naïveté are strongly and significantly correlated, even conditional on academic achievement and performance in the conditional expectation follow-up question, see columns (2) through (5). The raw Pearson correlation between the naïveté parameters is $\rho = 0.45$, $p = 0.0099$. Appendix 3.F.3 visualizes this relationship. Taken together, neglecting selection effects and neglecting redundancies is correlated within subject, and essentially the same treatment variation can be employed to switch both biases on and off. This provides a first indication that two important updating biases in social networks – though conceptually distinct – share common cognitive foundations based on limited attention, and might hence be fruitfully modeled in a unified way.

3.6 Conclusion

This paper has provided an analysis of how people form beliefs in the presence of homophily-driven selection effects, both individually and when interacting with others. I conclude by discussing two potential applications and extensions of the findings. First, the most straightforward implication of the neglect of homophily-driven selection effects is that it tends to reinforce the belief patterns upon which the original group entry decision was based. In this sense, the experimental results are consistent with popular concerns that belief-based segregation might produce increased polarization (Sunstein, 2009; Bishop, 2009; Pariser, 2011). Crucially, this paper shows that such a pattern can arise in the absence of motivated reasoning or wishful thinking, but rather only due to people's cognitive limitations in dealing with homophilous information samples.

Second, the present paper has also shown that the strong type heterogeneity in updating from selected sources may have predictable consequences for belief heterogeneity in society. In particular, two complementary sets of findings suggest that people need not necessarily learn from each other. First, people tend to select advisors or communication partners based on whether they process information in the same way as they do. Second, even when forced to consider the views of people with different updating rules do many people judge their own problem-solving strategy to be correct. While these results were obtained in the context of selection problems, they may nevertheless apply to boundedly rational belief formation more broadly. If true, this raises the intriguing conjecture that part of the large belief heterogeneity observed in field data may not be due to private information per se, but rather due to heterogeneous updating rules in combination with people's tendency to disproportionately trust their own assessment of the available evidence. An interesting question is under which conditions the result that people do not learn from others holds. For instance, it may be that naïfs start learning from rationals if the latter have a chance (and incentives) to explain to naïfs how and why their strategy is incorrect.

References

Acemoglu, Daron, Munther A. Dahleh, Ilan Lobel, and Asuman Ozdaglar (2011): "Bayesian Learning in Social Networks." *Review of Economic Studies*, 78 (4), 1201–1236. [82]

Acemoglu, Daron, Asuman Ozdaglar, and Ali ParandehGheibi (2010): "Spread of (Mis) Information in Social Networks." *Games and Economic Behavior*, 70 (2), 194–227. [82]

Andreoni, James and Tymofiy Mylovanov (2012): "Diverging Opinions." *American Economic Journal: Microeconomics*, 4 (1), 209–232. [84]

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy (2015): "Incentives in Experiments: A Theoretical Analysis." *Working Paper*. [87]

Bénabou, Roland (2013): "Groupthink: Collective Delusions in Organizations and Markets." *Review of Economic Studies*, 80 (2), 429–462. [80]

Bishop, Bill (2009): *The Big Sort: Why the Clustering of Like-Minded America is Tearing us Apart*. Houghton Mifflin Harcourt. [80, 107]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2015): "Stereotypes." *Working Paper*. [84]

Brenner, Lyle A., Derek J. Koehler, and Amos Tversky (1996): "On the Evaluation of One-Sided Evidence." *Journal of Behavioral Decision Making*, 9 (1), 59–70. [84]

Brocas, Isabelle, Juan D. Carrillo, Stephanie W. Wang, and Colin F. Camerer (2014): "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games." *Review of Economic Studies*, 81 (3), 944–970. [84]

Brunnermeier, Markus K. and Jonathan A. Parker (2005): "Optimal Expectations." *American Economic Review*, 95 (4), 1092–1118. [80]

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini (2013): "Overconfidence and Social Signalling." *Review of Economic Studies*, 80 (3), 949–983. [82, 98]

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong (2004): "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics*, 861–898. [81]

Camerer, Colin and Dan Lovallo (1999): "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review*, 306–318. [82, 98]

Caplin, Andrew and Mark Dean (forthcoming): "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review*. [84]

Caplin, Andrew, Mark Dean, and Daniel Martin (2011): "Search and Satisficing." *American Economic Review*, 101 (7), 2899–2922. [84]

Çelen, Bogaçhan, Shachar Kariv, and Andrew Schotter (2010): "An Experimental Test of Advice and Social Learning." *Management Science*, 56 (10), 1687–1701. [82]

Chandrasekhar, Arun G., Horacio Larreguy, and Juan P. Xandri (2015): "Testing Models of Social Learning on Networks: Evidence from a Framed Field Experiment." *Working Paper*. [83]

Charness, Gary, Edi Karni, and Dan Levin (2010): "On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda." *Games and Economic Behavior*, 68 (2), 551–556. [84]

Choi, Syngjoo, Edoardo Gallo, and Shachar Kariv (2015): "Networks in the Laboratory." *Working Paper*. [83]

Costa-Gomes, Miguel A. and Vincent P. Crawford (2006): "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *American Economic Review*, 1737–1768. [81]

Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri (2013): "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications." *Journal of Economic Literature*, 51 (1), 5–62. [81]

Currarini, Sergio, Matthew O Jackson, and Paolo Pin (2009): "An Economic Model of Friendship: Homophily, Minorities, and Segregation." *Econometrica*, 77 (4), 1003–1045. [83]

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel (2003): "Persuasion Bias, Social Influence, and Unidimensional Opinions." *Quarterly Journal of Economics*, 118 (3), 909–968. [83, 105]

Eil, David and Justin M Rao (2011): "The good news-bad news effect: asymmetric processing of objective information about yourself." *American Economic Journal: Microeconomics*, 3 (2), 114–138. [80]

Enke, Benjamin and Florian Zimmermann (2015): "Correlation Neglect in Belief Formation." *Working Paper*. [82, 83, 88, 105, 106, 126, 127]

Esponda, Ignacio and Emmanuel Vespa (2015): "Endogenous Sample Selection in Common Value Environments: A Laboratory Study." *Working Paper*. [84]

Eyster, Erik and Matthew Rabin (2010): "Naïve Herding in Rich-Information Settings." *American Economic Journal: Microeconomics*, 2 (4), 221–243. [83, 105]

Eyster, Erik and Matthew Rabin (2014): "Extensive Imitation is Irrational and Harmful." *Quarterly Journal of Economics*, 129 (4), 1861–1898. [83, 105]

Eyster, Erik, Matthew Rabin, and Georg Weizsäcker (2015): "An Experiment on Social Mislearning." *Working Paper*. [83]

Fischbacher, Urs (2007): "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10 (2), 171–178. [87]

Fragiadakis, Daniel E., Daniel T. Knoepfle, and Muriel Niederle (2013): "Identifying Predictable Players: Relating Behavioral Types and Predictable Subjects." *Working Paper*. [81]

Fudenberg, Drew (2006): "Advancing Beyond "Advances in Behavioral Economics"." *Journal of Economic Literature*, 694–711. [83, 106]

Gennaioli, Nicola and Andrei Shleifer (2010): "What Comes to Mind." *Quarterly Journal of Economics*, 125 (4), 1399–1433. [84]

Gentzkow, Matthew and Jesse M. Shapiro (2006): "Media Bias and Reputation." *Journal of Political Economy*, 114 (2). [79]

Gentzkow, Matthew and Jesse M. Shapiro (2011): "Ideological Segregation Online and Offline." *Quarterly Journal of Economics*, 126 (4), 1799–1839. [79]

Glaeser, Edward L. and Cass R. Sunstein (2009): "Extremism and Social Learning." *Journal of Legal Analysis*, 1 (1), 263–324. [83, 106]

Golub, Benjamin and Matthew O. Jackson (2010): "Naïve Learning in Social Networks and the Wisdom of Crowds." *American Economic Journal: Microeconomics*, 2 (1), 112–149. [82]

Golub, Benjamin and Matthew O. Jackson (2012): "How Homophily Affects the Speed of Learning and Best-Response Dynamics." *Quarterly Journal of Economics*, 127 (3), 1287–1338. [82, 83]

Grether, David M. (1980): "Bayes Rule as a Descriptive Model: The Representativeness Heuristic." *Quarterly Journal of Economics* (3), 537–557. [88]

Grimm, Veronika and Friederike Mengel (2014): "An Experiment on Belief Formation in Networks." *Working Paper*. [83]

Kahneman, Daniel (2011): *Thinking, Fast and Slow*. Macmillan. [103]

Koehler, Jonathan J. and Molly Mercer (2009): "Selection Neglect in Mutual Fund Advertisements." *Management Science*, 55 (7), 1107–1121. [84]

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001): "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology*, 27, 415–444. [79]

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat (2014): "Managing Self-Confidence: Theory and Experimental Evidence." *Working Paper*. [80, 83]

Möbius, Markus M., Thuan Phan, and Adam Szeidl (2013): "Treasure Hunt: Social Learning in the Field." *Working Paper*. [83]

Müller-Frank, Manuel (2013): "A General Framework for Rational Learning in Social Networks." *Theoretical Economics*, 8 (1), 1–40. [82]

Pariser, Eli (2011): *The Filter Bubble: What the Internet is Hiding From You*. Penguin UK. [80, 107]

Pew Research Center (2014): *Political Polarization in the American Public*. http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/. [Online, accessed July 9, 2015]. [79]

Rabin, Matthew and Joel L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics*, 114 (1), 37–82. [84]

Schkade, David, Cass R. Sunstein, and Reid Hastie (2007): "What Happened on Deliberation Day?" *California Law Review* (3), 915–940. [84]

Schotter, Andrew (2003): "Decision Making With Naive Advice." *American Economic Review: Papers & Proceedings* (2), 196–201. [82]

Schwartzstein, Joshua (2014): "Selective Attention and Learning." *Journal of the European Economic Association*, 12 (6), 1423–1452. [84]

Sunstein, Cass R. (2009): *Republic.com 2.0*. Princeton University Press. [80, 107]

Appendix 3.A Treatment Overview

Table 3.7. Treatment overview

Treatment	# of subjects	Session length (mins)	Ave earnings (euros)
Selected	48	50	10.10
Control	30	50	13.70
Salience	48	50	11.60
Advice	59	50	10.80
Advice only	60	50	11.40
Exogenous	96	70	11.60

Appendix 3.B Details for Individual Belief Formation Treatments

3.B.1 Kernel Density Estimates for each Task

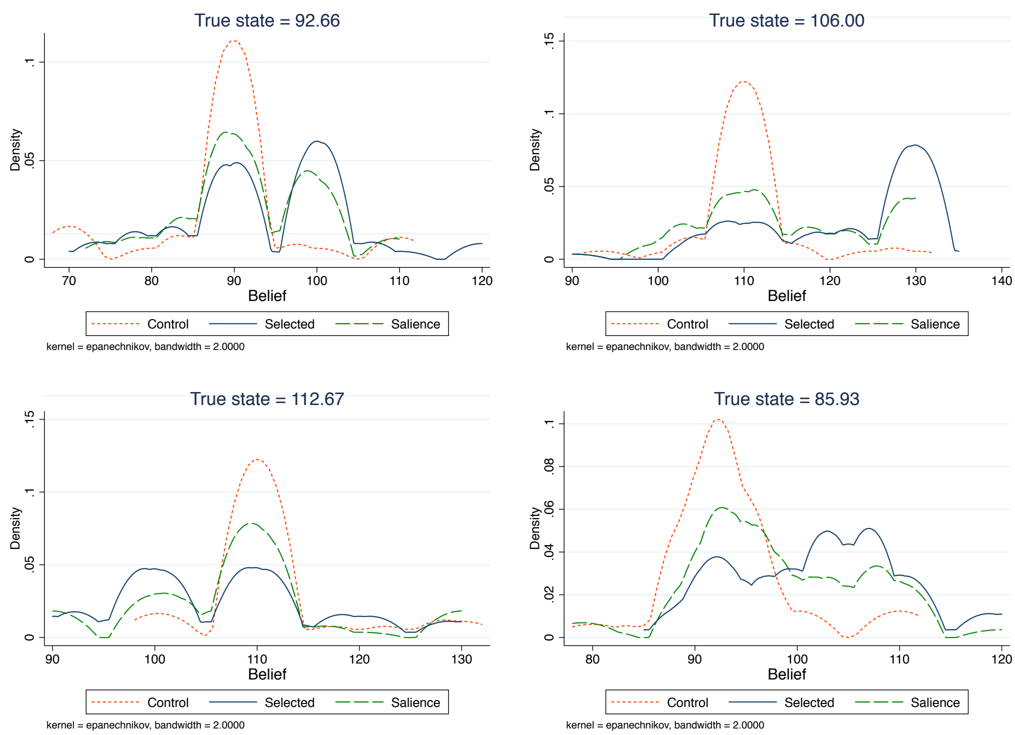


Figure 3.10. Distribution of beliefs by task (1/2). To ease readability, the plots exclude extreme outliers whose distance to both the fully naïve and rational benchmarks is larger than 20.

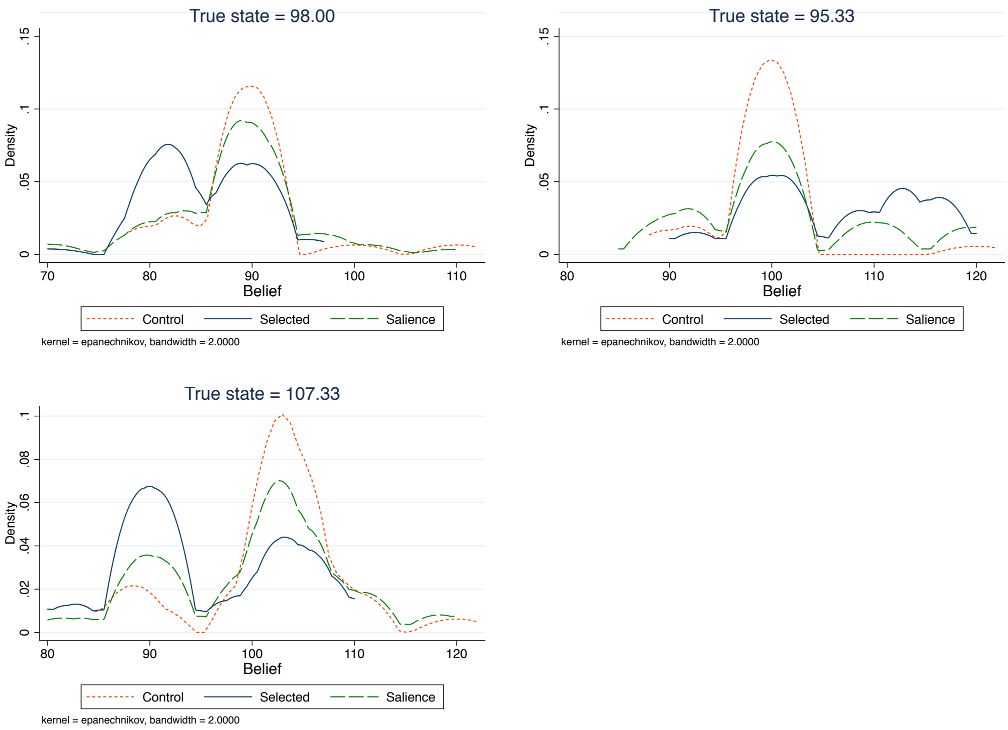


Figure 3.11. Distribution of beliefs by task (2/2). To ease readability, the plots exclude extreme outliers whose distance to both the fully naïve and rational benchmarks is larger than 20.

3.B.2 Consistency of Beliefs Across Tasks

This section investigates the consistency with which subjects in *Selected* exhibit a certain degree of naïveté across tasks. To this end, I define a set of potential types $\chi = -0.5, -0.4, -0.3, \ldots, 1.5$. Then, for each individual and each χ, I count the number of beliefs which reflect naïveté in $[\chi - 0.1, \chi + 0.1]$. Denote the number of beliefs in this interval as n_χ. Finally, I take the maximum over all n_χ, for each individual. This maximum represents the number of beliefs that exhibit a certain degree of consistency in the sense that they are within a rather small interval around some degree of naïveté. Figure 3.12 presents a histogram of this measure, which reveals that almost 70% of all subjects state at least three consistent beliefs. Thus, overall, subjects' responses reflect a considerable degree of consistency.

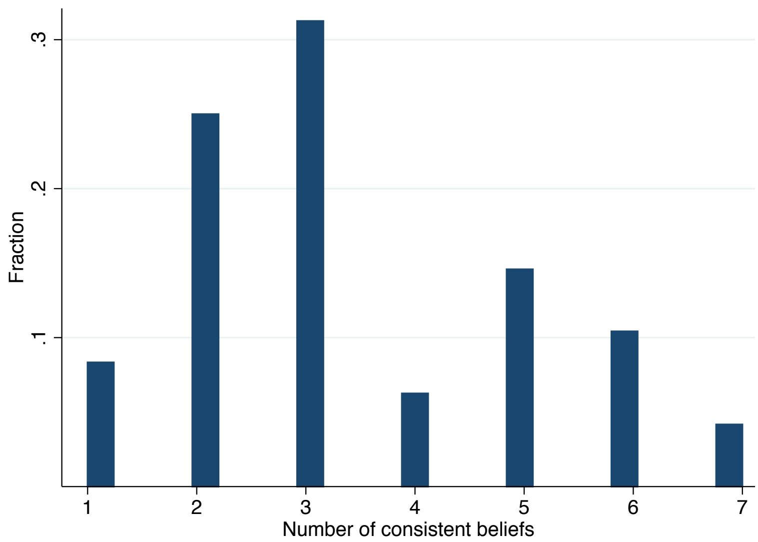


Figure 3.12. Number of consistent beliefs in treatment *Selected*.

Appendix 3.C Additional Results for Treatments *Advice* and *Advice only*

3.C.1 Distribution of Naïveté in First Three Tasks

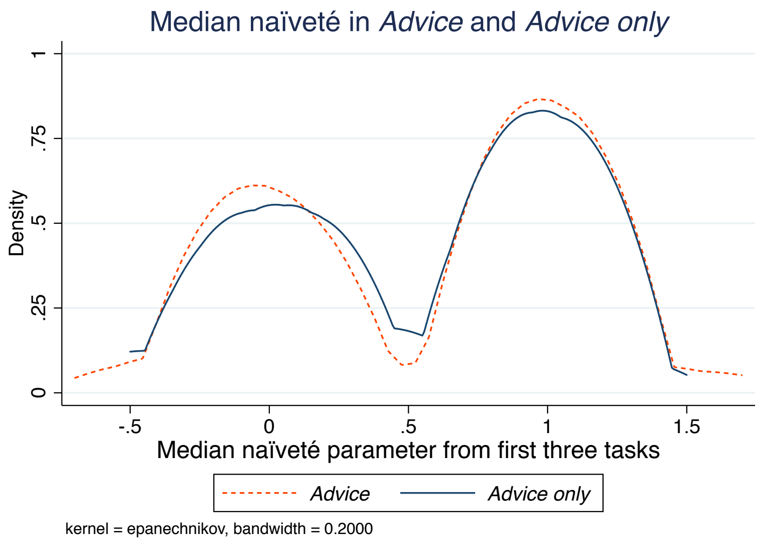


Figure 3.13. Distribution of median naïveté in the first three tasks (i.e., without advice). The densities excluce observations outside [-.5,1.5].

3.C.2 Distribution of Naïveté and Choice of Advisor

Table 3.8 analyzes beliefs in the tasks in which subjects had access to the advisor. Results show that subjects' implied naïveté in these tasks is strongly correlated with their choice of advisor, conditional on their inherent naïveté as measured in the first three tasks.

Figures 3.14 and 3.15 depict the distribution of naïveté implied in all beliefs in the tasks where subjects in *Advice* and *Advice only* had access to an advisor. The figures are partitioned by subjects' inherent naïveté type as determined in the first three tasks without advice.

Figures 3.16 and 3.17 provide an overview of the naïveté implied in subjects' beliefs with and without advice, conditional on their choice of advisor. That is, compared to the figures described in the preceeding paragraph, the figures are not conditional upon subjects' inherent naïveté, but instead conditional on their choice of advisor.

Table 3.8. Endogenous advice and naïveté

	Dependent variable: Median naiveté in last tasks					
	Treatment *Advice*			Treatment *Advice only*		
	(1)	(2)	(3)	(4)	(5)	(6)
Median naïveté in first three tasks	0.38***		0.24***	0.26***		0.055
	(0.09)		(0.08)	(0.04)		(0.06)
1 if chose naïve advisor		0.70***	0.53***		0.87***	0.86***
		(0.10)	(0.13)		(0.09)	(0.14)
Constant	0.26***	0.085	-0.00011	0.24***	-0.063	-0.090
	(0.08)	(0.09)	(0.45)	(0.08)	(0.04)	(0.41)
Additional controls	No	No	Yes	No	No	Yes
Observations	59	59	59	60	60	60
R^2	0.358	0.448	0.608	0.228	0.555	0.614

OLS estimates, robust standard errors in parentheses. Additional controls include age, gender, high school grades, marital status fixed effects, and log monthly income. See Table 3.5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

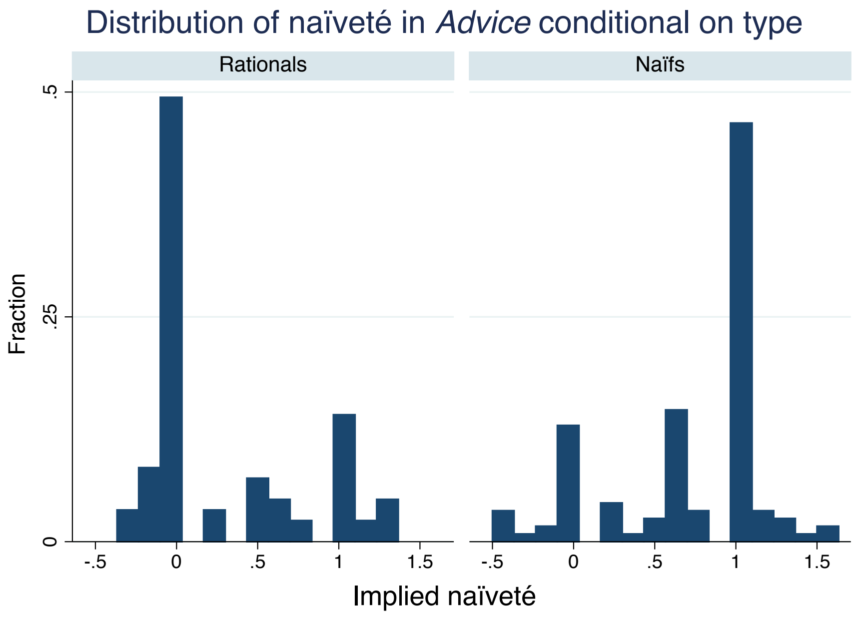


Figure 3.14. Distribution of decisions in the last four tasks (i.e., with advice), conditional on subjects' naïveté type in the first three tasks. Rationals are defined as $\chi \leq 0.5$ and naïfs as $\chi > 0.5$. The histograms exclude observations outside [-1,1.5].

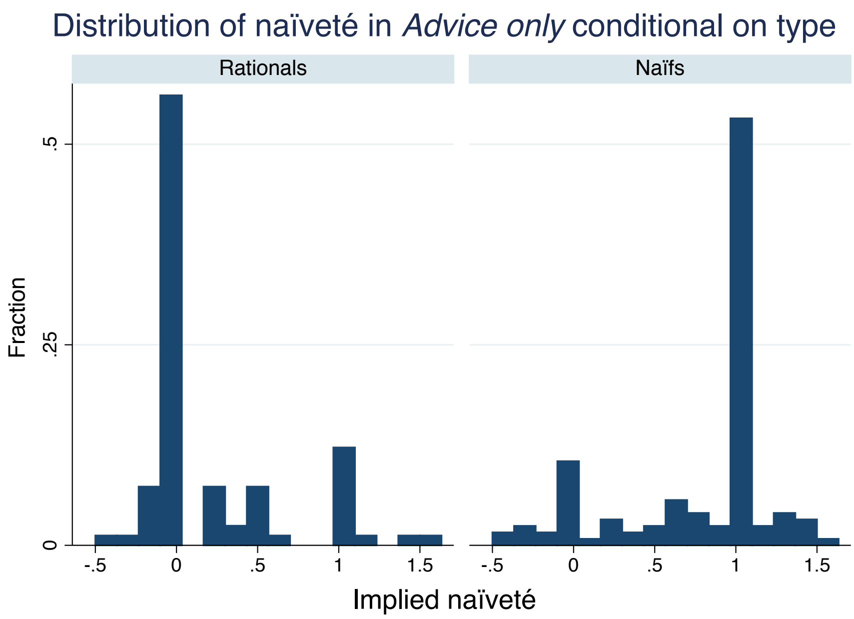


Figure 3.15. Distribution of decisions in the last four tasks (i.e., with advice), conditional on subjects' naïveté type in the first three tasks. Rationals are defined as $\chi \leq 0.5$ and naïfs as $\chi > 0.5$. The histograms exclude observations outside [-1,1.5].

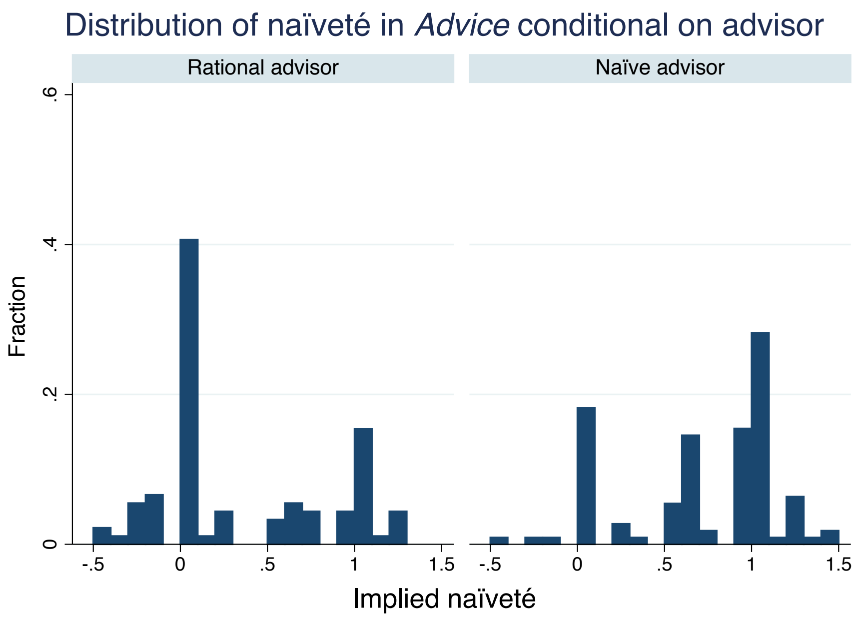


Figure 3.16. Distribution of decisions in the last four tasks (i.e., with advice), conditional on subjects' naïveté type in the first three tasks. Rationals are defined as $\chi \leq 0.5$ and naïfs as $\chi > 0.5$. The histograms exclude observations outside [-1,1.5].

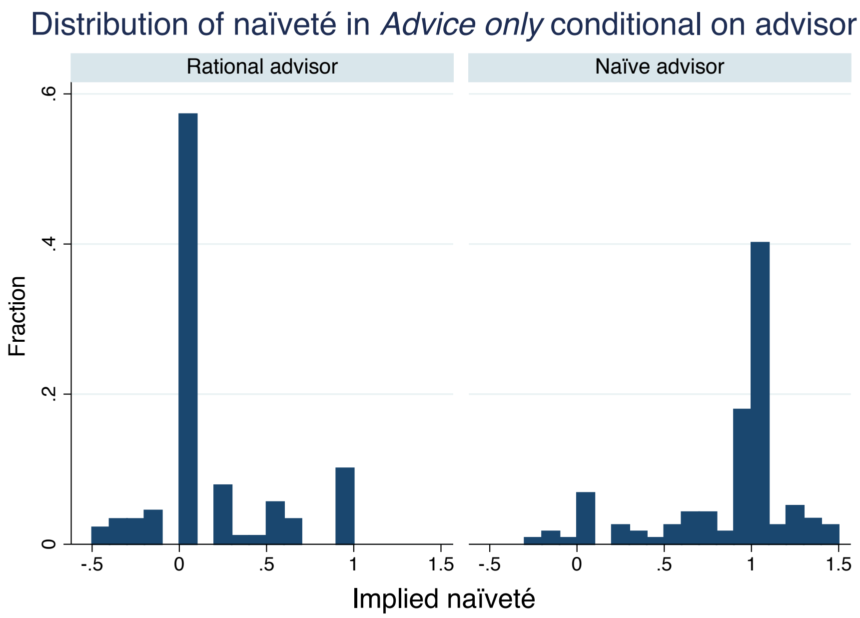


Figure 3.17. Distribution of decisions in the last four tasks (i.e., with advice), conditional on subjects' naïveté type in the first three tasks. Rationals are defined as $\chi \leq 0.5$ and naïfs as $\chi > 0.5$. The histograms exclude observations outside [-1,1.5].

Appendix 3.D Extensions and Robustness Checks for Treatment *Exogenous*

3.D.1 Distribution of Naïveté

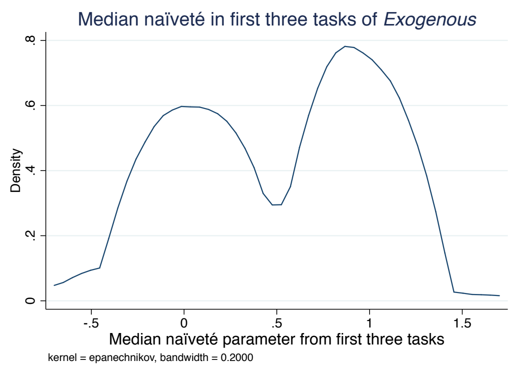


Figure 3.18. Distribution of median naïveté in the first three tasks (i.e., without seeing the beliefs of others). The density excluces observations outside [-.5,1.5].

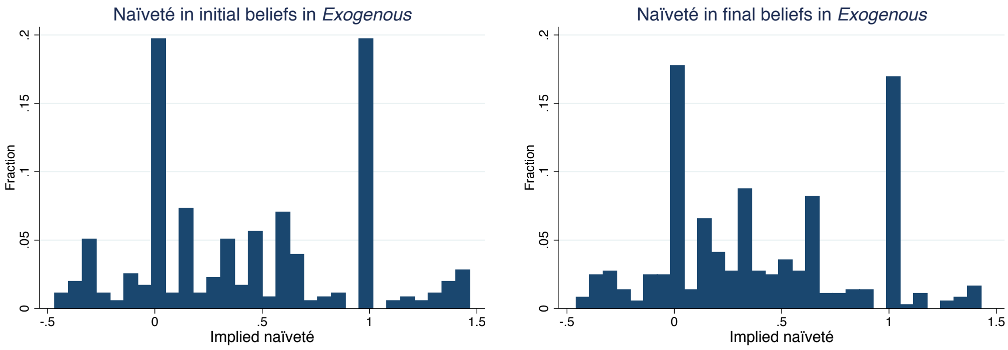


Figure 3.19. Distribution of decisions in the last four tasks (i.e., when seeing the beliefs of others). The left panel depicts the distribution of initial beliefs (before seeing the beliefs of the senders), and the right panel the distribution of post-communication beliefs. The histograms exclude observations outside [-.5,1.5].

3.D.2 Raw Correlation Between Pre- and Post-Communication Beliefs

Figure 3.20 presents the raw correlation between the first and second belief in the last four tasks in treatment *Exogenous*, regardless of whether the receiver's belief differs from that of at least one sender.

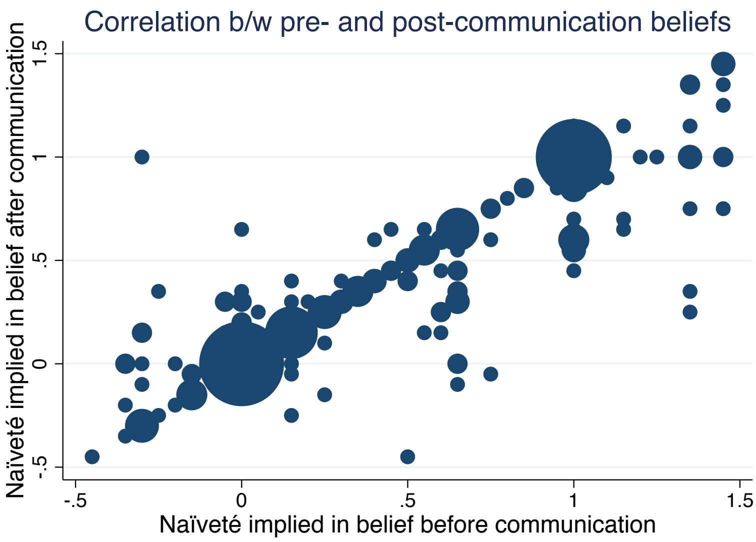


Figure 3.20. Raw correlation between the naïveté χ implied in first and second beliefs ($\rho = 0.90$). To construct this figure, subjects' pre- and post-communication naïveté is rounded to multiples of 0.05. The ball size then represents the number of observations in the respective bin. To ease readability, the scatter excludes 30 (out of 384) observations for which the implied naïveté of at least one belief is outside [-.5,1.5].

3.D.3 Robustness Checks for Adjustment Patterns

In the main text, rationals and naïfs were defined through a combination of the out-of-sample naïveté measure derived from the first three tasks as well as the first belief in the respective task. Figures 3.21 and 3.22 show that very similar patterns obtain when I classify subjects exclusively based on the out-of-sample measure or based on the first belief in the respective task.

3.D.4 Extensions and Robustness Checks for Regressions

Table 3.9 provides a robustness check on the effect of initial naïveté and naïveté of the senders in determining final naïveté (as implied in the second belief in each task). While the main text classified rationals and naïfs by making use of both the out-of-sample naïveté measure from the first three tasks and the naïveté implied in the respective first belief, I now classify subjects based on either of these naïveté measures. The results are unchanged.

Table 3.10 presents an extensions for this types of analysis. In particular, I investigate whether subjects tend to place lower weight on the beliefs of the senders if subjects are not very confident. As column (5) shows, however, no significant relationship emerges. Columns (6)-(15) show the robustness of this finding among the sub-samples of rationals and naïfs, respectively.

3.D.5 Do Subjects Who Revise Their Beliefs Learn?

It is conceivable that those naïve subjects who substantially revise their beliefs become less naïve in subsequent tasks. This could happen, for example, if subjects learn from the beliefs of more rational subjects. Table 3.11 presents the results of OLS regressions of subjects' naïveté in a given task on the degree of adjustment towards the rational belief in the previous task, conditional on the initial naïveté in the previous task. In these analyses, the sample is restricted to naïve subjects, i.e., to those participants whose out-of-sample median naïveté parameter from the first three tasks is larger than 0.5. Results show that those subjects who strongly revise their beliefs do not become more rational over time. This suggests that some subjects may feel that their own problem-solving is incorrect, but have no superior way of solving the problem themselves.

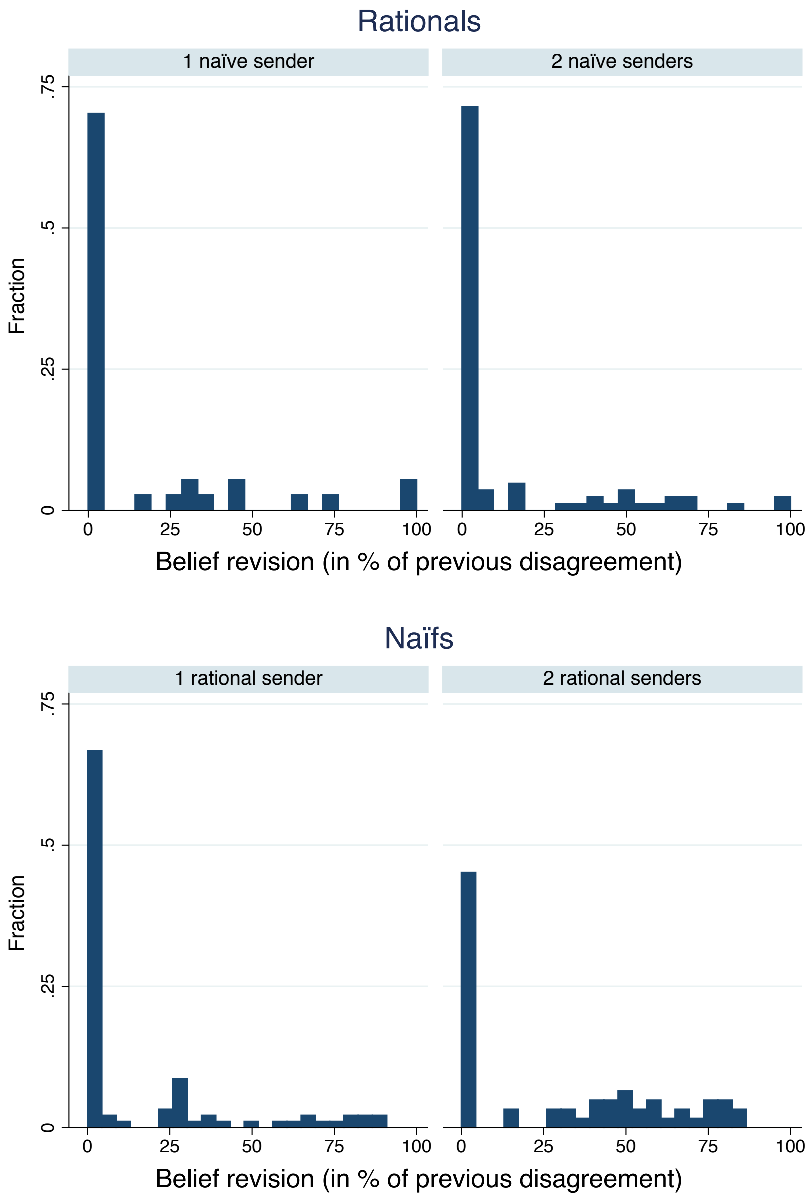


Figure 3.21. Magnitude of belief revisions. Each histogram denotes the belief revision between the first and second belief (expressed in terms of units of naïveté) conditional on the type of the subject (top / bottom panel) and on the composition of the two senders. The top left panel shows the adjustment of rational subjects who face one naïve and one rational belief, while the top right panel illustrates the rational types' belief revision if they faced two naïfs. The bottom left panel depicts the adjustment behavior of naïfs when they faced one rational and one naïve belief, while the right panel illustrates adjustment in case of two rational senders. For a given subject and task, a subject ("receiver") is classified as rational if the out-of-sample median naïveté parameter from the first three tasks is "rational" (i.e., $\chi \leq 0.5$), and analogously for naïfs ($\chi > 0.5$). Adjustments $> |1|$ are excluded to ease readability.

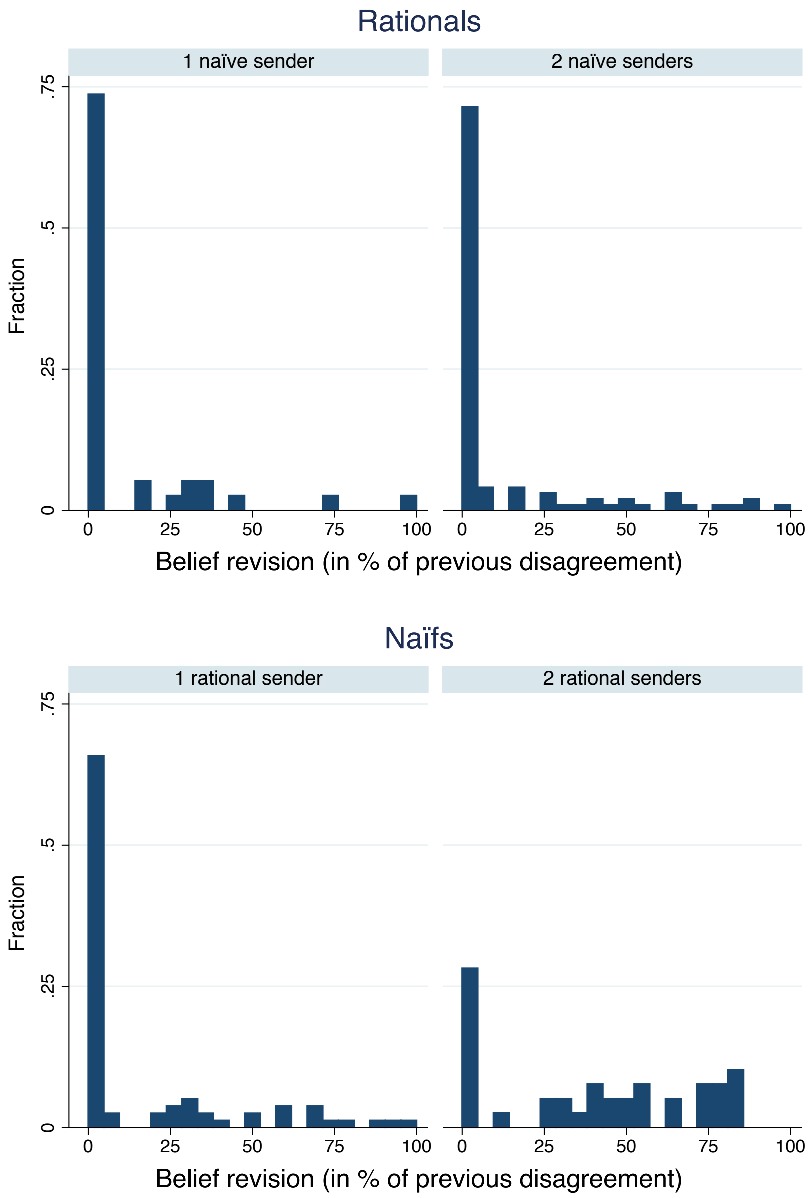


Figure 3.22. Magnitude of belief revisions. Each histogram denotes the belief revision between the first and second belief (expressed in terms of units of naïveté) conditional on the type of the subject (top / bottom panel) and on the composition of the two senders. The top left panel shows the adjustment of rational subjects who face one naïve and one rational belief, while the top right panel illustrates the rational types' belief revision if they faced two naïfs. The bottom left panel depicts the adjustment behavior of naïfs when they faced one rational and one naïve belief, while the right panel illustrates adjustment in case of two rational senders. For a given subject and task, a subject ("receiver") is classified as rational if the first belief statement in the respective task is "rational" (i.e., $\chi \leq 0.5$), and analogously for naïfs ($\chi > 0.5$). Adjustments $> |1|$ are excluded to ease readability.

Table 3.9. Influence of others: Robustness to classification of rationals and naiïfs

	Dependent variable: Naïveté implied in second belief														
				Classification based on out-of-sample naïvete parameter						Classification based on first belief in task					
	Full sample			Rationals			Naïfs			Rationals			Naïfs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Naïveté in first belief	0.75^{***}		0.73^{***}	0.73^{***}		0.71^{***}	0.74^{***}		0.70^{***}	0.68^{***}		0.69^{***}	0.66^{***}		0.60^{***}
	(0.03)		(0.04)	(0.06)		(0.07)	(0.05)		(0.05)	(0.07)		(0.07)	(0.11)		(0.09)
Avg. naïveté of senders		0.24^{***}	0.15^{***}		0.13^{***}	0.11^{***}		0.37^{***}	0.20^{***}		0.068	0.098^{***}		0.30^{***}	0.22^{***}
		(0.05)	(0.03)		(0.04)	(0.03)		(0.08)	(0.05)		(0.04)	(0.03)		(0.08)	(0.05)
Constant	0.072^{***}	0.29^{***}	0.040	0.059^{***}	0.13^{***}	-0.021	0.091^{***}	0.41^{***}	-0.012	0.055^{***}	0.028	0.048	0.19^{*}	0.71^{***}	-0.15
	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.03)	(0.03)	(0.06)	(0.05)	(0.01)	(0.03)	(0.05)	(0.10)	(0.05)	(0.10)
Task FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	382	382	382	176	176	176	206	206	206	219	219	219	163	163	163
R^2	0.770	0.048	0.796	0.707	0.026	0.739	0.751	0.100	0.787	0.531	0.009	0.555	0.432	0.162	0.600

OLS estimates, robust standard errors (clustered at subject level) in parentheses. In columns (4)-(9), aor a given subject and task, a subject ("receiver") is classified as rational if the out-of-sample median naïveté parameter from the first three tasks is "rational" (i.e., $\chi \leq 0.5$), and analogously for naïfs ($\chi > 0.5$). In columns (10)-(15), the classification is based on the first belief in the respective task. All regressions exclude extreme outliers with $|\chi| > 3$; the results are robust to including these outliers when employing median regressions. Disagreement among senders is defined as the absolute difference between the naïveté implied in the senders' beliefs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.10. Influence of others: Extensions

	Dependent variable: Naïveté implied in second belief														
	Full sample					Rationals					Naïfs				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Naïveté in first belief	0.75^{***}		0.73^{***}	0.73^{***}	0.71^{***}	0.74^{***}		0.76^{***}	0.77^{***}	0.76^{***}	0.63^{***}		0.58^{***}	0.58^{***}	0.57^{***}
	(0.03)		(0.04)	(0.04)	(0.04)	(0.09)		(0.10)	(0.11)	(0.10)	(0.13)		(0.11)	(0.11)	(0.10)
Avg. naïveté of senders		0.24^{***}	0.15^{***}	0.22^{***}	0.26^{***}		0.078^{*}	0.084^{**}	0.15^{**}	0.16		0.32^{***}	0.24^{***}	0.36^{***}	0.39^{**}
		(0.05)	(0.03)	(0.04)	(0.09)		(0.04)	(0.03)	(0.06)	(0.13)		(0.08)	(0.06)	(0.07)	(0.15)
Disagreement among senders				0.060^{*}					-0.023					0.17^{**}	
				(0.03)					(0.03)					(0.07)	
Avg. naïveté × disagreement of senders				-0.066^{***}					-0.054					-0.12^{***}	
				(0.02)					(0.03)					(0.04)	
Confidence					-0.0022					-0.0096					0.0042
					(0.01)					(0.01)					(0.02)
Avg. naïvete of senders × confidence					-0.018					-0.012					-0.032
					(0.01)					(0.02)					(0.03)
Constant	0.072^{***}	0.29^{***}	0.040	-0.013	0.059	0.057^{***}	-0.0024	0.11^{*}	0.12^{*}	0.18^{**}	0.23^{*}	0.76^{***}	-0.11	-0.17	-0.13
	(0.02)	(0.04)	(0.04)	(0.04)	(0.07)	(0.02)	(0.03)	(0.06)	(0.07)	(0.09)	(0.13)	(0.06)	(0.12)	(0.13)	(0.16)
Task FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Observations	382	382	382	382	382	134	134	134	134	134	121	121	121	121	121
R^2	0.770	0.048	0.796	0.801	0.800	0.543	0.015	0.589	0.608	0.603	0.423	0.176	0.595	0.642	0.605

OLS estimates, robust standard errors (clustered at subject level) in parentheses. For a given subject and task, a subject ("receiver") is classified as rational if both the out-of-sample median naïveté parameter from the first three tasks and the first belief statement in the respective task are "rational" (i.e., $\chi \leq 0.5$), and analogously for naïfs ($\chi > 0.5$). Very similar results obtain when I define rationals and naïfs exclusively based on the out-of-sample naïveté measure or exclusively based on the first belief in the respective task, see Table 3.9. All regressions exclude extreme outliers with $|\chi| > 3$; the results are robust to including these outliers when employing median regressions. Disagreement among senders is defined as the absolute difference between the naïveté implied in the senders' beliefs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.11. Belief adjustment and learning

	Dependent variable: Naïveté implied in first belief	
	(1)	(2)
Naïveté in previous task	0.26***	0.30***
	(0.08)	(0.09)
Adjustment in previous task	0.094	0.058
	(0.09)	(0.10)
Age		-0.016
		(0.01)
1 if female		0.34***
		(0.10)
Log [Monthly income]		-0.017
		(0.05)
Constant	0.41***	0.25
	(0.07)	(0.38)
Task FE	No	Yes
Observations	156	156
R^2	0.133	0.428

OLS estimates, robust standard errors (clustered at subject level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 3.E Subjects' Confidence

To investigate the relationship between subjects' decisions and their confidence in their own problem-solving strategy, I make use of a qualitative question that was asked after the first three tasks, i.e., before the choice of the advisor was introduced: "On a scale from 1 (not certain at all) to 10 (very certain), how certain are you that your previous estimates (and the underlying strategy) were correct?"

Table 3.12 presents the results of OLS estimations of subjects' confidence on their characteristics. The regressions pool subjects from treatments *Advice*, *Advice only*, and *Exogenous*, because these treatments proceeded in an essentially identical fashion before the confidence question was asked, i.e., subjects completed three tasks from the *Selected* condition by themselves. Results show that subjects' naïveté is not significantly correlated with their confidence, despite the relatively large sample size. Men and wealthier subjects are more likely to express higher confidence in their beliefs.

Table 3.13 analyzes the relationship between subjects' confidence and their choice of advisor, for both rationals and naïfs. Results show that measured confidence is only weakly related to the choice of advisor for both types. If anything, more confident naïfs are more likely to choose the rational advisor.

Table 3.12. Correlates of confidence

	Dependent variable: Confidence		
	(1)	(2)	(3)
Median naïveté in first three tasks	-0.36*	-0.40**	-0.31*
	(0.19)	(0.19)	(0.17)
Constant	6.24***	6.10***	4.12***
	(0.20)	(0.28)	(1.49)
Treatment FE	No	Yes	Yes
Additional controls	No	No	Yes
Observations	215	215	215
R^2	0.024	0.041	0.119

OLS estimates, robust standard errors in parentheses. Additional controls include age, gender, log monthly income, marital status fixed effects, and high school grades. See Table 3.5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.13. Confidence and choice of advisor

	Dependent variable: 1 if chose naïve advisor			
	Rationals		Naïfs	
	(1)	(2)	(3)	(4)
Confidence	-0.12	-0.13	-0.11	-0.27*
	(0.13)	(0.18)	(0.09)	(0.16)
Constant	-0.23	-0.58	1.77***	1.53
	(0.83)	(2.15)	(0.63)	(2.09)
Additional controls	No	Yes	No	Yes
Observations	52	52	67	67
Pseudo R^2	0.030	0.142	0.030	0.284

Probit estimates, robust standard errors in parentheses. Additional controls include age, gender, log monthly income, marital status fixed effects, and high school grades. See Table 3.5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.E.1 Details for "*Salience*" Treatment

Table 3.14. Overview of beliefs across tasks

True State	Median Belief *Control* Treatment	Median Belief *Selected* Treatment	Median Belief *Salience* Treatment	p-value (Ranksum) *Salience* vs *Control*	p-value (Ranksum) *Salience* vs *Selected*
92.66	90.00	100.00	90.00	0.2664	0.0522
106.00	110.00	128.00	114.165	0.0308	0.0031
112.67	110.00	108.00	110.00	0.0302	0.7844
85.93	93.15	105.00	95.00	0.0077	0.0149
98.00	90.00	85.00	90.00	0.7228	0.0584
95.33	100.00	107.50	100.00	0.4535	0.0052
107.33	103.00	91.50	103.15	0.8328	0.0125

Notes. Overview of the estimation tasks in order of appearance. See Table 3.1 for details on the signals in each task as well as the computation of the rational and the naïve benchmarks.

Appendix 3.F Details for Correlation Neglect Follow-Up Study

3.F.1 Experimental Design

The design is taken from Enke and Zimmermann (2015). Subjects were asked to estimate a hypothetical true state μ, where I induced a prior belief by informing subjects that μ would be drawn from $\mathcal{N}(0; 250,000)$. Computers A-D generated four unbiased iid signals about μ by drawing from $s_h \sim \mathcal{N}(\mu; 250,000)$.

Intermediary 1 observed the signal of Computer A and transmitted it to subjects. The intermediaries 2 to 4 observed both the signal of computer A and of computers B to D, respectively, and then reported the average of these two signals. Since subjects knew the signal of Computer A, they could extract the other independent signals from the intermediaries' reports.

As in the experiments designed to identify selection neglect, this treatment features an exogenous data-generating process wich is fully known to subjects. Control questions ensured that subjects understood the mechanics of this process. No feedback was provided between the five independent tasks. Earnings were computed through a quadratic scoring rule with maximum earnings of 12 euros: $\pi = max\{0; 12 - 0.01 \times (\text{Belief} - \text{True state})^2\}$. These experiments lasted 40 minutes on average, and subjects earned an average of 12.30 euros including a 7 euros show-up fee.

Table 3.15 presents details on the belief formation tasks as well as median beliefs in each task. As can be inferred from the rightmost column, median beliefs are always between the rational and the full correlation neglect benchmark.

3.F.2 Computation and Distribution of Naïveté Parameters

Given the known data-generating process, one can again define and measure an individual-level naïveté parameter. As in the case of selection neglect, I assume full base rate neglect for this purpose, which is bolstered by the findings in Enke and Zimmermann (2015). The individual-level naïveté parameter is then computed as follows:

Table 3.15. Overview of correlation neglect tasks

True State	Computer A	Computer B	Computer C	Computer D	Rational Belief	Correlation Neglect Belief	Median Belief
-241	249	-699	-139	70	-129.75	59.63	0.00
-563	-446	-1,374	-1,377	-1,475	-1,168	-807	-1,000
38	442	173	58	233	226.5	334.25	250.00
1,128	1,989	781	440	2,285	1,373.75	1,681.38	1373.75
-23	810	-822	-99	409	74.5	442.25	257

Notes. Overview of the correlation neglect estimation tasks in order of appearance. See Section 3.F.2 for the derivation of the rational and the full correlation neglect benchmarks.

Subjects observed s_1 and $\tilde{s}_h = (s_1 + s_h)/2$ for $h \in \{2,3,4\}$. When prompted to estimate μ, a rational decision maker would extract the underlying independent signals from the $\tilde{s}_h$ and compute the mean Bayesian posterior as $b_B = \sum_{h=1}^{4} s_h/4$. However, now suppose that the decision maker suffers from correlation neglect, i.e., he does not fully take into account the extent to which $\tilde{s}_h$ reflects s_1, but rather treats $\tilde{s}_h$ (to some extent) as independent. Call such a decision maker naïve and let his degree of naïveté be parameterized by $\chi \in [0,1]$ such that $\chi = 1$ implies full correlation neglect. A naïve agent extracts s_h from $\tilde{s}_h$ according to the rule

$$\hat{s}_h \;=\; \chi\tilde{s}_h + (1-\chi)s_h \;=\; s_h + \frac{1}{2}\chi(s_1 - s_h)$$

where $\hat{s}_h$ for $h \in \{2,3,4\}$ denotes the agent's (possibly biased) inference of s_h. He thus forms mean posterior beliefs according to

$$b_{CN} \;=\; \frac{s_1 + \sum_{h=1}^{3} \hat{s}_h}{4} \;=\; \bar{s} + \frac{3}{8}\chi(s_1 - \bar{s}_{-1})$$

where $\bar{s} = (\sum_{h=1}^{4} s_h)/4$ and $\bar{s}_{-1} = (\sum_{h=2}^{4} s_h)/3$.

Rearranging yields an individual- and task-specific naïveté parameter:

$$\chi \;=\; \frac{8 \times (b_{CN} - \bar{s})}{3 \times (s_1 - \bar{s}_{-1})}$$

For each individual, I then define their overall naïveté as the median χ across all tasks. Figure 3.23 plots the distribution of (median) naïveté in the follow-up study. As in Enke and Zimmermann (2015), this distribution exhibits a bimodal structure with some fraction of subjects fully accounting for the double-counting problem and others approximately fully ignoring the partial redundancy.

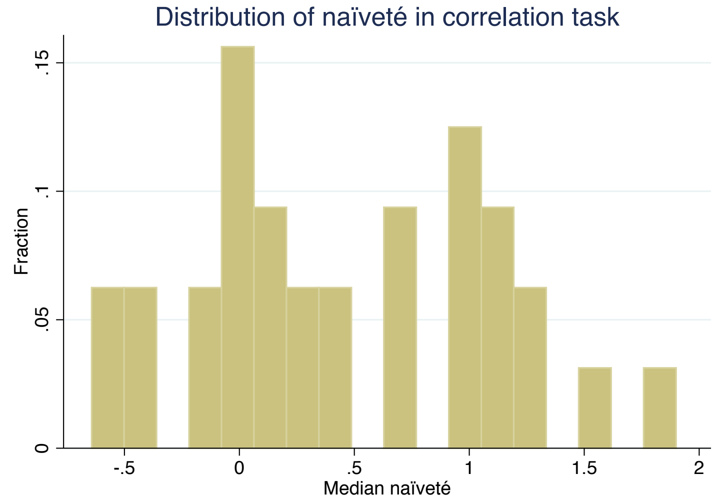


Figure 3.23. Distribution of median naïveté in correlation neglect task.

3.F.3 Relationship Between Correlation and Selection Neglect

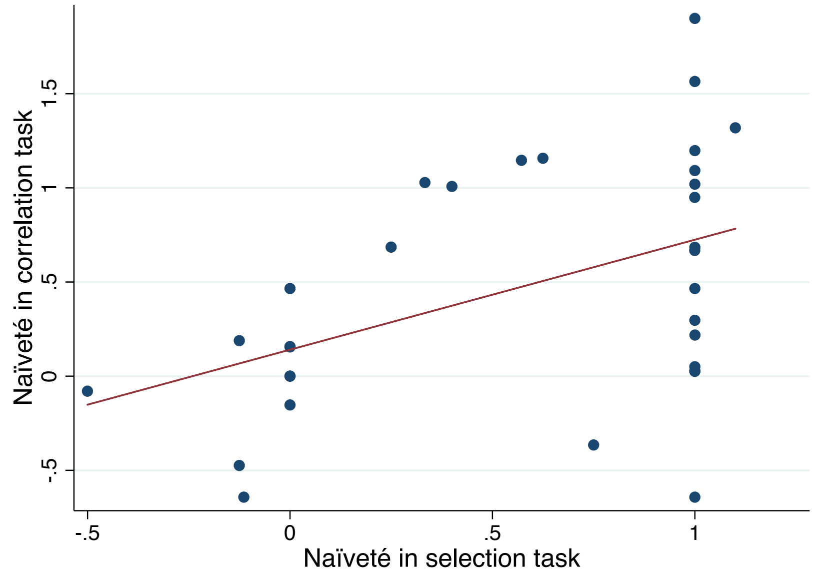


Figure 3.24. Raw correlation between selection and correlation neglect median naïveté parameters ($\rho = 0.44$).

4

The Precision of Expectations Data and the Explanatory Power of Economic Models*

4.1 Introduction

Stock market expectations are among the most important primitives of economic portfolio choice models. With the recent emergence of large-scale datasets including subjective expectations, researchers have begun to incorporate them into empirical models of investor behavior. While the results have been by and large encouraging, working with subjective beliefs data has proved challenging. First, many researchers are troubled by the apparent pervasiveness of measurement error in subjective expectations data. For example, stated beliefs often cluster at focal points (Kleinjans and van Soest, 2014) and many respondents' answers violate even the most basic laws of probability (Manski, 2004; Hurd et al., 2011). Second, the association between subjective beliefs and stockholding decisions tends to be statistically significant, but usually rather small in magnitude (**AmeriksEtAl2016**; Hurd, 2009).

In this paper, we propose a reconciliation of these two facts. Our point of departure is that different households are likely to employ different thought processes to arrive at their financial decisions. For some people, the canonical economic model of forming a choice rule by combining preferences and beliefs about future states of the world will be a good approximation. Others, however, could take their decisions very differently. For example, almost half of the Dutch population report that they mostly rely on the advice of family, friends, or professionals when it comes to important financial decisions (von Gaudecker, 2015). Likewise, as emphasized by large literatures in behavioral

* We would like to thank the team of CentERdata, especially Miquelle Marchand, for their help in the data collection process, Philipp Kloke and Lukas Wendlik for able research assistance, Bas van Heiningen for help in recording the introductory screencast, as well as Jürgen Maurer for sharing and explaining the code used in Maurer et al. (2010) that forms the basis of ours. Seminar participants at the Max Planck Institute for Research on Collective Goods, at Queen's University, at the Mannheim meeting of Young German Microeconometricians, at Tilburg University, and at the Strasbourg conference in honor of François Laisney provided helpful comments.

finance and cognitive psychology, households may take financial decisions intuitively (**BinswangerSalm2014**; Kahneman, 2011) or employ simple rules of thumb (Ameriks and Zeldes, 2004).

The presence of such alternative decision modes can produce the patterns observed in the data. First, individuals who do not base their decisions upon beliefs have little reason to frequently reflect upon the evolution of the stock market. Thus, they will likely maintain only rudimentary, diffuse, and unstable expectations. In consequence, when prodded to state these expectations in surveys, their answers will lack precision: they will be error-ridden, inconsistent, and exhibit large variation across survey instruments. Second, preferences and beliefs will have little explanatory power for portfolio decisions as they do not enter the decision-making process of all individuals. For example, neither preferences nor beliefs will explain variation in the behavior of people who exclusively rely on a rule of thumb to arrive at their financial decisions. In combination, these observations imply our research hypothesis. The responsiveness of financial decisions to variation in subjective expectations and other primitives of economic models should be high for individuals whose stated beliefs exhibit high precision. Beliefs and preferences should induce only very little variation in financial decisions for people with imprecise expectations measures.

To explore the channel of heterogeneous choice rules and motivate our empirical strategy, Section 4.3.1 presents a simple economic model of stock market participation that clarifies the roles of expectations, preferences, and transaction costs. In Section 4.3.2, we discuss in detail why a variety of alternative decision modes imply that individuals have low incentives to frequently reflect upon their beliefs about the future evolution of the stock market.

Section 4.3.3 lays out our econometric approach. The above arguments suggest that the explanatory power of our model of stock market participation will vary across individuals. To empirically incorporate this particular form of heteroskedasticity, we estimate a Klein and Vella (2009) semiparametric double index model. In this model, the first index contains the primitives of our theoretical model (such as beliefs and preferences), while the second index includes quantitative and qualitative indicators for the precision of measured beliefs. Both indices include further controls and may interact in a fully nonparametric fashion to obtain predicted stockholding probabilities.

Section 4.2 describes the dataset that we collected specifically for this study. The data contain individual-level information on stock market participation, subjective belief distributions, risk preferences, as well as a variety of quantitative and qualitative proxies for the precision of subjective expectations from a large probability sample of the Dutch population. Section 4.4 presents the results of our empirical application. We demonstrate that changes in primitives of the economic model induce large variation in stock market participation if expectations measures are precise. If their precision is low, however, the effect of changes in beliefs and preferences on stockholdings is close to zero. We perform a number of variations on this theme and show that the results hold up in several different specifications. We then demonstrate the usefulness of our modeling approach for the analysis of less detailed data by estimating a specification with variables that are commonly available or inexpensive to collect. In particular, we show that restricting ourselves to a simple measure of expectations and purely qualitative proxies for the precision of expectations measures yields a similar, yet less pronounced overall pattern.

Our findings suggest that imprecision in measured beliefs should not necessarily be treated as a standard case of measurement error, which needs to be corrected through, e.g., improved measurement devices or multiple measurements (Wansbeek and Meijer, 2000). While many of the symptoms of diffuse and unstable expectations are observationally equivalent to measurement error, they do not reflect erroneous reporting, but rather the structure of the expectations. Our results hence suggest that individual-level variation in the precision of measured expectations might be informative about economic mechanisms of interest. To bolster this interpretation, we conclude in Section 4.5 by discussing why our findings are unlikely to be driven by traditional notions of measurement error in subjective beliefs.

4.2 Data and Descriptive Statistics

Our data stem from the Dutch LISS study (Longitudinal Internet Studies for the Social Sciences), which regularly administers Internet surveys and experiments to a panel of households comprising a probability sample drawn from the population register kept by Statistics Netherlands.

Implementing our empirical strategy requires data on individual stock market participation, subjective beliefs and risk aversion, proxies for the degree of imprecision in individual responses, and a rich set of sociodemographic covariates. Only the latter are present in the LISS panel by default. In order to obtain measures for the main quantities of interest, we implemented a series of incentivized experiments and survey questions in August and September of 2013. We restricted our experiments to households with financial wealth in excess of 1,000 € to focus on respondents with substantial incentives to think about portfolio allocations. To increase turnout, we also included individuals who refused to answer questions about their exact amount of wealth. Within households, we selected the financial decision maker. In total, 2,125 individuals completed both survey waves. After dropping observations with missing data, we are left with a final sample of 2,072 observations.

4.2.1 Outcome Variable: Stock Market Participation

LISS routinely collects detailed data on respondents' financial background, including information on asset ownership. To ensure the relevance of elicited beliefs for current portfolio allocations, we asked respondents to update their information on asset holdings in August 2013. For this purpose, we asked them whether they had any type of bank or savings account and/or investments (stocks, bonds, funds, or options). Our outcome variable is a binary index that equals 1 if the respective respondent held any investments, and 0 otherwise. A quarter of the households in our sample holds risky assets (cf. Table 4.1). This is in the range of values reported for the Netherlands from other datasets and earlier periods (Alessie et al., 2004; van Rooij et al., 2011). In particular, using an administrative dataset from the Netherlands, **KnoefEtAl2015** report almost exactly the same rate of stock market participation, providing reassuring evidence for the data quality of our main outcome variable.

Table 4.1. Descriptive Statistics

	Statistic		Index	
	Mean	Std. Dev.	Model	Subj. Data Prec.
Holds risky assets	0.25			
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	-1.18	8.10	×	
Subjective beliefs: μ_{t+1}^{AEX}	2.01	6.19		
Subjective beliefs: $\mu_{t+1}^{\text{sav. acc.}}$	3.18	4.89		
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	6.25	4.01	×	
Risk aversion	0.00	1.00	×	
Absolute difference between belief measures	11.20	13.57		×
Lack of confidence in AEX return estimate	0.54	0.23		×
Lack of confidence in sav. acc. return estimate	0.36	0.24		×
Experimental tasks difficult	0.49	0.33		×
Experimental tasks obscure	0.31	0.25		×
Financial wealth $\in$ (10000 €, 30000 €]	0.27		×	×
Financial wealth $\in$ (30000 €, ∞)	0.27		×	×
Financial wealth missing	0.18		×	×
Net income $>$ 2500 €	0.46		×	×
Net income missing	0.07		×	×
High education	0.38		×	×
$30 < \text{Age} \leq 50$	0.30		×	×
$50 < \text{Age} \leq 65$	0.34		×	×
$\text{Age} > 65$	0.29		×	×

Sources: LISS panel and own calculations. Variables related to the confidence in return estimates, task difficulty, and task obscurity are scaled to range between 0 and 1. Risk aversion is the standardized average of 3 standardized risk aversion proxies. We omit standard deviations of binary variables. The number of observations is 2,072.

4.2.2 Variables Entering the Economic Model Index

Subjective Expectations. In August 2013, we asked respondents to describe their expectations about the one-year return of the Amsterdam Exchange Index (AEX). We employed a variation of the ball allocation procedure developed by Delavande and Rohwedder (2008), which was explicitly designed for usage in Internet experiments. For each individual, the procedure yields an 8-binned histogram for the expectation of the AEX's one-year return. Using the resulting 7 points on the cumulative distribution function, we follow Hurd et al. (2011) and fit a log-normal distribution to obtain individual-level measures for μ_{t+1}^{risky} and $\sigma_{t+1}^{\text{risky}}$. Because our theoretical framework requires expected excess returns, we also asked respondents for a point estimate of the return of a one-year investment into a standard savings account as the most prevalent safe asset. Section 4.A.1.1 of the Internet Appendix contains detailed descriptions of both procedures.

Recent research in the experimental economics literature has shown that financial incentives induce more truthful reporting of beliefs in tasks like ours (see, for example, Palfrey and Wang, 2009; Gächter and Renner, 2010; Wang, 2011). In order to incentivize subjects, we employed the binarized scoring rule of Hossain and Okui (2013) which is incentive-compatible for a wide range of utility functions. As is common practice with large samples like ours, we randomly selected one in ten subjects for actual payment. The maximum earnings per selected subject were 100 € and average earnings equaled 39.66 € conditional on being selected for payment in September 2014.

We relegate a detailed presentation of summary statistics of the belief measures to Section 4.A.1.1 of the Internet Appendix and only discuss some notable features at this point. First, the cross-sectional patterns in our data resemble findings in previous literature (e.g., Manski, 2004; Hurd, 2009; Hurd et al., 2011), e.g., we find that male, richer, and better educated respondents tend to hold more optimistic expectations. Second, though our respondents expect a positive AEX return on average, their expectations are rather pessimistic relative to the AEX's historical return distribution: the mean subjective expectation implied by the distribution is 2.01%, while the AEX returned 7.89% (5.93% inflation-adjusted) on average since 1993. This discrepancy between subjective expectations and historic returns aligns with existing results in the literature, in particular those in Hurd (2009) regarding the AEX. In addition, as Figure 4.9 in Section 4.A.1.1 of the Internet Appendix shows, our participants tend to place lower probabilities on extreme returns than what has historically been observed. Finally, and in contrast to the relative pessimism we observe for the AEX, the mean expected return for the savings account, 3.18%, exceeds the rates actually offered at the time of the survey (roughly 1%) by a substantial percentage. In our empirical analyses, we employ the difference between the expected mean return for the AEX and the expected return for the savings account as the empirical analog of the expected excess return.

Risk Preferences. In September 2013, we elicited risk preferences by asking respondents to complete a variant of the "Preference Survey Module", which was developed in Falk et al. (2014) to measure economic preference parameters in large-scale surveys. We further describe it in Section 4.A.1.3 of the Internet Appendix. Respondents first provided a qualitative self-assessment of their willingness to take risks in general and in the financial domain. They then made choices in a series of hypothetical binary lottery tasks. In our main analysis, we employ the average of the three measures' standardized values.

Transaction costs. We include several variables to empirically model the impact of transaction costs on stock market participation decisions. We focus on variables that proxy for variation in transaction costs in the form of either monetary or information costs. If monetary expenses of stock market participation are to some degree fixed—e.g., because banks charge a constant amount for setting up and keeping an investment account—then these costs will be less relevant for wealthy households. We therefore include net household income and financial wealth in the economic index to control for variation in the relevance of monetary transaction costs. If comprehension of the basic functioning of the stock market comes with information costs, then these costs will be lower for more numerate and cognitively able households. Both vary with educational attainment and age (McArdle et al., 2011), which we include as further controls.

4.2.3 Variables Entering the Subjective Data Precision Index

Several quantitative and qualitative measures serve to capture the precision of subjective expectations data. We employ variables for (i) the consistency with which participants report their expectations, (ii) the confidence they express in their own beliefs, and (iii) their self-assessment concerning both difficulty and clarity of our survey tasks. On top of such direct proxies, we also include the variables entering transaction costs. Indeed, it is difficult to argue for exclusion restrictions in one direction or another for education, income, financial wealth, or age.

In September 2013, one month after eliciting the distribution of beliefs, we asked the same set of respondents to provide a point estimate for the one-year ahead return of the AEX. As a quantitative proxy for the precision of households' expectations, we compute the absolute difference between the response to this question and the mean belief from the ball allocation task. We conjecture that large discrepancies between the two estimates indicate that a household entertains only diffuse expectations and is thus unlikely to employ them in actual decision-making.[1]

The first two qualitative proxies relate to the confidence respondents have in their own estimates. Following the elicitation of the point estimates for the expected returns of the AEX and the savings account, we asked respondents to use a slider interface to express their confidence in their own belief on a scale from 0 to 10, where larger values corresponded to more confidence. We conjecture that respondents maintaining only imprecise expectations will have little faith in their own estimates. For our analysis, we scale answers to both questions to the unit interval.

Both in August and September 2013, we asked subjects to use five-point scales to indicate how clear they found the task descriptions and how simple they considered the belief elicitation itself. We expect that respondents who do not have an elaborate belief distribution find it hard to understand and to complete the tasks. For both questions, we aggregate the responses for August and September and we scale the resulting variables to the unit interval to create two further proxies.

The Internet Appendix provides a more detailed description and further summary statistics of all proxies. The pairwise correlations between task simplicity, clarity, and the two confidence variables are all positive, whereas all of them are negatively correlated to the absolute difference between the two belief measures. Notably, all of the proxies' correlations to sociodemographic variables conform to our prior expectations. For example, the correlations suggest that highly educated households or households with higher net income entertain more precise expectations, resembling previously-found patterns regarding inconsistent survey responses or item non-response (Manski, 2004; Hurd, 2009).

[1] We are not aware of changes in the economic environment between the two surveys that could have induced people to systematically and substantially revise their beliefs. Between August and September 2013, the AEX varied little with closing prices between 362.93 and 382.58.

4.3 Motivation and Empirical Strategy

We develop our empirical strategy in three steps. First, we characterize a household's portfolio choice problem by means of a simple economic model. We then explain in detail why we conjecture that the degree to which this model serves as an adequate description of the decision-making process varies across households and why we expect that variation in the precision of subjective expectations can be exploited to capture this adequacy. In the third step, we present our econometric strategy to implement these ideas.

4.3.1 A Simple Economic Model of Stock Market Participation

Our depiction of households' portfolio choice behavior in an economic model follows Campbell and Viceira (2002). We assume that the household maximizes a power utility function defined over next period's expected financial wealth $E_t\left[W_{t+1}\right]$ by allocating fractions of period-t wealth to one safe and one risky asset. If the household can neither short the risky asset nor leverage his position in it, the optimal risky asset share θ^{opt} solves:

$$\theta^{\text{opt}} \;=\; \arg\max_{\theta} \left\{ \frac{E_t\big[W_{t+1}(\theta)^{1-\gamma}\big]}{1-\gamma} \right\} \qquad \text{s.t.} \quad 0 \;\leq\; \theta \;\leq\; 1$$

Risk aversion and a household's beliefs about the returns of the two assets determine the optimal decision. Denote a household's expected return for the safe asset by μ_{t+1}^{safe} and assume that the household's expectations for the risky asset's return can be described by a log-normal distribution with mean μ_{t+1}^{risky} and standard deviation $\sigma_{t+1}^{\text{risky}}$. When returns are log-normally distributed, so is W_{t+1}. For a log-normal variable it holds that $\log E[X] = E\big[\log X\big] + 1/2\,\text{Var}\big[\log X\big]$. Thus, the maximization problem can be rewritten as:

$$\theta^{\text{opt}} \;=\; \arg\max_{\theta} \left\{ (1-\gamma)E_t\big[w_{t+1}(\theta)\big] + \frac{1}{2}(1-\gamma)^2\,\text{Var}_t\big[w_{t+1}(\theta)\big] \right\} \qquad \text{s.t.} \quad 0 \;\leq\; \theta \;\leq\; 1$$

where lower case letters are logarithms. Using a first-order Taylor series approximation, next period's log wealth can be written as:

$$w_{t+1}(\theta) \;=\; w_t + (1-\theta)\mu_{t+1}^{\text{safe}} + \theta\mu_{t+1}^{\text{risky}} + \frac{1}{2}\theta(1-\theta)\left(\sigma_{t+1}^{\text{risky}}\right)^2$$

Substituting this into the expression for θ^{opt} and dividing by $1-\gamma$, we obtain for the maximand:

$$w_t + \theta\left(\mu_{t+1}^{\text{risky}} - \mu_{t+1}^{\text{safe}}\right) + \frac{1}{2}\theta\Big(1-\gamma\theta\Big)\left(\sigma_{t+1}^{\text{risky}}\right)^2$$

Solving the first-order condition of this problem for the optimal share θ^{opt} yields:

$$\theta^{\text{opt}} = \frac{\mu_{t+1}^{\text{risky}} - \mu_{t+1}^{\text{safe}} + \frac{1}{2}\left(\sigma_{t+1}^{\text{risky}}\right)^2}{\gamma\left(\sigma_{t+1}^{\text{risky}}\right)^2} \tag{4.1}$$

At plausible parameter values of γ, the optimal risky asset share will be positive when estimates based on historical return data are used to proxy households' expectations for $\mu^{\text{safe}}, \mu^{\text{risky}}$, and σ^{risky}. However, studies on stock ownership find that a large fraction of the population does not participate in the stock market (e.g., Haliassos and Bertaut, 1995). Arguably the most prominent explanation for why households abstain from participation is the existence of broadly defined transaction costs (Vissing-Jørgensen, 2002). These transaction costs are likely to vary with household characteristics. If participation comes with fixed monetary costs, for example, wealthy households will be more likely to invest in risky assets, since for them the fixed costs are spread over larger investments. If information costs play an important role, transaction costs will be lower for numerate respondents who are quicker to grasp the basic functioning of the stock market. We assume that the variables affecting transaction costs can be modeled by observable household characteristics X^{ta}; denote the resulting transaction costs by $f(X^{\text{ta}})$.

We now combine the optimal risky asset share (4.1), transaction costs, and random influences ε in a simple random utility model of stock market participation:

$$Y \equiv \mathbf{I}\{\theta > 0\} = \begin{cases} 1 & \text{if} \quad \theta^{\text{opt}}\left(\mu_{t+1}^{\text{risky}} - \mu_{t+1}^{\text{safe}}, \sigma_{t+1}^{\text{risky}}, \gamma\right) - f(X^{\text{ta}}) > \varepsilon \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

According to (4.2), the probability of participating in the stock market will depend on the mean and variance of beliefs over the risky asset, the expected risk-free rate, risk aversion, variables proxying transaction costs, and the stochastic properties of ε. If the latter was normally distributed, one could estimate (4.2) by means of a standard Probit model. Estimators that make minimal distributional assumptions but enable the researcher to recover marginal effects still require ε to either be homoskedastic or have a very particular form of heteroskedasticity (Klein and Vella, 2009). If our conjecture about a varying explanatory power of $\theta^{\text{opt}} - f(X^{\text{ta}})$ is correct, this will be reflected in a form of heteroskedasticity that violates these assumptions. In particular, the variance of ε will vary with the precision of beliefs in a form that is unknown a priori.

4.3.2 Putting the Precision of Subjective Data to Productive Use

The model combines effortful reasoning about future states of the world with personal risk tolerance to form a choice rule. While such behavior is at the heart of economic thinking, it will only adequately describe the decision process of a part of the population. The behavioral finance and cognitive psychology literatures have proposed a number of alternative decision modes. For example, almost half of the Dutch population report that they mostly rely on the advice of family, friends, or professionals when it comes to important financial decisions (von Gaudecker, 2015). Other households may take decisions intuitively (**BinswangerSalm2014**; Kahneman, 2011) or employ simple rules of thumb like holding an equity share of 100 minus age (see, e.g., the discussion in Ameriks and Zeldes, 2004).

Many of these alternative decision processes, however, do not require households to frequently reflect about the future evolution of the stock market. As a consequence, we suggest that households who rely on such decision processes will only maintain very rudimentary, possibly diffuse or even unstable expectations. Eliciting such expectations will lead to imprecise, inconsistent, and error-ridden measurements even when using the same survey instrument at different points in time. Likewise, such respondents should find tasks related to belief elicitation rather difficult and the confidence they express in their estimates should be low.

Indeed, these patterns closely resemble the measurement issues that have been documented in the vast literature on subjective expectations of stock market developments (see the excellent overviews in Manski (2004) and Hurd (2009)). For example, when asked for their expectations about the future of the stock market, respondents frequently violate basic laws of probability or they provide focal point answers such as 50:50 (**BinswangerSalm2014**; Bruine de Bruin et al., 2000; Manski, 2004; Hurd, 2009; Bruine de Bruin and Carman, 2012; Kleinjans and van Soest, 2014). In addition, non-response tends to be concentrated among sub-groups who do not follow the development of the stock market (Hurd, 2009), suggesting that stating beliefs requires significant cognitive effort for people who are not accustomed to reflecting upon the stock market.[2]

Previously, such patterns have frequently been interpreted as cases of measurement error (see, e.g., the discussion in Manski, 2004). However, while often observationally equivalent to measurement error, the semantics of imprecise expectations is very different from the contexts in which measurement error is usually studied. In the case of variables like past income, savings, or consumption, measurement error arises because of, e.g., imperfect recall (Hoderlein and Winter, 2010) or incongruent definitions of precisely defined “true” non-stochastic quantities. In the case of subjective expectations, however, we conjecture that the precision and meaningfulness of expectation measures reflects the structure of beliefs itself. In consequence, when attempting to predict household investment behavior, the degree of precision should be informative about the relevance of expectations in the decision process. Specifically, we hypothesize that measures indicative of more precise expectations should be associated with an increase in the explanatory power of expectations for variation in portfolio decisions.

In sum, different pieces of evidence suggest that part of the population holds only imprecise subjective stock market beliefs. We propose that this imprecision contains informational content that will allow us to uncover heterogeneity in choice behavior. In particular, we suggest that the degree of imprecision will allow us to evaluate to which extent households’ stock market participation decisions are adequately described by the simple model discussed above.

[2] Similar patterns of imprecise measurements have been documented for risk preferences. von Gaudecker et al. (2011) and Choi et al. (2014) show that for respondents with high socio-economic status, sequences of lottery decisions are much more consistent with flexible parametric utility functions and the generalized axiom of revealed preferences, respectively. Put differently, risk preference parameters are much more precisely measured for these subgroups.

4.3.3 Econometric Specification

In econometric terms, a consequence of varying precision in expectations measures is that ε in (4.2) will be heteroskedastic, i.e., its variance will increase as subjective expectations become noisier. Depending on the precise decision-making process, it may also have group-specific means different from zero. For example, the most prevalent advice by family and friends seems to be non-participation in the stock market (von Gaudecker, 2015). For the group of individuals who follow this advice, participation rates will be low even if $\theta^{\text{opt}} - f\,(X^{\text{ta}})$ takes on positive values on average. To capture these consequences, we require an econometric specification where the predictions of the choice model (4.2) interact with the extent of precision in subjective expectations data in a flexible way. The double index binary choice model of Klein and Vella (2009) is ideally suited for the structure of our problem. The model obtains an estimate of the probability of stock market participation by nonparametrically combining two linear indices.

We first aggregate $\mu_{t+1}^{\text{risky}} - \mu_{t+1}^{\text{safe}}, \sigma_{t+1}^{\text{risky}}, \gamma$, and X^{ta} into one vector X^{mod}; $X^{\text{mod}}\beta^{\text{mod}}$ approximates our choice model from 4.3.1.[3] We will refer to $X^{\text{mod}}\beta^{\text{mod}}$ as the economic model index in what follows. A second vector X^{sdp} contains the variables related to the subjective data's precision. These will be quantitative and qualitative indicators as well as covariates that we would expect to influence the "propensity to use economic reasoning"; we allow the latter to overlap with the transaction cost proxies included in the economic model index. Accordingly, we refer to $X^{\text{sdp}}\beta^{\text{sdp}}$ as the subjective data precision index. The Klein and Vella (2009) estimator models the relationship of both indices and risky asset holdings as:[4]

$$P\big(Y = 1 \mid X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\big) = h\big(X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\big) \tag{4.3}$$

This structure is directly related to (4.2) in that the subjective data precision index further parameterizes ε, i.e., the random component is systematic to some extent. The function $h(\cdot,\cdot)$ provides a nonparametric link mapping the indices for the economic model and subjective data precision into stock market participation probabilities.

To attain identification (up to location and scale) of the parameters β^{mod} and β^{sdp}, we require that at least one continuous variable per index is excluded from the other index. In each index, we normalize the coefficients on one of these variables. The resulting model satisfies the form in A5 of Klein and Vella (2009) without requiring reparameterization. Under assumptions given in Klein and Vella (2009)—mainly smoothness of $h(\cdot,\cdot)$ and compact support of the covariates—the probability to participate in the stock

[3] We also experimented with calculating (4.1) and including it alongside X^{ta}. This led to numerical difficulties as the covariance matrix of the two indices was near-singular for a wide range of parameter values. We attribute this to the lack of a quantitatively meaningful measure of γ (Rabin, 2000) and to a fat right tail of $\big(\sigma_{t+1}^{\text{risky}}\big)^2$. The latter is likely responsible for the numerical problems; it is also the reason why we use the standard deviation of beliefs instead of the variance.

[4] Klein and Vella (2009) frame their discussion in terms of an estimator for a single-equation binary response model with dummy endogenous variable when no instruments are present. A first application that applies it directly to two indices is given in Maurer (2009).

market can be expressed as a function of the densities conditional on participation:

$$P\left(Y = 1 \mid X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\right) = \frac{f_{Y=1}\left(X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\right) \cdot P\left(Y = 1\right)}{f\left(X^{\text{mod}}\beta^{\text{mod}},\ X^{\text{sdp}}\beta^{\text{sdp}}\right)}, \tag{4.4}$$

where $f(\cdot)$ denotes the unconditional density of the bivariate index and $f_{Y=1}(\cdot)$ its density conditional on participation in the stock market. Kernel density estimators for these quantities are obtained under a multi-stage local smoothing procedure to achieve a sufficiently low order of the bias. Denoting the resulting estimator for (4.4) as $\hat{P}_i\left(\beta^{\text{mod}}, \beta^{\text{sdp}}\right)$, we can write the semiparametric maximum likelihood estimator for $\beta^{\text{mod}}, \beta^{\text{sdp}}$ as:

$$\left(\hat{\beta}^{\text{mod}}_{\text{ml}},\ \hat{\beta}^{\text{sdp}}_{\text{ml}}\right) = \underset{\beta^{\text{mod}},\ \beta^{\text{sdp}}}{\arg\max} \sum_{i=1}^{N} \hat{\tau}_i \left[Y_i \cdot \log \hat{P}_i\left(\beta^{\text{mod}}, \beta^{\text{sdp}}\right) + (1 - Y_i) \cdot \log\left(1 - \hat{P}_i\left(\beta^{\text{mod}}, \beta^{\text{sdp}}\right)\right)\right], \tag{4.5}$$

where $\hat{\tau}_i$ denotes a smooth trimming function ensuring that densities do not become too small (Klein and Spady, 1993). Klein and Vella (2009) show that $\left(\hat{\beta}^{\text{mod}}_{\text{ml}},\ \hat{\beta}^{\text{sdp}}_{\text{ml}}\right)$ converges at rate $\sqrt{N}$ to its true value. While the parameter values do not allow for a direct interpretation, various quantities of interest like average partial effects can be computed with little effort.

In sum, when it comes to generating choice behavior, our empirical model allows for a flexible interplay between traditional economic parameters and proxies for their precision. In particular, it will allow an analysis of how marginal changes in model parameters translate into variation in stock market participation, and how this relationship varies across respondents.

4.4 Results

4.4.1 Main Specification

Table 4.2 presents parameter estimates for the coefficients of the main specification. In the economic model index, we normalize the coefficient on $\mu^{\text{AEX}}_{t+1} - \mu^{\text{sav. acc.}}_{t+1}$ to 1, thus expressing the remainder of β^{mod} relative to subjective excess return expectations. In the subjective data precision index, we set the coefficient on the absolute difference between the belief measures to -1. Larger values in this index would thus be interpreted as indicative of more precise data. As we will discuss in detail below, the link function $h(\cdot,\cdot)$ is (close to) monotonically increasing in the economic model index as well as in the subjective data precision index. This allows us to infer the direction of partial effects from the coefficient estimates.

The coefficients in both indices are estimated with reasonable precision; their signs and relative magnitudes are plausible given the aforementioned shape of the link function and the scaling of the variables (see Table 4.1). In particular, all variables with exclusion restrictions have the expected signs and most of them are significant. The economic

Table 4.2. Coefficient estimates for the economic model index and the subjective data precision index

	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.76	0.29	·	·
Risk aversion	-7.90	1.78	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	59.04	27.55
Lack of confidence in sav. acc. return estimate	·	·	29.23	21.97
Experimental tasks difficult	·	·	54.88	19.70
Experimental tasks obscure	·	·	15.25	18.21
Financial wealth ∈ (10000 €, 30000 €]	20.16	5.93	-19.00	21.21
Financial wealth ∈ (30000 €, ∞)	42.73	9.14	-91.51	36.79
Financial wealth missing	30.06	7.28	-58.24	27.80
Net income > 2500 €	7.32	2.65	28.48	11.48
Net income missing	-6.37	4.14	-4.85	12.86
High education	3.52	2.96	-63.59	19.18
30 < Age ≤ 50	11.78	5.55	22.90	16.86
50 < Age ≤ 65	7.24	5.53	-16.56	15.15
Age > 65	-0.45	5.23	-22.80	16.33

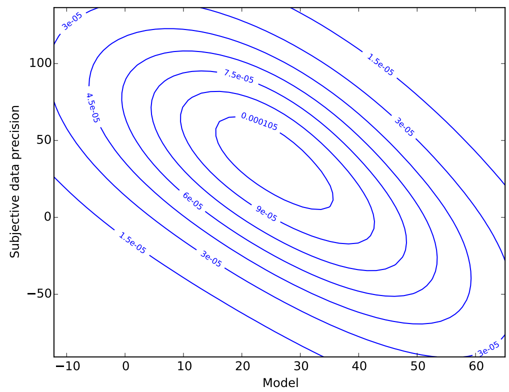
Sources: LISS panel and own calculations. The table shows coefficient estimates for the double index binary choice model of Klein and Vella (2009); see Section 4.3.3 for a detailed description. The dependent variable is a household's stock market participation decision, a binary variable equalling 1 in case the household reports holding any investments, and 0 otherwise. Columns 2 and 3 present estimates of the coefficients and standard errors for the variables contained in the economic model index. Columns 4 and 5 present estimates for the variables contained in the subjective data precision index. In the first index, we normalize the coefficient of the mean excess return to 1, whereas we normalize the coefficient on the absolute difference between the belief measures to -1 in the second index.

model index increases in the level of the expected excess returns; it decreases in the standard deviation of returns and in risk aversion. The subjective data precision index increases with all 4 qualitative proxies and, by construction, decreases with the absolute difference between the belief measures.

Both indices vary significantly with a number of the common covariates. For example, financial wealth is positively related to both indices. This is consistent with wealthy households facing lower transaction costs, while at the same time having stronger incentives to form an opinion about stock market developments. Interestingly, education seems to mostly work through the subjective data precision index, but it has little impact on the economic model index.

For presenting the results of semi- and nonparametric methods, it is particularly important to clarify the support of the data, which in our case refers to the two indices. Figure 4.1 shows a contour plot of the joint density of the estimated indices. We limit the area of Figure 4.1 and of all subsequent plots to the rectangle spanned by the 5% — 95% quantiles of the marginal distributions of both indices. With a correlation coefficient of 0.63, the indices are characterized by a pronounced positive correlation. Note that this correlation does not arise purely mechanically due to the previously noted influence of wealth on both indices – in a model that drops all variables common to both indices (described in the next section), we find the same pattern.

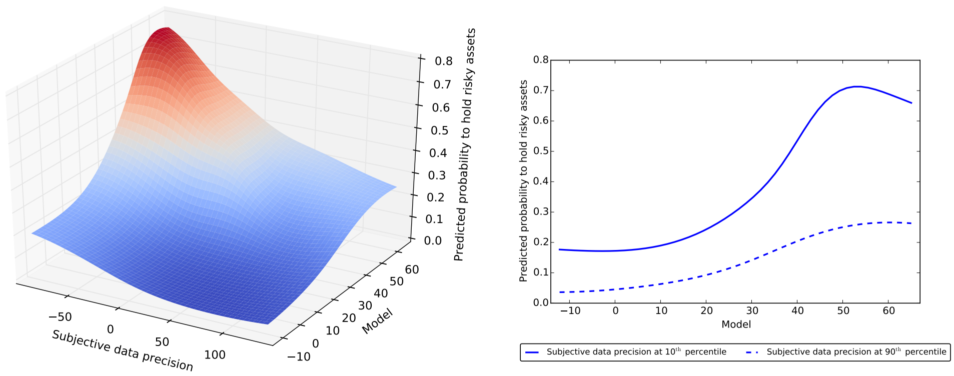
The left panel of Figure 4.2 plots the link function $h(\cdot,\cdot)$, i.e., the predicted probability of stock market participation, for varying levels of the economic model and subjective



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.1. Joint density of the two indices

data precision indices. Three features of the plot stand out: First, predicted stock market participation rates vary substantially, ranging from single-digit values to more than 70%. Second, participation rates in general increase monotonically in both the index for the economic model and the subjective data precision index. Third and most importantly, the effects are highly non-linear and interact strongly. In particular, stock market participation is much more responsive to changes in the economic model's ingredients at high levels of the subjective data precision index than at low levels.



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision index. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index (43 and 223). Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.2. Predicted probability to hold risky assets

To illustrate the last point more clearly, the second panel in Figure 4.2 extracts two slices from the first panel. The solid line shows the average response of stock market participation to variation in the model index at the 90%-quantile of the subjective data precision index. There is a pronounced gradient in the middle region, causing predicted risky asset participation to rise from just below 20% to 70%. The dashed line plots the same relation for the 10%-quantile of precision in subjective data. Again, predicted stock market participation varies in the economic model index as expected, but to a much lesser extent. In particular, even for the highest levels of the economic model index, the predicted probability of participation does not rise above 30%. The discrepancy in shapes of the two lines highlights the importance of precision in subjective data in understanding the relationship between the primitives of economic models and choices.

We calculate average partial effects to quantify the dependence between individual covariates and stock market participation probabilities. In Table 4.3, we show how changes in covariates affect participation through either the economic model or subjective data precision index. We also show the combined effect that operates through both indices simultaneously. To calculate average partial effects, we increase continuous variables by one standard deviation. For binary variables, we assign individuals in the left-out category a value of 1.

For the variables solely included in the economic model index, the average partial effects of expected excess return and risk aversion are somewhat larger than the effect of a change in the expected standard deviation of returns. An increase in the expected excess return by one standard deviation is associated with an increase of 3.4 percentage points in the probability to hold investments. Comparable increases in the expected standard deviation and risk aversion reduce the predicted participation rate by 1.4 and 3.8 percentage points, respectively. In the subjective data precision index, a one standard deviation increase in the absolute difference between the two belief measures reduces predicted participation by 1.4 percentage points. Increases in either of the 4 remaining proxies by one standard deviation increase the propensity to participate by between 0.4 and 2 percentage points. If one thinks of the different proxies in terms of a factor structure, varying the underlying factor would likely yield effects of the same order of magnitude as for beliefs or risk aversion.

The effects of financial wealth tend to work through both indices, increasing the propensity to participate in the stock market through the economic model index as well as the subjective data precision index. In contrast, education seems to affect participation mainly through the subjective data precision index.

In sum, this section indicates that respondents' beliefs and risk attitudes are indeed predictive of economic choices. However, the extent to which this is the case varies strongly in the population. Hence, precision in the primitives of the economic model can be used to uncover heterogeneity in its explanatory power.

4.4.2 Robustness

To illustrate the robustness of our results to alternative specifications of both the economic model and the subjective data precision index, we now present an overview of a number of additional analyses. Section 4.B of the Internet Appendix contains all tables, figures, and additional information.

Table 4.3. Average partial effects

	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.034	·	0.034
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.014	·	-0.014
Risk aversion	-0.038	·	-0.038
Absolute difference between belief measures	·	-0.014	-0.014
Lack of confidence in AEX return estimate	·	-0.014	-0.014
Lack of confidence in sav. acc. return estimate	·	-0.007	-0.007
Experimental tasks difficult	·	-0.019	-0.019
Experimental tasks obscure	·	-0.004	-0.004
Financial wealth ∈ (10000 €, 30000 €]	0.099	0.017	0.098
Financial wealth ∈ (30000 €, ∞)	0.247	0.119	0.373
Financial wealth missing	0.171	0.068	0.219
Net income > 2500 €	0.037	-0.028	0.009
Net income missing	-0.031	0.005	-0.027
High education	0.017	0.080	0.098
30 < Age ≤ 50	0.055	-0.025	0.025
50 < Age ≤ 65	0.035	0.020	0.054
Age > 65	-0.002	0.027	0.019

Sources: LISS panel and own calculations. The table presents average partial effects of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description. The effects are calculated for a change of 1 standard deviation in continuous variables. For binary variables, we calculate the effect of assigning individuals in the left-out category a value of 1.

No transaction cost proxies. Our main specification includes several covariates that proxy transaction costs. Some of them—financial wealth in particular—have strong effects on stock market participation through both the economic model index and the subjective data precision index. To investigate whether the predicted interactions between the economic model and imprecise measures are driven by these sociodemographics only, we estimate one specification without all of the corresponding proxies, i.e., we only include beliefs, risk preferences, and subjective data precision proxies. Except for lower predicted levels of stock market participation at high values of the model index, the overall results on $h(\cdot, \cdot)$ look very similar. Naturally, the partial effects increase.

Mean beliefs only. In this specification, we restrict the model index to consist of expected excess returns only, which gives it an interpretable scale. Section 4.B.2 of the Internet Appendix shows that the gist of our main results is present even in this stripped-down version. The relationship between beliefs and stock market participation is essentially flat at the 10th percentile of the subjective data precision index, while the probability to hold stocks doubles along the beliefs distribution at the 90th percentile of the subjective data precision index. This doubling is concentrated around expected excess returns of zero, whereas the relationship is flat at both extremes of the beliefs distributions. The pattern illustrates the usefulness of our semiparametric approach; typical parametric models such as Logit or Probit would yield the steepest gradient to lie at the right tail of the index' support instead of the center.

Additional covariates. We also check the other extreme and employ a “kitchen-sink”-type approach, including binary variables for gender, having children, and being married in both indices along with the variables from our main specification. It turns out, however, that none of these is significantly associated with either the index of the economic model or the subjective data precision index. In consequence, their inclusion does not affect our results.

Discarding individuals with missing data on financial wealth. In our main specification, we included dummies for financial wealth terciles and for whether information on financial wealth was missing. Since wealth is among the strongest drivers of stock market participation in our model, it is possible that inclusion of respondents with missing information on portfolio value affects our results. To address this concern, we estimate our main specification only with respondents who provided all components of financial wealth. The results are very similar. In particular, the shape of $h(\cdot,\cdot)$ is virtually unchanged. Some of the average partial effects of beliefs and preferences slightly change in magnitude, but all of them qualitatively confirm the main results.

Alternative belief measure. We showed our main results using stated beliefs over the future development of the Amsterdam Exchange Index (AEX). While it is plausible that expectations over a composite index with high media exposure are a good proxy for "the" risky asset in our model, it is still conceivable that our results are biased due to this specific choice. We therefore elicited the same set of belief variables for the future stock return of Philips N.V., one of the largest publicly traded companies of the Netherlands. As one would expect for a single stock with additional idiosyncratic risk, average partial effects relating to the moments of the belief distribution become weaker. The general shape of the link function and the essence of the remaining results, however, is unchanged.

Disaggregated risk aversion measures. By averaging over three distinct variables, we employed a particularly simple aggregation procedure for the risk aversion measure used in our main analysis. When including the three variables separately in the model index, aversion to risk in financial matters emerges as its most important component (Section 4.B.10 of the Internet Appendix). The remainder of our results is not affected.

Interaction between risk aversion and subjective uncertainty. The main specification contains risk aversion and the standard deviation of the subjective belief distribution as separate variables. To investigate whether increased subjective uncertainty is more important for relatively risk averse subjects, we estimate an additional specification including their interaction in the economic model index. The results in Section 4.B.13 of the Internet Appendix closely resemble those for the main model, and they indicate that the effect of subjective uncertainty does not vary with risk aversion.

Raw returns instead of excess returns. Our theoretical framework suggests employing subjective expected excess returns to predict stock market participation. As discussed in Section 4.2, our subjects are simultaneously rather pessimistic about the future returns of the market and relatively optimistic about those of a standard savings account. In consequence, a large fraction of our sample expects negative excess returns. While this feature of our data is in line with previous literature, we estimate an additional specification replacing expected excess returns with expected returns to assess the robustness of our results. They are essentially unaffected.

Financial literacy. As mentioned above, a lack of financial literacy may lead subjects to base their participation decision not on expectations about risk and return but on alternative rationales. To assess how our results relate to variation in the respondents' levels of financial literacy, we ran an additional survey in October 2014. In this survey, we asked subjects a set of questions to determine their familiarity with basic financial concepts (Section 4.B.15 of the Appendix contains the exact wording). We then used their responses to create binary variables (1 = false answer, 0 = correct answer) and included them in a new specification as additional covariates in both indices. As Section 4.B.15 of the Appendix shows, our results remain robust. In addition and in confirmation of our results, most of the average partial effects of the precision proxies are of similar magnitude as in our main specification, suggesting that the precision proxies we employ do not merely pick up a lack of financial literacy.

Alternative ways of calculating the moments of belief distributions. We arrived at our individual-level measures of μ_{t+1}^{AEX} and $\sigma_{t+1}^{\text{AEX}}$ by fitting log-normal distributions to respondents' stated cumulative distribution functions. We obtain very similar results when we estimate the moments assuming uniformly distributed expectations within bins (Section 4.B.11 of the Internet Appendix) or when we follow Bellemare et al. (2012) in approximating each respondent's distribution using a spline interpolation method (Section 4.B.12).

Alternative ways of calculating the absolute difference between belief measures. We constructed a quantitative proxy for imprecise measures as the absolute difference between the point estimate and the mean of the subjective belief distribution. Some subjects, however, may have had the mode or median in mind when providing a point estimate (Delavande and Rohwedder, 2011). Sections 4.B.3 and 4.B.4 of the Internet Appendix show that we obtain quantitatively and qualitatively very similar results when we define the absolute difference based on the median or mode of the belief distribution. To give respondents the benefit of the doubt, Section 4.B.5 estimates one specification where we pick the moment (mean, median, mode) of the belief distribution that minimizes the absolute difference to the point estimate. Again, our findings are not affected.

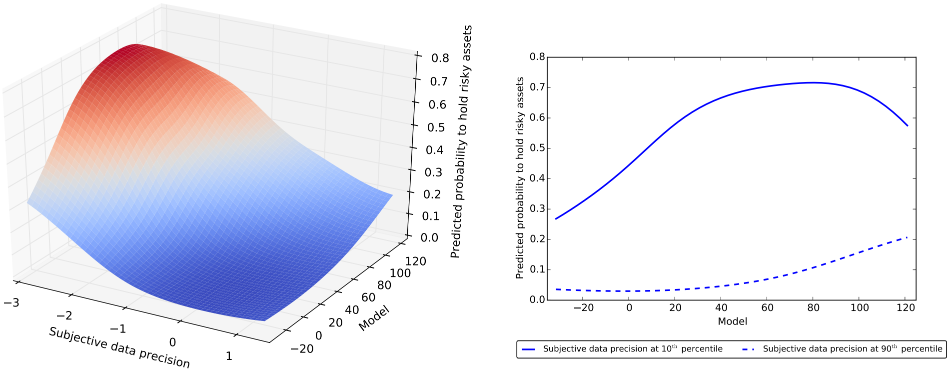
4.4.3 Specification with Less Customized Data

Our analyses employ very detailed data on respondents' stock market expectations based on an incentivized Online Experiment. Our proxies for the precision of expectations include a quantitative variable derived from repeated belief measurements and several qualitative indicators. In many surveys, asking for information this detailed is either impossible or impractical. We now evaluate the applicability of our empirical approach to situations with less customized data.

In the model index, we replace the mean of the log-normal belief distribution derived from the ball allocation task by individuals' point estimates. We drop the standard deviation of beliefs and use aversion towards risks in general instead of our composite variable (see Section 4.A.1 of the Internet Appendix for a detailed description of all measures). In the subjective data precision index, we only keep the answers to the qualitative questions which asked respondents about the difficulty and clarity of our survey. We retain all sociodemographic covariates. We then re-run our main analyses using this limited set of variables.

Figure 4.3 illustrates that the main results for this model are broadly similar to those of our main specification.[5] As the left panel indicates, the predicted probability of holding risky assets strongly varies with both model indices. Importantly, we find strong variation

[5] Section 4.C of the Internet Appendix provides the full set of figures and tables for this model with reduced data requirements.



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. These estimations are based on a limited set of variables. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.3. Predicted probability to hold risky assets, specification with less customized data

in the gradient of the economic model even with these much coarser data: While the probability of investing in the stock market is sensitive to changes in the economic model index at high values of the data precision index, the relationship is essentially flat for low levels. The average partial effects in Table 4.4 again suggest that beliefs and willingness to take risks positively affect stock market participation. The same holds for the precision proxies. All magnitudes are roughly similar to our main specification.

Table 4.4. Average partial effects, specification with less customized data

	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs (direct question): Log expected excess return	0.033	·	0.033
Aversion to risks in general	-0.029	·	-0.029
Experimental tasks difficult	·	-0.034	-0.034
Experimental tasks obscure	·	-0.009	-0.009
Financial wealth ∈ (10000 €, 30000 €]	0.086	0.029	0.102
Financial wealth ∈ (30000 €, ∞)	0.063	0.338	0.396
Financial wealth missing	0.105	0.100	0.204
Net income > 2500 €	0.026	-0.009	0.017
Net income missing	-0.112	0.067	-0.057
High education	-0.004	0.119	0.115
30 < Age ≤ 50	0.102	-0.091	0.014
50 < Age ≤ 65	0.054	0.013	0.071
Age > 65	-0.039	0.068	0.025

Sources: LISS panel and own calculations. The table presents average partial effects of the Klein and Vella (2009) model with a limited number of variables. The effects are calculated for a change of 1 standard deviation in continuous variables. For binary variables, we calculate the effect of assigning individuals in the left-out category a value of 1.

These results entail two consequences: On the one hand, they suggest that imprecise measures will also interfere with our understanding of stock market participation decisions when working with simple measures of beliefs and risk preferences. On the other hand, they suggest that our empirical approach to making productive use of imprecise measures of this kind does not seem to rely on very detailed data to work.

4.5 Discussion and Conclusions

Attempts to measure subjective stock market expectations have dramatically increased over the last two decades. By and large, the results have been encouraging, but obvious signs of poor data quality remain for large fractions of the population regardless of particular survey devices (Manski, 2004; Hurd, 2009; Kleinjans and van Soest, 2014). When these measures have been employed to predict portfolio choice behavior (e.g., Hurd and Rohwedder, 2011; Hurd et al., 2011; Kézdi and Willis, 2011; Hudomiet et al., 2011; Huck et al., 2014), significant correlations in the expected direction have emerged. Nevertheless, it seems fair to say that these are not of the magnitude economists might have hoped for. For example, the abstract of **AmeriksEtAl2016** notes that "estimated risk tolerance, expected return, and perceived risk have economically and statistically significant explanatory power for the distribution of stock shares. Relative to each other, the magnitudes are in proportion with the predictions of benchmark theories, but they are all substantially attenuated." In this paper, we have explored a mechanism that can explain both facts. We have argued that differences in the "propensity to use economic reasoning" may drive heterogeneity in the precision of subjective expectations data and explain why the explanatory power of portfolio choice models has been moderate on average.

While the idea of heterogeneous decision rules is certainly not new (**BinswangerSalm2014**; e.g., Ameriks and Zeldes, 2004; Kahneman, 2011, among many others), we are the first to suggest that the degree of precision in subjective expectations data can be used to uncover such heterogeneity. To explore this link empirically, we have used a semiparametric double index model due to Klein and Vella (2009) on a dataset specifically collected for this purpose. Our results show that stock market participation reacts strongly to the primitives of an economic model (preferences, beliefs, and transaction costs) when subjective data are measured with high precision. When measurement precision is low, there is hardly any reaction at all. This pattern obtains in a wide variety of specification choices, including a setting where we restrict ourselves to variables that are available in many datasets.

A key implication of our findings is that "low quality" of subjective beliefs data should not be treated as a standard measurement error problem, because the strong variation in the precision or meaningfulness of expectations measures actually reflects behaviorally relevant heterogeneity in choice behavior, rather than erroneous reporting. Three pieces of evidence lend further support to our interpretation of the results as reflecting heterogeneous decision modes rather than attenuation bias resulting from standard measurement error. First, if we were dealing with standard versions of measurement error in beliefs (e.g., due to carelessness of some respondents in filling out the survey), taking averages of multiple measurements with uncorrelated idiosyncratic variation should increase the predictive power of expectations. A simple exercise shows that such a pattern does not obtain in our data. We run OLS regressions of stock market participation on convex combinations of our two belief measures (the results are unchanged if we add controls). In Section 4.D of the Internet Appendix, we show that the maximum R^2 is reached close to the point where all the weight is on the mean from the ball allocation task. Hence, adding the second measure hardly helps at all. Second, in all our specifications the likelihood to participate in the stock market was lower for households entertaining imprecise expectations. This suggests that the patterns we found do not merely reflect attenuation bias due to respondents' carelessness or differential effort when responding to the belief questions. If some subjects gave random answers which were uncorrelated with portfolio allocations, participation rates should be the same on average. Third, **ArmantierEtAl2015** show related patterns for subjective inflation expectations in an experimental setting – financially literate individuals react much more strongly to their expectations than others. Similarly, in an experimental portfolio choice problem, Huck et al. (2014) show that the investment behavior of less sophisticated households is less responsive to exogenous changes in incentives.

Our method is applicable to a wide range of settings where subjective data are used, as long as the dataset contains some individual-level information on the precision or meaningfulness of the respective variables. For example, we noted above that the precision of individual-level risk preference parameters obtained from experiments via revealed-preference paradigms varies tremendously in heterogeneous populations (von Gaudecker et al., 2011; Choi et al., 2014). We have shown how the individual-level precision in data on structural parameters can be used when these parameters are employed to explain economically interesting outcomes. Doing so should help dampen the hostility of economists to subjective data (Manski, 2004) that has arisen largely because of perceived data quality. We have turned this argument around and shown that once there is direct information on data precision at the individual level, it can be used to learn about the economic mechanism of interest.

References

Alessie, Rob, Stefan Hochgürtel, and Arthur van Soest (2004): "Ownership of Stocks and Mutual Funds: A Panel Data Analysis." *Review of Economics and Statistics*, 86 (3), 783–796. [131]

Ameriks, John and Stephen P. Zeldes (2004): "How Do Household Portfolio Shares Vary with Age?" Mimeo, Columbia Graduate School of Business. [130, 137, 147]

Bellemare, Charles, Luc Bissonnette, and Sabine Kröger (2012): "Flexible Approximation of Subjective Expectations Using Probability Questions." *Journal of Business & Economic Statistics*, 30 (1), 125–131. [145, 188]

Bruine de Bruin, Wändi and Katherine G. Carman (2012): "Measuring Risk Perceptions: What Does the Excessive Use of 50% Mean?" *Medical Decision Making*, 32 (2), 232–236. [137]

Bruine de Bruin, Wändi, Baruch Fischhoff, Susan G. Millstein, and Bonnie L. Halpern-Felsher (2000): "Verbal and Numerical Expressions of Probability: ıit's a Fifty–fifty Chanceȷ." *Organizational Behavior and Human Decision Processes*, 81 (1), 115–131. [137]

Campbell, John Y. and Luis M. Viceira (2002): *Strategic Asset Allocation*. Clarendon Lectures in Economics. Oxford University Press. [135]

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman (2014): "Who is (More) Rational?" *American Economic Review*, 104 (6), 1518–1550. [137, 148]

Delavande, Adeline and Susann Rohwedder (2008): "Eliciting Subjective Probabilities in Internet Surveys." *Public Opinion Quarterly*, 72 (5), 866–891. [132, 151]

Delavande, Adeline and Susann Rohwedder (2011): "Individuals' Uncertainty about Future Social Security Benefits and Portfolio Choice." *Journal of Applied Econometrics*, 26 (3), 498–519. [145]

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011): "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences." *Journal of the European Economic Association*, 9 (3), 522–550. [161]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde (2014): "An Experimentally Validated Preference Survey Module." Mimeo, Universität Bonn. [133, 161]

Gächter, Simon and Elke Renner (2010): "The Effects of (Incentivized) Belief Elicitation in Public Goods Experiments." *Experimental Economics*, 13, 364–377. [132]

Haliassos, Michael and Carol C. Bertaut (1995): "Why Do so Few Hold Stocks?" *Economic Journal*, 105 (432), 1110–1129. [136]

Hoderlein, Stefan and Joachim Winter (2010): "Structural Measurement Errors in Nonseparable Models." *Journal of Econometrics*, 157 (2), 432–440. [137]

Hossain, Tanjim and Ryo Okui (2013): "The Binarized Scoring Rule." *Review of Economic Studies*, 80 (3), 984–1001. [133, 153]

Huck, Steffen, Tobias Schmidt, and Georg Weizsäcker (2014): "The Standard Portfolio Choice Problem in Germany." Mimeo, Humboldt Universität Berlin. Available at http://dx.doi.org/10.2139/ssrn.2432045. [147, 148]

Hudomiet, Péter, Gábor Kézdi, and Robert J. Willis (2011): "Stock Market Crash and Expectations of American Households." *Journal of Applied Econometrics*, 26 (3), 393–415. [147]

Hurd, Michael D. (2009): "Subjective Probabilities in Household Surveys." *Annual Review of Economics*, 1 (1), 543–562. [129, 133, 134, 137, 147]

Hurd, Michael D. and Susann Rohwedder (2011): "Stock Price Expectations and Stock Trading." RAND Labor and Population Working Paper 938. [147]

Hurd, Michael D., Maarten C. J. van Rooij, and Joachim Winter (2011): "Stock Market Expectations of Dutch Households." *Journal of Applied Econometrics*, 26 (3), 416–436. [129, 132, 133, 147, 152]

Kahneman, Daniel (2011): *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux. [130, 137, 147]

Kézdi, Gábor and Robert J. Willis (2011): "Household Stock Market Beliefs and Learning." NBER Working Paper 17614. [147]

Klein, Roger W. and Richard H. Spady (1993): "An Efficient Semiparametric Estimator for Binary Response Models." *Econometrica*, 61 (2), 387–421. [139]

Klein, Roger W. and Francis Vella (2009): "A Semiparametric Model for Binary Response and Continuous Outcomes under Index Heteroscedasticity." *Journal of Applied Econometrics*, 24 (5), 735–762. [130, 136, 138–141, 143, 146, 147, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 194, 197, 199]

Kleinjans, Kristin J. and Arthur van Soest (2014): "Rounding, Focal Point Answers and Nonresponse to Subjective Probability Questions." *Journal of Applied Econometrics*, 29 (4), 567–585. [129, 137, 147]

Manski, Charles F. (2004): "Measuring Expectations." *Econometrica*, 72 (5), 1329–1376. [129, 133, 134, 137, 147, 148, 152]

Maurer, Jürgen (2009): "Who Has a Clue to Preventing the Flu? Unravelling Supply and Demand Effects on the Take-up of Influenza Vaccinations." *Journal of Health Economics*, 28 (3), 704–717. [138]

Maurer, Jürgen, Roger W. Klein, and Francis Vella (2010): "Subjective Health Assessments and Active Labor Market Participation of Older Men: Evidence from a Semiparametric Binary Choice Model with Nonadditive Correlated Individual-specific Effects." *Review of Economics and Statistics*, 93 (3), 764–774. [129]

McArdle, John J., James P. Smith, and Robert J. Willis (2011): "Cognition and Economic Outcomes in the Health and Retirement Survey." In *Explorations in the Economics of Aging*. Ed. by David A. Wise. University of Chicago Press. Chap. 7, 209–233. [133]

Palfrey, Thomas R. and Stephanie W. Wang (2009): "On Eliciting Beliefs in Strategic Games." *Journal of Economic Behavior & Organization*, 71 (2), 98–109. [132]

Rabin, Matthew (2000): "Risk Aversion and Expected Utility Theory: A Calibration Theorem." *Econometrica*, 68 (5), 1281–1292. [138]

Van Rooij, Maarten C. J., Annamaria Lusardi, and Rob Alessie (2011): "Financial Literacy and Stock Market Participation." *Journal of Financial Economics*, 101 (2), 449–472. [131]

Vissing-Jørgensen, Annette (2002): "Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures." NBER Working Paper 8884. [136]

Von Gaudecker, Hans-Martin (2015): "How Does Household Portfolio Diversification Vary with Financial Sophistication and Financial Advice?" *Journal of Finance*, 70 (2). [129, 136, 138]

Von Gaudecker, Hans-Martin, Arthur van Soest, and Erik Wengström (2011): "Heterogeneity in Risky Choice Behaviour in a Broad Population." *American Economic Review*, 101 (2), 664–694. [137, 148]

Wang, Stephanie W. (2011): "Incentive Effects: The Case of Belief Elicitation from Individuals in Groups." *Economics Letters*, 111 (1), 30–33. [132]

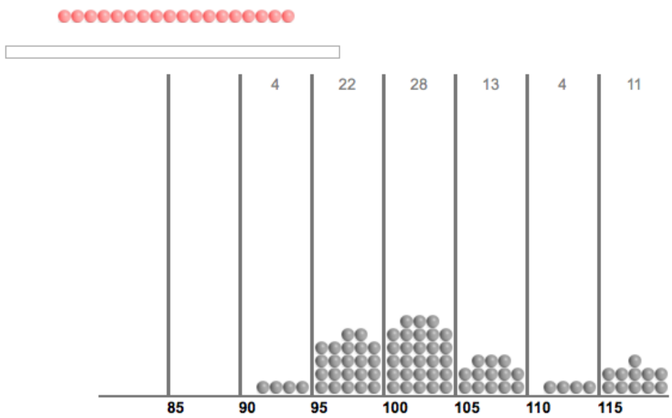
Wansbeek, Tom J. and Erik Meijer (2000): *Measurement Error and Latent Variables in Econometrics*. Amsterdam: Elsevier North Holland. [131]

Appendix 4.A Extended data description

4.A.1 Variable definitions and descriptives

4.A.1.1 Subjective expectations of stock market returns

AEX return - Ball allocation task. In August 2013, we asked respondents to describe their expectations for the one-year return of the Amsterdam Exchange Index (AEX). To elicit the full distribution of individual expectations, we employed a variation of the procedure presented in Delavande and Rohwedder (2008), which was explicitly developed for usage in Internet experiments and pays particular attention to the cognitive burden placed on heterogeneous subject pools. We asked respondents to imagine that they invested 100 € into an exchange traded AEX index fund today and to think about the likely value of this investment in one year. To aid respondents' thinking process and ensure comprehension of the task, the instructions clarified what an index fund is and provided an explicit formula for the value of the investment in one year (*value in a year = 100 €- 0.30 € (fees) + change in the AEX index*).

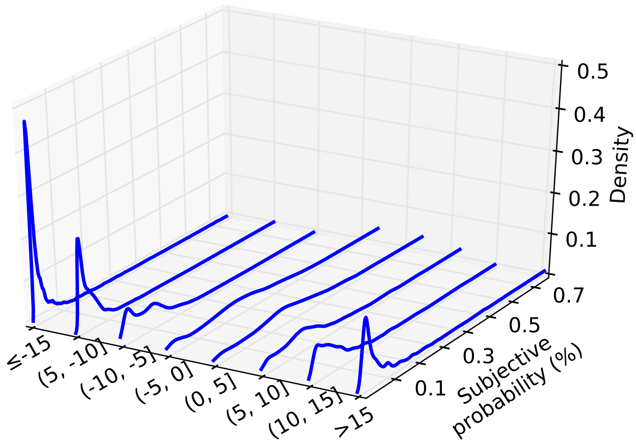


The figure shows the final step of the belief eliciation procedure. Respondents used the slider above to allocate 100 balls to the 8 bins below. The figure shows both the remaining balls and the number of balls assigned to each return interval in the previous steps.

Figure 4.4. Visual interface to elicit belief distribution (final step)

We then provided respondents with a visual interface that employed an iterative procedure to allow them to state their beliefs as accurately as possible (see Figure 4.4). To familiarize subjects with the visual interface, we showed them an introductory video before asking them for their beliefs about the stock market. The video used the example of expected annual rainy days in London to describe the intuition behind the ball allocation procedure and guided subjects through the controls of the interface.

In the first step of the iterative procedure, the interface presented all possible values of the investment as two intervals, $[0, 100]$ and $(100, \infty)$. We asked participants to use a slider to allocate 100 balls to indicate their relative confidence that the final value of the investment would fall into either of these intervals. We then split up the interval $(100, \infty)$ into $(100, 105]$ and $(105, \infty)$, and we asked subjects to re-allocate the balls from the previous interval to this finer grid. This procedure continued successively until subjects had distributed all balls into 6 interior bins covering intervals of 5 € each and two exterior bins covering the intervals $[0, 85]$ and $(115, \infty)$. Figure 4.5 shows the resulting distribution of balls for each interval expressed in terms of expected returns. While the exterior bins contained only a small number of balls for the large majority of respondents, the distribution of balls in the interior bins was substantially more dispersed.



Sources: LISS panel and own calculations. The picture shows Kernel density estimates of the distribution of probabilites for each of the 8 return intervals.

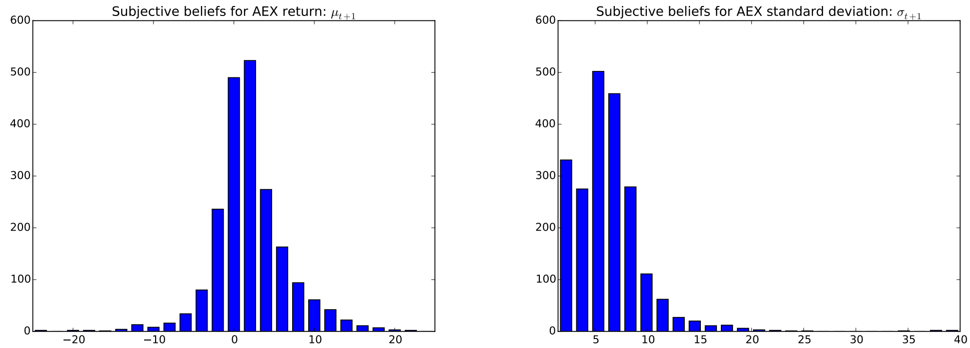
Figure 4.5. Distribution of probabilities within bins

The iterative procedure provides an intuitively simple way of eliciting beliefs and the resulting distribution of balls lends itself to a straightforward interpretation as a histogram. One of its desirable properties is that it does not ask respondents for cumulative probabilities. In contrast, standard survey questions based on the elicitation of points on a cumulative probability distribution often yield logically inconsistent responses due to frequent monotonicity violations. This regularly forces researchers to discard large amounts of data, thereby potentially introducing severe selection effects into the empirical analyses (see, e.g., Manski, 2004; Hurd et al., 2011).

To obtain estimates of the mean and variance of individual belief distributions, we employ a procedure similar to Hurd et al. (2011). We first cumulated the number of balls each respondent assigned to the bins to arrive at a discrete cumulative distribution function. We then used the 7 interior boundary points (b) and the associated values of the CDF (p) to minimize

$$\sum_{i=1}^{7}\left(p_i - \Phi\left(\frac{log(b_i/100) - \mu}{\sigma}\right)\right)^2$$

over μ and σ, our estimates of the mean and standard deviation of the repondent's belief distribution. On average, respondents expect a mean return of 2.01% and a standard deviation of 6.25%. Figure 4.6 shows the distribution of estimated mean returns and the distribution of estimated standard deviations. As is evident from the two distributions, subjects have very heterogenuous expectations regarding both the expected return of the AEX as well as its expected standard deviation.



Sources: LISS panel and own calculations.

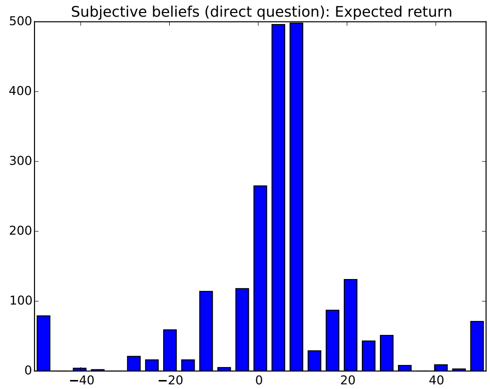
Figure 4.6. Distribution of expected mean and standard deviation of returns

To financially incentive the task, we used the binarized scoring rule of Hossain and Okui (2013). Subjects could either earn 100 € or 0 €, depending on their stated beliefs, the actually realized value of a 100 € investment into the AEX after one year, and the outcome of a random draw. For each subject, we computed the sum of the squared deviations of the belief distribution from the actual value of a 100 € investment after 12 months, $\sum_{i=1}^{8}(b_i - 100 \times 1_i)^2$, where 1_i equalled 1 if the realized value of the investment fell into bin i and 0 otherwise. We then drew a random number from $U[1, 20.000]$. If that random number turned out to be larger (smaller) than the sum of squared deviations, the participant received 100 (0) €.

AEX return - One-shot estimate. In September 2013, we asked our full set of respondents for a second, this time non-incentivized, estimate of the one-year return of the AEX using a one-shot question similar to those commonly employed in large-scale surveys:

> *Please consider the Dutch stock market. The AEX index aggregates the stock prices of many of the largest Dutch companies. Now consider an investment fund tracking the AEX index, i.e. this investment exactly moves up and down with the AEX after subtracting rather small fees. If you invested 100 € in such a fund today, the amount of money you would have in a year from now will be:*
>
> $$\text{value in a year} = 100 \text{ €} - 0.30 \text{ € (fees)} + \text{change in the AEX index}$$
>
> *What do you think will be this value in a year from now? Please type in your estimate (in Euros).*

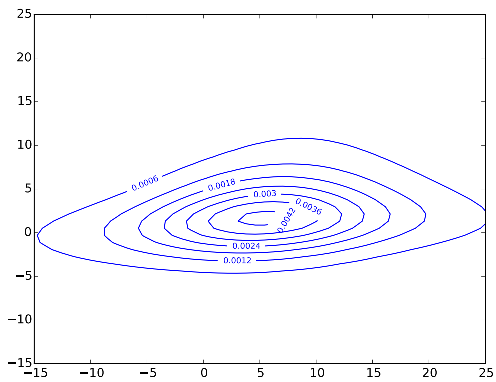
Figure 4.7 shows the distribution of expected returns implied by subjects' responses to this question. With an average expected return of 4.76%, subjects' point estimates are more optimistic than the mean estimates from the visual task. As is often the case in large-scale representative surveys, we observe a number of outliers in the unrestricted point estimates. Many of these are likely due to typing mistakes or lack of comprehension. Thus, before calculating returns, we winsorize the point estimates at the values of a 100 € investment into the AEX at the 2.5% and 97.5% percentiles of its historical return distribution (49.6 € and 151.3 €). This affected 99 responses.



Sources: LISS panel and own calculations.

Figure 4.7. Distribution of one-shot estimates for return of AEX

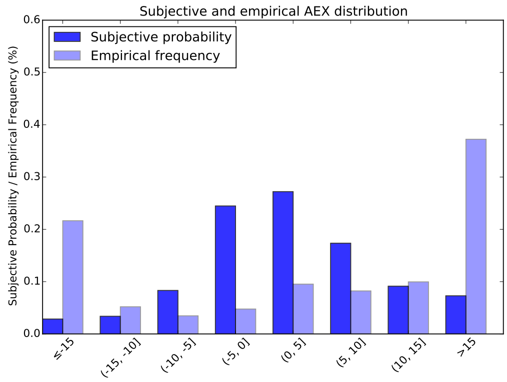
Joint distribution. Figure 4.8 shows the joint distribution of the mean estimate from the visual task and the direct estimate from the one shot question. With standard deviations of 6.19% and 17.47%, respectively, the distribution of mean estimates from the visual task is substantially less dispersed than the distribution of direct estimates.



Sources: LISS panel and own calculations.

Figure 4.8. Joint distribution of both average belief measures

Comparison to historical distribution of AEX returns. Figure 4.9 plots the historical distribution of (inflation-adjusted) AEX returns alongside the average probabilities expected by our sample respondents. Respondents considered returns at both ends of the spectrum of the intervals we provided, i.e., in excess of +15% as well as below −15%, far less likely than what has historically been observed. For example, while our average repondent expects less than a 1 in 20 chance of observing returns below −15%, the historical probability of this happening exceeded 20%.



Sources: LISS panel and own calculations.

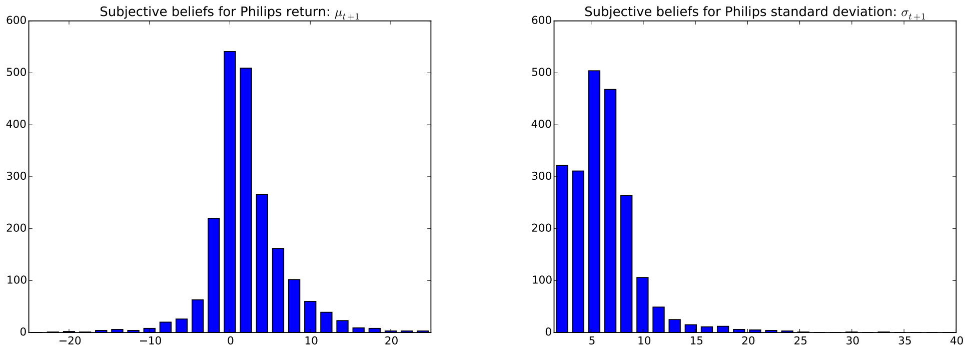
Figure 4.9. Expected and historical distribution of AEX

Alternative belief measure - Philips N.V. As part of the survey in August 2013, we also asked our respondents to use the visual interface to express their beliefs for the future development of Philips N.V., one of the largest publicly traded companies of the Netherlands. Figure 4.10 shows the distributions of the mean and standard deviation our respondents expect, calculated in the same manner as the moments of the belief distribution for the AEX. The median respondent expects a mean return of 1.534% for Philips, only minimally different from the median expectation of 1.562% for the AEX. The joint density in Figure 4.12 shows that the correlation between the mean beliefs for both assets is fairly high ($\rho = 0.36$). The correlation between the expected standard deviations is of similar magnitude ($\rho = 0.35$).

Figure 4.11 compares the average probabilities expected by our sample respondents to the historical distribution of (inflation-adjusted) Philips returns. Similar to the results presented in Figure 4.9 for the expected returns of the AEX, we see that respondents consider extreme returns for Philips much less likely than what has historically been observed.

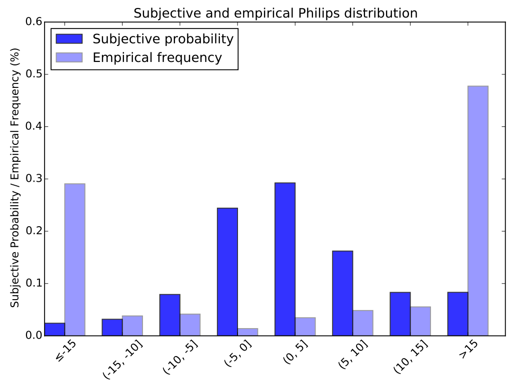
In September, we also asked respondents for a one-shot estimate for the return of Philips alongside their one-shot estimate for the return of the AEX. Figure 4.13 shows the distribution of their answers.

Return to savings account - One-shot estimate. In August 2013, we asked respondents for an estimate of the return of a one-year investment into a standard savings account:



Sources: LISS panel and own calculations.

Figure 4.10. Distribution of expected mean and standard deviation of returns - Philips



Sources: LISS panel and own calculations.

Figure 4.11. Expected and historical distribution of philips

Suppose you invested 100 € into a standard savings account with a large Dutch bank. Then, in a year from now, the total amount of money you would have will be:

$$\textit{value in a year} \;=\; 100\ \text{€} + \textit{interest payments}$$

What do you think will be this value in a year from now? Please type in your estimate (in Euros).

To ensure comprehension of the question, the computer screen also contained a link with more detailed information and the example of a savings account with Rabobank (Rabo SpaarRekening). Figure 4.14 shows the distribution of savings estimates. Somewhat surprisingly, subjects' average return estimate for the savings account is 3.35% and thus larger than their average estimate for the AEX in the visual task, though it is smaller than the average point estimate for the AEX. Similar to the one-shot AEX estimates, we

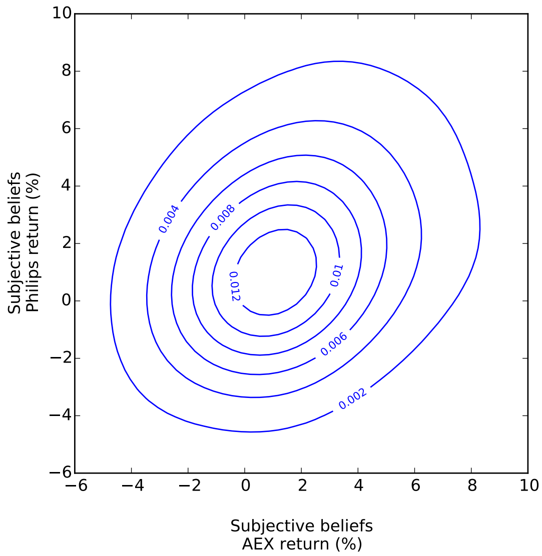
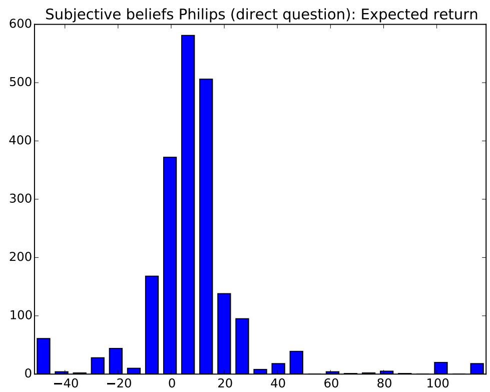


Figure 4.12. Joint density of mean beliefs for AEX and Philips



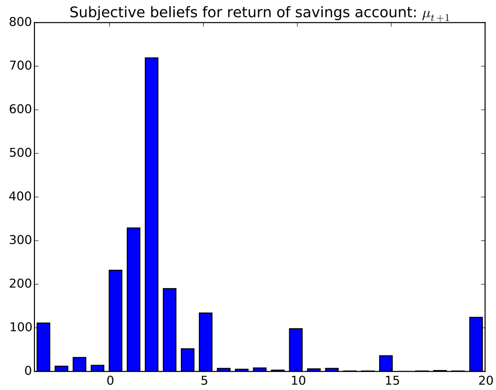
Sources: LISS panel and own calculations.

Figure 4.13. Distribution of one-shot estimates for return of philips

winsorize point estimates for the savings account at the 5 and 95% percentiles of the sample distribution before calculating returns.

4.A.1.2 Proxies for the precision of subjective data

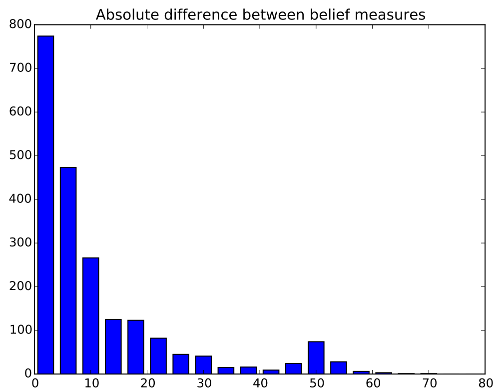
Our rich data allow us to employ a number of different variables to proxy for the precision of subjective data. We use 5 proxies in total, 1 based on the consistency in stated beliefs, 2 based on subjects' confidence in their estimates, and 2 based on the subjects' perception of our survey.



Sources: LISS panel and own calculations.

Figure 4.14. Distribution of one-shot estimates for savings account

Consistency in beliefs. As discussed in Section 4.A.1.1, we used the survey in September 2014 to ask our full set of respondents for a second estimate of the one-year return of the AEX. We use the absolute difference between the response to this question and the mean belief from the visual task as a quantitative proxy for the precision of subjective data. Figure 4.15 shows a histogram of the absolute differences. On average, subjects' second estimate deviates from the mean estimate from the visual task by a considerable margin, 11.20 percentage points. This seems particularly large when compared to the average expected standard deviation of returns from the ball allocation task (6.25%). Note that these differences are not artifacts of the method we employ to estimate mean beliefs. Other methods, which we describe in Sections 4.B.11 and 4.B.12 of this appendix, yield very similar results.



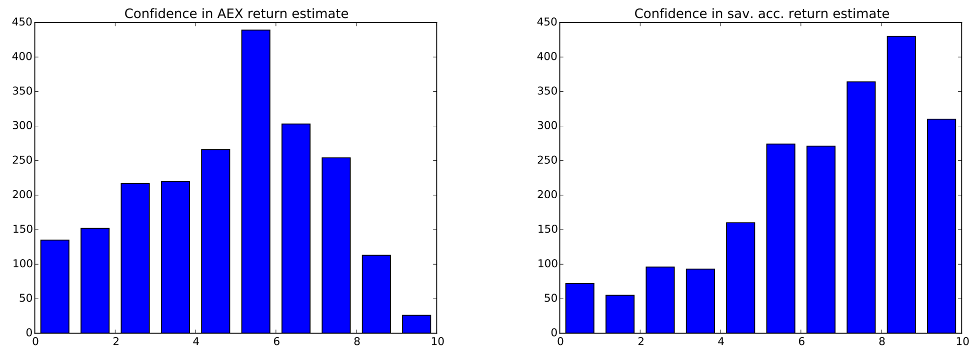
Sources: LISS panel and own calculations.

Figure 4.15. Distribution of absolute differences between mean belief in visual task and point estimate

Confidence in estimates. Following the elicitation of the point estimates for the expected returns of the AEX and the savings account, we asked respondents how certain they felt about their responses:

> *Please use the slider to indicate how certain you are that the value in a year will equal your estimate. 0 indicates "not certain at all" and 10 means "absolutely certain".*

We conjecture that respondents with little confidence in their own estimates (e.g., because they know that they did not expend much cognitive effort into developing their prediction) provide estimates that are noisy and hence not very predictive of actual choices. Figure 4.16 shows histograms for the answers to both questions. Respondents seem to be on average less confident in their estimates for the return of the AEX as compared to their estimates for the saving account. For the empirical analyses, we invert the responses so that larger values correspond to a lack of confidence and scale the resulting variables to range between 0 and 1.

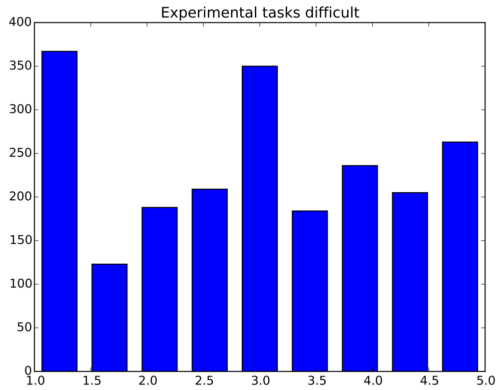


Sources: LISS panel and own calculations.

Figure 4.16. Distribution of slider values for confidence in estimates

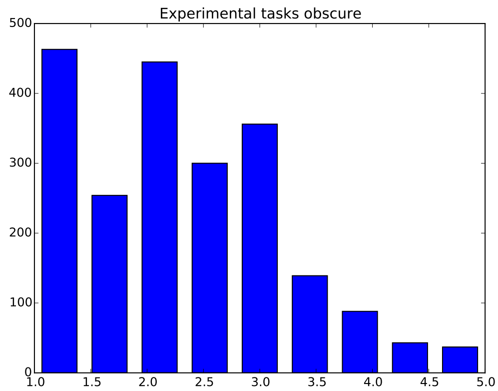
Difficulty. Following the survey in August 2013 and September 2013, we asked subjects to use five-point scales to indicate how difficult they considered the preceding belief elicitation task. We conjecture that answers by respondents who found it very hard to detail their stock market expectations are likely to exhibit a high variability. Figure 4.17 shows the distribution of the average of the responses in both surveys. Respondents vary greatly in their assessment of the tasks' difficulties. While some considered it simple, others seemed to find the task very demanding. We scale the average to range between 0 and 1 for our empirical analysis.

Clarity. In August 2013 and September 2013, we also asked subjects to use five-point scales to indicate how vague/obscure they found our questions. We expect that limited comprehension of the task on the side of respondents will lead to noisier measures of expectations. Figure 4.18 shows a histogram of the average response to this question in both surveys. For the empirical analysis, we also scale the average to range between 0 and 1.



Sources: LISS panel and own calculations.

Figure 4.17. Distribution of assessments of difficulty



Sources: LISS panel and own calculations.

Figure 4.18. Distribution of assessments of obscurity

4.A.1.3 Risk preferences

We use a composite variable to measure risk aversion. To construct this variable, we ask respondents two questions on their self-assessed willingness to take risks and we elicit one quantitative measure based on hypothetical lottery choices. In our empirical analyses, we use the average of the standardized values of all three measures to proxy for risk aversion, suitably coded so that larger values of individual variables as well as well as the composite variable correspond to larger values of risk aversion.

Risk questions. The subjective self-assessments directly ask for an individual's willingness to take risks, both in general terms and in financial matters:

> *"Different people have different opinions and characteristics. We are interested in how you describe yourself. In general, to what extent are you willing to take risks? You can answer this question by clicking somewhere on the slider (0-10)."*

"And, in general, to what extent are you willing to take risks in financial matters? You can answer this question by clicking somewhere on the slider (0-10)."

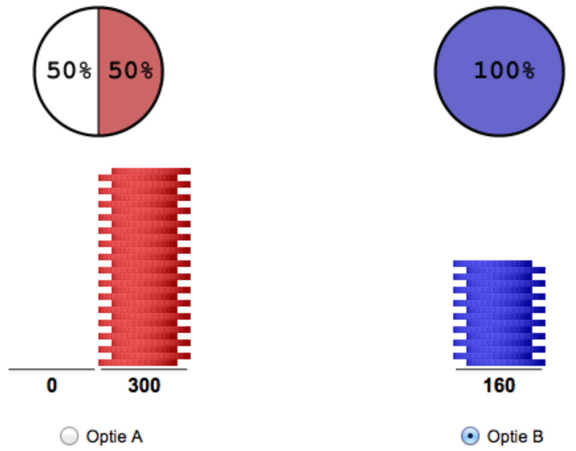
Risk lottery. We derive a quantitative measure of risk aversion from a series of five interdependent hypothetical binary lottery choices, a format commonly referred to as the "staircase procedure". In each of the questions, participants had to decide between a 50/50 lottery to win 300 €or nothing and a varying safe payment. The questions were interdependent in the sense that the choice of a lottery resulted in an increase of the safe amount being offered in the next question, while the choice of the safe payment resulted in a decrease of the safe amount in the next question. For instance, the fixed payment in the first question was 160 €. In case the respondent chose the lottery, the safe payment increased to 240 €in the second question. In case the respondent chose the safe payment, the next question's fixed payment was reduced to 80 €. By adjusting the fixed payment according to previous choices, the questions allow for a relatively fine quantitative assessment of an individual's attitudes towards risk. With 32 possible outcomes evenly spaced between 0 and 320 €, the procedure can in principle pin down a respondent's certainty equivalent to a range of 10 euros. Because of the task's abstract nature and our heterogeneous subject pool, we accompanied each lottery decision with a visual representation of the current lottery to ensure comprehension, see Figure 4.19.

The above variables resemble the variables developed for the "Preference Survey Module" in Falk et al. (2014) to measure economic preference parameters in large-scale surveys. Falk et al. (2014) use an experimental validation procedure to select behaviorally valid survey items to measure economic preferences. Dohmen et al. (2011) show that responses to our qualitative survey items correlate with many risky field choices, including stockholdings. Thus, even though the questions we asked were not financially incentivized, they are known to be behaviorally valid and were explicitly developed for the purpose of large-scale studies like ours.

In Figure 4.20, we show histograms of the indiviual components as well the composite variable. There is substantial variation in the answers to all three questions. In the lottery task, most of our subjects end up with estimated certainty equivalents below 160 €, suggesting that the majority of our subjects is risk averse.

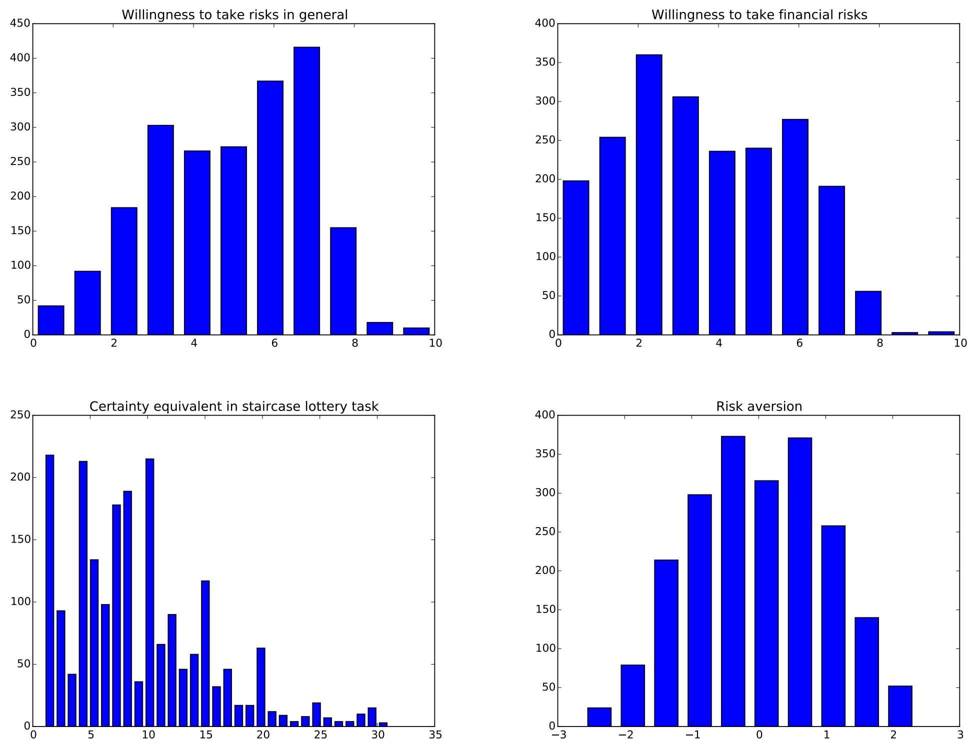
4.A.1.4 Transaction cost proxies / sociodemographics

Portfolio value. LISS collects detailed information on the value of a respondent's financial assets. To calculate an estimate of the total value of a respondent's portfolio, we sum the amounts held as investments and those in the bank, which we set to 0 in case the household reported negative values. LISS allows respondents to provide either continuous or interval statements for each category of assets. To calculate the overall portfolio value, we replace categorical answers by the midpoint of the respective interval. For example, we set an answer like "7.500 to 10.000 €" to 8.750 €. For all respondents, we use the most detailed level of information available. For investments, LISS asks both for the aggregate value of investments as well as for the value of the subcategories (stocks, funds, and other investments). We use the more detailed data if available, and we use the answer to the aggregate question otherwise.



The figure shows the visual interface accompanying one of the lottery decisions.

Figure 4.19. Graphical illustration of hypothetical lottery choice



Sources: LISS panel and own calculations.

Figure 4.20. Distribution of risk aversion components and aggregate variable

Employing the resulting estimate of a respondent's portfolio value, we create categorical variables for each of the sample's portfolio value terciles. Some respondents prefer not to answer the questions concerning their financial situation, so we create one more binary variable for missing portfolio values.

Net household income. Using LISS's information, we create a binary variable for net household income in excess of 2.500 €, the median income of households providing an answer to the income question. We create a further dummy for households with missing values for income ($\approx$ 7% of the sample).

Education. LISS asks respondents for the highest educational degree. In our main estimation, we include a dummy variable for respondents who either report having a university degree or higher vocational education.

Age. Using LISS's data on birthyears, we create binary variables for several different age groups (31 to 50, 51 to 65, and for respondents older than 65).

4.A.2 Correlations

Table 4.5 shows the correlation matrix for all main variables.

Table 4.5. Correlation matrix

	Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	Risk aversion	Abs. diff. between belief measures	Lack of confidence in AEX return estimate	Lack of confidence in sav. acc. return estimate	Experimental tasks difficult	Experimental tasks obscure	Financial wealth ∈ (10000 €, 30000 €]	Financial wealth ∈ (30000 €, ∞)	Net income > 2500 €	High education	30 < Age ≤ 50	50 < Age ≤ 65	Age > 65
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1	**-0.19**	**-0.11**	**-0.21**	**-0.16**	**-0.24**	**-0.17**	**-0.18**	0.00	**0.18**	**0.13**	**0.15**	-0.02	**0.09**	-0.03
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	·	1	0.01	**0.25**	**0.09**	**0.09**	0.04	**0.08**	-0.03	**-0.08**	-0.04	**-0.06**	0.01	-0.04	0.00
Risk aversion	·	·	1	**0.06**	**0.33**	**0.22**	**0.20**	**0.13**	-0.00	**-0.07**	**-0.12**	**-0.16**	**-0.09**	0.01	**0.14**
Abs. diff. between belief measures	·	·	·	1	**0.13**	**0.23**	**0.08**	**0.11**	-0.05	**-0.20**	**-0.14**	**-0.19**	-0.01	**-0.06**	0.05
Lack of confidence in AEX return estimate	·	·	·	·	1	**0.52**	**0.23**	**0.20**	-0.04	**-0.08**	**-0.13**	**-0.07**	0.00	-0.04	0.04
Lack of confidence in sav. acc. return estimate	·	·	·	·	·	1	**0.24**	**0.28**	**-0.07**	**-0.20**	**-0.20**	**-0.20**	**-0.07**	-0.04	**0.12**
Experimental tasks difficult	·	·	·	·	·	·	1	**0.48**	0.01	-0.04	**-0.13**	**-0.07**	**-0.07**	**-0.08**	**0.18**
Experimental tasks obscure	·	·	·	·	·	·	·	1	-0.01	**-0.10**	**-0.09**	**-0.10**	-0.05	**-0.07**	**0.13**
Financial wealth ∈ (10000 €, 30000 €]	·	·	·	·	·	·	·	·	1	**-0.37**	0.02	0.03	0.01	**-0.07**	0.05
Financial wealth ∈ (30000 €, ∞)	·	·	·	·	·	·	·	·	·	1	**0.20**	**0.17**	**-0.13**	**0.14**	0.05
Net income > 2500 €	·	·	·	·	·	·	·	·	·	·	1	**0.23**	**0.10**	0.03	**-0.10**
High education	·	·	·	·	·	·	·	·	·	·	·	1	0.05	-0.01	**-0.10**
30 < Age ≤ 50	·	·	·	·	·	·	·	·	·	·	·	·	1	**-0.47**	**-0.42**
50 < Age ≤ 65	·	·	·	·	·	·	·	·	·	·	·	·	·	1	**-0.46**
Age > 65	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1

Significant correlations ($p < 0.01$) printed in bold.

4.A.3 Correlates of beliefs

Table 4.6 presents regressions of various measures of expectations on sociodemographic covariates. In column (1), the dependent variable is the mean belief from the ball allocation task, in column (2) it is the corresponding standard deviation, and column (3) employs the point estimate of the return of a savings account.

Table 4.6. Beliefs and sociodemographics

	(1)	(2)	(3)
Constant	2.066***	6.811***	5.739***
	(0.504)	(0.350)	(0.610)
Financial wealth ∈ (10000 €, 30000 €]	-0.018	-0.536***	-0.476
	(0.313)	(0.199)	(0.319)
Financial wealth ∈ (30000 €, ∞)	1.035***	-0.739***	-0.927***
	(0.330)	(0.212)	(0.290)
Financial wealth missing	-1.058***	-0.122	0.099
	(0.379)	(0.256)	(0.410)
Net income > 2500 €	0.476*	0.040	-0.357
	(0.254)	(0.161)	(0.249)
Net income missing	0.284	0.015	-1.281***
	(0.445)	(0.331)	(0.472)
High education	0.695***	-0.314**	-1.131***
	(0.237)	(0.155)	(0.218)
30 < Age ≤ 50	0.357	-0.044	-1.092*
	(0.475)	(0.336)	(0.624)
50 < Age ≤ 65	0.224	-0.363	-2.332***
	(0.485)	(0.342)	(0.596)
Age > 65	-0.618	-0.129	-1.762***
	(0.498)	(0.342)	(0.619)
Female	-1.397***	0.262*	1.251***
	(0.238)	(0.157)	(0.237)
Married	-0.034	-0.041	-0.561**
	(0.253)	(0.165)	(0.249)
Has children	0.230	-0.244	0.078
	(0.272)	(0.185)	(0.280)
Observations	2,108	2,108	2,125
Adj. (pseudo) R^2 (%)	5.6	1.2	6.6

The left-hand variable in column (1) is the mean return from the visual task. In column (2), it is the standard deviation of returns in the visual task. Column (3) includes the estimate for the return of the savings account as the left-hand variable. Robust standard errors in parentheses.

Appendix 4.B Robustness checks

4.B.1 No transaction cost proxies

Table 4.7. Coefficient estimates for the economic model index and the subjective data precision index, model without transaction cost proxies

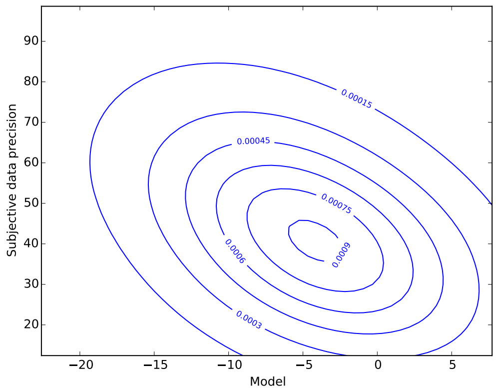
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu^{\text{AEX}}_{t+1} - \mu^{\text{sav. acc.}}_{t+1}$	1.00	·	·	·
Subjective beliefs: $\sigma^{\text{AEX}}_{t+1}$	-0.75	0.24	·	·
Risk aversion	-4.56	1.00	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	-4.19	11.51
Lack of confidence in sav. acc. return estimate	·	·	56.04	20.68
Experimental tasks difficult	·	·	34.09	10.31
Experimental tasks obscure	·	·	10.06	10.68

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text. The model excludes all transaction cost proxies (financial wealth, net income, education, age).

Table 4.8. Average partial effects, model without transaction cost proxies

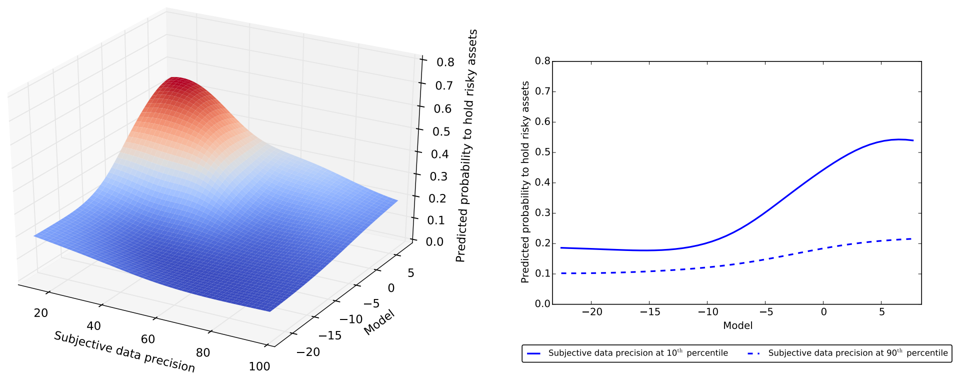
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu^{\text{AEX}}_{t+1} - \mu^{\text{sav. acc.}}_{t+1}$	0.065	·	0.065
Subjective beliefs: $\sigma^{\text{AEX}}_{t+1}$	-0.031	·	-0.031
Risk aversion	-0.046	·	-0.046
Absolute difference between belief measures	·	-0.034	-0.034
Lack of confidence in AEX return estimate	·	0.002	0.002
Lack of confidence in sav. acc. return estimate	·	-0.035	-0.035
Experimental tasks difficult	·	-0.029	-0.029
Experimental tasks obscure	·	-0.006	-0.006

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text. The model excludes all transaction cost proxies (financial wealth, net income, education, age).



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.21. Joint density of the two indices, model without transaction cost proxies



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.22. Predicted probability to hold risky assets, model without transaction cost proxies

4.B.2 Mean beliefs only

Table 4.9. Coefficient estimates for the economic model index and the subjective data precision index, model with mean beliefs and proxies for the subjective data precision only

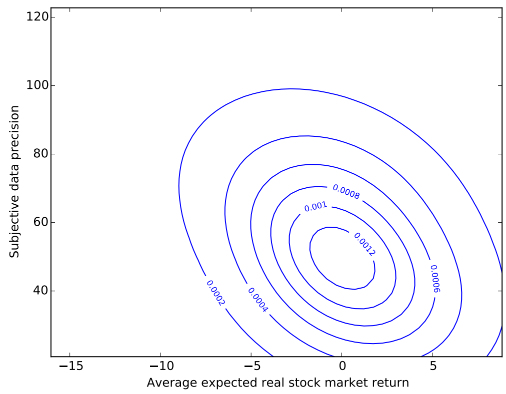
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	11.54	10.07
Lack of confidence in sav. acc. return estimate	·	·	80.10	25.04
Experimental tasks difficult	·	·	26.35	8.74
Experimental tasks obscure	·	·	12.40	10.75

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text. The model excludes the standard deviation in beliefs, risk preferences, and all transaction cost proxies (financial wealth, net income, education, age).

Table 4.10. Average partial effects, model with mean beliefs and proxies for the precision of subjective data only

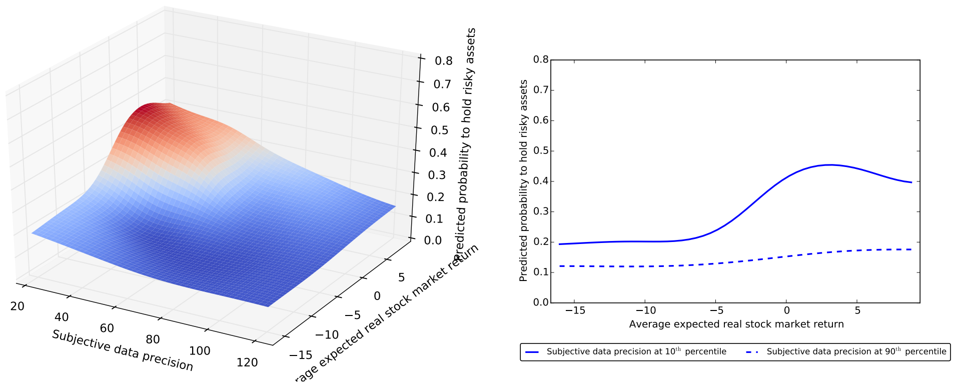
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.036	·	0.036
Absolute difference between belief measures	·	-0.036	-0.036
Lack of confidence in AEX return estimate	·	-0.007	-0.007
Lack of confidence in sav. acc. return estimate	·	-0.051	-0.051
Experimental tasks difficult	·	-0.023	-0.023
Experimental tasks obscure	·	-0.008	-0.008

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text. The model excludes the standard deviation in beliefs, risk preferences, and all transaction cost proxies (financial wealth, net income, education, age).



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.23. Joint density of the two indices, model with mean beliefs and proxies for the precision of subjective data only



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.24. Predicted probability to hold risky assets, model with mean beliefs and proxies for the precision of subjective data only

4.B.3 Redefining errors as absolute difference between modal belief in visual task and point estimate

Table 4.11. Coefficient estimates for the economic model index and the subjective data precision index, errors as absolute difference between modal belief and point estimate

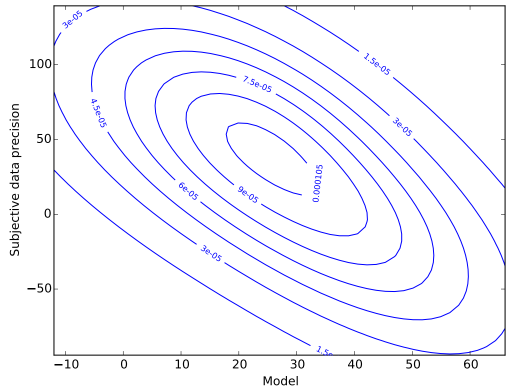
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.77	0.28	·	·
Risk aversion	-7.87	1.78	·	·
Abs. difference between mode and point estimate	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	60.84	28.24
Lack of confidence in sav. acc. return estimate	·	·	30.90	23.03
Experimental tasks difficult	·	·	55.71	20.27
Experimental tasks obscure	·	·	16.45	18.93
Financial wealth ∈ (10000 €, 30000 €]	20.67	5.90	-21.50	22.73
Financial wealth ∈ (30000 €, ∞)	43.61	9.13	-96.15	39.00
Financial wealth missing	30.58	7.22	-62.04	29.49
Net income > 2500 €	7.32	2.64	29.37	11.87
Net income missing	-6.61	4.20	-4.51	13.37
High education	3.82	2.99	-65.17	19.85
30 < Age ≤ 50	11.66	5.62	23.98	17.60
50 < Age ≤ 65	7.47	5.60	-16.23	15.69
Age > 65	-0.46	5.28	-22.68	16.91

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, except that we include the absolute difference between the modal belief in the visual task and the point estimate in the subjective data precision index.

Table 4.12. Average partial effects, errors as absolute difference between modal belief and point estimate

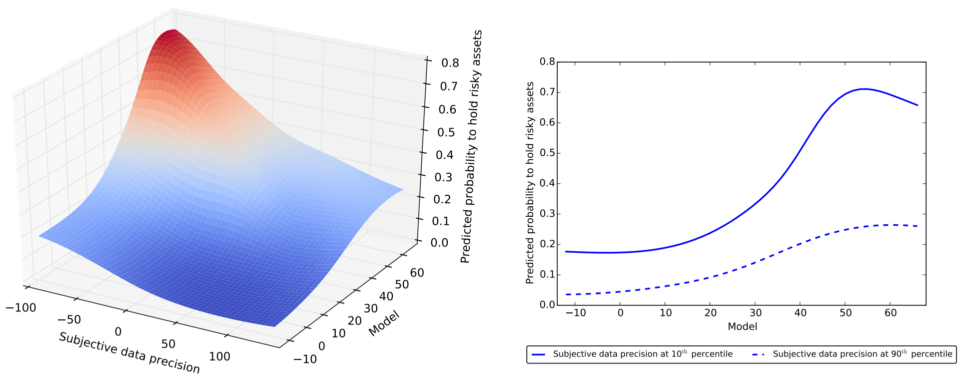
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.034	·	0.034
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.014	·	-0.014
Risk aversion	-0.037	·	-0.037
Abs. difference between mode and point estimate	·	-0.013	-0.013
Lack of confidence in AEX return estimate	·	-0.013	-0.013
Lack of confidence in sav. acc. return estimate	·	-0.007	-0.007
Experimental tasks difficult	·	-0.018	-0.018
Experimental tasks obscure	·	-0.004	-0.004
Financial wealth ∈ (10000 €, 30000 €]	0.100	0.018	0.100
Financial wealth ∈ (30000 €, ∞)	0.248	0.120	0.375
Financial wealth missing	0.171	0.070	0.220
Net income > 2500 €	0.037	-0.028	0.009
Net income missing	-0.032	0.005	-0.028
High education	0.018	0.079	0.098
30 < Age ≤ 50	0.054	-0.025	0.024
50 < Age ≤ 65	0.036	0.018	0.053
Age > 65	-0.002	0.026	0.018

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, except that we include the absolute difference between the modal belief in the visual task and the point estimate in the subjective data precision index.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.25. Joint density of the two indices, errors as absolute difference between modal belief and point estimate



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.26. Predicted probability to hold risky assets, errors as absolute difference between modal belief and point estimate

4.B.4 Redefining errors as absolute difference between median belief in visual task and point estimate

Table 4.13. Coefficient estimates for the economic model index and the subjective data precision index, errors as absolute difference between median belief and point estimate

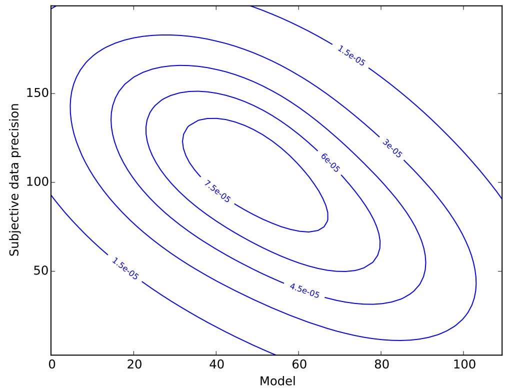
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.32	0.37	·	·
Risk aversion	-9.72	2.28	·	·
Abs. difference between median and point estimate	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	115.67	34.37
Lack of confidence in sav. acc. return estimate	·	·	51.37	25.51
Experimental tasks difficult	·	·	17.71	14.73
Experimental tasks obscure	·	·	29.37	19.82
Financial wealth ∈ (10000 €, 30000 €]	27.28	7.96	-16.97	18.30
Financial wealth ∈ (30000 €, ∞)	56.30	12.23	-83.30	34.01
Financial wealth missing	39.10	9.92	-70.00	29.08
Net income > 2500 €	8.39	3.27	30.60	11.39
Net income missing	-7.66	5.66	-0.88	14.65
High education	21.97	5.42	12.66	11.56
$30 < \text{Age} \leq 50$	21.86	7.76	34.14	18.12
$50 < \text{Age} \leq 65$	18.10	7.02	-2.87	15.61
Age > 65	9.16	6.51	2.34	16.60

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, except that we include the absolute difference between the median belief in the visual task and the point estimate in the subjective data precision index.

Table 4.14. Average partial effects, errors as absolute difference between median belief and point estimate

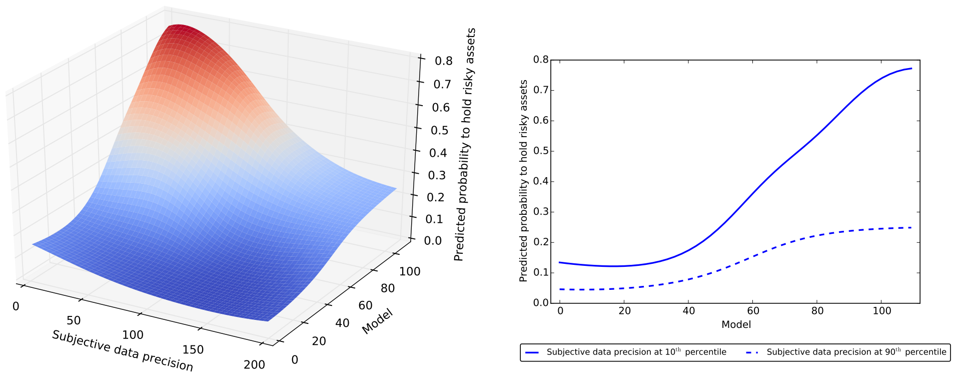
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.031	·	0.031
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.005	·	-0.005
Risk aversion	-0.037	·	-0.037
Abs. difference between median and point estimate	·	-0.013	-0.013
Lack of confidence in AEX return estimate	·	-0.025	-0.025
Lack of confidence in sav. acc. return estimate	·	-0.012	-0.012
Experimental tasks difficult	·	-0.005	-0.005
Experimental tasks obscure	·	-0.006	-0.006
Financial wealth ∈ (10000 €, 30000 €]	0.100	0.020	0.095
Financial wealth ∈ (30000 €, ∞)	0.254	0.116	0.368
Financial wealth missing	0.163	0.100	0.241
Net income > 2500 €	0.035	-0.023	0.012
Net income missing	-0.030	0.001	-0.029
High education	0.101	-0.011	0.090
$30 < \text{Age} \leq 50$	0.086	-0.037	0.049
$50 < \text{Age} \leq 65$	0.071	0.003	0.078
Age > 65	0.035	-0.002	0.036

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, except that we include the absolute difference between the median belief in the visual task and the point estimate in the subjective data precision index.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.27. Joint density of the two indices, errors as absolute difference between median belief and point estimate



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.28. Predicted probability to hold risky assets, errors as absolute difference between median belief and point estimate

4.B.5 Redefining errors as minimum of absolute differences between mean, median, and modal belief in visual task and point estimate

It is possible that respondents differ in their understanding of our question for a point estimate of the AEX. While some may think that this corresponds to a question concerning the expected mean, others may think we are asking for the expected mode or median. To give respondents the benefit of the doubt when calculating the absolute error, we also estimate one specification where we base the latter calculation on the moment that minimizes the absolute difference. That is, for each respondent we select the moment (mean, mode, median) that is absolutely closest to the mean from the ball allocation task. Based on this moment, we then calculate the absolute difference in beliefs that enters the subjective data precision index.

Table 4.15. Coefficient estimates for the economic model index and the subjective data precision index, errors as absolute difference between median belief and point estimate

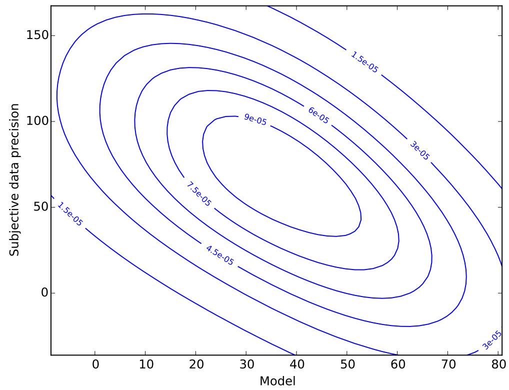
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.70	0.31	·	·
Risk aversion	-8.83	1.89	·	·
Minimal abs. diff. between point estimate and lognormal mean/mode/median	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	73.33	24.74
Lack of confidence in sav. acc. return estimate	·	·	57.76	20.41
Experimental tasks difficult	·	·	34.64	16.66
Experimental tasks obscure	·	·	16.76	16.51
Financial wealth $\in$ (10000 €, 30000 €]	24.60	6.77	-12.17	17.26
Financial wealth $\in$ (30000 €, ∞)	51.21	10.39	-66.89	30.30
Financial wealth missing	35.75	8.09	-50.18	23.42
Net income > 2500 €	7.32	2.80	32.85	10.70
Net income missing	-7.17	4.10	-2.59	12.33
High education	5.80	3.16	-52.62	17.77
$30 < \text{Age} \leq 50$	14.84	6.41	34.28	6.87
$50 < \text{Age} \leq 65$	11.74	5.91	0.02	nan
Age > 65	3.44	5.62	-11.36	nan

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, except that we include the minimum of the absolute differences between the mean, median, or modal belief in the visual task and the point estimate in the subjective data precision index.

Table 4.16. Average partial effects, errors as absolute difference between median belief and point estimate

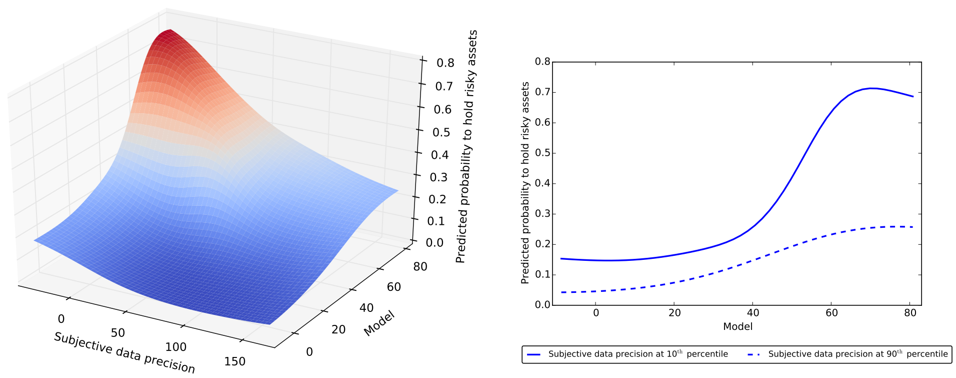
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.035	·	0.035
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.013	·	-0.013
Risk aversion	-0.041	·	-0.041
Minimal abs. diff. between point estimate and lognormal mean/mode/median	·	-0.012	-0.012
Lack of confidence in AEX return estimate	·	-0.016	-0.016
Lack of confidence in sav. acc. return estimate	·	-0.013	-0.013
Experimental tasks difficult	·	-0.011	-0.011
Experimental tasks obscure	·	-0.004	-0.004
Financial wealth $\in$ (10000 €, 30000 €]	0.105	0.012	0.105
Financial wealth $\in$ (30000 €, ∞)	0.278	0.081	0.367
Financial wealth missing	0.184	0.060	0.225
Net income > 2500 €	0.036	-0.027	0.010
Net income missing	-0.034	0.002	-0.032
High education	0.027	0.058	0.086
$30 < \text{Age} \leq 50$	0.071	-0.036	0.033
$50 < \text{Age} \leq 65$	0.057	-0.000	0.060
Age > 65	0.017	0.011	0.026

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, except that we include the minimum of the absolute differences between the mean, median, or modal belief in the visual task and the point estimate in the subjective data precision index.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.29. Joint density of the two indices, errors as absolute difference between median belief and point estimate



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.30. Predicted probability to hold risky assets, errors as absolute difference between median belief and point estimate

4.B.6 Additional covariates

Table 4.17. Coefficient estimates for the economic model index and the subjective data precision index, model with additional covariates

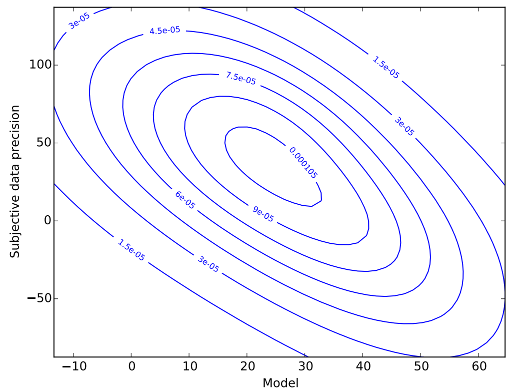
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.80	0.28	·	·
Risk aversion	-7.83	2.07	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	51.54	26.19
Lack of confidence in sav. acc. return estimate	·	·	37.05	24.08
Experimental tasks difficult	·	·	52.54	18.38
Experimental tasks obscure	·	·	16.92	17.82
Financial wealth ∈ (10000 €, 30000 €]	19.96	6.24	-18.44	18.62
Financial wealth ∈ (30000 €, ∞)	42.37	9.73	-87.41	31.09
Financial wealth missing	30.13	7.82	-57.80	24.63
Net income > 2500 €	8.70	2.81	23.25	10.90
Net income missing	-6.65	4.03	-6.53	12.74
High education	2.31	3.16	-62.53	17.41
30 < Age ≤ 50	12.06	6.05	18.98	18.35
50 < Age ≤ 65	7.78	6.21	-21.00	16.74
Age > 65	1.20	6.25	-27.94	18.28
Female	-0.57	2.69	0.12	8.34
Married	-4.21	2.53	11.37	8.83
Has children	3.65	3.24	5.12	9.45

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, except for the female, marriage, and having children dummies.

Table 4.18. Average partial effects, model with additional covariates

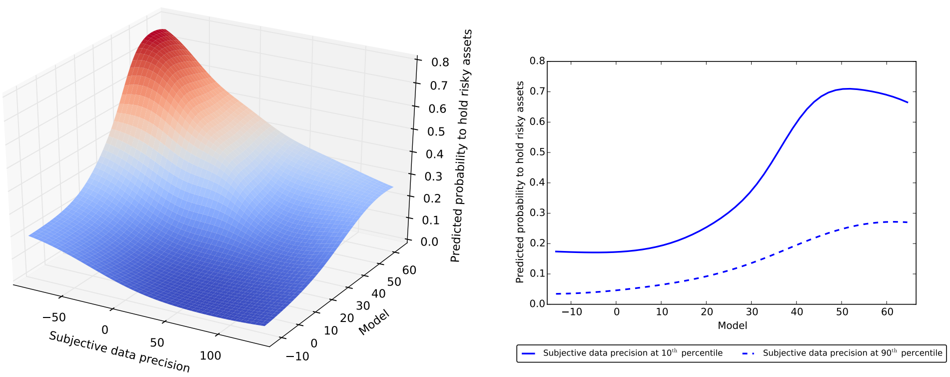
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.034	·	0.034
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.015	·	-0.015
Risk aversion	-0.037	·	-0.037
Absolute difference between belief measures	·	-0.014	-0.014
Lack of confidence in AEX return estimate	·	-0.012	-0.012
Lack of confidence in sav. acc. return estimate	·	-0.010	-0.010
Experimental tasks difficult	·	-0.019	-0.019
Experimental tasks obscure	·	-0.004	-0.004
Financial wealth ∈ (10000 €, 30000 €]	0.099	0.017	0.097
Financial wealth ∈ (30000 €, ∞)	0.245	0.119	0.369
Financial wealth missing	0.172	0.073	0.224
Net income > 2500 €	0.044	-0.025	0.019
Net income missing	-0.032	0.007	-0.026
High education	0.011	0.083	0.095
30 < Age ≤ 50	0.057	-0.022	0.029
50 < Age ≤ 65	0.038	0.026	0.062
Age > 65	0.006	0.035	0.035
Female	-0.003	-0.000	-0.003
Married	-0.019	-0.013	-0.032
Has children	0.017	-0.006	0.011

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, except for the female, marriage, and having children dummies.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.31. Joint density of the two indices, model with additional covariates



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.32. Predicted probability to hold risky assets, model with additional covariates

4.B.7 Expected return instead of expected excess return

Table 4.19. Coefficient estimates for the economic model index and the subjective data precision index, expected returns instead of expected excess returns

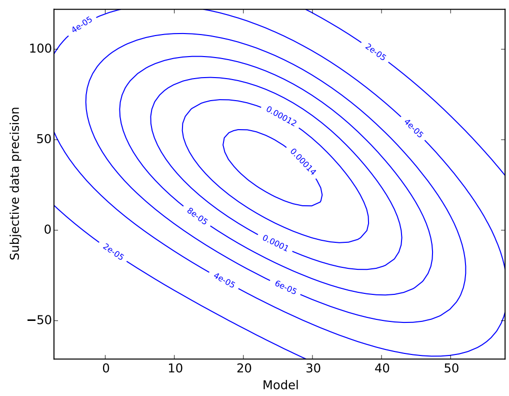
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: μ_{t+1}^{AEX}	1.00	·	·	·
Subjective beliefs: σ_{t+1}^{AEX}	-0.78	0.31	·	·
Risk aversion	-6.75	1.90	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	45.85	23.05
Lack of confidence in sav. acc. return estimate	·	·	31.25	22.13
Experimental tasks difficult	·	·	47.26	15.93
Experimental tasks obscure	·	·	12.89	15.44
Financial wealth ∈ (10000 €, 30000 €]	17.94	5.66	-6.43	18.80
Financial wealth ∈ (30000 €, ∞)	37.84	9.73	-66.29	28.44
Financial wealth missing	26.49	7.67	-39.04	22.69
Net income > 2500 €	6.59	2.64	21.50	9.34
Net income missing	-5.26	4.08	-5.65	12.90
High education	2.36	3.34	-55.43	15.69
30 < Age ≤ 50	10.69	4.97	19.31	13.88
50 < Age ≤ 65	5.36	4.67	-16.10	13.15
Age > 65	-1.69	4.77	-24.45	15.26

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, except that we replace the expected excess return with the expected return.

Table 4.20. Average partial effects, expected returns instead of expected excess returns

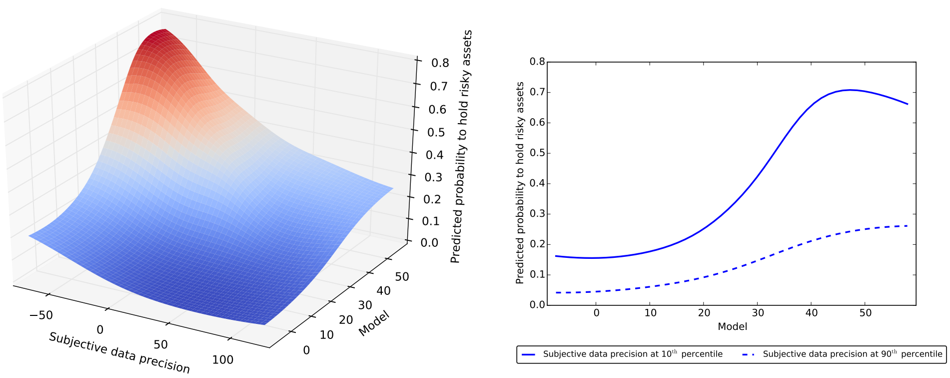
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: μ_{t+1}^{AEX}	0.029	·	0.029
Subjective beliefs: σ_{t+1}^{AEX}	-0.016	·	-0.016
Risk aversion	-0.036	·	-0.036
Absolute difference between belief measures	·	-0.017	-0.017
Lack of confidence in AEX return estimate	·	-0.013	-0.013
Lack of confidence in sav. acc. return estimate	·	-0.010	-0.010
Experimental tasks difficult	·	-0.020	-0.020
Experimental tasks obscure	·	-0.004	-0.004
Financial wealth ∈ (10000 €, 30000 €]	0.105	0.008	0.096
Financial wealth ∈ (30000 €, ∞)	0.251	0.109	0.370
Financial wealth missing	0.176	0.059	0.217
Net income > 2500 €	0.036	-0.028	0.009
Net income missing	-0.028	0.008	-0.022
High education	0.012	0.089	0.103
30 < Age ≤ 50	0.054	-0.026	0.020
50 < Age ≤ 65	0.029	0.024	0.052
Age > 65	-0.010	0.037	0.019

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, except that we replace the expected excess return with the expected return.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.33. Joint density of the two indices, expected returns instead of expected excess returns



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.34. Predicted probability to hold risky assets, expected returns instead of expected excess returns

4.B.8 Discarding individuals with missing data on financial wealth

Table 4.21. Coefficient estimates for the economic model index and the subjective data precision index, sample restricted to individuals with available information on financial wealth

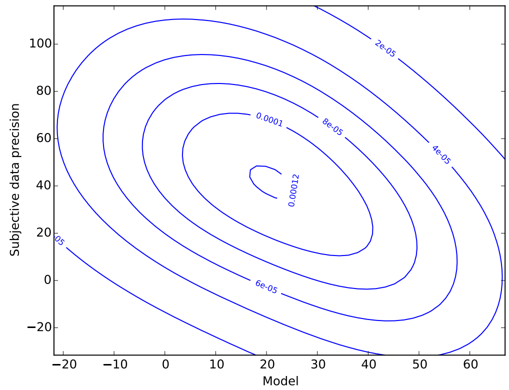
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.88	0.42	·	·
Risk aversion	-10.58	2.81	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	46.82	30.49
Lack of confidence in sav. acc. return estimate	·	·	14.32	24.68
Experimental tasks difficult	·	·	33.68	18.29
Experimental tasks obscure	·	·	17.07	19.59
Financial wealth ∈ (10000 €, 30000 €]	24.74	8.33	-5.17	19.45
Financial wealth ∈ (30000 €, ∞)	49.20	13.27	-40.53	28.97
Net income > 2500 €	6.61	3.98	24.29	13.32
Net income missing	-10.30	9.50	-19.90	14.46
High education	-2.11	4.85	-41.00	16.02
30 < Age ≤ 50	11.71	11.40	29.88	25.23
50 < Age ≤ 65	1.32	9.20	-7.56	18.05
Age > 65	-6.13	8.52	-19.20	18.63

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text. The model excludes respondents with missing information on financial wealth.

Table 4.22. Average partial effects, sample restricted to individuals with available information on financial wealth

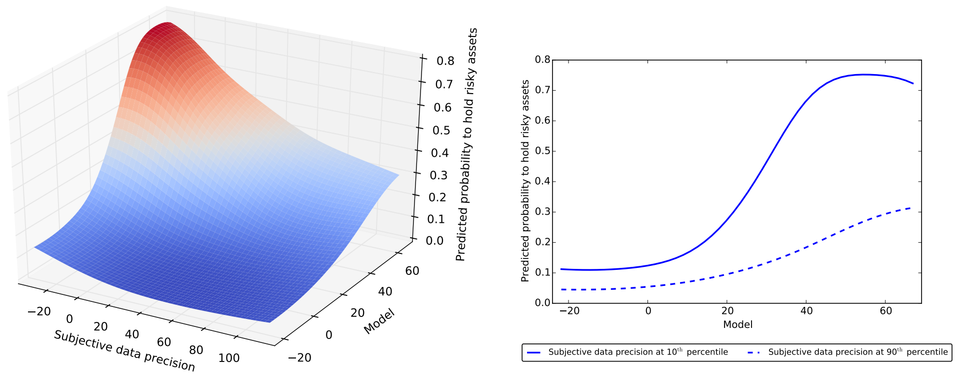
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.032	·	0.032
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.015	·	-0.015
Risk aversion	-0.046	·	-0.046
Absolute difference between belief measures	·	-0.021	-0.021
Lack of confidence in AEX return estimate	·	-0.017	-0.017
Lack of confidence in sav. acc. return estimate	·	-0.005	-0.005
Experimental tasks difficult	·	-0.018	-0.018
Experimental tasks obscure	·	-0.006	-0.006
Financial wealth ∈ (10000 €, 30000 €]	0.062	0.008	0.069
Financial wealth ∈ (30000 €, ∞)	0.281	0.088	0.376
Net income > 2500 €	0.029	-0.039	-0.010
Net income missing	-0.045	0.033	-0.018
High education	-0.009	0.081	0.071
30 < Age ≤ 50	0.045	-0.051	-0.011
50 < Age ≤ 65	0.006	0.015	0.020
Age > 65	-0.027	0.038	0.005

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text. The model excludes respondents with missing information on financial wealth.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.35. Joint density of the two indices, sample restricted to individuals with available information on financial wealth



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.36. Predicted probability to hold risky assets, sample restricted to individuals with available information on financial wealth

4.B.9 Alternative belief measure

Table 4.23. Coefficient estimates for the economic model index and the subjective data precision index, Philips instead of AEX

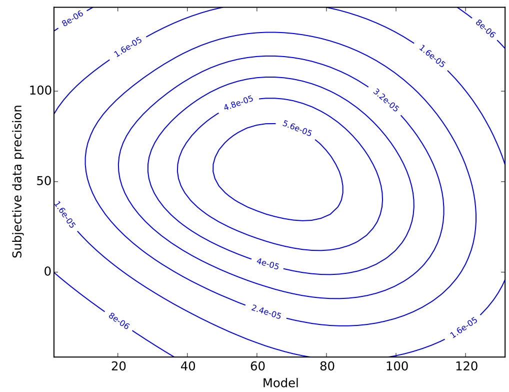
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{Philips}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{Philips}}$	-0.46	0.66	·	·
Risk aversion	-13.34	4.28	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in Philips return estimate	·	·	-4.23	19.85
Lack of confidence in sav. acc. return estimate	·	·	30.53	21.08
Experimental tasks difficult	·	·	53.43	22.43
Experimental tasks obscure	·	·	23.76	18.50
Financial wealth ∈ (10000 €, 30000 €]	35.09	15.64	-15.97	28.14
Financial wealth ∈ (30000 €, ∞)	47.02	18.59	-93.15	40.58
Financial wealth missing	47.19	17.60	-42.95	30.99
Net income > 2500 €	29.20	10.29	54.77	19.75
Net income missing	-6.90	10.85	14.75	21.40
High education	19.24	8.62	-12.21	14.50
30 < Age ≤ 50	36.39	16.06	47.14	24.06
50 < Age ≤ 65	26.42	12.10	3.42	19.88
Age > 65	7.96	9.47	-5.80	19.26

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, except for the belief measures pertaining to Philips N.V..

Table 4.24. Average partial effects, Philips instead of AEX

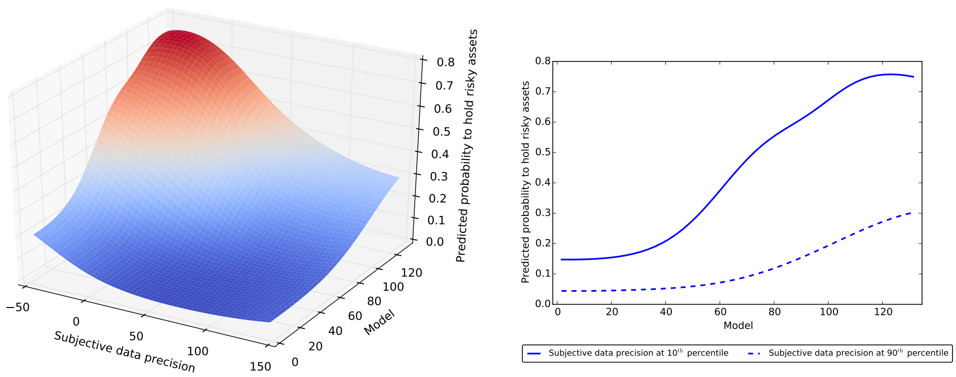
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{Philips}} - \mu_{t+1}^{\text{sav. acc.}}$	0.025	·	0.025
Subjective beliefs: $\sigma_{t+1}^{\text{Philips}}$	-0.005	·	-0.005
Risk aversion	-0.042	·	-0.042
Absolute difference between belief measures	·	-0.026	-0.026
Lack of confidence in Philips return estimate	·	0.002	0.002
Lack of confidence in sav. acc. return estimate	·	-0.011	-0.011
Experimental tasks difficult	·	-0.027	-0.027
Experimental tasks obscure	·	-0.009	-0.009
Financial wealth ∈ (10000 €, 30000 €]	0.119	0.020	0.106
Financial wealth ∈ (30000 €, ∞)	0.168	0.214	0.377
Financial wealth missing	0.168	0.074	0.221
Net income > 2500 €	0.108	-0.076	0.021
Net income missing	-0.022	-0.023	-0.045
High education	0.068	0.021	0.090
30 < Age ≤ 50	0.121	-0.077	0.043
50 < Age ≤ 65	0.089	-0.006	0.088
Age > 65	0.026	0.011	0.037

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, except for the belief measures pertaining to Philips N.V..



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.37. Joint density of the two indices, Philips instead of AEX



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.38. Predicted probability to hold risky assets, Philips instead of AEX

4.B.10 Disaggregated risk aversion measures

Table 4.25. Coefficient estimates for the economic model index and the subjective data precision index, separate risk measures

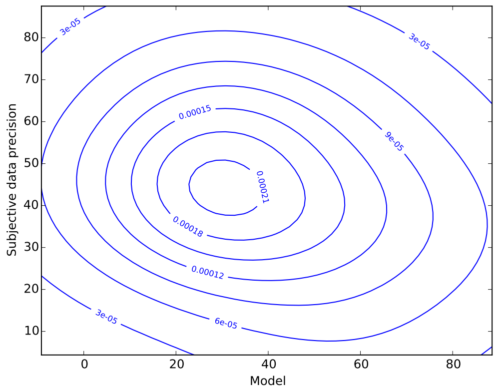
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.82	0.43	·	·
Aversion to risks in general	4.69	2.12	·	·
Aversion to financial risks	-15.09	3.50	·	·
Risk aversion index based on staircase lottery task	-0.34	1.38	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	9.41	13.00
Lack of confidence in sav. acc. return estimate	·	·	20.12	12.18
Experimental tasks difficult	·	·	0.68	7.37
Experimental tasks obscure	·	·	21.56	9.24
Financial wealth $\in$ (10000 €, 30000 €]	23.88	6.74	6.29	9.62
Financial wealth $\in$ (30000 €, ∞)	44.25	11.30	-28.19	13.78
Financial wealth missing	35.36	8.47	-10.65	10.83
Net income > 2500 €	7.08	3.23	7.34	3.92
Net income missing	-6.27	5.27	5.52	5.75
High education	17.67	5.67	26.77	5.55
$30 < \text{Age} \leq 50$	15.37	7.04	12.49	13.51
$50 < \text{Age} \leq 65$	15.08	6.87	13.12	13.84
$\text{Age} > 65$	4.92	6.08	1.92	13.09

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, except for the disaggregated risk aversion measure.

Table 4.26. Average partial effects, separate risk measures

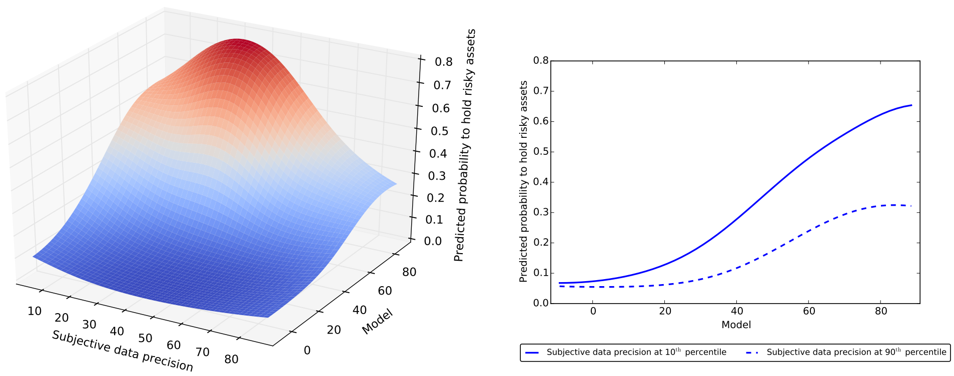
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.041	·	0.041
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.016	·	-0.016
Aversion to risks in general	0.024	·	0.024
Aversion to financial risks	-0.070	·	-0.070
Risk aversion index based on staircase lottery task	-0.001	·	-0.001
Absolute difference between belief measures	·	-0.024	-0.024
Lack of confidence in AEX return estimate	·	-0.003	-0.003
Lack of confidence in sav. acc. return estimate	·	-0.008	-0.008
Experimental tasks difficult	·	-0.000	-0.000
Experimental tasks obscure	·	-0.009	-0.009
Financial wealth $\in$ (10000 €, 30000 €]	0.121	-0.018	0.091
Financial wealth $\in$ (30000 €, ∞)	0.273	0.091	0.367
Financial wealth missing	0.205	0.035	0.234
Net income > 2500 €	0.037	-0.010	0.027
Net income missing	-0.031	-0.009	-0.039
High education	0.104	-0.016	0.079
$30 < \text{Age} \leq 50$	0.076	-0.017	0.061
$50 < \text{Age} \leq 65$	0.074	-0.018	0.058
$\text{Age} > 65$	0.023	-0.002	0.022

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, except for the disaggregated risk aversion measure.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.39. Joint density of the two indices, separate risk measures



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.40. Predicted probability to hold risky assets, separate risk measures

4.B.11 Moments of the belief distribution calculated using uniformly distributed expectations within bins

The simplest way to approximate the individual-specific distribution of beliefs is to assume that respondents' expectations are uniformly distributed within bins. To calculate moments under this assumption, we need to assign values to the outer bounds of the exterior bins. We fix these bounds at the value a 100 € investment would have had at the 2.5% and 97.5% percentile of the AEX's historical return distribution, 49.6 € and 151.3 €. We then compute the moments of the distribution assuming that the balls are uniformly distributed within each of the resulting 8 intervals.

Table 4.27. Coefficient estimates for the economic model index and the subjective data precision index, moments of beliefs calculated assuming uniform distributions within bins

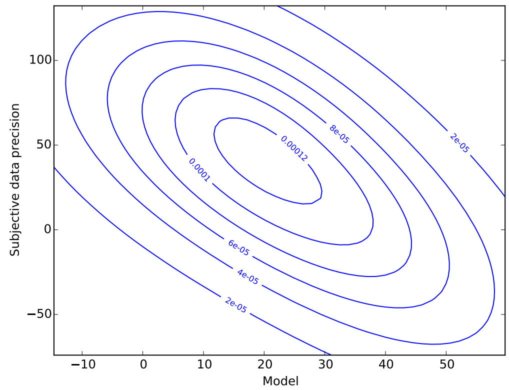
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs (uniform): Expected excess return	1.00	·	·	·
Subjective beliefs (uniform): Expected standard deviation	-0.74	0.23	·	·
Risk aversion	-7.05	1.51	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	58.17	25.87
Lack of confidence in sav. acc. return estimate	·	·	23.64	20.32
Experimental tasks difficult	·	·	53.21	18.67
Experimental tasks obscure	·	·	13.07	16.20
Financial wealth $\in$ (10000 €, 30000 €]	17.85	5.29	-13.48	19.58
Financial wealth $\in$ (30000 €, ∞)	39.32	8.07	-77.99	31.32
Financial wealth missing	26.61	6.31	-49.33	24.11
Net income > 2500 €	6.81	2.40	27.39	10.55
Net income missing	-5.45	3.98	-6.40	13.14
High education	3.76	2.81	-57.88	17.96
$30 < \text{Age} \leq 50$	11.07	5.21	22.30	16.27
$50 < \text{Age} \leq 65$	7.16	5.19	-15.00	14.77
Age > 65	-0.28	4.91	-22.52	15.93

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, except for the way of calculating moments of beliefs.

Table 4.28. Average partial effects, moments of beliefs calculated assuming uniform distributions within bins

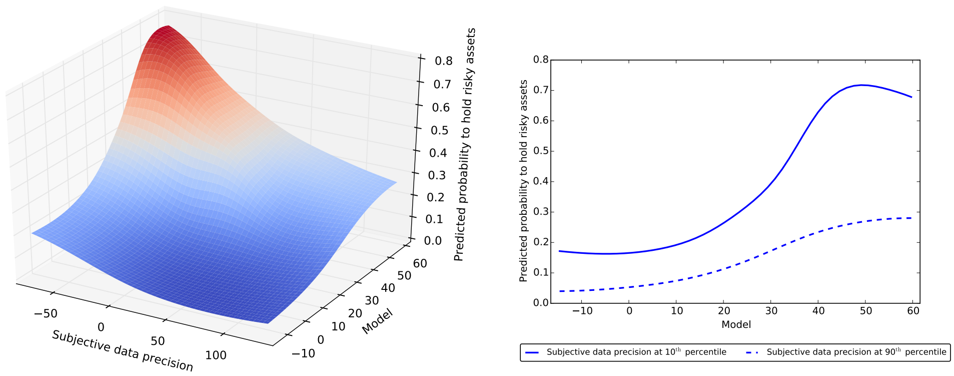
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs (uniform): Expected excess return	0.040	·	0.040
Subjective beliefs (uniform): Expected standard deviation	-0.016	·	-0.016
Risk aversion	-0.037	·	-0.037
Absolute difference between belief measures	·	-0.014	-0.014
Lack of confidence in AEX return estimate	·	-0.014	-0.014
Lack of confidence in sav. acc. return estimate	·	-0.006	-0.006
Experimental tasks difficult	·	-0.018	-0.018
Experimental tasks obscure	·	-0.003	-0.003
Financial wealth $\in$ (10000 €, 30000 €]	0.096	0.013	0.095
Financial wealth $\in$ (30000 €, ∞)	0.252	0.105	0.367
Financial wealth missing	0.166	0.061	0.209
High education	0.020	0.074	0.094
Net income > 2500 €	0.038	-0.027	0.012
Net income missing	-0.029	0.007	-0.023
$30 < \text{Age} \leq 50$	0.059	-0.024	0.028
$50 < \text{Age} \leq 65$	0.039	0.018	0.056
Age > 65	-0.002	0.026	0.019

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, except for the way of calculating moments of beliefs.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.41. Joint density of the two indices, moments of beliefs calculated assuming uniform distributions within bins



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.42. Predicted probability to hold risky assets, moments of beliefs calculated assuming uniform distributions within bins

4.B.12 Moments of the belief distribution calculated using piecewise cubic Hermite interpolating splines

We also approximate individual belief distributions using piecewise cubic Hermite interpolating splines, very similar to the method proposed in Bellemare et al. (2012). For each respondent, we first calculate a discrete cumulative distribution function by successively summing the probabilities assigned to each of the 8 bins. The method is less sensitive to the assumptions concerning the support of the exterior bins, so we fix these at more conservative values (the minimum and maximum of the AEX's historical return distribution over a calendar year, i.e., 47.0 € and 176.9 €). We then use a Hermite spline to connect the 9 points on the resulting CDF. The spline interpolates the CDF between each pair of neighboring points by a monotonically increasing cubic polynomial, whose first derivative at each of the 7 interior points coincides with the respective first derivative of the polynomial in the next-higher interval. We employ the resulting estimate of an indivual's belief distribution to calculate the mean and standard deviation of the individual's return estimate.[6]

Table 4.29. Coefficient estimates for the economic model index and the subjective data precision index, moments of beliefs calculated by approximating the distribution using splines

	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs (Splines): Expected excess return	1.00	·	·	·
Subjective beliefs (Splines): Expected standard deviation	-0.72	0.17	·	·
Risk aversion	-7.06	1.45	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	58.86	26.59
Lack of confidence in sav. acc. return estimate	·	·	24.09	22.15
Experimental tasks difficult	·	·	53.36	18.37
Experimental tasks obscure	·	·	11.37	16.88
Financial wealth ∈ (10000 €, 30000 €]	19.71	5.19	-3.03	21.04
Financial wealth ∈ (30000 €, ∞)	41.11	7.29	-67.26	31.56
Financial wealth missing	28.65	6.01	-38.22	25.19
Net income > 2500 €	6.92	2.45	27.05	10.76
Net income missing	-6.25	3.94	-8.91	13.18
High education	4.02	2.92	-58.32	18.18
30 < Age ≤ 50	11.09	5.53	22.70	16.60
50 < Age ≤ 65	7.82	5.53	-12.51	14.04
Age > 65	0.39	5.28	-21.39	15.08

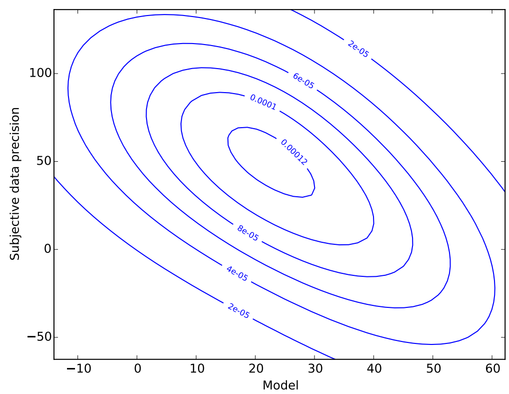
Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, except for the way of calculating moments of beliefs.

[6] We use the SciPy functions scipy.interpolate.PchipInterpolator to fit the splines and scipy.integrate.quad to calculate their moments.

Table 4.30. Average partial effects, moments of beliefs calculated by approximating the distribution using splines

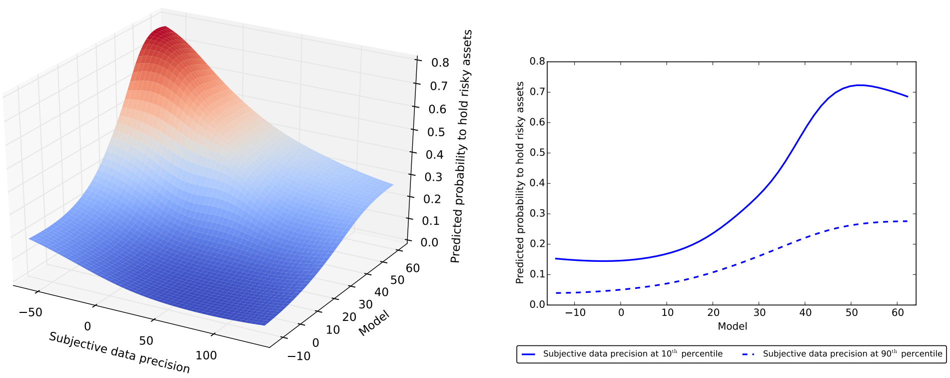
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs (Splines): Expected excess return	0.042	·	0.042
Subjective beliefs (Splines): Expected standard deviation	-0.020	·	-0.020
Risk aversion	-0.037	·	-0.037
Absolute difference between belief measures	·	-0.014	-0.014
Lack of confidence in AEX return estimate	·	-0.013	-0.013
Lack of confidence in sav. acc. return estimate	·	-0.006	-0.006
Experimental tasks difficult	·	-0.018	-0.018
Experimental tasks obscure	·	-0.003	-0.003
Financial wealth $\in$ (10000 €, 30000 €]	0.107	0.003	0.099
Financial wealth $\in$ (30000 €, ∞)	0.264	0.091	0.369
Financial wealth missing	0.179	0.048	0.212
High education	0.021	0.074	0.096
Net income > 2500 €	0.039	-0.026	0.013
Net income missing	-0.034	0.009	-0.026
$30 < \text{Age} \leq 50$	0.059	-0.024	0.029
$50 < \text{Age} \leq 65$	0.043	0.015	0.056
Age > 65	0.002	0.025	0.021

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, except for the way of calculating moments of beliefs.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.43. Joint density of the two indices, moments of beliefs calculated by approximating the distribution using splines



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.44. Predicted probability to hold risky assets, moments of beliefs calculated by approximating the distribution using splines

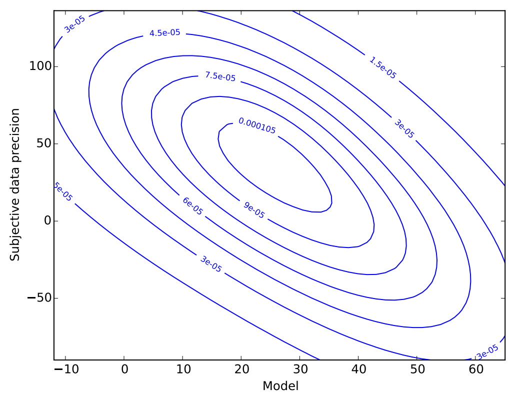
4.B.13 Including interaction between risk aversion and subjective uncertainty

Expected subjective uncertainty may be more relevant for stock market participation decisions of respondents who are more risk averse. To assess this possibility, we add the interaction between $\sigma_{t+1}^{\text{AEX}}$ and the standardized measure of risk aversion to the economic model index.

Table 4.31. Coefficient estimates for the economic model index and the subjective data precision index, including interaction between risk aversion and subjective uncertainty

	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.74	0.28	·	·
Risk aversion	-8.44	2.62	·	·
Interaction: $\sigma_{t+1}^{\text{AEX}} *$ Risk Aversion	0.10	0.32	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	59.66	27.71
Lack of confidence in sav. acc. return estimate	·	·	27.91	22.05
Experimental tasks difficult	·	·	54.12	19.71
Experimental tasks obscure	·	·	15.22	18.49
Financial wealth $\in$ (10000 €, 30000 €]	20.55	5.93	-18.61	21.44
Financial wealth $\in$ (30000 €, ∞)	42.92	9.10	-90.32	36.99
Financial wealth missing	30.47	7.23	-57.56	28.02
Net income > 2500 €	7.38	2.67	28.63	11.42
Net income missing	-6.88	4.23	-5.31	12.76
High education	3.23	3.06	-63.40	19.11
$30 < \text{Age} \leq 50$	11.83	5.54	23.37	16.80
$50 < \text{Age} \leq 65$	6.95	5.43	-17.24	15.04
Age > 65	-0.71	5.13	-23.19	16.18

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text, adding the interaction between the standard deviation of subjective beliefs and the standardized measure of risk aversion.



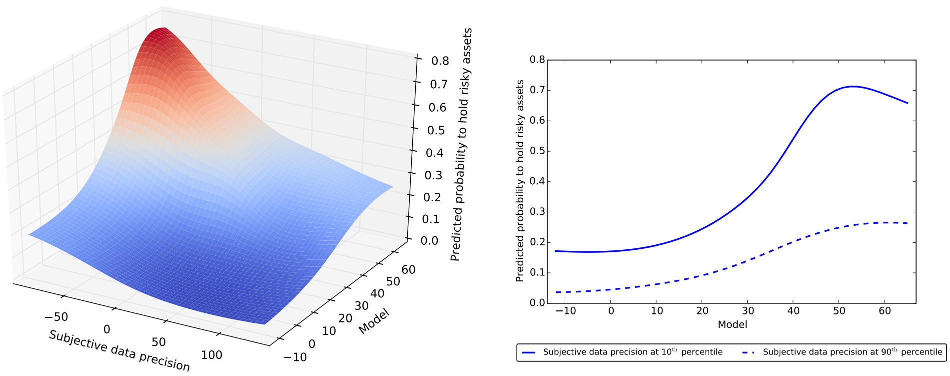
Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.45. Joint density of the two indices, including interaction between risk aversion and subjective uncertainty

Table 4.32. Average partial effects, including interaction between risk aversion and subjective uncertainty

	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu^{\text{AEX}}_{t+1} - \mu^{\text{sav. acc.}}_{t+1}$	0.034	·	0.034
Subjective beliefs: $\sigma^{\text{AEX}}_{t+1}$	-0.014	·	-0.014
Risk aversion	-0.041	·	-0.041
Interaction: $\sigma^{\text{AEX}}_{t+1} *$ Risk Aversion	0.004	·	0.004
Absolute difference between belief measures	·	-0.014	-0.014
Lack of confidence in AEX return estimate	·	-0.014	-0.014
Lack of confidence in sav. acc. return estimate	·	-0.007	-0.007
Experimental tasks difficult	·	-0.019	-0.019
Experimental tasks obscure	·	-0.004	-0.004
Financial wealth ∈ (10000 €, 30000 €]	0.101	0.017	0.099
Financial wealth ∈ (30000 €, ∞)	0.248	0.119	0.374
Financial wealth missing	0.173	0.069	0.222
Net income > 2500 €	0.037	-0.029	0.009
Net income missing	-0.034	0.006	-0.029
High education	0.015	0.081	0.097
30 < Age ≤ 50	0.055	-0.025	0.024
50 < Age ≤ 65	0.034	0.021	0.053
Age > 65	-0.004	0.028	0.019

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text, adding the interaction between the standard deviation of subjective beliefs and the standardized measure of risk aversion.



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.46. Predicted probability to hold risky assets, including interaction between risk aversion and subjective uncertainty

4.B.14 Dropping confidence, task obscurity, and task difficulty

Table 4.33. Coefficient estimates for the economic model index and the subjective data precision index, dropping confidence, task obscuiry, and task difficulty

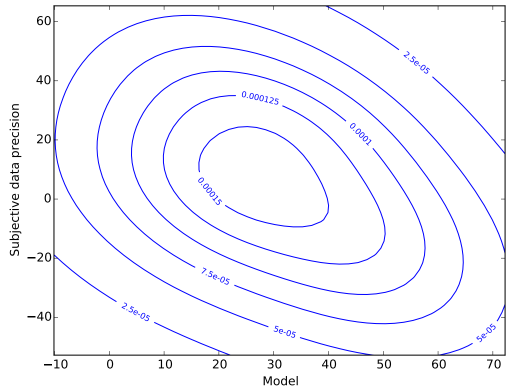
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.64	0.36	·	·
Risk aversion	-7.42	1.98	·	·
Absolute difference between belief measures	·	·	1.00	·
Financial wealth ∈ (10000 €, 30000 €]	21.36	6.50	-2.02	20.38
Financial wealth ∈ (30000 €, ∞)	45.66	11.12	-31.95	33.81
Financial wealth missing	34.53	8.80	9.26	27.58
Net income > 2500 €	4.31	3.36	18.46	9.43
Net income missing	-7.14	5.04	-17.76	10.46
High education	10.62	5.68	-34.22	15.14
30 < Age ≤ 50	8.05	7.19	33.45	14.45
50 < Age ≤ 65	9.89	5.32	5.30	10.49
Age > 65	0.91	4.85	-16.13	13.16

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text.

Table 4.34. Average partial effects, dropping confidence, task obscuiry, and task difficulty

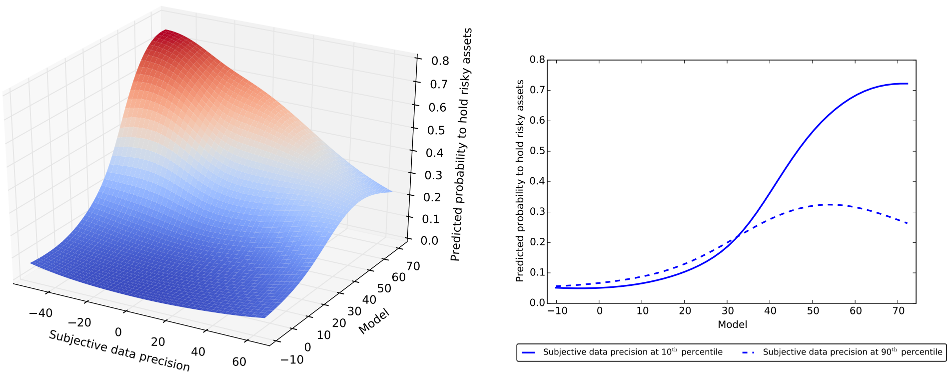
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.047	·	0.047
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.015	·	-0.015
Risk aversion	-0.044	·	-0.044
Absolute difference between belief measures	·	-0.008	-0.008
Financial wealth ∈ (10000 €, 30000 €]	0.115	0.002	0.109
Financial wealth ∈ (30000 €, ∞)	0.320	0.036	0.377
Financial wealth missing	0.230	-0.008	0.205
Net income > 2500 €	0.027	-0.009	0.017
Net income missing	-0.043	0.009	-0.037
High education	0.067	0.029	0.097
30 < Age ≤ 50	0.050	-0.023	0.027
50 < Age ≤ 65	0.061	-0.005	0.058
Age > 65	0.006	0.012	0.015

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.47. Joint density of the two indices, dropping confidence, task obscuiry, and task difficulty



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.48. Predicted probability to hold risky assets, dropping confidence, task obscuiry, and task difficulty

4.B.15 Including financial numeracy questions in both indices

In October 2014, we asked respondents three questions to determine their familiarity with basic financial concepts:

> **Question 1 - Simplest numeracy:** Suppose you have 100 euros on a savings account with an annual interest rate of 2 per cent. How much will you have on the savings account after five years, assuming you leave the money in this account?
>
> - More than 102 Euros
> - Less than 102 Euros
> - Exactly 102 Euros
> - Do not know
>
> **Question 2 - Interest compounding:** Suppose you have 100 euros on a savings account with an annual interest rate of 20 per cent and you never withdraw any money or interest. How much will you have after five years in total?
>
> - More than 200 Euros
> - Less than 200 Euros
> - Exactly 200 Euros
> - Do not know
>
> **Question 3 - Inflation:** Suppose the interest rate on your savings account is 1 per cent per year and inflation is 2 per cent per year. After one year, how much will you be able to buy with the money in the account?
>
> - Less than today
> - More than today
> - Exactly the same as today
> - Do not know

For each question, we create a binary variable and set it to 1 in case the subject provided the correct response, and to 0 otherwise. We include all variables as additional covariates in both indices.

Table 4.35. Coefficient estimates for the economic model index and the subjective data precision index, including financial numeracy questions as additional covariates

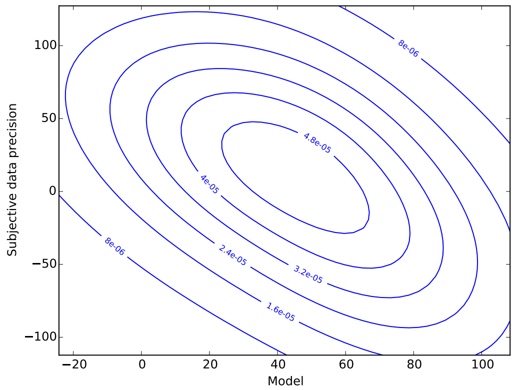
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	1.00	·	·	·
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.41	0.62	·	·
Risk aversion	-12.90	3.90	·	·
Absolute difference between belief measures	·	·	1.00	·
Lack of confidence in AEX return estimate	·	·	53.53	38.39
Lack of confidence in sav. acc. return estimate	·	·	30.25	34.70
Experimental tasks difficult	·	·	44.95	28.22
Experimental tasks obscure	·	·	32.41	24.52
Financial numeracy: Simplest numeracy question false	6.89	7.57	15.33	23.21
Financial numeracy: Interest compounding question false	-14.20	6.63	-8.30	16.91
Financial numeracy: Inflation question false	-21.70	9.36	5.24	19.05
Financial wealth ∈ (10000 €, 30000 €]	37.62	12.86	-30.47	37.33
Financial wealth ∈ (30000 €, ∞)	69.97	21.22	-98.42	69.58
Financial wealth missing	56.29	17.89	-66.61	57.08
Net income > 2500 €	14.28	6.07	36.44	19.32
Net income missing	-17.80	8.69	-12.58	16.55
High education	-0.20	7.96	-68.85	32.35
30 < Age ≤ 50	24.28	11.30	21.81	23.82
50 < Age ≤ 65	12.54	11.19	-23.89	21.58
Age > 65	-2.80	9.51	-43.84	26.63

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text. We add binary variables describing whether subjects correctly answered 3 distinct questions related to basic financial numeracy.

Table 4.36. Average partial effects, including financial numeracy questions as additional covariates

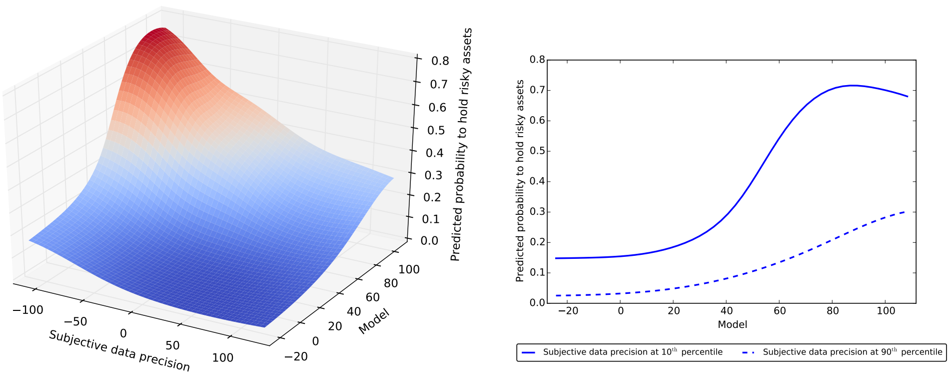
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs: $\mu_{t+1}^{\text{AEX}} - \mu_{t+1}^{\text{sav. acc.}}$	0.023	·	0.023
Subjective beliefs: $\sigma_{t+1}^{\text{AEX}}$	-0.004	·	-0.004
Risk aversion	-0.038	·	-0.038
Absolute difference between belief measures	·	-0.015	-0.015
Lack of confidence in AEX return estimate	·	-0.013	-0.013
Lack of confidence in sav. acc. return estimate	·	-0.008	-0.008
Experimental tasks difficult	·	-0.017	-0.017
Experimental tasks obscure	·	-0.009	-0.009
Financial numeracy: Simplest numeracy question false	0.019	-0.018	0.001
Financial numeracy: Interest compounding question false	-0.045	0.010	-0.036
Financial numeracy: Inflation question false	-0.069	-0.006	-0.074
Financial wealth ∈ (10000 €, 30000 €]	0.114	0.028	0.112
Financial wealth ∈ (30000 €, ∞)	0.249	0.132	0.378
Financial wealth missing	0.196	0.079	0.251
Net income > 2500 €	0.044	-0.039	0.004
Net income missing	-0.054	0.015	-0.041
High education	-0.001	0.095	0.094
30 < Age ≤ 50	0.069	-0.024	0.040
50 < Age ≤ 65	0.037	0.030	0.066
Age > 65	-0.009	0.056	0.041

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text. We add binary variables describing whether subjects correctly answered 3 distinct questions related to financial numeracy.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.49. Joint density of the two indices, including financial numeracy questions as additional covariates



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.50. Predicted probability to hold risky assets, including financial numeracy questions as additional covariates

Appendix 4.C Specification with less customized data

This section reports the results for the specification with less customized data described in Section 4.4.3 of the main text. As discussed in there, we restrict the specification to (i) the point estimate of AEX returns, (ii) one qualitative question to elicit risk attitudes, (iii) two simple qualitative proxies for the precision of subjective data, and (iv) sociodemographics.

Table 4.37. Coefficient estimates for the economic model index and the subjective data precision index, specification with less customized data

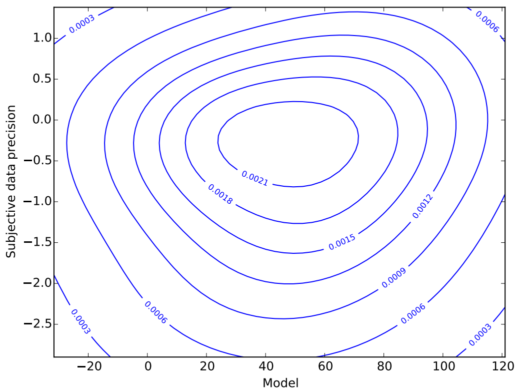
	Model		Imprecision of Measured Expectations	
	Estimate	Std. Err.	Estimate	Std. Err.
Subjective beliefs (direct question): Log expected excess return	1.00	·	·	·
Aversion to risks in general	-14.58	3.95	·	·
Experimental tasks difficult	·	·	1.00	·
Experimental tasks obscure	·	·	0.35	0.32
Financial wealth ∈ (10000 €, 30000 €]	39.87	16.70	-0.36	0.47
Financial wealth ∈ (30000 €, ∞)	28.69	28.48	-2.20	0.65
Financial wealth missing	50.20	20.85	-0.94	0.54
Net income > 2500 €	12.87	11.47	0.08	0.24
Net income missing	-55.81	18.94	-0.61	0.39
High education	-1.86	16.34	-0.96	0.32
30 < Age ≤ 50	56.42	20.71	0.94	0.49
50 < Age ≤ 65	26.02	15.74	-0.12	0.32
Age > 65	-19.28	16.09	-0.57	0.34

Sources: LISS panel and own calculations. All variables as described in Table 4.2 in the main text. The model only includes the point estimate as measure of beliefs, a qualitative question to elicit risk attitudes, and two qualitative proxies for the precision of subjective data.

Table 4.38. Average partial effects, specification with less customized data

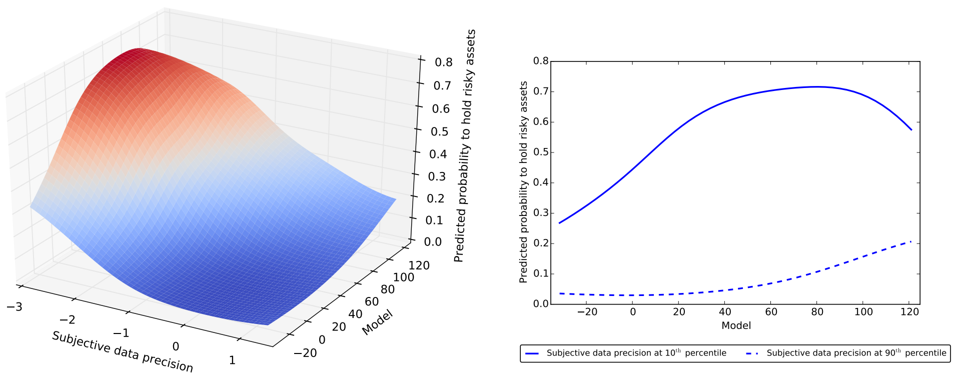
	Model	Imprecision of Measured Expectations	Combined
Subjective beliefs (direct question): Log expected excess return	0.033	·	0.033
Aversion to risks in general	-0.029	·	-0.029
Experimental tasks difficult	·	-0.034	-0.034
Experimental tasks obscure	·	-0.009	-0.009
Financial wealth ∈ (10000 €, 30000 €]	0.086	0.029	0.102
Financial wealth ∈ (30000 €, ∞)	0.063	0.338	0.396
Financial wealth missing	0.105	0.100	0.204
Net income > 2500 €	0.026	-0.009	0.017
Net income missing	-0.112	0.067	-0.057
High education	-0.004	0.119	0.115
30 < Age ≤ 50	0.102	-0.091	0.014
50 < Age ≤ 65	0.054	0.013	0.071
Age > 65	-0.039	0.068	0.025

Sources: LISS panel and own calculations. All variables as described in Table 4.3 in the main text. The model only includes the point estimate as measure of beliefs, a qualitative question to elicit risk attitudes, and two qualitative proxies for the precision of subjective data.



Sources: LISS panel and own calculations. The figure plots the joint density of the estimated indices of the Klein and Vella (2009) model; see Section 4.3.3 for a detailed description.

Figure 4.51. Joint density of the two indices, specification with less customized data



Sources: LISS panel and own calculations. The left panel presents the predicted probability of stock market participation for varying levels of the economic model and subjective data precision indices. The right panel plots the relation between the predicted probability of participation and the economic model index for the 10 and 90% quantiles of the subjective data precision index. Ranges are limited to the interval between the 5% and 95% quantiles of the marginal distributions.

Figure 4.52. Predicted probability to hold risky assets, specification with less customized data

Appendix 4.D Can we correct for measurement error using multiple measures?

This section provides some tentative evidence that attempting to correct for measurement error in subjective beliefs through multiple measures is of little help. To this end, Figure 4.53 presents the R^2 of an OLS regression of a stock market participation dummy on various linear combinations of two belief measures: (i) the mean belief constructed from the ball allocation task and (ii) the point estimate. The figure shows that – contrary to what one would expect if repeated measurements reduce measurement error – the variance explained is maximized by putting almost maximal weight on the belief from the ball allocation task. This suggests that traditional methods of correcting for measurement error do not apply in the case of subjective beliefs.

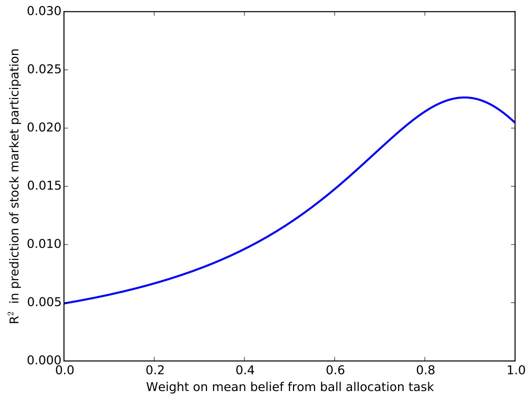


Figure 4.53. Variance explained in stockholdings by different linear combinations of two belief measures

5

The Nature and Predictive Power of Preferences: Global Evidence*

5.1 Introduction

This paper presents the Global Preference Survey (GPS), a novel and unique globally representative dataset. The data include measures of risk preference, time preference, positive and negative reciprocity, altruism, and trust that we collected for 80,000 individuals, drawn as representative samples in each of 76 countries. The coverage of countries spans all continents, a broad set of cultures, a wide range of development levels, and represents about 90 percent of both the world's population and global income, making the data also representative across countries. The underlying survey measures were selected and tested through a rigorous ex ante experimental validation procedure involving real monetary stakes, so that the survey items have a demonstrated ability to capture actual heterogeneity in state-of-the-art experiments with financial incentives (Falk et al., 2015). To ensure comparability of preference measures across countries, the elicitation followed a standardized protocol that was implemented through the professional infrastructure of the Gallup World Poll. Monetary stakes related to the elicitation involved comparable values in terms of purchasing power across countries, and the survey items were culturally neutral and translated using state-of-the-art procedures. In addition, pre-tests in 22 countries of various cultural heritage revealed the broad applicability of our survey items. In consequence, the resulting dataset provides an ideal basis for the first systematic investigation of the distribution, determinants, and predictive power of preferences around the world.

Using these data, we provide evidence for several novel findings, both at the country and at the individual level. First, for each of the six traits, we document a substantial variation not just across individuals, but also across entire countries. Second, we show that this cross-country heterogeneity is at least partly systematic and follows pronounced economic, geographic and cultural patterns. All preferences are significantly associated

* For valuable comments and discussions we are grateful to Doug Bernheim, Johannes Hermle, Benedikt Herrmann, Fabian Kosse, Andrei Shleifer, and seminar participants at Caltech, Konstanz, and UC San Diego. Ammar Mahran provided oustanding research assistance.

with important country-level variables including per capita income, democracy, inequality, redistributive policies, religion, and geographic or climatic variables. Third, the various preference measures are correlated, giving rise to distinct "preference profiles" of groups of countries. Fourth, in spite of the substantial between-country variation, most of the total individual-level variation in all preferences is due to within-country heterogeneity. Fifth, investigating the structure of this individual-level variation, we find that in the world population as a whole, all of the preferences are systematically related to individual characteristics. For instance, women tend to be less patient and more risk averse, and exhibit stronger social predispositions, than men. Patience is hump-shaped in age, while risk taking as well as positive and negative reciprocity are lower for older people. Self-reported cognitive skills positively correlate with patience, risk taking, and all social preferences. Sixth, we provide evidence of heterogeneity across countries underlying the strong *average* patterns of the individual-level correlates of preferences in the world population as a whole. We show that while some relationships between preferences and sociodemographics (such as between risk aversion and gender) are common to almost all cultures, others appear more culturally or institutionally specific. For example, patience and positive reciprocity exhibit a hump-shaped relationship with age in developed countries that is almost entirely absent in developing nations. Seventh, we show that individual-level preferences are also significantly correlated with household income, subjective perceptions of safety and health, as well as religious affiliation. Eighth, we examine the predictive power of preference heterogeneity for economic behaviors. Around the world, patient individuals are more likely to save and have higher educational attainment; more risk tolerant individuals are more likely to become self-employed and to be smokers; and social preferences are highly predictive of a broad range of prosocial behaviors and outcomes such as donating, volunteering time, assisting strangers, helping friends and relatives, or family structure. Finally, we shed light on the cultural origins of the global preference variation by making use of information on language structure: people who speak languages that do not require an explit coding of the future are more patient, positively reciprocal, trusting, and altruistic, both across and within countries.

Our analysis provides the first systematic assessment of the nature and explanatory power of preference heterogeneity around the world. The underlying data are, however, well-suited for a much broader research agenda on the determinants and implications of certain preference profiles. Going forward, the data lend themselves to investigations both at the micro- and the macro-level. At the micro level, several studies have examined individual-level preference heterogeneity and the corresponding correlates, like gender, in specific samples and cultures (see, e.g., Barsky et al., 1997; Frederick, 2005; Croson and Gneezy, 2009; Dohmen et al., 2008, 2010, 2011). However, the previous lack of data has prevented systematic investigations of the cultural specificity of such findings, an issue that is relevant for understanding the cultural or biological mechanisms through which individual characteristics like age or gender might shape preferences. Our results highlight some cases in which generalizing beyond single countries can be particularly misleading, because it ignores the country and population specificity of such effects. At the same time, the data show how some relationships are close to universal. Likewise, while previous work has provided evidence that preferences are predictive of important economic decisions, it has been an open question whether preferences are uniformly predictive of behaviors across cultures and institutional backgrounds, and to which extent they shape heterogeneity in life outcomes.[1]

The GPS data may also prove valuable for research in cultural economics and political economy (Guiso et al., 2006; Fernández, 2011; Alesina and Giuliano, forthcoming; Giuliano and Nunn, 2013). To date, empirical research into the roots of cross-country variation in preferences has been impeded by a lack of appropriate measures and representative sampling; contributions on the cross-country heterogeneity in preferences have typically made use of small and non-representative samples in a limited set of countries (Roth et al., 1991; Henrich et al., 2001; Herrmann et al., 2008). Accordingly, researchers interested in the determinants and implications of cultural variation have considered variables such as female labor force participation, fertility, individualism, and future-orientation (Giuliano, 2007; Fernández and Fogli, 2009; Gorodnichenko and Roland, 2011; Alesina et al., 2013; Chen, 2013; Alesina et al., forthcoming; Galor and Özak, 2014), but have not studied the preference component of culture. The data of the GPS, which feature 80,000 individuals from various cultural backgrounds, are likely to produce new insights in this direction.

Apart from such micro-level analyses, the representative cross-country nature of our data also permits an investigation of the relationships of preferences to aggregate economic and social outcomes across countries, which to date is uncharted territory.[2] Motivated by the strong and systematic correlations reported in this paper, the preference data may be used both in an attempt to explain cross-country differences in aggregate outcomes, and in controlling for preference differences when interest lies in identifying other relationships.

The remainder of the paper proceeds as follows. In the next section, we present the Global Preference Survey dataset. In Section 5.3, we describe the nature of cross-country variation in preferences. Section 5.4 studies the relationship between preferences and individual characteristics, while Section 5.5 investigates the relationships between preferences and behaviors. In Section 5.6, we analyze the relationship between preferences and language structure. Section 5.7 concludes.

5.2 Dataset

5.2.1 General Data Characteristics

The Global Preference Survey (GPS) is a new globally representative survey designed to measure respondents' time preferences, risk preferences, social preferences, and trust.

[1] Time preference correlates with outcomes ranging from savings to Body Mass Index (Ventura, 2003; Kirby and Petry, 2004; Borghans and Golsteyn, 2006; Eckel et al., 2005; Chabris et al., 2008; Tanaka et al., 2010; Meier and Sprenger, 2010; Sutter et al., 2013; Golsteyn et al., 2014). Risk preferences are related to various risky decisions, including being self-employed, migrating, and holding risky assets (See, e.g., Barsky et al., 1997; Bonin et al., 2007; Guiso and Paiella, 2008; Dohmen et al., 2011). Social preferences are correlated with cooperative behaviors in various aspects of life including in the workplace (Dohmen et al., 2009; Rustagi et al., 2010; Carpenter and Seki, 2011; Kosfeld and Rustagi, 2015).

[2] An exception is the burgeneoing literature on the importance of trust, see, e.g., Knack and Keefer (1997), Guiso et al. (2009), and Algan and Cahuc (2010).

The GPS data were collected within the framework of the Gallup World Poll, which surveys representative population samples in a large number of countries about social and economic issues on an annual basis. In 2012, we added the GPS to the World Poll's questionnaire in 76 countries, so that the survey items were fielded through the existing professional infrastructure of one of the world's leading global survey companies. Four noteworthy features characterize the preference data: (i) representative population samples within each country, (ii) geographical and economic representativeness in terms of countries covered, (iii) a rigorous experimental validation and selection procedure of the underlying survey items, and (iv) a standardized data collection protocol across countries. We discuss these features in the following; in addition, Appendix 5.A contains an extensive documentation of the data-collection process as well as additional details on the survey measures.

First, we measure preferences in large representative population samples in each country.[3] The median sample size was 1,000 participants per country, in 76 countries all over the world.[4] In total, we collected preference measures for more than 80,000 participants worldwide. Respondents were selected through probability sampling; ex-post representativeness of the data can be achieved using weights provided by Gallup.[5] In sum, our data allow for valid inferences about the distribution of preferences in each country as well as about between-country differences in preferences.

Second, the data are characterized by geographical representativeness in terms of the countries being covered. The sample of 76 countries is not restricted to Western industrialized nations, but covers all continents, various cultures, and different levels of development. Specifically, our sample includes 15 countries from the Americas, 25 from Europe, 22 from Asia and Pacific, as well as 14 African countries, 11 of which are Sub-Saharan. This set of countries covers about 90% of both the world population and global income.

Third, we designed, tested, and selected the survey items of the GPS using a rigorous ex-ante experimental validation and selection procedure (for details see Falk et al., 2015). While items in international surveys are frequently designed based on introspective arguments of plausibility or relevance, our items are the result of an explicit formal selection procedure, which also ensures that the resulting measures are predictive of actual preferences as measured through state-of-the-art experiments. Arguably, such an ex-ante validation of survey items constitutes a significant methodological advance over the ad-hoc selection of questions for surveys. As detailed in Falk et al. (2015), in the validation procedure, experimental subjects completed incentivized choice experiments to measure their preference parameters, and also answered a large battery of candidate survey questions. For each preference, those survey items that jointly perform best in predicting the financially incentivized behavior were selected to form the preference survey

[3] Data sets that contain preference measures for several countries typically come from small- or medium-scale surveys or experiments and are based on student or other convenience samples (e.g., Wang (2011), Rieger et al. (forthcoming), Vieider et al. (2015), Vieider et al. (2014).

[4] Notable exceptions include China (2,574 obs.), Haiti (504 obs.), India (2,539 obs.), Iran (2,507 obs.), Russia (1,498 obs.), and Suriname (504 obs.).

[5] These weights are constructed to render the observations representative in terms of age, gender, income, education, and geographic location.

module.[6] Thus, the module does not only consist of survey questions that predict behavior, but is composed of the best behavioral predictors out of a large set of candidate measures.

In a next step, the GPS was developed for implementation in the Gallup World Poll. To this end, Gallup conducted pre-tests in 22 countries of various cultural heritage, in order to ensure the implementability of the module in the available survey time of 7 to 8 minutes, and to test whether respondents of culturally and economically heterogeneous background understand and interpret the items adequately (see Appendix 5.A.3 for details). Other measures taken to ensure that the survey items were comparable across cultures included: (i) translation of all items back and forth in an iterative process using Gallup's regular translation scheme, and (ii) calibration of monetary values used in the survey questions according to median household income for each country.[7] Finally, the interviews for the World Poll 2012 took place face-to-face or via telephone by professional interviewers. Thus, the survey items were fielded in a comparable way using a standardized procedure across countries.

5.2.2 Preference Measures

For each preference, we obtain a final individual-level measure by weighing responses to multiple survey items using the weights obtained from the experimental validation procedure. These weights are based on an OLS regression of observed behavior in the financially incentivized experiments on the respective survey measures Falk et al. (see 2015, for details). We first standardize individual-level responses to all items (i.e., compute z-scores) and then weigh these standardized responses using the OLS weights to derive the best predictor of observed experimental behavior. Finally, for ease of interpretation, each preference measure is again standardized at the individual level, so that, by construction, each preference has a mean of zero and a standard deviation of one in the individual-level world sample.

The GPS contains twelve items which are summarized in Table 5.1. For most preferences, the set of questions consists of a combination of qualitative items, which are more abstract, and quantitative questions, which put the respondent into precisely defined hypothetical choice scenarios.[8]

[6] We excluded quantitative measures that require long and complex instructions, or which had shorter alternative quantitative measures that were close substitutes, from the set of candidate measures before the item selection procedure was conducted.

[7] As a benchmark, we used the monetary amounts in Euro that were offered in the validation study in Germany. Since monetary amounts used in the validation study with the German sample were round numbers to facilitate easy calculations (e.g., the expected return of a lottery with equal chances of winning and losing) and to allow for easy comparisons (e.g., 100 Euro today versus 107.50 in 12 months), we also rounded monetary amounts in all other countries to the next "round" number. While this necessarily resulted in some (very minor) variations in the real stake size between countries, it minimized cross-country differences in the understanding the quantitative items due to difficulties in assessing the involved monetary amounts.

[8] Under certain assumptions, the quantitative items allow the computation of quantitative measures such as a CRRA coefficient or an internal rate of return.

5.2.2.0.1 Patience. Our measure of patience is derived from the combination of responses to two survey measures, one with a quantitative and one with a qualitative format. The quantitative survey measure consists of a series of five interdependent hypothetical binary choices between immediate and delayed financial rewards, a format commonly referred to as "staircase" (or "unfolding brackets") procedure (Cornsweet, 1962). In each of the five questions, participants had to decide between receiving a payment today or larger payments in 12 months:

> *Suppose you were given the choice between receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which one you would choose. Please assume there is no inflation, i.e., future prices are the same as today's prices. Please consider the following: Would you rather receive amount x today or y in 12 months?*

The immediate payment x remained constant in all subsequent four questions, but the delayed payment y was increased or decreased depending on previous choices (see Appendix 5.A.6.1 for an exposition of the entire sequence of binary decisions). In essence, by adjusting the delayed payment according to previous choices, the questions "zoom in" around the respondent's point of indifference between the smaller immediate and the larger delayed payment and make efficient use of limited and costly survey time. The sequence of questions has 32 possible ordered outcomes. In the international survey, monetary amounts x and y were expressed in the respective local currency, scaled relative to median household income in the given country. Notably, this measure not only resembles standard experimental procedures of eliciting time preferences, but it is also precisely defined, arguably making it less prone to culture-dependent interpretations. This makes the quantitative patience measure well-suited for a multinational study like the present one.

The qualitative measure of patience is given by the respondents' self-assessment regarding their willingness to wait on an 11-point Likert scale, asking "how willing are you to give up something that is beneficial for you today in order to benefit more from

Table 5.1. Survey items of the GPS

Preference	Item Description	Weight
Patience	Intertemporal choice sequence using staircase method	0.71
	Self-assessment: Willingness to wait	0.29
Risk taking	Lottery choice sequence using staircase method	0.47
	Self-assessment: Willingness to take risks in general	0.53
Positive reciprocity	Self-assessment: Willingness to return a favor	0.48
	Gift in exchange for help	0.52
Negative reciprocity	Self-assessment: Willingness to take revenge	0.37
	Self-assessment: Willingness to punish unfair behavior towards self	0.265
	Self-assessment: Willingness to punish unfair behavior towards others	0.265
Altruism	Donation decision	0.54
	Self-assessment: Willingness to give to good causes	0.46
Trust	Self-assessment: People have only the best intentions	1

Notes. See Appendix 5.A.6 for the wording of the questions and Appendix 5.A.7.2 for a discussion of the weights.

that in the future?" As discussed above, the two items were first standardized and then combined linearly to form the final measure of patience, which was then standardized again at the individual level in the world sample. The quantitative measure obtained a weight of 71%.

5.2.2.0.2 Risk Taking. Risk preferences were also elicited through a series of related quantitative questions as well as one qualitative question. Just as with patience, the quantitative measure consists of a series of five binary choices between a fixed lottery and varying sure payments, hence making use of the advantages of precisely defined, quantitative survey items in culturally and economically heterogeneous samples:

> *Please imagine the following situation. You can choose between a sure payment of a particular amount of money, or a draw, where you would have an equal chance of getting amount x or getting nothing. We will present to you five different situations. What would you prefer: a draw with a 50 percent chance of receiving amount x, and the same 50 percent chance of receiving nothing, or the amount of y as a sure payment?*

The questions are again interdependent in the sense that the choice of the lottery results in an increase of the sure amount being offered in the next question, and vice versa. Appendix 5.A.6.2 contains an exposition of the entire sequence of survey items. The qualitative item asks for the respondents' self-assessment of their willingness to take risks on an eleven-point scale ("In general, how willing are you to take risks?"). This qualitative subjective self-assessment has previously been shown to be predictive of risk-taking behavior in the field in a representative sample (Dohmen et al., 2011) as well as of incentivized experimental risk-taking across countries in student samples (Vieider et al., 2014). The qualitative item and the outcome of the quantitative staircase measure were combined through roughly equal weights.

5.2.2.0.3 Positive Reciprocity. People's propensity to act in a positively reciprocal way was also measured using one qualitative item and one question with a quantitative component. First, respondents were asked to provide a self-assessment about how willing they are to return a favor on an 11-point Likert scale. Second, participants were presented a choice scenario in which they were asked to imagine that they got lost in an unfamiliar area and that a stranger – when asked for directions – offered to take them to their destination. Participants were then asked which out of six presents (worth between 5 and 30 euros in 5 euros intervals) they would give to the stranger as a "thank you". These two items receive roughly equal weights.

5.2.2.0.4 Negative Reciprocity. Negative reciprocity was elicited through three self-assess-ments. First, people were asked how willing they are to take revenge if they are treated very unjustly, even if doing so comes at a cost (0-10). The second and third item probed respondents about their willingness to punish someone for unfair behavior, either towards *themselves* or towards a *third person*.[9] This last item captures prosocial punishment and hence a concept akin to norm enforcement. These three items receive weights of about one third each.

5.2.2.0.5 Altruism. Altruism was measured through a combination of one qualitative and one quantitative item, both of which are related to donation. The qualitative question asked people how willing they would be to give to good causes without expecting anything in return on an 11-point scale. The quantitative scenario depicted a situation in which the respondent unexpectedly received 1,000 euros and asked them to state how much of this amount they would donate. These two items were weighted about equally.

5.2.2.0.6 Trust. To measure trust, we used one item, which asked people whether they assume that other people only have the best intentions (Likert scale, 0-10).[10]

5.2.3 Further Variables of Interest

The GPS data include a wide range of individual-level background variables which can be linked to the preference measures. These background variables include the core items of the Gallup World Poll such as (i) extensive sociodemographic information (e.g., age, gender, family structure, country of birth, religious affiliation, location of residence, or migration background including country of origin), (ii) a variety of self-reported behaviors and economic outcome variables including income, educational attainment, savings, labor market decisions, health, and behavior in social interactions, and (iii) opinions and attitudes about issues such as local and global politics, local institutional quality, economic prospects, safety, or happiness. We also elicited a self-reported proxy for cognitive skills by asking people to assess themselves regarding the statement “I am good at math” on an 11-point Likert scale. Finally, the data contain regional identifiers (usually at the state or province level), hence allowing for cross-regional analyses within countries.

5.3 Cross-Country Analysis

The analysis begins with an investigation of the heterogeneity of preferences around the world. Figure 6.1 shows how the country averages for each (standardized) preference compare to the world average. The figure reveals that preferences vary substantially across countries, by at least one standard deviation for each preference (see figure notes on color coding).[11] Most country differences displayed in Figure 5.1 are statisti-

[9] In the original validation study, the second and third item were collapsed into one question which asked people how willing they are to punish others, without specifying *who* was treated unfairly (Falk et al., 2015). However, in the cross-country pre-test, a number of respondents indicated that this lack of specificity confused them, so that we broke this survey item up into two questions. Accordingly, the weights for deriving an individual-level index of negative reciprocity are determined by dividing the OLS weight for the original item by two.

[10] Given the existence of the World Values Survey data, we can perform a first plausibility check on our data by showing that our trust measure is correlated with the WVS data ($\rho = 0.53$, $p < 0.01$).

[11] Appendix 5.A.8 provide an alternative way to visualize the heterogeneity, with histograms of preferences at the country and individual levels.

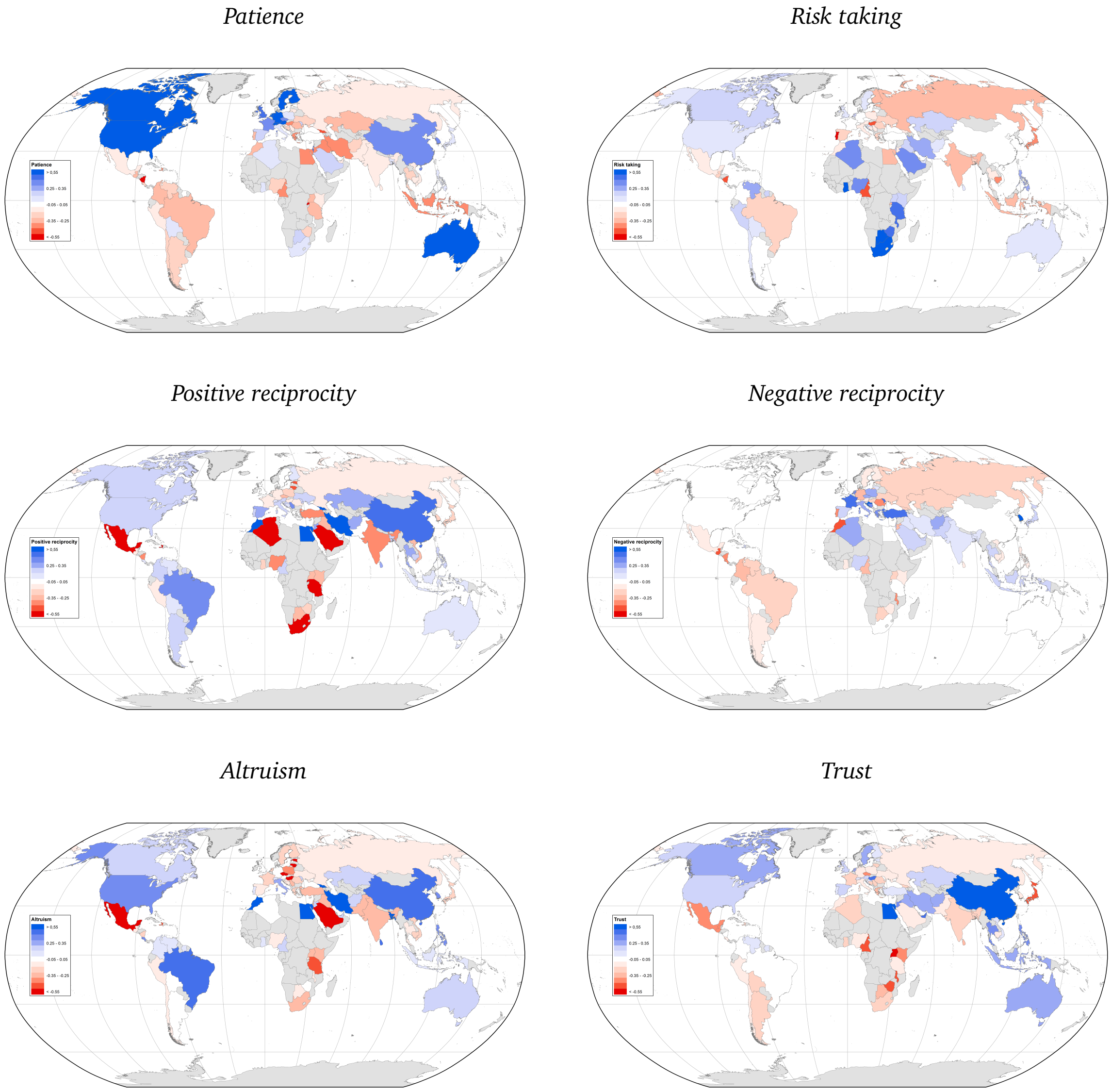


Figure 5.1. World maps of preferences. In each figure, white denotes the world average. Darker blue indicates higher values of a given trait, while darker red colors indicate lower values, all of which are measured in standard deviations from the world mean. Grey indicates missings.

cally significant. Calculating t-tests of all possible (2,850) pairwise comparisons for each preference, the fraction of significant (1-percent level) country differences are: 78% for risk, 83% for patience, 80% for altruism, 81% for positive reciprocity, 79% for negative reciprocity, and 78% for trust, respectively.

To provide a complementary perspective on the geographic and cultural variation in aggregate preferences, Figures 5.9a and 5.9b in Appendix 5.C group countries into six world regions: Western and "Neo" Europe (i.e., the US, Canada, and Australia), Former Communist Eastern Europe, Asia, North Africa and Middle East, Sub-Saharan Africa, and Southern America. For each region, we present two scatter plots which illustrate the distribution of patience, risk taking, negative reciprocity, and “prosociality”[12] within each region, relative to the world mean of the respective preference. Populations in Western and “Neo” Europe tend to be substantially more patient than the world mean. In fact, all of the ten most patient countries in the world are either located in Western Europe or part of the English-speaking world, with the Northern European countries exhibiting particularly high levels of patience. Western European countries are also notable for negative reciprocity; eight out of the ten most negatively reciprocal countries are located in Europe. Three of the five most negatively reciprocal countries in our sample – Turkey, Greece, and South Korea – have been found in previous research (Herrmann et al., 2008) to be particularly prone to retaliatory (anti-social) punishments in incentivized social dilemma games.

To the East, the former communist Eastern European countries are on average rather risk averse and not very patient, but the patterns are less clear compared to their Western European counterparts. Similar patterns obtain for East and South Asia, where most populations except the Confucian ones (China, Japan, South Korea) are relatively impatient.

Middle Eastern and North African populations have in common relatively high levels of risk tolerance and low levels of patience. Prosociality and negative reciprocity of this group of countries are fairly diverse. Notably, all of the ten most risk tolerant countries in our sample are located in the Middle East or Africa; in addition, *all* sub-Saharan populations are on average less prosocial than the world mean and are rather impatient.

Finally, in the Southern Americas, most populations appear impatient. They also have low levels of negative reciprocity and intermediate values in risk taking and prosociality. In sum, these results highlight that different types of preferences are spatially and culturally concentrated.

To begin to open the black box of cross-country variation in preferences, we proceed by relating preferences to country-level characteristics. In a first step, we seek to understand which fraction of the between-country variation can be explained by commonly employed variables that are plausibly exogenous to preferences. To this end, Figure 5.2 plots the R-squared of OLS regressions of each preference on a set of variables which proxy for geography, climate, diversity, and religion. Geography includes distance to the equator, longitude, and the fraction of the population at risk of contracting malaria; climate includes average precipitation and average temperature as well as the fraction of the population living in the (sub)-tropics; diversity includes ethnic and religious fractionalization as well as linguistic diversity (Fearon, 2003); and religion includes the share of the population adhering to catholic, protestant, muslim, buddhist, hinduist, or jewish beliefs.

As Figure 5.2 shows, the variance explained by categories differs, and it differs by preference. For example, geography explains 20% of the variation in patience, but is largely unrelated to risk. Climate, by contrast, is strongly related to both preferences. The diversity indices explain more than a quarter of the varience in patience, but virtually no variation in altruism or positive reciprocity. Religion shares explain a considerable

[12] Given the high correlations between altruism, positive reciprocity, and trust (see below), we define prosociality as the unweighted average of these three measures. Very similar results obtain if we run a factor analysis and use the first factor of the three measures.

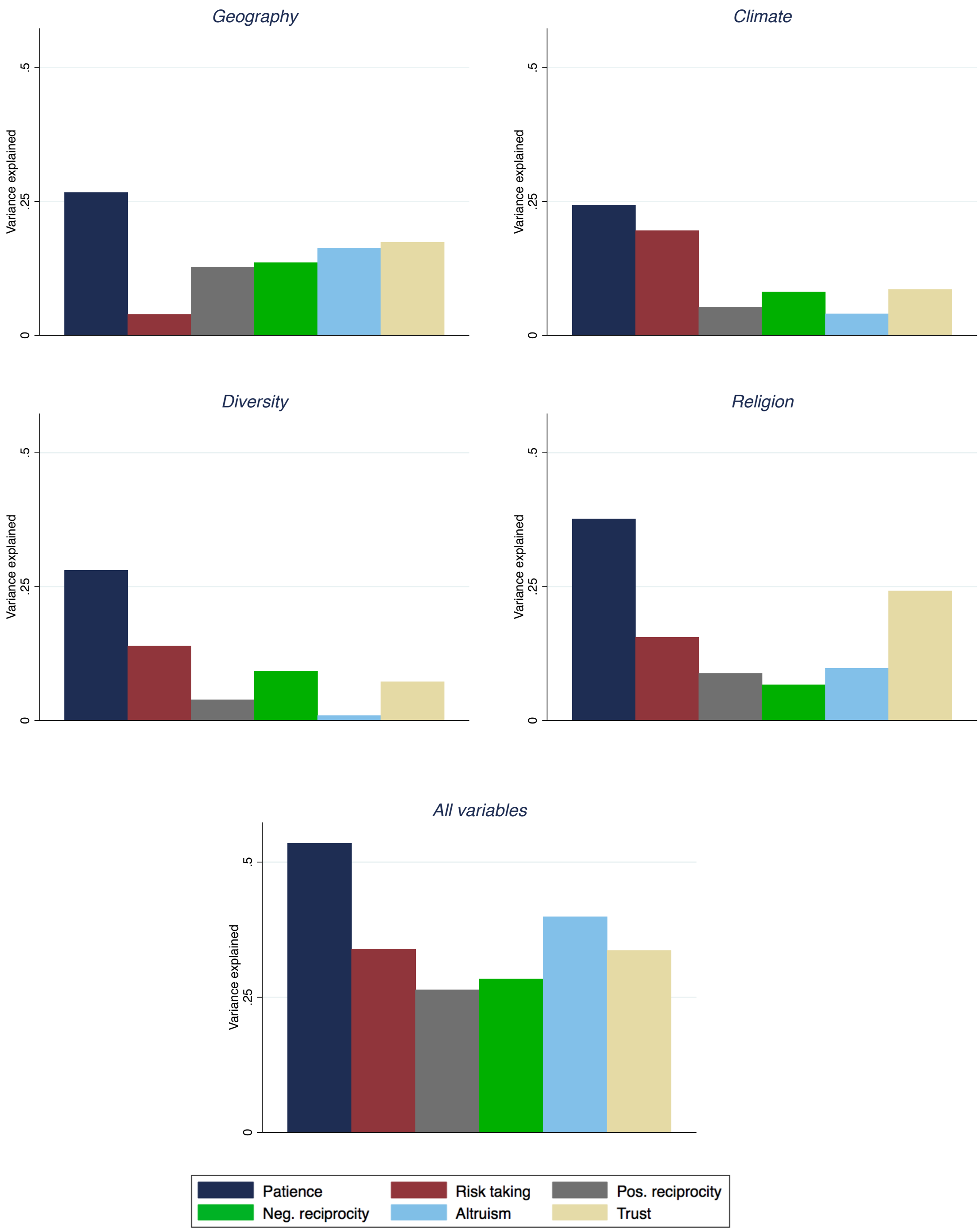


Figure 5.2. Cross-country variance explained. The figures plot the R^2 of an OLS regression of the respective preference on absolute latitude, longitude, and the fraction of the population at risk of contracting malaria (geography), average temperature, average precipitation and the fraction of the population living in the (sub)-tropics (climate), religious and ethnic fractionalization as well as linguistic diversity (diversity), and the share of protestants, catholics, muslims, jews, hinduists, and buddhists (religion), as well as all of these variables. All figures contain 73 countries. See Appendix 5.D for the regression coefficients and Appendix 5.I for details on all covariates.

fraction of the variation in patience and trust, but little of the other social preferences. All variables in combination can account for between 26% and 53% of the between-country variation in preferences, with patience generally exhibiting the highest share of variance explained. Appendix 5.D presents the regression coefficients underlying each figure.

In a next step, we study the correlations between preferences and country-level outcomes or characteristics which may be endogenous to preferences. Table 5.2 presents Pearson correlations between all preferences and a set of economic, political, and health variables.[13] Overall, the country-level preference measures vary systematically with important aggregate economic and social outcomes. In particular, we find that a population's average patience is strongly correlated with the degree of economic and institutional development, as exemplified by high income per capita, a high democracy index, and high average life expectancy at birth. Average risk taking exhibits significant correlations with a set of variables which are connected to the "riskiness" of the respective economic, social, and health environment. Risk tolerant populations are those with low life expectancy, high economic inequality, low redistribution, low labor protection, and more intentional homicides. Thus, overall, risk averse populations seem to be more protected and have stronger social safety nets, consistent with the view that the riskiness of the overall environment is influenced by, or influences, people's risk attitudes.

Regarding social preferences, positive reciprocity tends to be higher in environments with high life expectancy and in countries with low crime. Populations with strong negatively reciprocal inclinations are richer, have lower inequality, redistribute more, and exhibit lower numbers of homicides. Finally, trust is positively correlated with national income, but not with democracy. Higher inequality and a higher homicide rate are associated with lower trust. These intriguing correlations raise the question of whether preferences might shape, be shaped by, or co-evolve with cross-country variation in development, inequality, or institutions, pointing towards an important direction for future research.

While various preferences exhibit economic and geographic patterns, preferences may also be correlated amongst each other, giving rise to country-level preference profiles. To investigate the relationship among different preferences, Table 5.3 shows Pearson correlations of preferences together with levels of significance.[14] The significant correlations indicate that preferences are not distributed independently of one another. One set of traits that goes together is risk tolerance and patience, as shown by the positive and statistically significant correlation at the country level. This is in spite of the special case of Sub-Saharan African countries, which tend to be risk seeking and impatient, as discussed above.[15] Another grouping of positively correlated traits involves prosociality, i.e., the traits of positive reciprocity, altruism and trust. While trust constitutes a belief rather than a preference, all of these traits share in common that they describe positive behavioral dispositions towards others. The correlation between altruism and positive

[13] Computing Spearman rank correlations yields very similar results.

[14] The results are similar when computing Spearman correlations.

[15] Excluding African countries, the positive correlation between risk taking and patience increases to 0.30, while other correlations remain largely the same. The correlation between the staircase risk and patience items is 0.19, while that between the two qualitative risk and patience items is 0.55.

Table 5.2. Pairwise correlations between preferences and country-level variables

	Log [GDP p/c PPP]	Democracy index	Life expectancy	Gini coefficient	Redistribution (% of GDP)	Labor regulation	Log [# of homicides]
Patience	0.630***	0.449***	0.425***	0.064	0.512***	-0.003	-0.459***
Risk taking	-0.103	-0.167	-0.394***	0.318**	-0.228*	-0.367**	0.210*
Positive reciprocity	0.078	-0.051	0.284**	-0.217	0.111	0.066	-0.231**
Negative reciprocity	0.230**	0.091	0.212*	-0.291**	0.264**	0.002	-0.400***
Altruism	-0.091	-0.227*	0.075	0.011	-0.204*	-0.058	-0.051
Trust	0.282***	-0.089	0.382***	-0.297**	0.144	0.188	-0.266**
Observations	76	74	76	50	70	56	76

Notes. Pairwise Pearson correlations between average preferences and other variables at country level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

reciprocity is particularly high, and trust also tends to be higher where people are positively reciprocal. This is intuitive as it is hard to imagine stable and high levels of trust in environments absent positive reciprocity, i.e., trust rewarding behaviors.[16] Despite being related to the social domain, negative reciprocity is not at all correlated with prosociality. We report the correlation structure among preferences at the individual level in Appendix 5.B.

Evidence that preference dispositions vary substantially across countries does not imply that cross-country or cultural differences are the primary source of preference variation in the world. Table 5.4 shows results from a total variance decomposition, which reveals that the within-country variation in preferences is actually larger than the between-country variation, an observation that varies only minimally by preference. Part of the within-country variation might reflect measurement error, so that the variation in true preferences is overstated.[17] However, the available evidence on the size of test-retest correlations and measurement error suggests that it is highly unlikely that measurement error alone produces the fact that within-country variation dominates between-country variation, see Appendix 5.H for details.

The relative importance of within-country variation does not imply that country differences are negligible or irrelevant. It does, however, suggest that individual characteristics contribute relatively more to the formation of human preferences than national borders.

Table 5.3. Pairwise correlations between preferences at country level

	Patience	Risk taking	Pos. reciprocity	Neg. reciprocity	Altruism	Trust
Patience	1					
Risk taking	0.231**	1				
Positive reciprocity	0.0202	-0.256**	1			
Negative reciprocity	0.262**	0.193*	-0.154	1		
Altruism	-0.00691	-0.0155	0.711***	-0.132	1	
Trust	0.186	-0.0613	0.363***	0.160	0.272**	1

Notes. Pairwise Pearson correlations between average preferences at country level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[16] Given that our survey item for trust measures only the belief-component of trust (as opposed to first-mover behavior in trust games, which is also affected by risk preferences), the low correlation between trust and risk taking is consistent with previous within-country findings.

[17] In fact, comparing the between- and within-country variation across survey items reveals that the between-country variation tends to be relatively larger for the quantitative survey items. For example, in the case of patience, the quantitative staircase procedure exhibits a between-country variation of 15.7%, while the qualitative patience measure has a between-country variation of 7.3%.

Table 5.4. Between- vs. within-country variation

Preference	Between-country variation (%)	Within-country variation (%)
Patience	13.5	86.5
Risk taking	9.0	91.0
Positive reciprocity	12.0	88.0
Negative reciprocity	7.0	93.0
Altruism	12.3	87.7
Trust	8.2	91.8

Notes. Results from a variance decomposition in which the total individual-level variation in the respective preference is decomposed into the variance of the average preference across countries and the average of the within-country variance. Formally, the between-country variation corresponds to the R^2 of an OLS regression of all individual-level observations on a set of country dummies in which all observations are weighted by the sampling weights provided by Gallup to achieve (ex post) representativeness.

5.4 Preferences and Individual Characteristics

5.4.1 Age, Gender, and Cognitive Ability

The pronounced within-country heterogeneity calls for a better understanding of the individual-level determinants of preferences. The analysis focuses on three main characteristics: age, gender and cognitive ability, taking self-reported math skills as a proxy for the latter.[18] Gender, age, and cognitive ability are interesting to study, and have received particular attention in previous research on preferences (e.g., Barsky et al., 1997; Donkers et al., 2001; Croson and Gneezy, 2009; Frederick, 2005; Sutter and Kocher, 2007; Dohmen et al., 2010, 2011; Benjamin et al., 2013), for two main reasons. First, they are associated with important differences in economic outcomes. If preferences vary with these traits, they could be part of the explanation. Second, these traits are plausibly exogenous to preferences. The previous literature has proposed various mechanisms through which gender, age, and cognitive ability might be related to preferences, ranging from biological to purely social (Croson and Gneezy, 2009; Dohmen et al., 2011; Benjamin et al., 2013). There is limited knowledge, however, about the relative importance of these different types of mechanisms. For example, alternative explanations for age effects include an influence of idiosyncratic historical and cultural environments on the one hand, to biological aspects of the aging process on the other hand. The ability to examine how different preferences vary with characteristics across countries with diverse historical experiences can shed light on such questions.

Table 5.5 reports regressions of each preference on age, age squared, gender and math skills in the full world sample, conditional on country fixed effects. For each preference, the second column contains additional covariates to be discussed below. The

[18] This proxy may tend to capture the numeracy aspect of cognitive skills. Subjective assessments of ability are correlated with measured cognitive ability, and have predictive power for academic achievement (Spinath et al., 2006). While such relative self-assessment might be interpreted in different ways across countries, we only use self-reported cognitive skills in within-country analyses.

variables are standardized, so the coefficients show the change in the dependent variable (respective preference) in standard deviation units, for a one unit change in an independent variable (individual characteristics).

The estimates in table 5.5 reveal that, in the world population as a whole, preferences vary significantly with gender, age, and cognitive ability. Specifically, for gender, the strongest relationship is for risk preference: women are relatively more risk averse than men. Women are also significantly more prosocial, i.e., they tend to be more altruistic, positively reciprocal, and trusting, and are less negatively reciprocal. Women are slightly more impatient than men. In terms of age, the regression results indicate that, on average: Young individuals are relatively more willing to take risks, and punish; the middle aged are especially positively reciprocal and patient; the elderly have the strongest risk aversion and are relatively trusting. Preferences are also significantly related to self-reported cognitive ability: high cognitive ability individuals are more patient, less risk averse, more positively and negatively reciprocal, more trusting, and more altruistic.

We next exploit the ability to study the relationship of preferences to characteristics separately, for 76 different countries, to understand the extent to which the relationship of preferences to characteristics is culturally specific. Figure 5.3 addresses this question for gender differences. For each country, we regress each preference on age, age squared, gender, and cognitive ability. Figure 5.3 plots the resulting (conditional) gender coefficients. Each dot represents a country, the respective coefficient and the respective level of significance (green not significantly different from zero, pink at 10, blue at 5, and red at 1 percent level of significance, respectively). To ease reading, each panel also contains a horizontal line at zero.

Figure 5.3 shows that greater risk aversion among women is common to most countries. In 95 percent of countries, the gender coefficient is non-zero and in the direction of greater risk aversion among women. Of these negative coefficients, 82 percent are statistically significant at least at the 10-percent level. The gender difference in negative reciprocity is next in terms of universality: in 89 percent of countries is the female population on average less willing to reciprocate negatively than men. The impression that the majority of countries have a similar qualitative relationship between gender and preferences extends to the other social preferences, although there is more heterogeneity. For altruism, positive reciprocity, and trust about 79 percent, 71 percent, and 71 percent of countries have women being more pro-social than men, respectively. Patience is the most variable in terms of gender differences, but still, 68 percent of countries have non-zero coefficients of the same sign. These findings show that there are striking commonalities in terms of how gender and preferences are related, across a large and diverse set of cultural backgrounds. At the same time, there is substantial heterogeneity in the magnitude of the relationships across countries.

Figure 5.4 explores how the relationship between age and preferences varies across countries. This figure is divided by whether countries are OECD-members or not, in order to show some of the most salient cross-country differences: OECD-members exhibit a hump-shaped pattern for both positive reciprocity and patience, that is almost entirely absent in non-OECD countries. In Appendix 5.E.1, we also provide the age profiles for the six world regions defined above, i.e., Western and “Neo” Europe, Former Communist

Table 5.5. Correlates of preferences at individual level

	Dependent variable:											
	Patience		Risk taking		Pos. reciprocity		Neg. reciprocity		Altruism		Trust	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age	0.72***	0.69***	-0.083	0.46**	1.02***	0.79***	-0.36*	-0.18	-0.0060	0.024	0.37*	0.071
	(0.17)	(0.18)	(0.20)	(0.19)	(0.17)	(0.18)	(0.19)	(0.19)	(0.14)	(0.14)	(0.21)	(0.16)
Age squared	-1.45***	-1.34***	-1.20***	-1.75***	-1.17***	-0.90***	-0.45**	-0.69***	0.015	-0.13	0.032	0.24
	(0.20)	(0.21)	(0.21)	(0.19)	(0.18)	(0.20)	(0.18)	(0.20)	(0.15)	(0.15)	(0.20)	(0.17)
1 if female	-0.056***	-0.033***	-0.17***	-0.16***	0.049***	0.057***	-0.13***	-0.12***	0.10***	0.10***	0.066***	0.059***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Subj. math skills	0.028***	0.025***	0.046***	0.043***	0.038***	0.037***	0.040***	0.043***	0.044***	0.042***	0.056***	0.060***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log [Household income p/c]		0.038***		0.063***		0.042***		0.019*		0.051***		-0.013*
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
Subj. law and order index		0.058***		0.044*		0.021		-0.066**		0.037*		0.22***
		(0.02)		(0.02)		(0.03)		(0.03)		(0.02)		(0.03)
Subj. health index		0.097***		0.12***		0.12***		-0.056**		0.10***		0.052**
		(0.02)		(0.03)		(0.04)		(0.03)		(0.02)		(0.03)
1 if christian		-0.082**		-0.12***		-0.012		-0.13***		0.13***		0.10***
		(0.03)		(0.02)		(0.03)		(0.03)		(0.03)		(0.02)
1 if muslim		-0.13**		-0.098**		-0.0000071		-0.11**		0.19***		0.18***
		(0.05)		(0.05)		(0.05)		(0.05)		(0.05)		(0.05)
1 if hinduist		-0.12*		-0.16***		0.0091		-0.18***		0.13**		0.16***
		(0.06)		(0.04)		(0.05)		(0.05)		(0.05)		(0.05)
1 if buddhist		0.10		-0.19***		0.014		-0.062		0.16**		-0.061
		(0.09)		(0.05)		(0.06)		(0.05)		(0.06)		(0.12)
1 if jew		0.40***		-0.11**		-0.024		0.20***		0.29***		0.12**
		(0.06)		(0.04)		(0.05)		(0.06)		(0.04)		(0.05)
1 if other religion		-0.00022		-0.18***		-0.028		-0.060		0.14**		0.0053
		(0.06)		(0.04)		(0.08)		(0.05)		(0.05)		(0.09)
Constant	-0.37***	-0.58***	0.21***	-0.32***	-0.079**	-0.40***	0.37***	0.41***	-0.064**	-0.67***	-0.078**	-0.29***
	(0.04)	(0.09)	(0.04)	(0.09)	(0.04)	(0.09)	(0.05)	(0.10)	(0.03)	(0.06)	(0.04)	(0.08)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	78501	58479	78445	58437	78869	58733	77521	57731	78632	58532	77814	57952
R^2	0.165	0.154	0.167	0.185	0.128	0.121	0.112	0.114	0.135	0.130	0.111	0.100

Notes. OLS estimates, standard errors (clustered at country level) in parentheses. Coefficients are in terms of units of standard deviations of the respective preference (relative to the individual world mean). For the purposes of this table, age is divided by 100. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

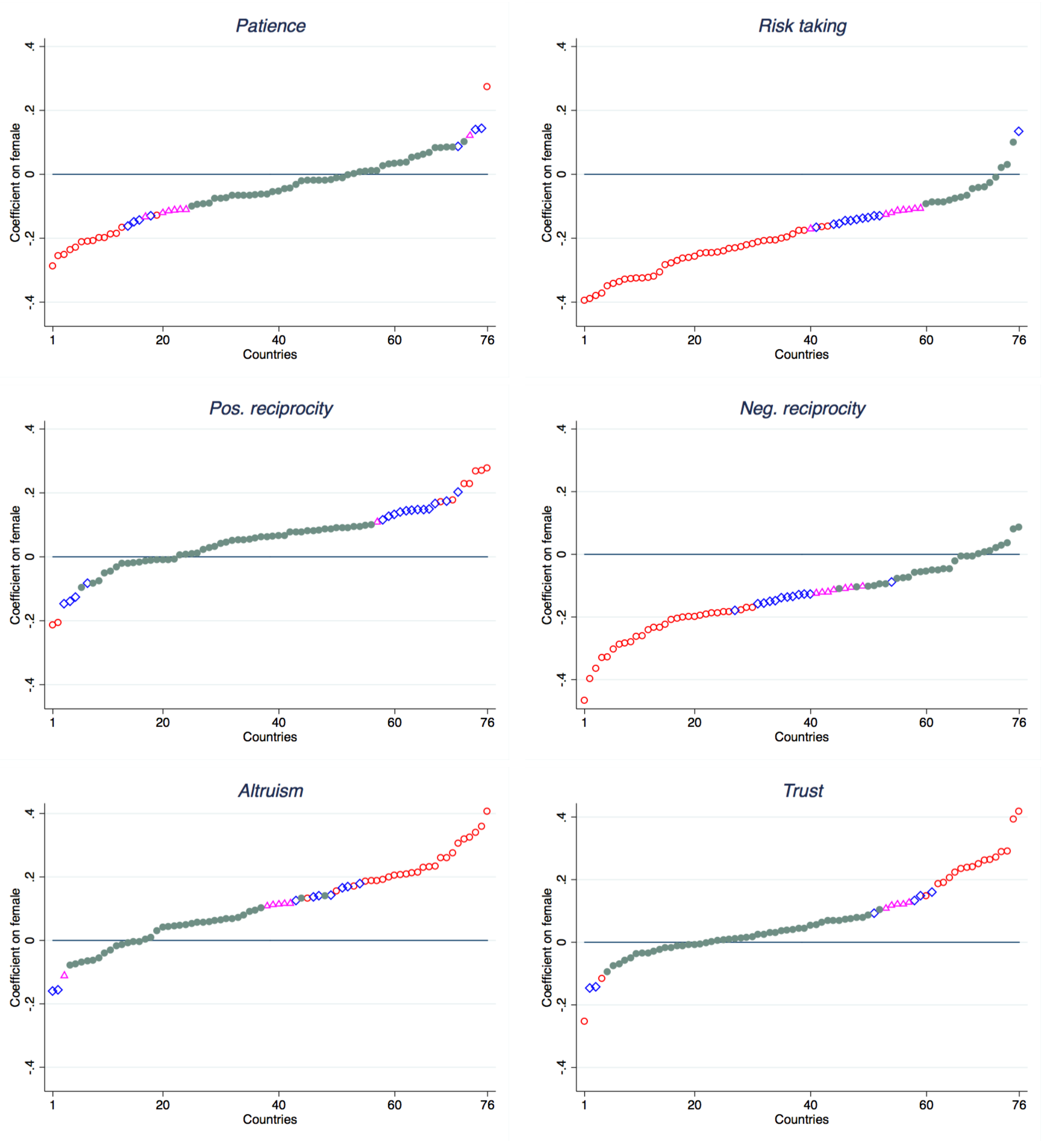


Figure 5.3. Gender correlations separately by country. Each panel plots the distribution of gender correlations. That is, for each country, we regress the respective preference on gender, age and its square, and subjective math skills, and plot the resulting gender coefficients as well as their significance level. In order to make countries comparable, each preference was standardized (z-scores) within each country before computing the coefficients. Green dots indicate countries in which the gender correlation is not statistically different from zero at the 10% level, while red / blue / pink dots denote countries in which the effect is significant at the 1% / 5% / 10% level, respectively. Positive coefficients imply that women have higher values in the respective preference.

Eastern Europe, Asia, North Africa and Middle East, Sub-Saharan Africa, and Southern America. These figures show that patience and positive reciprocity have hump-shaped age profiles mainly in Western and Neo Europe, but also reveal that age profiles for altruism are variable across world regions. By contrast, the age profiles for risk taking, negative reciprocity, and trust are more universal.[19]

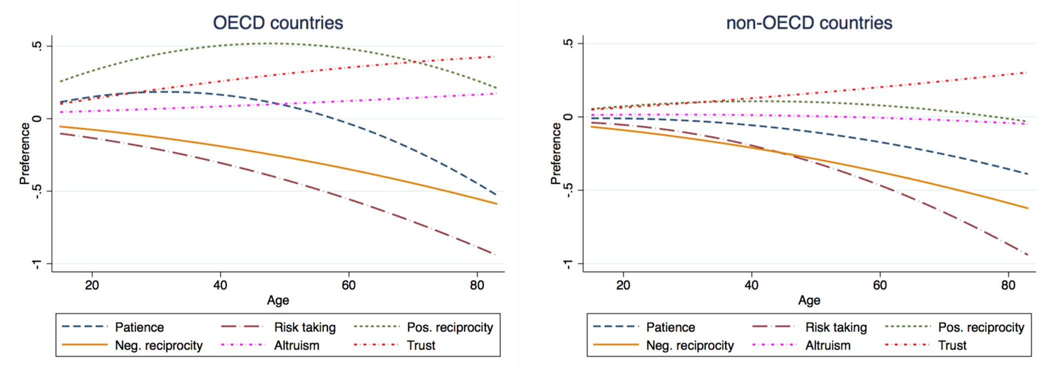


Figure 5.4. Age profiles by OECD membership. The figures depict the relationship between preferences and age conditional on country fixed effects, gender, and subjective math skills. These are augmented component plus residuals plots, in which the vertical axis represents the component of the preference that is predicted by age and its square plus the residuals from the regression in the second column of Table 5.5. The horizontal axis represents age, winsorized at 83 (99th percentile).

Intuitively, the relationship between cognitive ability and preferences might not be as strongly tied to socially constructed roles as gender, or age, and might hence have an even more universal relationship to preferences independent of cultural background. Indeed, we find that self-reported cognitive ability is related to each preference in a strikingly similar way across countries. For every preference, at least 93 percent of countries have the same qualitative relationship between the preference and cognitive ability, for all preferences. Similarly to Figure 5.3, Appendix 5.E.2 shows graphs of the cognitive ability coefficients, by preference and country.

In sum, the findings reveal that there are at least some mechanisms linking preferences to gender, age, and cognitive ability that are universal features of humans, possibly due to biological or psychological mechanisms. At the same time, the results also show a significant amount of variation in the quantitative magnitude and sometimes even the direction of relationships with individual characteristics. Thus, the results provide an important caveat to studies on more specific samples, which sometimes produce highly variable or even contradictory results.

5.4.2 Further Correlates of Preferences

The analysis proceeds by investigating the relationship between preferences on the one hand and income, religious affiliation, physical health, and subjective safety perceptions,

[19] Similar results on the (partially non-linear) age profiles and their relationship to the degree of development obtain when we depict the age profile for all 76 countries separately; these figures are available upon request.

on the other hand. While these characteristics and attitudes are not as exogenous to preferences as age, gender, and cognitive ability, they might nevertheless plausibly affect the formation of preferences.

Health and safety perceptions are evaluated using two indices that Gallup provides by aggregating several survey items. For example, the subjective law and order index includes a question on whether respondents recently had money or property stolen. The subjective health index makes use of an item asking people whether they have health problems that are atypical given their age, among others (see Appendix 5.I for details).

The results are reported in columns (2), (4), (6), (8), (10), and (12) of Table 5.5. The table shows that household income is positively correlated with all preferences, but not with trust. Subjective perceptions of law and order are positively correlated with patience and trust, and negatively associated to negative reciprocity. One possible interpretation is that if people feel safe, they trust others more and are more willing to postpone financial rewards. Negative reciprocity might be less prevalent if an individual believes institutions already provide mechanisms for strong formal sanctions. Previous work finds that retaliatory punishment is more pronounced in countries where people perceive the rule of law as weak (Herrmann et al., 2008).

Subjective physical health perceptions are positively correlated with all traits, except for negative reciprocity. Turning to religious affiliation, the regressions investigate the correlation between six religion dummies (christian, muslim, hinduist, buddhist, jew, other), taking seculars as baseline category. Again note that the religious affiliation coefficients are entirely identified from within-country variation in religious denomination and preferences. We find a very consistent pattern: relative to seculars, christians, muslims, and hinduists are less patient, more risk averse, less negatively reciprocal as well as more altristic and trusting. Similar patterns hold for buddhists. Jews, on the other hand, are more patient, altruistic, and trusting, and have strong negatively reciprocal inclinations. Overall, these results suggest that, relative to atheists, being religious of any denomination tends to be associated with relatively high risk aversion, high values on the prosocial dimensions altruism and trust, and low negative reciprocity (albeit with a few notable exceptions).

5.5 Preferences and Individual Behaviors

We now turn to investigating the relationships of preferences to individual behaviors and outcomes. Understanding the relationship between our preference measures and individual-level economic and social decisions is important in two respects. First, such analyses provide insights into the role of heterogeneity in underlying preference parameters for generating observed choice behavior, on a global scale. Second, it allows us to evaluate the meaningfulness and behavioral relevance of our items in a culturally and economically highly heterogeneous sample.[20]

[20] Throughout this section, the respective dependent variables are sometimes only available for a subset of countries because the respective question was not part of Gallup's core questionnaire.

5.5.1 Accumulation Decisions

We evaluate the explanatory power of the GPS patience measure by relating it to the accumulation of physical and human capital. Table 5.6 presents estimates of OLS regressions of different outcomes on patience. Columns (1) and (2) display the results of a linear probability model, in which we employ as dependent variable a binary indicator for whether the respondent saved in the previous year. Patience is correlated with savings behavior both with and without country fixed effects, and conditional on socioeconomic covariates such as age, gender, income, cognitive ability, and religion. The point estimate implies that a one standard deviation increase in patience is associated with a roughly 20% increase of the probability of saving relative to the baseline probability of 26.7%. Columns (3) and (4) establish that patience is also significantly related to educational attainment; these estimates are based on a three-step categorical variable (roughly: primary, seondary, and tertiary education).[21]

In Appendix 5.F.1, we show that the significant relationship between our patience variable and accumulation processes is not driven by only a few countries. Specifically, by plotting the distribution of point estimates and their significance level across countries, we show that the coefficient of patience is positive in the more than 90% of countries for both savings and education, and mostly statistically significant. For instance, the correlation between patience and education is in most cases statistically significant at least at the 5% level in 74% of all countries.

5.5.2 Risky Choices

To investigate whether risk preferences are related to important risky decisions in life, we build on previous within-country findings, which have found a relationship of risk attitudes to self-employment and health behavior (Dohmen et al., 2011). As columns (5) and (6) of Table 5.6 establish, our preference measure predicts actual self-employment both across and within countries. The same pattern holds when considering individuals' intention to start their own business, conditional on not being self-employed (columns (7)-(8)).

Columns (9) and (10) relate risk preferences to the respondent's smoking intensity, measured on a three-point scale (never, occasionally, and frequently). We find that more risk-tolerant people are more likely to smoke, both with and without country fixed effects, and conditional on a large set of covariates. Appendix 5.F.1 shows that the correlations between risk preferences and labor market or health decisions are not restricted to a particular set of countries. Rather, risk preferences are related to risky behaviors in a qualitatively similar way around the world, although quantitative magnitudes of the relationships do vary. For example risk taking is significantly positively related to planned self-employment at least at the 10% level in about 90% of countries in the sample.

Table 5.6. Patience and accumulation decisions, risk preferences and risky choices

	Dependent variable:									
	Accumulation decisions				Risky choices					
	Saved last year		Education level		Own business		Plan to start business		Smoking intensity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Patience	0.050***	0.025***	0.12***	0.033***						
	(0.01)	(0.01)	(0.02)	(0.00)						
Risk taking					0.031***	0.022***	0.033***	0.017***	0.050***	0.023*
					(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Age		0.025		0.99***		1.55***		0.56***		2.55***
		(0.23)		(0.26)		(0.12)		(0.09)		(0.31)
Age squared		-0.21		-1.83***		-1.54***		-0.69***		-2.86***
		(0.24)		(0.25)		(0.12)		(0.10)		(0.31)
1 if female		-0.0010		-0.016		-0.053***		-0.018***		-0.58***
		(0.01)		(0.01)		(0.01)		(0.00)		(0.03)
Subj. math skills		0.012***		0.041***		0.0056***		0.0030***		-0.011***
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Log [Household income p/c]		0.11***		0.14***		0.020***		-0.0072**		-0.010
		(0.01)		(0.01)		(0.00)		(0.00)		(0.01)
Constant	0.27***	-0.37***	1.87***	0.24***	0.14***	-0.38***	0.11***	0.038	0.38***	0.39***
	(0.03)	(0.09)	(0.03)	(0.07)	(0.01)	(0.05)	(0.01)	(0.03)	(0.03)	(0.07)
Country FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Religion FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	15260	14459	79357	69272	72839	62985	57072	51489	15309	14490
R^2	0.011	0.132	0.030	0.329	0.008	0.104	0.011	0.120	0.005	0.198

OLS estimates, standard errors (clustered at country level) in parentheses. For the purposes of this table, age is divided by 100. Saved last year is a binary indicator, while education level is measured in three categories (roughly elementary, secondary, and tertiary education, see Appendix 5.I). Self-employment and planned self-employment are binary, while smoking intensity is measured in three categories (never, occasionally, frequently). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.7. Social preferences and social interactions

	Dependent variable:													
	Donated money		Volunteered time		Helped stranger		Sent money / goods to other individual		Voiced opinion to official		Have friends / relatives I can count on		In a relationship	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Altruism	0.064***	0.061***	0.046***	0.038***	0.064***	0.052***	0.035***	0.033***	0.028***	0.025***	0.0079	0.016***	0.0048	0.0013
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)
Positive reciprocity	0.0055	-0.00037	-0.000074	0.0049	0.035***	0.033***	0.018***	0.017***	-0.0021	-0.0025	0.015*	0.017***	0.022***	0.0072***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)
Negative reciprocity	0.0040	-0.0042	-0.0033	-0.0035	-0.0026	-0.0032	0.011**	0.0051	0.016***	0.016***	0.012***	0.0037	-0.0034	-0.00074
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age		0.62***		0.42***		0.73***		0.26***		1.02***		-0.90***		5.58***
		(0.08)		(0.08)		(0.07)		(0.07)		(0.08)		(0.09)		(0.16)
Age squared		-0.47***		-0.47***		-0.90***		-0.29***		-1.00***		0.76***		-5.42***
		(0.09)		(0.08)		(0.07)		(0.08)		(0.09)		(0.09)		(0.17)
1 if female		0.013**		-0.017***		-0.014**		-0.00057		-0.045***		0.013***		-0.024***
		(0.01)		(0.01)		(0.01)		(0.00)		(0.01)		(0.00)		(0.01)
Subj. math skills		0.0091***		0.0074***		0.0091***		0.0070***		0.0091***		0.0053***		0.0024**
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Log [Household income p/c]		0.031***		0.0042		0.021***		0.034***		0.012***		0.036***		-0.036***
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Constant	0.32***	-0.022	0.22***	0.15***	0.49***	0.37***	0.24***	-0.17***	0.22***	-0.084**	0.82***	0.48***	0.58***	-0.32***
	(0.02)	(0.04)	(0.01)	(0.05)	(0.02)	(0.06)	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.05)
Country FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Religion FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	58229	53439	58213	53430	55991	53226	56253	53559	55944	53174	65986	59209	77881	68176
R^2	0.020	0.192	0.012	0.089	0.028	0.093	0.012	0.124	0.006	0.062	0.003	0.118	0.002	0.218

OLS estimates, standard errors (clustered at country level) in parentheses. For the purposes of this table, age is divided by 100. See Appendix 5.I for details on all dependent variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.5.3 Social Interactions

We analyze the relationships of the social preference measures to behaviors and outcomes in the social domain.[22] Table 5.7 summarizes the results. Columns (1)-(8) show that altruism is significantly related to a broad range of prosocial behaviors including donating, volunteering time, helping strangers, or sending money or goods to other people in need. Across the different behavioral categories, the point estimate is very consistent and implies that an increase in altruism by one standard deviation is correlated with an increase in the probability of engaging in prosocial activities of 3.5–6.5 percentage points, which corresponds to an increase of roughly 15–20% compared to the respective baseline probabilities.[23] Positive reciprocity is a significant correlate of helping people in need (columns (5) through (8)), perhaps a manifestation of generalized reciprocity in the sense that reciprocal people who have been helped before are also willing to help others. In contrast, the negative reciprocity variable is virtually uncorrelated with all of the prosocial activities in the first eight columns. As columns (9) and (10) show, however, negative reciprocity is a significant predictor of whether people are willing to voice their opinion to a public official.

Columns (11) through (14) examine the relationship between social preferences and respondents' family and friendship status. We find that more altruistic and more positively reciprocal people are more likely to have friends they can count on when in need, and that positive reciprocity correlates with being in a relationship.[24]

The overall pattern in Table 5.7 highlights that preferences are predictive of a wide range of behaviors in the social domain and that the preference measures are of sufficiently high quality to discriminate between different types of social behavior. As Appendix 5.F.1 shows, these relationships are not restricted to a small set of countries, but instead hold for most countries separately. For instance, the correlation between altruism and donating is statistically significant at the 5% level in 80% of all countries.

In sum, all of the GPS preference measures are significantly related to a broad range of economic and social behaviors in a way one would intuitively expect. This indicates that preference heterogeneity is important for understanding variation in economic outcomes worldwide. In addition, the fact that the correlations are qualitatively similar across cultural backgrounds and development levels provides reassuring evidence that the GPS survey items do indeed capture the relevant underlying preferences even in a very heterogeneous sample in terms of cultural background and economic development. In this sense, the correlations provide an important out-of-context validation check for the survey module.

---

21 Appendix 5.F.2 presents robustness checks on all results in this section using (ordered) probit estimations.

22 Since trust constitutes a belief rather than a preference, we do not incorporate it in the discussion here. However, all results are robust to controlling for trust.

23 These baseline probabilities are 31.8%, 21.6%, 48.3%, and 23.7%, repectively (see Table 5.7 for the order of variables).

24 Also see Dohmen et al. (2009).

5.6 The Cultural Origins of Preference Variation

The previous analyses have documented a large amount of preference heterogeneity both across and within countries. An immediate question is whether this variation has deep cultural origins. While heterogeneity in different preferences may have myriad of potentially different historical or cultural sources, we conclude the paper by drilling deeper into the potential cultural origins of preference differences. We focus on one particular proxy for cultural variation, i.e., linguistic features. Language has been used as proxy for cultural variation in many previous studies (e.g., Fearon, 2003; Desmet et al., 2009, 2012; Spolaore and Wacziarg, 2015). In deriving specific testable hypotheses on the relationship between preferences and linguistic or cultural variation, we follow the work of Chen (2013) and investigate the relationship between economic preferences and a structural feature of languages called future-time reference (FTR). As discussed in detail by Chen (2013), languages differ in whether or not they require people to grammatically mark future events: some languages require people to explicitly distinguish between present and future by making use of constructions such as "I *will* go to school tomorrow" (strong FTR), while others allow speakers to talk about the future in present tense (weak FTR). Chen (2013) argues that strong FTR languages may make the future feel more distant, potentially resulting in less future-oriented behavior. In empirically testing this proposition, he develops a binary FTR classification of languages and shows that – both across and within countries – people who speak weak FTR languages save more, are less likely to smoke, and have better health.[25]

Building on this insight, we investigate the relationship between preferences and future-time reference. Our analysis serves two purposes. First, our patience measure allows for a direct replication and extension of Chen's results on future-oriented behavior. In particular, our patience measure arguably constitutes a more fundamental and direct proxy for how people trade off current and future rewards than, e.g., medical obesity. Second, our data allow for a systematic investigation of whether the cultural trait captured by FTR is also related to other preferences rather than just time preference. In particular, it is conceivable that people for whom the future seems less distant are more likely to invest resources today to reap social benefits in the future. Thus, traits that are related to cooperation, repeated interaction, and reputation building should be more pronounced in weak FTR languages. Our data on positive reciprocity, trust, and altruism provide natural candidates for such an investigation. Our analysis proceeds in two steps. First, we investigate the relationship between average preferences and FTR at the country level. Then, we exploit within-country variation in preferences and FTR.

5.6.1 Cross-Country Evidence

To study the relationship between preferences and FTR, we employ Gallup's interview language as a proxy for the language respondents speak in their daily lives.[26] We apply Chen's classification of languages to our dataset, which results in a set of 55 coded lan-

[25] Sutter et al. (2014) show that the same relationship exists for children in a bilingual city in Italy.

[26] Correspondence with Gallup suggests that, naturally, in some countries interview language is an imperfect proxy for the language people are most familiar with. Thus, proxying people's daily language with

guages. In addition, we were able to code an additional four languages ourselves using the methodology outlined in Chen (2013).[27] In sum, we have access to 59 classified languages for a total of 75,224 respondents.[28] All results are robust to only making use of the languages coded by Chen.

After classifying each respondent, we compute the country-level fraction of people whose language corresponds to weak as opposed to strong FTR. Then, we regress average preferences in a given country on this fraction. To take into account that we can classify only a subset of respondents in some countries (making the fraction speaking weak FTR languages a less precise estimate of the true population counterpart), our regressions weigh all observations by the fraction of people whose language can be classified. Thus, countries in which we can classify a larger fraction of respondents receive higher weight, as should be the case from a measurement error perspective.[29]

Table 5.8 presents the results. For each preference, we report two specifications, one without covariates and one with control variables commonly employed in cross-country regressions, i.e., continent fixed effects, (log) per capita income, distance to the equator, longitude, the fraction of the population that is at risk of contracting malaria, and average precipitation. Results show that, across countries, weak FTR is significantly correlated with average patience (columns (1)-(2)). As columns (5)-(6) and (11)-(12) show, similar patterns obtain for positive reciprocity and trust. In contrast, altruism, risk taking, and negative reciprocity are not significantly correlated with the fraction speaking weak FTR languages.

5.6.2 Within-Country Evidence

In a second step of the analysis, we exploit within-country variation in preferences and FTR. Such analyses are arguably better suited to identify cultural origins of preferences because they can account for unobserved heterogeneity at the country level.

In many countries in our sample, we observe some variation in interview languages. However, variation in language does not necessarily mean variation in FTR. In fact, only in Estonia, Nigeria, and Switzerland (2,925 respondents in total) do interview languages vary across respondents such that we observe within-country variation in FTR. Thus, we proceed by regressing individual-level preferences on a dummy for whether a respondent speaks a weak or strong FTR language, conditional on country fixed effects and age, age squared, gender, and our cognitive skills proxy. Columns (1), (3), (5), (7), (9), and (11) of Table 5.9 present the results. Consistent with the cross-country evidence, we find that individuals speaking weak FTR languages are more patient, more positively reciprocal,

their interview language results in measurement error and hence attenuation bias, which works against finding statistically significant effects in our analyses.

[27] These languages are: Fulfulde (weak FTR), Khmer (strong FTR), Moroccan Arabic (weak FTR), and Dari (strong FTR). In addition, we changed one of Chen's classifications after corresponding with him. He classified Persian as strong FTR, while it is in fact weak FTR. None of our results depend on how we code Persian.

[28] We could not classify 23 languages, which are mostly spoken by small minorities (5,113 respondents in total).

[29] Appendix 5.G confirms that virtually identical results are obtained when running unweighted OLS regressions, suggesting that measurement error in the fraction speaking weak FTR languages is weak.

Table 5.8. Preferences and FTR: Cross-country results

	Dependent variable:											
	Patience		Risk taking		Pos. reciprocity		Neg. reciprocity		Altruism		Trust	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Fraction of population speaking weak FTR	0.37***	0.24***	-0.14*	-0.0080	0.13*	0.16**	-0.024	-0.088	0.043	0.099	0.17**	0.18**
	(0.13)	(0.09)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.09)	(0.09)	(0.10)	(0.08)	(0.08)
Log [GDP p/c PPP]		0.15***		0.032		-0.072*		0.058		-0.078*		-0.00072
		(0.04)		(0.03)		(0.04)		(0.04)		(0.04)		(0.03)
Distance to equator		0.010*		0.0017		-0.0069		-0.0081		-0.0022		-0.0072
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)		(0.00)
Longitude		-0.0019		0.0023		0.0022		0.00091		0.0025		-0.000039
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
% at risk of malaria		0.25		-0.14		-0.27		-0.13		-0.71**		-0.16
		(0.21)		(0.23)		(0.29)		(0.17)		(0.27)		(0.19)
Average precipitation		-0.00024		-0.00092		0.00065		-0.0011		0.0031**		-0.0011
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Constant	-0.055	-1.42***	0.034	-0.041	-0.047	1.39***	0.020	-0.13	-0.049	1.27**	-0.047	0.56
	(0.04)	(0.51)	(0.04)	(0.36)	(0.05)	(0.43)	(0.04)	(0.50)	(0.05)	(0.50)	(0.04)	(0.38)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	75	74	75	74	75	74	75	74	75	74	75	74
R^2	0.146	0.636	0.031	0.442	0.022	0.253	0.001	0.271	0.002	0.334	0.053	0.408

WLS estimates, robust standard errors in parentheses. All observations are weighted by the fraction of respondents whose language can be classified as weak or strong FTR. The regressions exclude Haiti for which no respondent could be classified. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and more trusting. In addition, these people are also significantly more altruistic.[30] We do not find significant relationships between FTR and risk taking or negative reciprocity. For each preference, a second column adds further controls, i.e., regional (state or province) fixed effects, religion fixed effects, household income, health, and subjective institutional quality. Despite this comprehensive set of covariates, and only exploiting within-region variation in FTR and preferences, we obtain almost identical results.

In sum, the results at the subnational level closely mirror those obtained in cross-country analyses, the one exception being altruism.[31] Thus, across levels of aggregation, weak FTR is predictive of higher patience and higher levels of the prosocial traits positive reciprocity, altruism, and trust. These results are supportive of the hypothesis that cultural variation might coevolve with preferences. In particular, just as in the case of the work by Chen (2013), our results lend themselves to two interesting interpretations. First, speaking a weak FTR language may actually *cause* patience and cooperation-enhancing prosociality; second, the historical evolution of linguistic features and the formation preferences may both be a product of some other very deep cultural trait. Regardless of the precise interpretation adopted, our results highlight that the contemporary preference variation may have very deep historical roots,[32] and that the GPS data are well-suited to identify such effects.

5.7 Discussion, Applications, and Outlook

Many theories about human behavior in economics and other fields assume that a fundamental set of preferences drives decision-making of individual agents. These include preferences about risk, timing of rewards, and in the social domain, reciprocity, altruism, and trust. Despite their importance, empirical evidence on the extent and nature of preference heterogeneity has been restricted to varying measures available for a limited set of countries, and typically non-representative samples. This paper has presented the first assessment of the distribution and nature of these fundamental traits on a globally representative basis using a novel dataset, which includes behaviorally validated survey measures of preferences. The findings in this paper are clearly only a first step towards tapping the potential of the GPS. The cross-cultural dimension of the data and the representative sampling design allow entirely new perspectives and levels of analysis. We illustrate this by discussing three broad directions for future research: the mechanisms underlying the relationship between preferences and individual characteristics, the deeper causes of cross-country variation in preferences, and the potential consequences of certain country-level preference profiles.

First, the data vastly expand the amount of information available for understanding the relationship between individual-level characteristics and preferences. The precise

[30] When we restrict the sample to those countries with within-country variation in FTR and regress the respective preference only on the FTR indicator as well as country fixed effects, the resulting coefficient is always positive and statistically significat at the 10% level for patience and at the 1% level for positive reciprocity, trust, and altruism. In Appendix 5.G, we report the coefficient on FTR separately for each country in which we observe within-country variation.

[31] Note that the correspondence between within- and across-country results is in no way mechanical, i.e., it need not necessarily be the case that individual- and country-level correlations are aligned.

[32] As discussed by Chen (2013), variation in future-time reference is at least several hundred years old.

Table 5.9. Preferences and FTR: Individual-level results

	Dependent variable:											
	Patience		Risk taking		Pos. reciprocity		Neg. reciprocity		Altruism		Trust	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1 if weak FTR	0.095**	0.053***	0.067	0.0079	0.18***	0.11***	-0.073*	-0.0037	0.24***	0.19**	0.33***	0.32**
	(0.05)	(0.02)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)	(0.09)	(0.04)	(0.09)	(0.04)	(0.13)
Age	0.76***	0.80***	-0.098	0.49**	1.07***	0.92***	-0.39***	-0.25	0.032	0.074	0.42***	0.045
	(0.10)	(0.18)	(0.10)	(0.21)	(0.10)	(0.17)	(0.10)	(0.20)	(0.10)	(0.16)	(0.10)	(0.16)
Age squared	-1.51***	-1.48***	-1.21***	-1.80***	-1.22***	-1.09***	-0.44***	-0.63***	-0.017	-0.21	-0.0044	0.27
	(0.11)	(0.21)	(0.11)	(0.21)	(0.11)	(0.18)	(0.11)	(0.20)	(0.11)	(0.17)	(0.11)	(0.16)
1 if female	-0.061***	-0.040***	-0.18***	-0.17***	0.045***	0.054***	-0.13***	-0.13***	0.100***	0.093***	0.069***	0.053***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Subj. math skills	0.028***	0.023***	0.045***	0.040***	0.039***	0.039***	0.039***	0.036***	0.044***	0.040***	0.056***	0.056***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log [Household income p/c]		0.037***		0.052***		0.049***		0.012		0.046***		-0.0083
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)
Subj. health index		0.10***		0.14***		0.062***		-0.028		0.084***		0.053***
		(0.02)		(0.03)		(0.02)		(0.03)		(0.02)		(0.02)
Subj. law and order index		0.065***		0.054**		0.0041		-0.075***		0.022		0.19***
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
Constant	-0.49***	-0.88***	0.15***	-0.79***	-0.13**	-0.51***	0.60***	0.16	-0.17***	-0.40***	-0.46***	-0.74***
	(0.06)	(0.08)	(0.06)	(0.10)	(0.06)	(0.09)	(0.07)	(0.13)	(0.06)	(0.13)	(0.06)	(0.16)
Country FE	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Region FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Religion FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	73460	52628	73414	52610	73811	52862	72501	52003	73580	52675	72811	52159
R^2	0.166	0.215	0.172	0.254	0.127	0.230	0.117	0.200	0.137	0.199	0.113	0.167

OLS estimates, robust standard errors (clustered at country level) in parentheses. For the purposes of this table, age is divided by 100. See Table 5.5 for the religion categories.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

ways in which the strength and direction of preference differences vary across different environments, locations, and institutions, may shed further light on the mechanisms underlying preference differences. For example, if gender differences in preferences are correlated within countries, this would suggest some deep mechanisms that extend across preference domains.

Second, there is much more that can be done to investigate the ultimate origins of the cross-country variation in preference. While our analysis of the relationship between FTR and preferences has provided a first step in this direction, other cultural proxies or historical events might likewise generate differences in preferences. For example, other linguistic structures, such as politeness in pronoun usage (Helmbrecht, 2003), or the role of gender in the language might be related to preference differences (Corbett, 1991). The correlation structure of preferences may also be informative for understanding the ultimate sources of preference differences. Traits may also coevolve, to the extent that they are complementary in contributing to evolutionary fitness. In this regard, it is suggestive that the groupings of positively correlated preferences that we find are plausibly complementary, in the context of theories about the human ability to sustain cooperation (e.g., high patience and strong negative reciprocity).

A third direction for future research is a detailed investigation of the link between aggregate outcomes and preferences at the country level. Given the previous lack of representative preference data, this is completely uncharted territory. Exploring in detail the many ramifications of preference differences for explaining outcomes is beyond the scope of this paper. However, to illustrate the potential power of the GPS data in understanding cross-country variation in the economic and social domain, we conclude with two examples. First, a tendency to retaliate could exacerbate conflicts, so that negative reciprocity may be relevant for explaining the occurrence of conflict or war. Second, a large body of dynamic theories of comparative development and growth highlight the crucial role of time preference for aggregate accumulation processes. Consistent with such theories, in a follow-up paper, Dohmen et al. (2015), we find that patience is not only predictive of future-oriented decisions and income at the individual level, but also across regions within countries, and even across entire populations: more patient countries have higher savings rates, invest more into education as well as into the stock of ideas and knowledge, grow faster, and are wealthier.

References

Alesina, Alberto, Yann Algan, Pierre Cahuc, and Paola Giuliano (forthcoming): “Family Values and the Regulation of Labor.” *Journal of the European Economic Association*. [203]

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003): “Fractionalization.” *Journal of Economic Growth*, 8 (2), 155–194. [266]

Alesina, Alberto and Paola Giuliano (forthcoming): “Culture and Institutions.” *Journal of Economic Literature*. [203]

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013): “On the Origins of Gender Roles: Women and the Plough.” *Quarterly Journal of Economics*, 128 (2), 469–530. [203]

Algan, Yann and Pierre Cahuc (2010): “Inherited Trust and Growth.” *American Economic Review*, 2060–2092. [203]

Ashraf, Quamrul and Oded Galor (2013): "The Out of Africa Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review*, 103 (1), 1–46. [266]

Barro, Robert (2003): *Religion Adherence Data*. http://scholar.harvard.edu/barro/publications/religion-adherence-data. [267]

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro (1997): "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics*, 112 (2), 537–579. [202, 203, 215]

Beauchamp, Jonathan, David Cesarini, and Magnus Johannesson (2011): "The Psychometric Properties of Measures of Economic Risk Preferences." *Working Paper*. [263]

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro (2013): "Who is' Behavioral'? Cognitive Ability and Anomalous Preferences." *Journal of the European Economic Association*, 11 (6), 1231–1255. [215]

Bonin, Holger, Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde (2007): "Cross-Sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes." *Labour Economics*, 14 (6), 926–937. [203]

Borghans, Lex and Bart H.H. Golsteyn (2006): "Time Discounting and the Body Mass Index: Evidence from the Netherlands." *Economics & Human Biology*, 4 (1), 39–61. [203]

Botero, Juan, Simeon Djankov, Rafael LaPorta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2004): "The Regulation of Labor." *Quarterly Journal of Economics*, 119 (4), 1339–1382. [267]

Carpenter, Jeffrey and Erika Seki (2011): "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay." *Economic Inquiry*, 49 (2), 612–630. [203]

Chabris, Christopher F., David Laibson, Carrie L. Morris, Jonathan P. Schuldt, and Dmitry Taubinsky (2008): "Individual Laboratory-Measured Discount Rates Predict Field Behavior." *Journal of Risk and Uncertainty*, 37 (2-3), 237–269. [203]

Chen, M. Keith (2013): "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets." *American Economic Review*, 103 (2), 690–731. [203, 225, 226, 228]

Corbett, Greville G. (1991): *Gender*. Cambridge University Press. [230]

Cornsweet, Tom N. (1962): "The Staircase-Method in Psychophysics." *American Journal of Psychology*, 75 (3), 485–491. [206]

Croson, Rachel and Uri Gneezy (2009): "Gender Differences in Preferences." *Journal of Economic Literature*, 47 (2), 448–474. [202, 215]

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg (2012): "The Political Economy of Linguistic Cleavages." *Journal of development Economics*, 97 (2), 322–338. [225]

Desmet, Klaus, Shlomo Weber, and Ignacio Ortuño-Ortín (2009): "Linguistic Diversity and Redistribution." *Journal of the European Economic Association*, 7 (6), 1291–1318. [225]

Dohmen, Thomas, Benjamin Enke, Armin Falk, David Huffman, and Uwe Sunde (2015): "Patience and the Wealth of Nations." *Working Paper*. [230]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2008): "Representative Trust and Reciprocity: Prevalence and Determinants." *Economic Inquiry*, 46 (1), 84–90. [202]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2009): "Homo Reciprocans: Survey Evidence on Behavioural Outcomes." *Economic Journal*, 119 (536), 592–612. [203, 224]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, 100 (3), 1238–1260. [202, 215]

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011): "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association*, 9 (3), 522–550. [202, 203, 207, 215, 221]

Donkers, Bas, Bertrand Melenberg, and Arthur Van Soest (2001): "Estimating Risk Attitudes Using Lotteries: A Large Sample Approach." *Journal of Risk and Uncertainty*, 22 (2), 165–195. [215]

Eckel, Catherine, Cathleen Johnson, and Claude Montmarquette (2005): "Savings Decisions of the Working Poor: Short- and Long-Term Horizons." In *Research in Experimental Economics. Vol. 10: Field Experiments in Economics*. Ed. by Jeff Carpenter, Glenn Harrison, and John List. Amsterdam: Elsevier, 219–260. [203]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde (2015): "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Working Paper*. [201, 204, 205, 208, 234, 242]

Fearon, James D (2003): "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, 8 (2), 195–222. [210, 225, 267]

Fernández, Raquel (2011): "Does Culture Matter?" In *Handbook of Social Economics*. Ed. by Jesse Benhabib, Matthew O. Jackson, and Alberto Bisin. North Holland. [203]

Fernández, Raquel and Alessandra Fogli (2009): "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics*, 1 (1), 146–177. [203]

Frederick, Shane (2005): "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives*, 19 (4), 25–42. [202, 215]

Gallup, John L., Jeffrey D. Sachs, et al. (2000): "The Economic Burden of Malaria." *American Journal of Tropical Medicine and Hygiene*, 64 (1-2 Suppl), 85–96. [266]

Galor, Oded and Ömer Özak (2014): "The Agricultural Origins of Time Preference." *Working Paper*. [203]

Giuliano, Paola (2007): "Living Arrangements in Western Europe: Does Cultural Origin Matter?" *Journal of the European Economic Association*, 5 (5), 927–952. [203]

Giuliano, Paola and Nathan Nunn (2013): "The Transmission of Democracy: From the Village to the Nation-State." *American Economic Review: Papers & Proceedings*, 103 (3), 86. [203]

Golsteyn, Bart H.H., Hans Grönqvist, and Lena Lindahl (2014): "Adolescent Time Preferences Predict Lifetime Outcomes." *Economic Journal*, 120 (580). [203]

Gorodnichenko, Yuriy and Gerard Roland (2011): "Individualism, Innovation, and Long-Run Growth." *Proceedings of the National Academy of Sciences*, 108 (Supplement 4), 21316–21319. [203]

Guiso, Luigi and Monica Paiella (2008): "Risk Aversion, Wealth, and Background Risk." *Journal of the European Economic Association*, 6 (6), 1109–1150. [203]

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2006): "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20 (2), 23–48. [203]

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2009): "Cultural Biases in Economic Exchange?" *Quarterly Journal of Economics*, 124 (3), 1095–1131. [203]

Helmbrecht, Johannes (2003): "Politeness Distinctions in Second Person Pronouns." *Deictic Conceptualisation of Space, Time and Person*, 112, 168. [230]

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath (2001): "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review*, 91 (2), 73–78. [203]

Herrmann, Benedikt, Christian Thöni, and Simon Gächter (2008): "Antisocial Punishment Across Societies." *Science*, 319 (5868), 1362–1367. [203, 210, 220]

Kirby, Kris N. and Nancy M. Petry (2004): "Heroin and Cocaine Abusers Have Higher Discount Rates for Delayed Rewards than Alcoholics or Non-Drug-Using Controls." *Addiction*, 99 (4), 461–471. [203]

Knack, Stephen and Philip Keefer (1997): "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics*, 1251–1288. [203]

Kosfeld, Michael and Devesh Rustagi (2015): "Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia." *American Economic Review*, 105 (2), 747–783. [203]

Meier, Stephan and Charles Sprenger (2010): "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics*, 193–210. [203]

Michalopoulos, Stelios (2012): "The Origins of Ethnolinguistic Diversity." *American Economic Review*, 102 (4), 1508–1539. [265]

Nordhaus, William D. (2006): "Geography and Macroeconomics: New Data and New Findings." *Proceedings of the National Academy of Sciences of the United States of America*, 103 (10), 3510–3517. [266]

Rieger, Marc-Oliver, Mei Wang, and Thorsten Hens (forthcoming): "Risk Preferences Around the World." *Management Science*. [204]

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991): "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *American Economic Review*, 81 (5), 1068–1095. [203]

Rustagi, Devesh, Stefanie Engel, and Michael Kosfeld (2010): "Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management." *Science*, 330 (6006), 961–965. [203]

Spinath, Birgit, Frank M. Spinath, Nicole Harlaar, and Robert Plomin (2006): "Predicting School Achievement From General Cognitive Ability, Self-Perceived Ability, and Intrinsic Value." *Intelligence*, 34 (4), 363–374. [215]

Spolaore, Enrico and Romain Wacziarg (2015): "Ancestry, Language and Culture." *Working Paper*. [225]

Sutter, Matthias, Silvia Angerer, Daniela Glätzle-Rützler, and Philipp Lergetporer (2014): "The Effects of Language on Children's Intertemporal Choices." *Working Paper*. [225]

Sutter, Matthias and Martin G. Kocher (2007): "Trust and Trustworthiness Across Different Age Groups." *Games and Economic Behavior*, 59 (2), 364–382. [215]

Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rützler, and Stefan T. Trautmann (2013): "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *American Economic Review*, 103 (1), 510–531. [203]

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen (2010): "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review*, 100 (1), 557–571. [203]

Ventura, Luigi (2003): "Direct Measures of Time Preference." *Economic and Social Review*, 34 (3), 293–310. [203]

Vieider, Ferdinand M., Thorsten Chmura, Tyler Fisher, Takao Kusakawa, Peter Martinsson, Frauke Mattison-Thompson, and Adewara Sunday (2015): "Within- versus Between-Country Differences in Risk Attitudes: Implications for Cultural Comparisons." *Theory and Decision*, 78 (2), 209–218. [204]

Vieider, Ferdinand M., Mathieu Lefebvre, Ranoua Bouchouicha, Thorsten Chmura, Rustamdjan Hakimov, Michal Krawczyk, and Peter Martinsson (2014): "Common Components of Risk and Uncertainty Attitudes Across Contexts and Domains: Evidence from 30 Countries." *Journal of the European Economic Association*, 13 (3). [204, 207]

Wang, Stephanie W. (2011): "Incentive Effects: The Case of Belief Elicitation from Individuals in Groups." *Economics Letters*, 111 (1), 30–33. [204]

Appendix 5.A Global Preference Survey

5.A.1 Overview

The cross-country dataset measuring risk aversion, patience, positive and negative reciprocity, altruism, and trust, was collected through the professional infrastructure of the Gallup World Poll 2012. The data collection process consisted of three steps. First, an experimental validation procedure was conducted to select the survey items. Second, there was a pre-test of the selected survey items in a variety of countries to ensure implementability in a culturally diverse sample. Third, the final data set was collected through the regular professional data collection efforts in the framework of the World Poll 2012.

5.A.2 Experimental Validation

To ensure the behavioral validity of the preference measures, all underlying survey items were selected through an experimental validation procedure (see Falk et al. (2015) for details). To this end, a sample of 409 German undergraduates completed standard state-of-the-art financially incentivized laboratory experiments designed to measure risk aversion, patience, positive and negative reciprocity, altruism, and trust. The same sample of subjects then completed a large battery of potential survey items. In a final step, for each preference, those survey items were selected which jointly performed best in explaining the behavior under real incentives observed in the choice experiments.

5.A.3 Pre-Test and Adjustment of Survey Items

Prior to including the preference module in the Gallup World Poll 2012, it was tested in the field as part of the World Poll 2012 pre-test, which was conducted at the end of 2011 in 22 countries. The main goal of the pre-test was to receive feedback on each item from various cultural backgrounds in order to assess potential difficulties in understanding and differences in the respondents' interpretation of items. Based on respondents' feedback and suggestions, minor modifications were made to several items before running the survey as part of the World Poll 2012.

The pre-test was run in 10 countries in central Asia (Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Russia, Tajikistan, Turkmenistan, Uzbekistan) 2 countries in South-East Asia (Bangladesh and Cambodia), 5 countries in Southern and Eastern Europe (Croatia, Hungary, Poland, Romania, Turkey), 4 countries in the Middle East and North Africa (Algeria, Jordan, Lebanon, and Saudi-Arabia), and 1 country in Eastern Africa (Kenya). In each country, the sample size was 10 to 15 people. Overall, more than 220 interviews were conducted. In most countries, the sample was mixed in terms of gender, age, educational background, and area of residence (urban / rural).

Participants in the pre-test were asked to state any difficulties in understanding the items and to rephrase the meaning of items in their own words. If they encountered difficulties in understanding or interpreting items, respondents were asked to make suggestions on how to modify the wording of the item in order to attain the desired meaning.

Overall, the understanding of both the qualitative items and the quantitative items was satisfactory. In particular, no interviewer received any complaints regarding difficulties in assessing the quantitative questions or understanding the meaning of the probability used in the hypothetical risky choice items. When asked about rephrasing the qualitative items in their own words, most participants seemed to have understood the items in exactly the way that was intended. Nevertheless, some (sub-groups of) participants suggested adjustments to the wording of some items. This resulted in minor changes to four items, relative to the "original" experimentally validated items:

1. The use of the term "lottery" in hypothetical risky choices was troubling to some Muslim participants. As a consequence, we dropped the term "lottery" and replaced it with "draw".

2. The term "charity" caused confusion in Eastern Europe and Central Asia, so it was replaced it with "good cause".

3. Some respondents asked for a clarification of the question asking about one's willingness to punish unfair behavior. This feedback lead to splitting the question into two separate items, one item asking for one's willingness to punish unfair behavior towards others, and another asking for one's willingness to punish unfair behavior towards oneself.

4. When asked about hypothetical choices between monetary amounts today versus larger amounts one year later, some participants, especially in countries with current or relatively recent phases of volatile and high inflation rates, stated that their answer would depend on the rate of inflation, or said that they would always take the immediate payment due to uncertainty with respect to future inflation. Therefore, we decided to add the following phrase to each question involving hypothetical choices between immediate and future monetary amounts: "Please assume there is no inflation, i.e., future prices are the same as today's prices."

5.A.4 Selection of Countries

The goal when selecting countries was to ensure representative coverage of the global population. Thus, countries from each continent and each region within continents were chosen. Another goal was to maximize variation with respect to observables, such as GDP per capita, language, historical and political characteristics, or geographical location and climatic conditions. Accordingly, the selection process favored non-neighboring and culturally dissimilar countries. This procedure resulted in the following sample of 76 countries:
East Asia and Pacific: Australia, Cambodia, China, Indonesia, Japan, Philippines, South Korea, Thailand, Vietnam
Europe and Central Asia: Austria, Bosnia and Herzegovina, Croatia, Czech Republic, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Italy, Kazakhstan, Lithuania, Moldova, Netherlands, Poland, Portugal, Romania, Russia, Serbia, Spain, Sweden, Switzerland, Turkey, Ukraine, United Kingdom
Latin America and Caribbean: Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Guatemala, Haiti, Mexico, Nicaragua, Peru, Suriname, Venezuela

Middle East and North Africa: Algeria, Egypt, Iran, Iraq, Israel, Jordan, Morocco, Saudi Arabia, United Arab Emirates
North America: United States, Canada
South Asia: Afghanistan, Bangladesh, India, Pakistan, Sri Lanka
Sub-Saharan Africa: Botswana, Cameroon, Ghana, Kenya, Malawi, Nigeria, Rwanda, South Africa, Tanzania, Uganda, Zimbabwe

5.A.5 Sampling and Survey Implementation

5.A.5.1 Background

Since 2005, the international polling company Gallup has conducted an annual World Poll, in which it surveys representative population samples in almost every country around the world on, e.g., economic, social, political, and environmental issues. The collection of our preference data was embedded into the regular World Poll 2012 and hence made use of the pre-existing polling infrastructure of one of the largest professional polling institutes in the world.[33]

Selecting Primary Sampling Units
In countries in which face-to-face interviews are conducted, the first stage of sampling is the identification of primary sampling units (PSUs), consisting of clusters of households. PSUs are stratified by population size and / or geography and clustering is achieved through one or more stages of sampling. Where population information is available, sample selection is based on probabilities proportional to population size. If population information is not available, Gallup uses simple random sampling.

In countries in which telephone interviews are conducted, Gallup uses a random-digit-dialing method or a nationally representative list of phone numbers. In countries with high mobile phone penetration, Gallup uses a dual sampling frame.

Selecting Households and Respondents
Gallup uses random route procedures to select sampled households. Unless an outright refusal to participate occurs, interviewers make up to three attempts to survey the sampled household. To increase the probability of contact and completion, interviewers make attempts at different times of the day, and when possible, on different days. If the interviewer cannot obtain an interview at the initially sampled household, he or she uses a simple substitution method.

In face-to-face and telephone methodologies, random respondent selection is achieved by using either the latest birthday or Kish grid methods.[34] In a few Middle East and Asian countries, gender-matched interviewing is required, and probability

[33] See http://www.gallup.com/strategicconsulting/156923/worldwide-research-methodology.aspx
[34] The latest birthday method means that the person living in the household whose birthday among all persons in the household was the most recent (and who is older than 15) is selected for interviewing. With the Kish grid method, the interviewer selects the participants within a household by using a table of random numbers. The interviewer will determine which random number to use by looking at, e.g., how many households he or she has contacted so far (e.g., household no. 8) and how many people live in the household (e.g., 3 people, aged 17, 34, and 36). For instance, if the corresponding number in the table is 7, he or she will interview the person aged 17.

sampling with quotas is implemented during the final stage of selection. Gallup implements quality control procedures to validate the selection of correct samples and that the correct person is randomly selected in each household.

Sampling Weights
Ex post, data weighting is used to ensure a nationally representative sample for each country and is intended to be used for calculations within a country. These sampling weights are provided by Gallup. First, base sampling weights are constructed to account for geographic oversamples, household size, and other selection probabilities. Second, post-stratification weights are constructed. Population statistics are used to weight the data by gender, age, and, where reliable data are available, education or socioeconomic status.

5.A.5.2 Translation of Items

The items of the preference module were translated into the major languages of each target country. The translation process involved three steps. As a first step, a translator suggested an English, Spanish or French version of a German item, depending on the region. A second translator, being proficient in both the target language and in English, French, or Spanish, then translated the item into the target language. Finally, a third translator would review the item in the target language and translate it back into the original language. If differences between the original item and the back-translated item occurred, the process was adjusted and repeated until all translators agreed on a final version.

5.A.5.3 Adjustment of Monetary Amounts in Quantitative Items

All items involving hypothetical monetary amounts were adjusted for each country in terms of their real value. Monetary amounts were calculated to represent the same share of a country's median income in local currency as the share of the amount in Euro of the German median income since the validation study had been conducted in Germany. Monetary amounts used in the validation study with the German sample were "round" numbers to facilitate easy calculations (e.g., the expected return of a lottery with equal chances of winning and losing) and to allow for easy comparisons (e.g., 100 Euro today versus 107.50 in 12 months). To proceed in a similar way in all countries, monetary amounts were always rounded to the next "round" number. For example, in the quantitative items involving choices between a lottery and varying safe options, the value of the lottery was adjusted to a round number. The varying safe options were then adjusted proportionally as in the original version. While this necessarily resulted in some (very minor) variations in the real stake size between countries, it minimized cross-country differences in the understanding the quantitative items due to difficulties in assessing the involved monetary amounts.

5.A.6 Wording of Survey Items

In the following, "willingness to act" indicates the following introduction: *We now ask for your willingness to act in a certain way in four different areas. Please again indicate your answer on a scale from 0 to 10, where 0 means you are "completely unwilling to do so" and a 10 means you are "very willing to do so". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.*

Similarly, "self-assessments" indicate that the respective statement was preceded by the following introduction: *How well do the following statements describe you as a person? Please indicate your answer on a scale from 0 to 10. A 0 means "does not describe me at all" and a 10 means "describes me perfectly". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.*

5.A.6.1 Patience

1. (Sequence of five interdependent quantitative questions:) *Suppose you were given the choice between receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which you would choose. Please assume there is no inflation, i.e, future prices are the same as today's prices. Please consider the following: Would you rather receive 100 Euro today or x Euro in 12 months?*

   The precise sequence of questions was given by the "tree" logic in Figure S5.5.

2. (Willingness to act:) *How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?*

5.A.6.2 Risk Taking

1. *(Similar to self-assessment:) Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.*

2. (Sequence of five interdependent quantitative questions:) *Please imagine the following situation. You can choose between a sure payment of a particular amount of money, or a draw, where you would have an equal chance of getting amount x or getting nothing. We will present to you five different situations. What would you prefer: a draw with a 50 percent chance of receiving amount x, and the same 50 percent chance of receiving nothing, or the amount of y as a sure payment?* The precise sequence of questions was given by the "tree" logic in Figure S5.6.

5.A.6.3 Positive Reciprocity

1. (Self-assessment:) *When someone does me a favor I am willing to return it.*

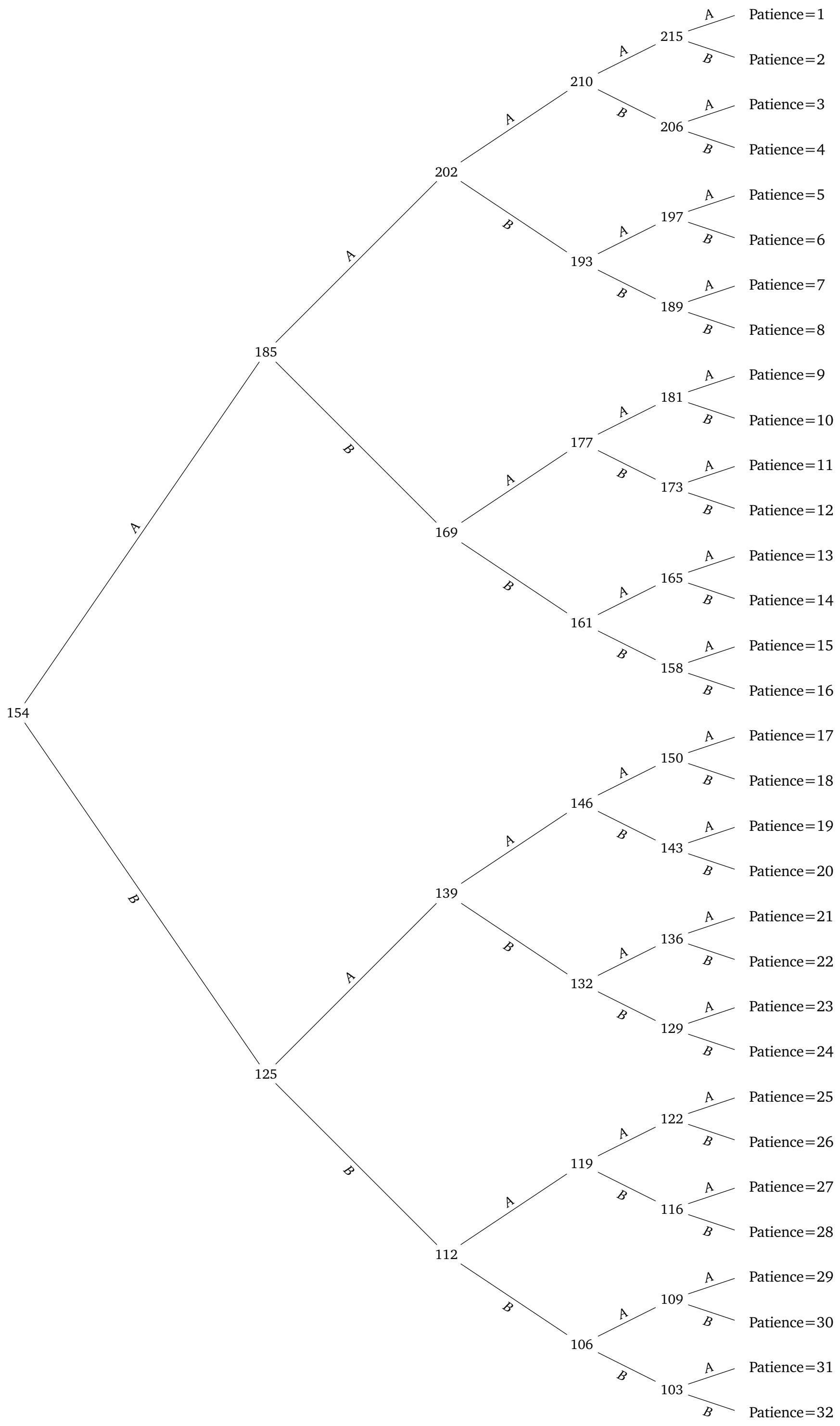


Figure 5.5. Tree for the staircase time task (numbers = payment in 12 months, A = choice of "100 euros today", B = choice of "x euros in 12 months". The staircase procedure worked as follows. First, each respondent was asked whether they would prefer to receive 100 euros today or 154 euros in 12 months from now (leftmost decision node). In case the respondent opted for the payment today ("A"), in the second question the payment in 12 months was adjusted upwards to 185 euros. If, on the other hand, the respondent chose the payment in 12 months, the corresponding payment was adjusted down to 125 euros. Working further through the tree follows the same logic.

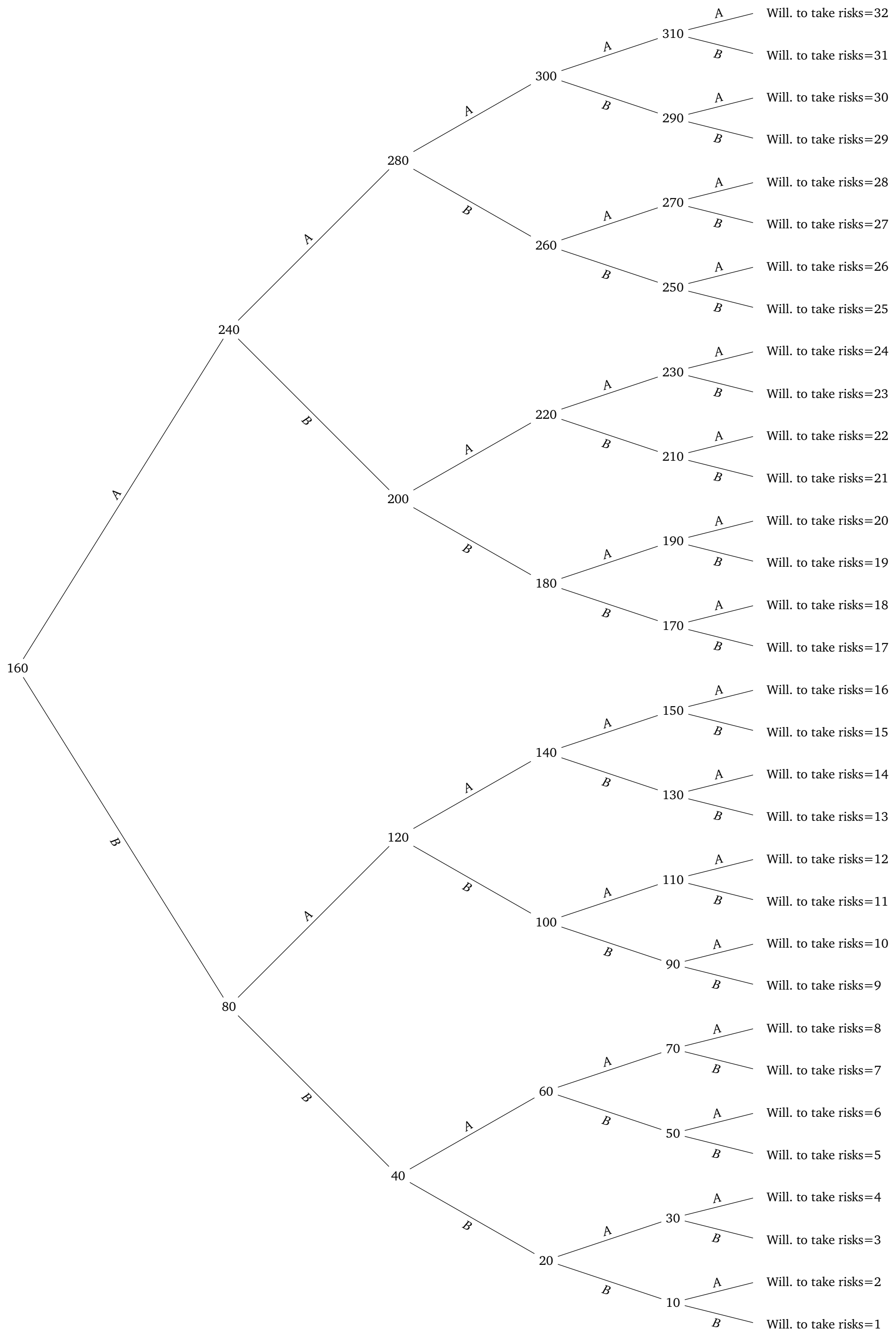


Figure 5.6. Tree for the staircase risk task (numbers = sure payment, A = choice of sure payment, B = choice of lottery). The staircase procedure worked as follows. First, each respondent was asked whether they would prefer to receive 160 euros for sure or whether they preferred a 50:50 chance of receiving 300 euros or nothing. In case the respondent opted for the safe choice (“B”), the safe amount of money being offered in the second question decreased to 80 euros. If, on the other hand, the respondent opted for the gamble (“A”), the safe amount was increased to 240 euros. Working further through the tree follows the same logic.

2. *(Hypothetical situation:) Please think about what you would do in the following situation. You are in an area you are not familiar with, and you realize you lost your way. You ask a stranger for directions. The stranger offers to take you to your destination. Helping you costs the stranger about 20 Euro in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs 5 Euro, the most expensive one costs 30 Euro. Do you give one of the presents to the stranger as a "thank-you"-gift? If so, which present do you give to the stranger? No present / The present worth 5 / 10 / 15 / 20 / 25 / 30 Euro.*

5.A.6.4 Negative Reciprocity

1. (Self-assessment:) *If I am treated very unjustly, I will take revenge at the first occasion, even if there is a cost to do so.*
2. (Willingness to act:) *How willing are you to punish someone who treats you unfairly, even if there may be costs for you?*
3. (Willingness to act:) *How willing are you to punish someone who treats others unfairly, even if there may be costs for you?*

5.A.6.5 Altruism

1. *(Hypothetical situation:) Imagine the following situation: Today you unexpectedly received 1,000 Euro. How much of this amount would you donate to a good cause? (Values between 0 and 1000 are allowed.)*
2. (Willingness to act:) *How willing are you to give to good causes without expecting anything in return?*

5.A.6.6 Trust

(Self-assessment:) *I assume that people have only the best intentions.*

5.A.7 Computation of Preference Measures

5.A.7.1 Cleaning and Imputation of Missings

In order to efficiently use all available information in our data, missing survey items were imputed based on the following procedure:

- If one (or more) survey items for a given preference were missing, then the missing items were predicted using the responses to the available items. The procedure was as follows:
  - Suppose the preference was measured using two items, call them a and b. For those observations with missing information on a, the procedure was to predict its value based on the answer to b and its relationship to a, which was estimated by regressing b on a for the sub-sample of subjects who had nonmissing information on both, a and b (on the world sample).

- For the unfolding-brackets time and risk items, the imputation procedure was similar, but made additional use of the informational content of the responses of participants who started but did not finish the sequence of the five questions. Again suppose that the preference is measured using two items and suppose that a (the staircase measure) is missing. If the respondent did not even start the staircase procedure, then imputation was done using the methodology described above. On the other hand, if the respondent answered between one and four of the staircase questions, a was predicted using a different procedure. Suppose the respondent answered four items such that his final staircase outcome would have to be either x or y. A probit was run of the "x vs. y" decision on b, and the corresponding coefficients were used to predict the decision for all missings (note that this constitutes a predicted probability). The expected staircase outcome was then obtained by applying the predicted probabilities to the respective staircase endpoints, i.e., in this case x and y. If the respondent answered three (or less) questions, the same procedure was applied, the only difference being that in this case the obtained predicted probabilities were applied to the expected values of the staircase outcome conditional on reaching the respective node. Put differently, the procedure outlined above was applied recursively by working backwards through the "tree" logic of the staircase procedure, resulting in an expected value for the outcome node.
- If all survey items for a given preference were missing, then no imputation took place.

- Across the 12 survey items, between 0% and 8% of all responses had to be imputed.

5.A.7.2 Computation of Preference Indices at the Individual Level

For each of the traits (risk preferences, time preferences, positive reciprocity, negative reciprocity, altruism, and trust), an individual-level index was computed that aggregated responses across different survey items. Each of these indices was computed by (i) computing the z-scores of each survey item at the individual level and (ii) weighing these z-scores using the weights resulting from the experimental validation procedure of Falk et al. (2015). Formally, these weights are given by the coefficients of an OLS regression of observed behavior in the experimental validation study on responses to the respective survey items, such that the weights sum to one. In practice, for almost all preferences, the coefficients assign roughly equal weight to all corresponding survey items. The weights are given by:

- Patience:

$$\text{Patience} = 0.7115185 \times \text{ Staircase time } + 0.2884815 \times \text{ Will. to give up sth. today}$$

- Risk taking:

$$\text{Risk taking} = 0.4729985 \times \text{ Staircase risk } + 0.5270015 \times \text{ Will. to take risks}$$

- Positive reciprocity:

$$\text{Pos. reciprocity} = 0.4847038 \times \text{Will. to return favor} + 0.5152962 \times \text{Size of gift}$$

- Negative reciprocity:

$$\begin{aligned}\text{Neg. reciprocity} &= 0.5261938/2 \times \text{Will. to punish if oneself treated unfairly} \\ &+ 0.5261938/2 \times \text{Will. to punish if other treated unfairly} \\ &+ 0.3738062 \times \text{Will. to take revenge}\end{aligned}$$

  As explained above, in the course of the pre-test, the negative reciprocity survey item asking people for their willingness to punish others was split up into two questions, one asking for the willingness to punish if oneself was treated unfairly and one asking for the willingness to punish if someone was treated unfairly. In order to apply the weighting procedure from the validation procedure to these items, the weight of the original item was divided by two and these modified weights were assigned to the new questions.

- Altruism:

$$\text{Altruism} = 0.5350048 \times \text{Will. to give to good causes} + 0.4649952 \times \text{Hypoth. donation}$$

- Trust: The survey included only one corresponding item.

5.A.7.3 Computation of Country Averages

In order to compute country-level averages, individual-level data were weighted with the sampling weights provided by Gallup, see Section S 1.5.1. These sampling weights ensure that our measures correctly represent the population at the country level.

5.A.8 Histograms by Preference

5.A.8.1 Individual Level

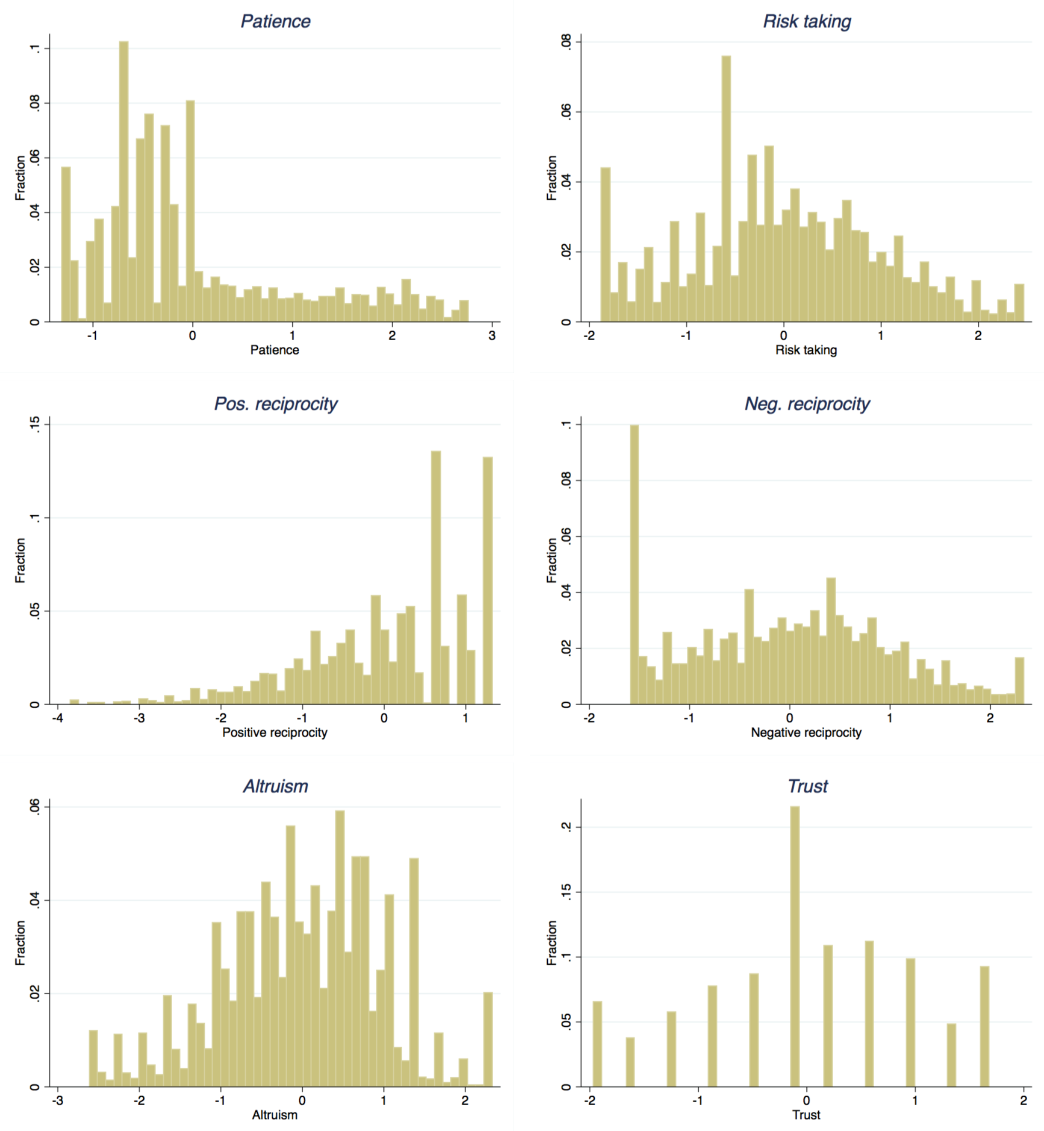


Figure 5.7. Distribution of preferences at individual level. The figure plots the distribution of standardized preference measures at the individual level. All data are standardized at the level of the individual in the full sample.

5.A.8.2 Country Level

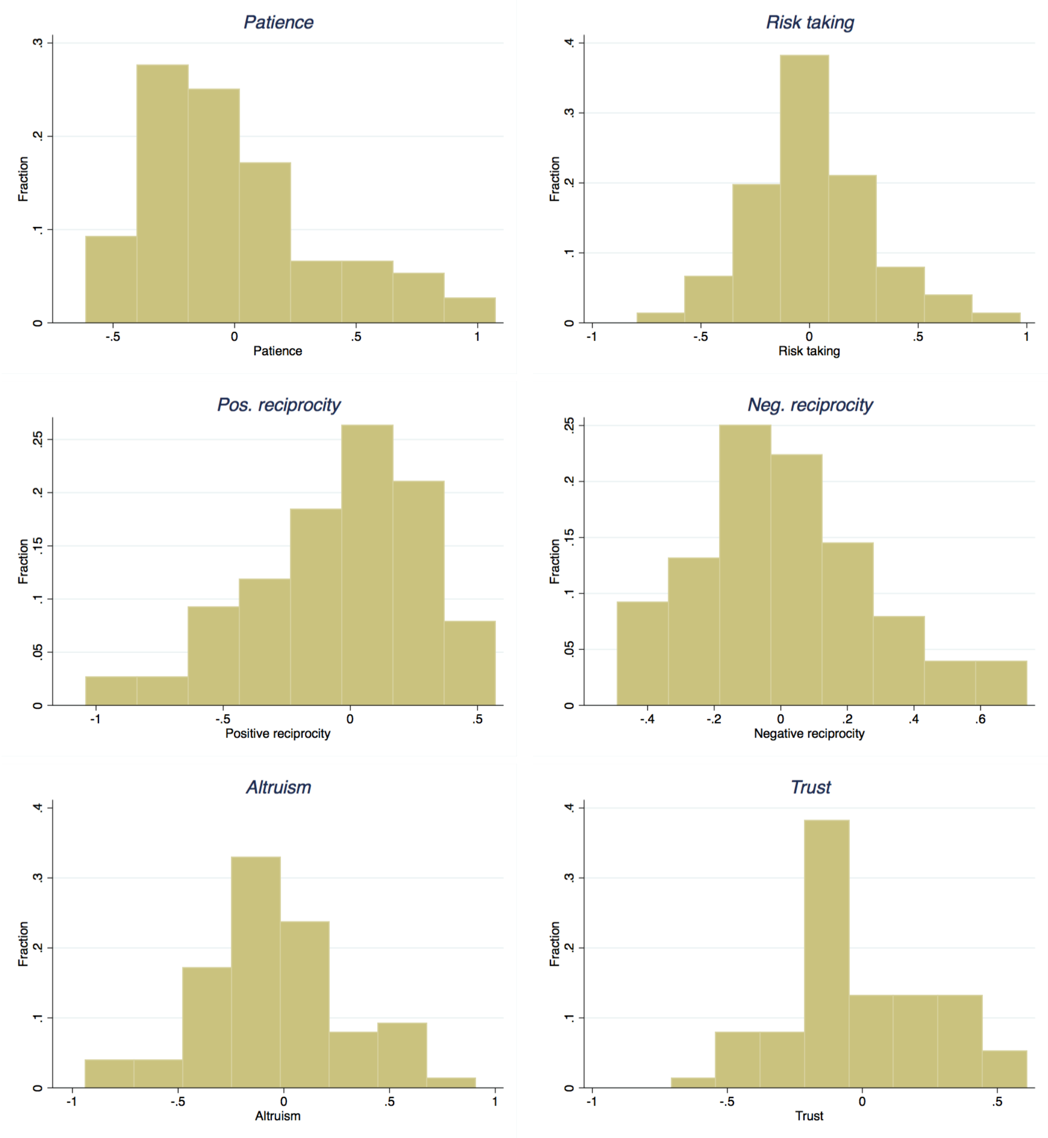


Figure 5.8. Distribution of preferences at country level. The figure plots the distribution of country averages of standardized preferences. All data are standardized at the level of the individual using the full sample.

Appendix 5.B Correlations Among Individual-Level Preferences

Table 5.10 reports the correlation structure among preferences at the individual level. The correlations are computed conditional on country fixed effects to ensure that level differences in preferences across countries do not spuriously generate the results. At the same time, the correlation structure without country fixed effects is quantitatively very similar and is available upon request.

Table 5.10. Partial correlations between preferences at individual level conditional on country fixed effects

	Patience	Risk taking	Pos. reciprocity	Neg. reciprocity	Altruism	Trust
Patience	1					
Risk taking	0.210^{***}	1				
Pos. reciprocity	0.084^{***}	0.068^{***}	1			
Neg. reciprocity	0.112^{***}	0.228^{***}	0.010^{***}	1		
Altruism	0.098^{***}	0.106^{***}	0.329^{***}	0.067^{***}	1	
Trust	0.044^{***}	0.047^{***}	0.114^{***}	0.075^{***}	0.151^{***}	1

Notes. Pairwise partial correlations between preferences at individual level, conditional on country fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The next step in the analysis shows that the significant individual-level correlations among preferences in the world sample are not driven by a few outlier countries only. To this end, Table 5.11 shows the number of countries in which each pair of preferences is significantly correlated at the 1% level. The results show that in most cases the correlations are significant in a large fraction of the 76 countries.

Table 5.11. Number of countries in which preferences are significantly correlated

	Patience	Risk taking	Pos. reciprocity	Neg. reciprocity	Altruism	Trust
Patience						
Risk taking	71					
Pos. reciprocity	40	30				
Neg. reciprocity	53	73	19			
Altruism	47	50	76	32		
Trust	21	24	54	37	62	

Notes. Number of countries for which a given pair of preferences is significantly correlated at the 1% level.

Appendix 5.C Scatter Plots of Preferences by World Region

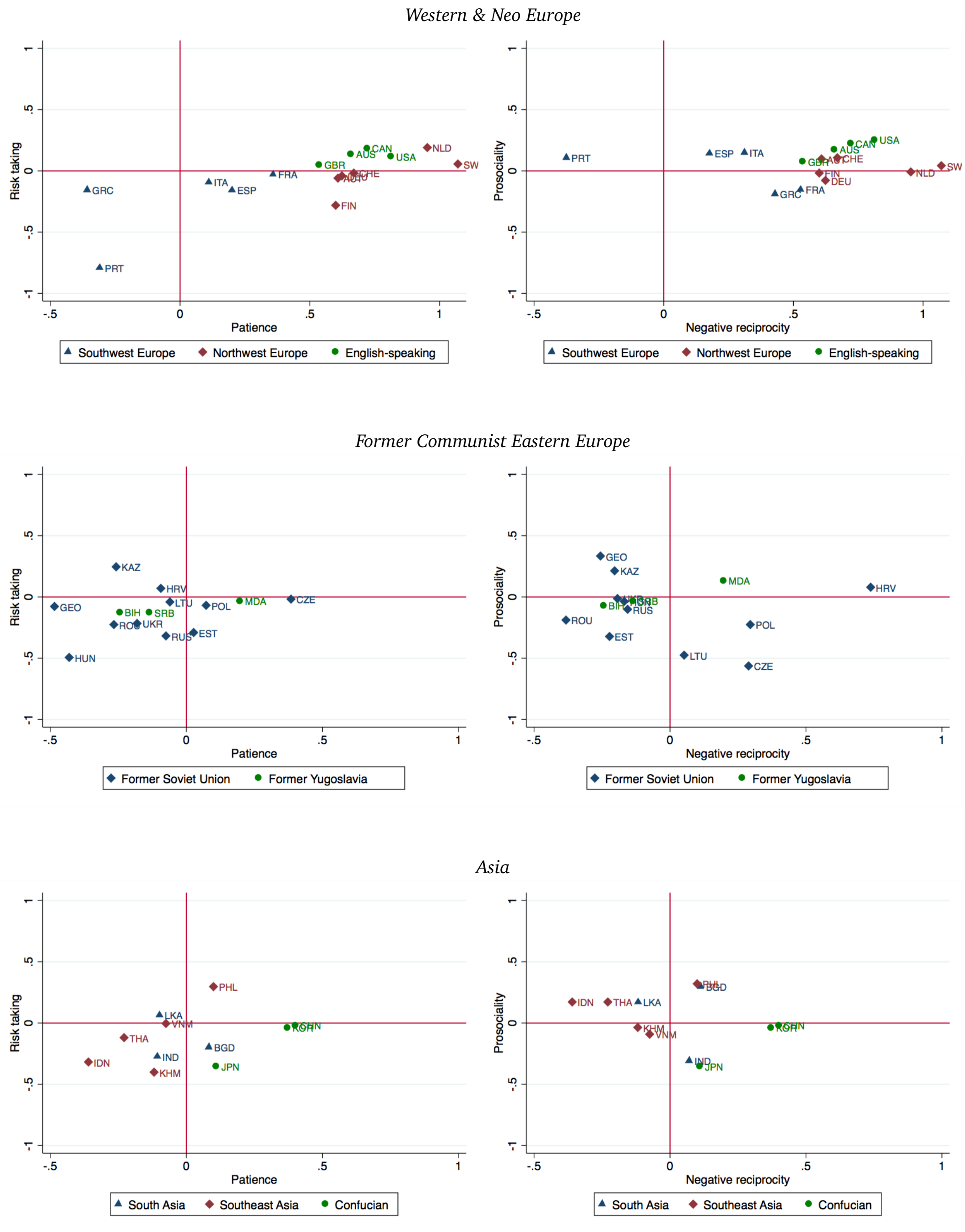


Figure 5.9a. Risk, time, and social preferences by world region (1/2). Each subpanel (row) plots risk taking, patience, negative reciprocity, and prosociality of all countries within a given world region. The prosociality score is computed as the average of altruism, positive reciprocity, and trust.

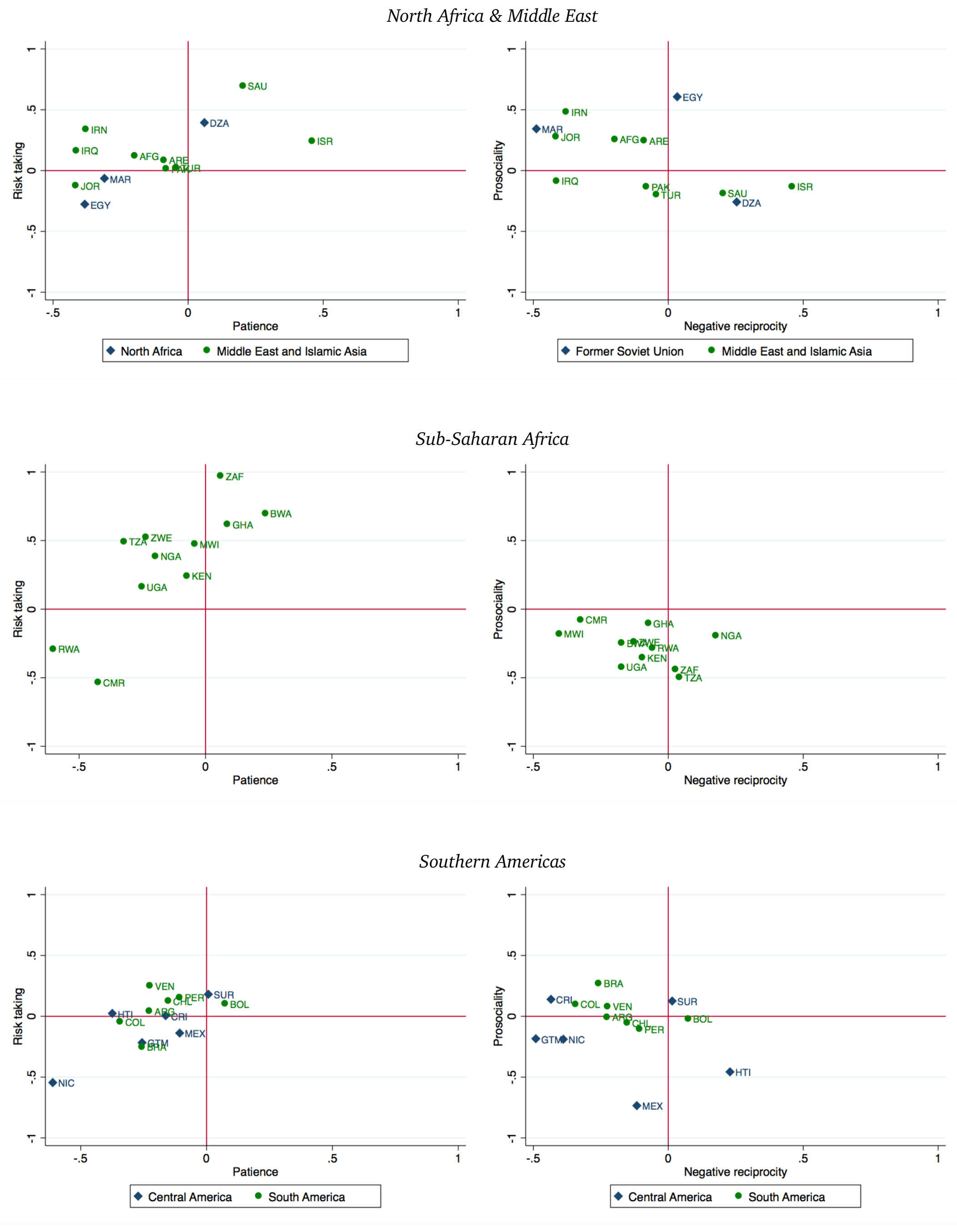


Figure 5.9b. Risk, time, and social preferences by world region (2/2). Each subpanel (row) plots risk taking, patience, negative reciprocity, and prosociality of all countries within a given world region. The prosociality score is computed as the average of altruism, positive reciprocity, and trust.

Appendix 5.D Regressions for Variance Explained Across Countries

Table 5.12. Preferences and geographic, climatic, religious, and diversity variables (1/2)

	Dependent variable:														
	Patience					Risk taking					Pos. reciprocity				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Distance to equator	0.013***				0.0047	-0.0014				-0.0020	-0.0040				-0.0024
	(0.00)				(0.01)	(0.00)				(0.01)	(0.00)				(0.01)
Longitude	0.000048				-0.00015	-0.000082				-0.00016	0.0013*				0.0017*
	(0.00)				(0.00)	(0.00)				(0.00)	(0.00)				(0.00)
% at risk of malaria	0.093				-0.17	0.12				-0.15	-0.44***				-0.43**
	(0.13)				(0.15)	(0.15)				(0.19)	(0.14)				(0.20)
Average precipitation		0.00050			0.00035		-0.0033***			-0.0027**		0.0019			0.00095
		(0.00)			(0.00)		(0.00)			(0.00)		(0.00)			(0.00)
Average temperature		-0.019**			-0.0010		0.0072			0.0067		-0.0022			-0.014
		(0.01)			(0.01)		(0.01)			(0.01)		(0.01)			(0.01)
% living in (sub-)tropical zones		-0.11			-0.069		0.27			0.26		-0.26			0.13
		(0.17)			(0.23)		(0.16)			(0.24)		(0.21)			(0.32)
Linguistic diversity			0.50**		0.22			0.081		0.20			-0.31		-0.53*
			(0.22)		(0.25)			(0.28)		(0.34)			(0.25)		(0.29)
Religious fractionalization			0.57***		0.43*			0.30*		0.41**			-0.22		-0.066
			(0.15)		(0.24)			(0.16)		(0.18)			(0.19)		(0.22)
Ethnic fractionalization			-0.95***		-0.39			0.26		-0.12			0.22		0.62**
			(0.21)		(0.24)			(0.23)		(0.29)			(0.18)		(0.24)
Constant	-0.41***	0.31**	0.00024	-0.31**	-0.45	0.036	0.089	-0.24**	0.25	0.085	0.14	-0.080	0.049	-0.12	0.068
	(0.11)	(0.13)	(0.09)	(0.13)	(0.44)	(0.10)	(0.08)	(0.10)	(0.18)	(0.40)	(0.11)	(0.10)	(0.10)	(0.10)	(0.47)
Religion shares	No	No	No	Yes	Yes	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Observations	73	73	73	73	73	73	73	73	73	73	73	73	73	73	73
R^2	0.267	0.243	0.280	0.377	0.535	0.039	0.196	0.139	0.155	0.339	0.128	0.053	0.039	0.088	0.264

Notes. OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.13. Preferences and geographic, climatic, religious, and diversity variables (2/2)

	Dependent variable:														
	Neg. reciprocity					Altruism					Trust				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Distance to equator	0.0029				0.0051	-0.010***				-0.012**	0.00013				0.00024
	(0.00)				(0.00)	(0.00)				(0.01)	(0.00)				(0.01)
Longitude	0.0013***				0.0012	0.0016**				0.0011	0.0012*				0.0012
	(0.00)				(0.00)	(0.00)				(0.00)	(0.00)				(0.00)
% at risk of malaria	-0.052				0.060	-0.51***				-0.74***	-0.30**				-0.29**
	(0.13)				(0.16)	(0.12)				(0.15)	(0.13)				(0.14)
Average precipitation		-0.00023			-0.00034		0.0016			0.00051		0.00046			-0.00033
		(0.00)			(0.00)		(0.00)			(0.00)		(0.00)			(0.00)
Average temperature		0.0040			-0.0014		0.0032			-0.020**		0.0035			0.0012
		(0.00)			(0.01)		(0.01)			(0.01)		(0.01)			(0.01)
% living in (sub-)tropical zones		-0.21			-0.034		-0.12			0.51**		-0.29			0.069
		(0.15)			(0.21)		(0.17)			(0.25)		(0.19)			(0.29)
Linguistic diversity			0.43**		0.36			-0.21		-0.31			0.16		-0.14
			(0.18)		(0.26)			(0.25)		(0.31)			(0.22)		(0.24)
Religious fractionalization			-0.024		0.20			-0.069		0.34			-0.23		-0.028
			(0.13)		(0.20)			(0.18)		(0.21)			(0.14)		(0.18)
Ethnic fractionalization			-0.51***		-0.39			0.13		-0.073			-0.27		0.074
			(0.17)		(0.28)			(0.21)		(0.29)			(0.20)		(0.23)
Constant	-0.099	0.028	0.11	-0.0023	-0.13	0.35***	-0.18	-0.00074	0.073	0.66*	0.016	-0.023	0.14*	-0.41***	-0.32
	(0.09)	(0.08)	(0.08)	(0.15)	(0.37)	(0.12)	(0.12)	(0.10)	(0.12)	(0.38)	(0.10)	(0.09)	(0.08)	(0.07)	(0.36)
Religion shares	No	No	No	Yes	Yes	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Observations	73	73	73	73	73	73	73	73	73	73	73	73	73	73	73
R^2	0.136	0.081	0.092	0.067	0.284	0.163	0.040	0.009	0.097	0.399	0.174	0.086	0.072	0.242	0.337

Notes. OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 5.E Individual-Level Characteristics and Preferences

5.E.1 Age Profiles Separately by World Region

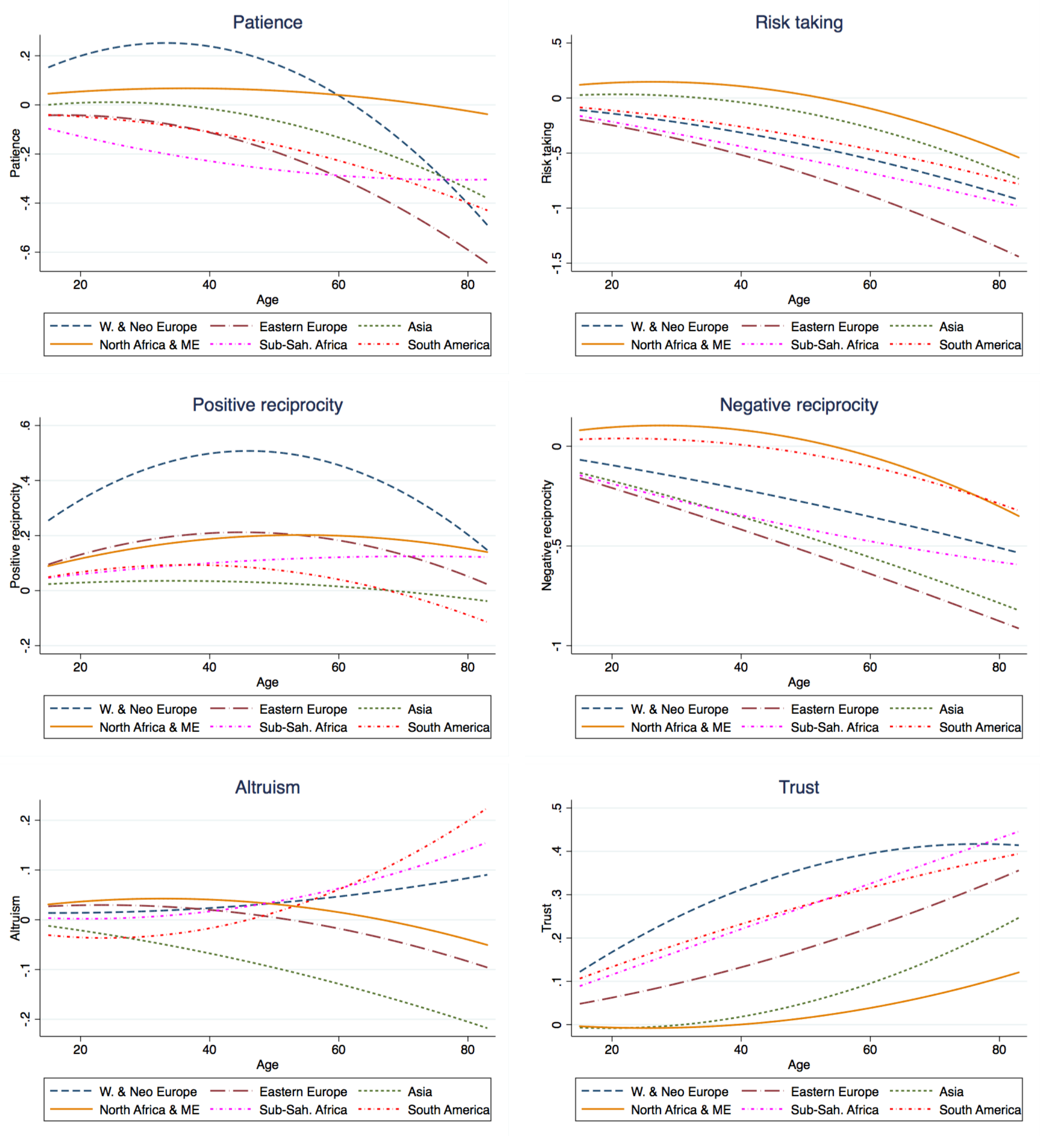


Figure 5.10. Age profiles separately by continent.

5.E.2 Cognitive Ability and Preferences by Country

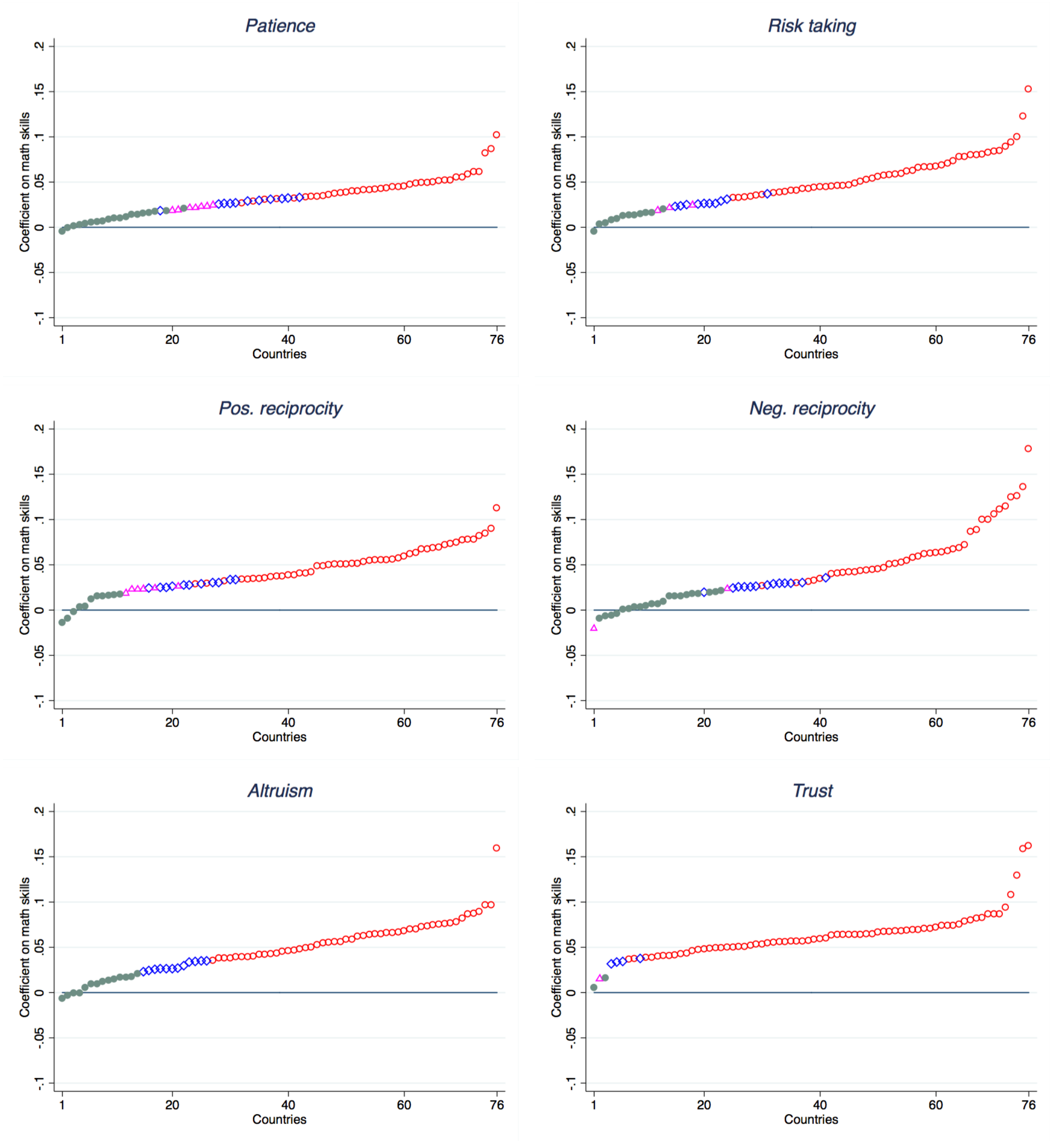


Figure 5.11. Cognitive ability correlations separately by country. Each panel plots the distribution of cognitive ability correlations. That is, for each country, we regress the respective preference on gender, age and its square, and subjective math skills, and plot the resulting math skill coefficients as well as their significance level. In order to make countries comparable, each preference was standardized (z-scores) within each country before computing the coefficients. Green dots indicate countries in which the cognitive ability effect is not statistically different from zero at the 10% level, while red / blue / pink dots denote countries in which the effect is significant at the 1% / 5% / 10% level, respectively. Positive coefficients imply that higher cognitive ability people have higher values in the respective preference.

Appendix 5.F Individual-Level Behaviors

5.F.1 Distribution of Coefficients Across Countries

This section shows that the conditional correlations on the relationships between preferences and individual-level behaviors that we reported on the global level in the main text, are not due to a few outlier countries only. Instead, the results suggest that our preference measures predict behavior across a broad set of countries. To show this, we regress the behaviors discussed in Section 5.5 on the respective preference, separately for each country, and then plot the distribution and statistical significance of the resulting coefficients. For instance, the top left panel in Figure 5.12 shows that the positive correlation between patience and savings holds in virtually all countries in our sample.

While Figure 5.12 reports the results for patience and risktaking, Figure 5.13 visualizes the relationships between altruism and behaviors. Finally, Figure 5.14 presents the correlations between positive and negative reciprocity and the behaviors discussed in Section 5.5 of the main text.

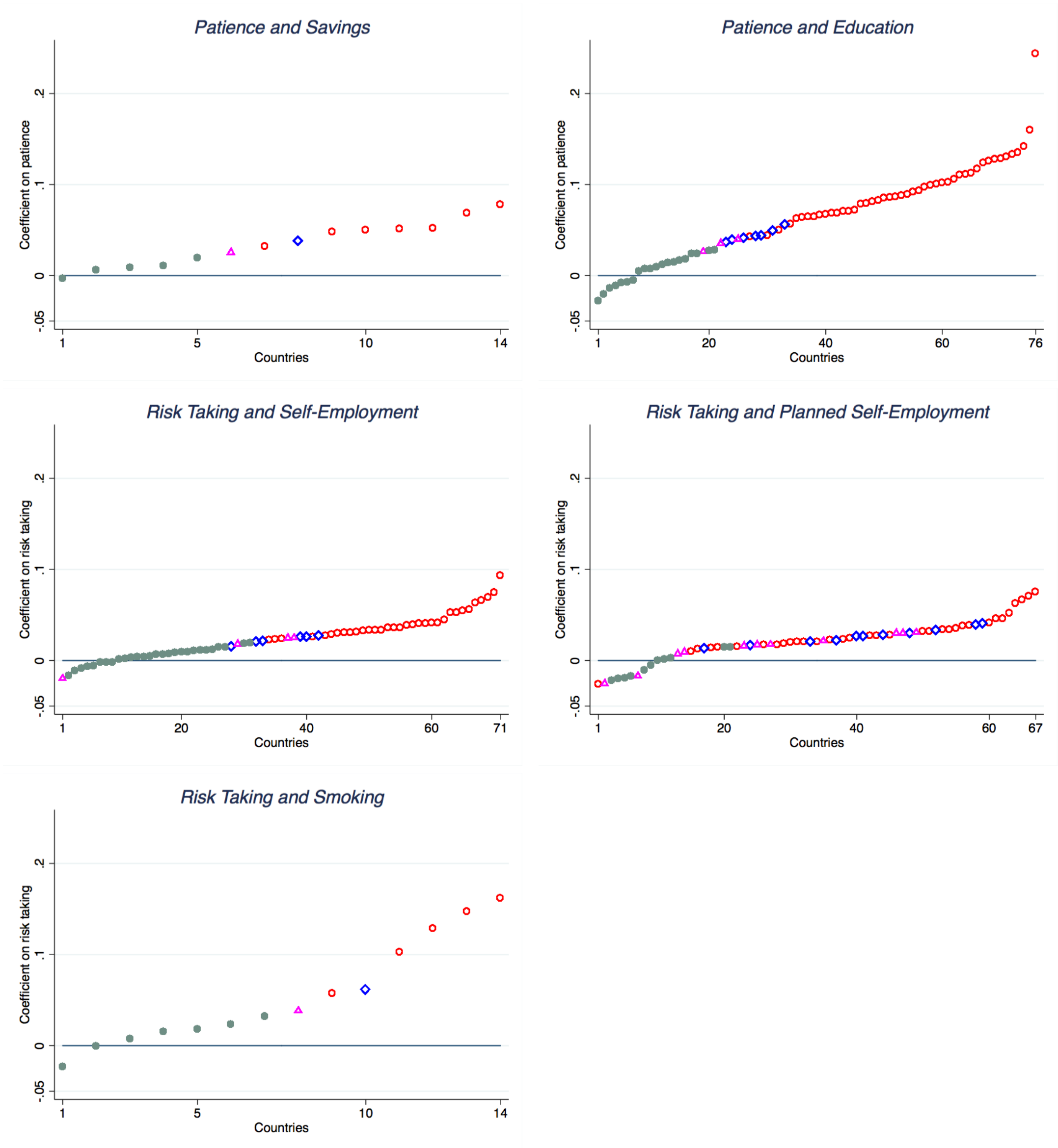


Figure 5.12. Correlations separately by country. Each panel plots the distribution of correlations across countries. That is, for each country, we regress the respective outcome on a preference and plot the resulting coefficients as well as their significance level. In order to make countries comparable, each preference was standardized (z-scores) within each country before computing the coefficients. Green dots indicate countries in which the correlation is not statistically different from zero at the 10% level, while red / blue / pink dots denote countries in which the correlation is significant at the 1% / 5% / 10% level, respectively. Positive coefficients imply that a higher preference measure is related to a higher outcome measure.

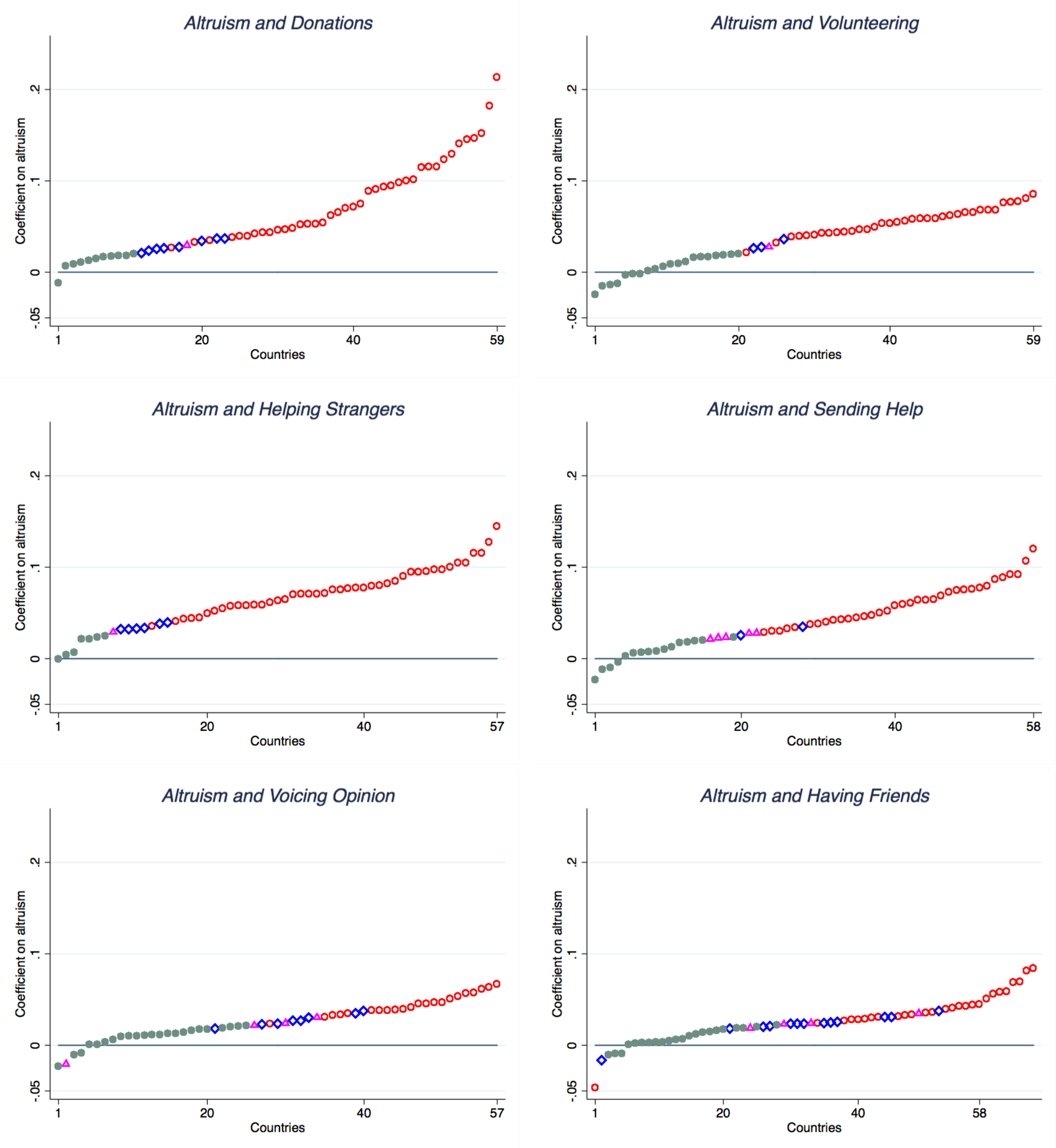


Figure 5.13. Correlations separately by country. Each panel plots the distribution of correlations across countries. That is, for each country, we regress the respective outcome on a preference and plot the resulting coefficients as well as their significance level. In order to make countries comparable, each preference was standardized (z-scores) within each country before computing the coefficients. Green dots indicate countries in which the correlation is not statistically different from zero at the 10% level, while red / blue / pink dots denote countries in which the correlation is significant at the 1% / 5% / 10% level, respectively. Positive coefficients imply that a higher preference measure is related to a higher outcome measure.

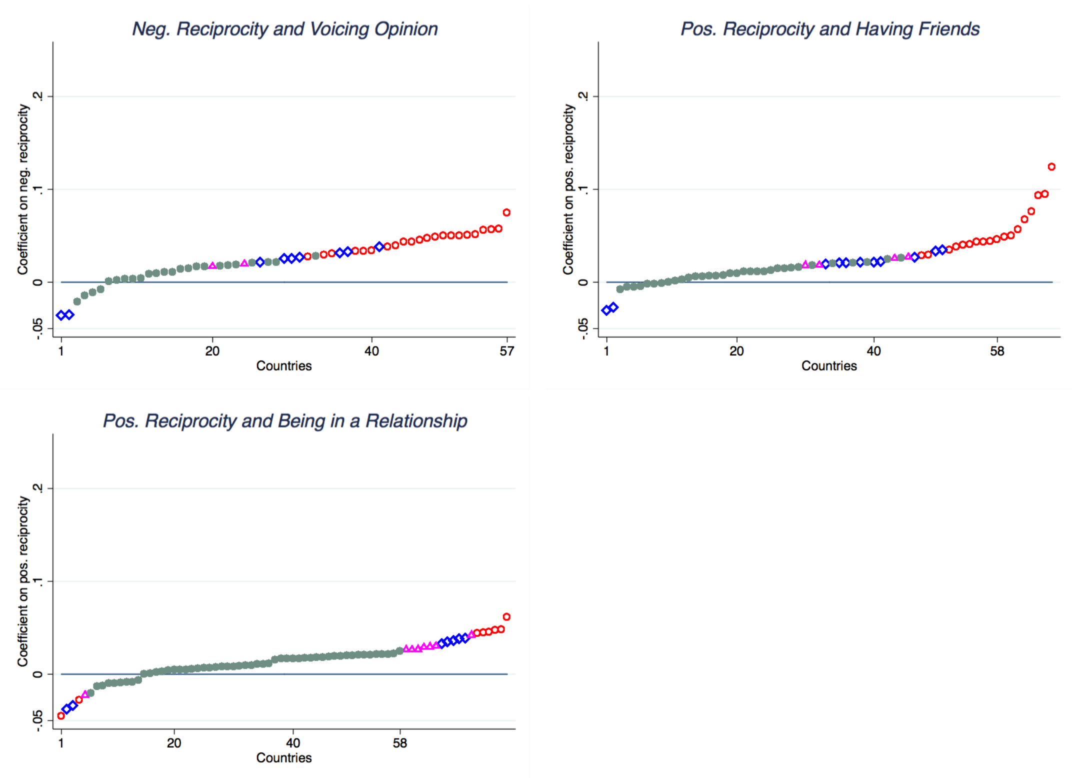


Figure 5.14. Correlations separately by country. Each panel plots the distribution of correlations across countries. That is, for each country, we regress the respective outcome on a preference and plot the resulting coefficients as well as their significance level. In order to make countries comparable, each preference was standardized (z-scores) within each country before computing the coefficients. Green dots indicate countries in which the correlation is not statistically different from zero at the 10% level, while red / blue / pink dots denote countries in which the correlation is significant at the 1% / 5% / 10% level, respectively. Positive coefficients imply that a higher preference measure is related to a higher outcome measure.

5.F.2 Robustness Checks

This appendix reports robustness checks on the relationship between preferences and behaviors at the individual level. Specifically, while Section 5.5 of the main text reported the results of OLS estimations, we now re-estimate all specifications using probit or ordered probit regressions. As Tables 5.14 and 5.15 show, the results are unchanged.

Table 5.14. Patience and accumulation decisions, risk preferences and risky choices

	Dependent variable:									
	Accumulation decisions				Risky choices					
	Saved last year		Education level		Own business		Plan to start business		Smoking intensity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Patience	0.15***	0.079***	0.20***	0.070***						
	(0.03)	(0.02)	(0.03)	(0.01)						
Risk taking					0.14***	0.11***	0.17***	0.11***	0.097***	0.050**
					(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Age		0.021		2.19***		9.23***		5.71***		6.30***
		(0.76)		(0.56)		(0.53)		(0.43)		(0.75)
Age squared		-0.63		-4.08***		-9.50***		-7.50***		-7.13***
		(0.78)		(0.54)		(0.57)		(0.52)		(0.72)
1 if female		-0.0036		-0.036		-0.27***		-0.13***		-1.32***
		(0.04)		(0.03)		(0.05)		(0.03)		(0.13)
Subj. math skills		0.043***		0.091***		0.029***		0.018***		-0.022***
		(0.01)		(0.00)		(0.00)		(0.00)		(0.01)
Log [Household income p/c]		0.36***		0.31***		0.092***		-0.040**		-0.026
		(0.04)		(0.02)		(0.01)		(0.02)		(0.02)
Constant	-0.61***	-2.86***	1.00***	4.94***	-1.07***	-4.18***	-1.24***	-2.07***	1.06***	1.37***
	(0.09)	(0.28)	(0.05)	(0.18)	(0.04)	(0.21)	(0.06)	(0.14)	(0.07)	(0.17)
Country FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Religion FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	15260	14459	79357	69272	72839	62985	57072	51489	15309	14490
Pseudo R^2	0.010	0.122	0.015	0.202	0.009	0.134	0.016	0.170	0.004	0.167

(Ordered) probit estimates, standard errors (clustered at country level) in parentheses. For the purposes of this table, age is divided by 100. Saved last year is a binary indicator, while education level is measured in three categories (roughly elementary, secondary, and tertiary education, see Appendix 5.I). Self-employment and planned self-employment are binary, while smoking intensity is measured in three categories (never, occasionally, frequently). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.15. Social preferences and social interactions

	Dependent variable:													
	Donated money		Volunteered time		Helped stranger		Sent money / goods to other individual		Voiced opinion to official		Have friends or relatives I can count on in need		In a relationship	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Altruism	0.19***	0.20***	0.16***	0.14***	0.16***	0.14***	0.12***	0.12***	0.097***	0.093***	0.029	0.066***	0.012	0.0040
	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Positive reciprocity	0.015	0.0029	-0.00046	0.017	0.091***	0.091***	0.062***	0.064***	-0.0071	-0.011	0.053*	0.067***	0.056***	0.023***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)
Negative reciprocity	0.013	-0.014	-0.0097	-0.0089	-0.0062	-0.0080	0.036**	0.022*	0.056***	0.060***	0.045***	0.011	-0.0088	-0.00076
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Age		2.13***		1.57***		2.01***		0.94***		3.85***		-3.76***		16.4***
		(0.26)		(0.29)		(0.20)		(0.24)		(0.27)		(0.38)		(0.59)
Age squared		-1.63***		-1.76***		-2.47***		-1.04***		-3.84***		3.11***		-15.9***
		(0.29)		(0.30)		(0.20)		(0.28)		(0.30)		(0.36)		(0.59)
1 if female		0.044**		-0.064***		-0.040**		0.0018		-0.16***		0.051***		-0.086***
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
Subj. math skills		0.030***		0.027***		0.025***		0.026***		0.033***		0.022***		0.0067**
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Log [Household income p/c]		0.10***		0.016*		0.056***		0.12***		0.042***		0.14***		-0.11***
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)		(0.02)		(0.01)
Constant	-0.46***	-1.71***	-0.78***	-1.09***	-0.018	-0.36**	-0.70***	-2.19***	-0.77***	-1.91***	0.90***	-0.11	0.21***	-2.37***
	(0.07)	(0.12)	(0.05)	(0.15)	(0.04)	(0.16)	(0.06)	(0.14)	(0.04)	(0.12)	(0.06)	(0.15)	(0.04)	(0.15)
Country FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Religion FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	58229	53439	58213	53430	55991	53226	56253	53559	55944	53174	65986	59209	77881	68176
Pseudo R^2	0.016	0.161	0.012	0.086	0.020	0.070	0.011	0.115	0.006	0.061	0.004	0.121	0.002	0.174

Probit estimates, standard errors (clustered at country level) in parentheses. For the purposes of this table, age is divided by 100.See Appendix 5.I for details on all dependent variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 5.G Details for Relationship Between Preferences and FTR

5.G.1 Individual-Level Regressions Separately by Country

Table 5.16. Preferences and FTR: Within-country results

Country	Weak FTR	Strong FTR	Patience	Pos. reciprocity	Trust	Altruism
Estonia	Estonian	Russian	0.05	0.13^{*}	0.38^{***}	0.45^{***}
Nigeria	Yoruba	English, Hausa, Igbo	-0.08	0.54^{***}	0.63^{***}	-0.11
Switzerland	German	French, Italian	0.17^{**}	0.14^{**}	0.28^{***}	0.30^{***}

OLS estimates, robust standard errors. The regressions report the coefficient on FTR in univariate regressions for each country in which we observe within-country variation in FTR. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.G.2 Country-Level Regressions: Robustness

While the main text reported WLS estimates, Table 5.17 reports OLS estimates.

Table 5.17. Preferences and FTR: Cross-country results

	Dependent variable:											
	Patience		Risk taking		Pos. reciprocity		Neg. reciprocity		Altruism		Trust	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Fraction of population speaking weak FTR	0.36***	0.23**	-0.11	0.015	0.15*	0.17**	-0.018	-0.082	0.061	0.11	0.19**	0.19**
	(0.13)	(0.09)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.09)	(0.09)	(0.09)	(0.08)	(0.07)
Log [GDP p/c PPP]		0.16***		0.032		-0.073*		0.052		-0.077*		-0.0055
		(0.04)		(0.03)		(0.04)		(0.04)		(0.04)		(0.03)
Distance to equator		0.010*		0.0015		-0.0078		-0.0054		-0.0033		-0.0057
		(0.01)		(0.00)		(0.01)		(0.01)		(0.01)		(0.00)
Longitude		-0.0018		0.0024		0.0021		0.00043		0.0028		0.000065
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
% at risk of malaria		0.25		-0.15		-0.33		-0.089		-0.72***		-0.092
		(0.19)		(0.24)		(0.28)		(0.17)		(0.26)		(0.19)
Average precipitation		-0.00013		-0.00081		0.00055		-0.0010		0.0031***		-0.0011
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Constant	-0.067	-1.42**	0.034	-0.51	-0.053	0.56	0.014	-0.019	-0.043	0.25	-0.058	0.39
	(0.04)	(0.56)	(0.04)	(0.45)	(0.05)	(0.67)	(0.04)	(0.50)	(0.05)	(0.61)	(0.04)	(0.48)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	75	74	75	74	75	74	75	74	75	74	75	74
R^2	0.141	0.641	0.021	0.381	0.029	0.280	0.001	0.246	0.005	0.356	0.072	0.420

OLS estimates, robust standard errors in parentheses. The regressions exclude Haiti for which no respondent could be classified. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 5.H Discussion of Measurement Error and Within- versus Between-Country Variation

In the presence of measurement error, a simple variance decomposition as shown in Table 2 tends to overstate the relative importance of within-country variation in preferences. This is because measurement error would be part of the within-country variation, whereas the aggregation to country averages mitigates measurement error and thus removes this source of variation. This section provides evidence that measurement error is unlikely to be large enough to drive the result.

To illustrate the impact of measurement error, recall that a simple regression of an individual-level preference measure M on a matrix of country dummies D yields

$$M = D'\gamma + \epsilon$$

In a setting without measurement error ϵ would be interpreted as individual specific effects that are not explained by the variation between countries. The total variance of M is given by

$$Var(M) = Var(\delta) + Var(\epsilon) + 2cov(\delta, \epsilon)$$

where $\delta = D'\gamma$. Note that the R^2 from a regression of M on the country dummies (i.e., $Var(\delta)/Var(M)$) could be interpreted as the between country-variation, i.e., the fraction of total variation explained by country dummies, if individual effects are unrelated to country effects.

If, however, the preference measure M measures the true preference parameter P with error, denoted e, the residual variation of the regression above does not only capture individual effects. Assume that M is a linear function of P and e, i.e.,

$$M = P + e,$$

such that we can rewrite

$$P + e = \delta + \epsilon$$

The total variance of the preference is hence

$$Var(P) = Var(\delta) + Var(\epsilon) - Var(e),$$

assuming that $\epsilon \perp \delta$ and $e \perp \epsilon, \delta, P$.

The regression model still allows identifying $Var(\delta)$, but the share of preference variation that is truly explained by the between-country variation is no longer given by the R^2, $Var(\delta)/Var(M)$, but rather by $Var(\delta)/Var(P)$. To assess whether between-country or within-country effects explain a larger share of total variation, one needs to compare $Var(\delta)/Var(P)$ to $Var(\epsilon)/Var(P)$. Since $Var(P) = Var(M) - Var(e)$, $Var(e)$ needs to be determined.

The variance of measurement error, $Var(e)$, is not directly observable, but estimates of test-retest correlations of relevant preference measures are available which can be used to gauge the size of $Var(e)$. Based on arguments of plausibility, the variance of the measurement error does not appear to be large enough to invalidate the claim that the within-country variation is smaller than the between-country variation. Consider how large the proportion of measurement error in the total variation of M can be, with between-country effects still explaining a smaller share of variation than individual-specific effects. Note that between- and within-country variation add up to total variation in preferences absent measurement error: $Var(\delta)/Var(P)$ = 1- $Var(\epsilon)/Var(P)$. Thus, between-country effects explain a relatively smaller share of total variation if $Var(\delta)/Var(P) < 0.50$. Letting q with $0 < q \leq 1$ be the fraction of measurement error in M, this condition can be evaluated by scaling up the R^2 from a regression of M on the set of country dummies by $1/(1-q)$. I.e., if $Var(\delta)/(Var(M)(1-q)) < 0.5$, the between-country variation is smaller than the within-country variation, even accounting for measurement error.

Take, as an example, the estimate for risk-taking in Table 5.4, for which the regression of the risk measure on the set of country dummies yields an R^2 of 0.09. Solving $R^2 < 0.5(1-q)$ for q shows that as long as $q < 0.828$, the within country variation exceeds the between country variation. Previous work has shown that the test-retest correlation of the single components of this particular risk measure is around 0.6 (Beauchamp et al., 2011). This implies that, in order for measurement error alone to be able to explain the greater variation of preferences within-country than between-country, measurement error would have to be twice as large as existing evidence suggests.

Appendix 5.I Description and Data Sources of Outcome Variables

5.I.1 Individual-Level Variables

5.I.1.0.1 Subjective law and order index. Included in Gallup's Background data (0-1). Derived from responses to three questions: "In the city or area where you live, do you have confidence in the local police force?"; "Do you feel safe walking alone at night in the city or area where you live?"; "Within the last 12 months, have you had money or property stolen from you or another household member?".

5.I.1.0.2 Subjective physical health index. Included in Gallup's Background data (0-1). Derived from responses to five questions: "Do you have any health problems that prevent you from doing any of the things people your age normally can do?"; "Now, please think about yesterday, from the morning until the end of the day. Think about where you were, what you were doing, who you were with, and how you felt. Did you feel well-rested yesterday?"; "Did you experience the following feelings during a lot of the day yesterday? How about physical pain?"; "Did you experience the following feelings during a lot of the day yesterday? How about worry?"; "Did you experience the following feelings during a lot of the day yesterday? How about sadness?".

5.I.1.0.3 Household income per capita. Included in Gallup's background data. To calculate income, respondents are asked to report their household income in local currency. Those respondents who have difficulty answering the question are presented a set of ranges in local currency and are asked which group they fall into. Income variables are created by converting local currency to International Dollars (ID) using purchasing power parity (PPP) ratios. Log household income is computed as log (1+ household income).

5.I.1.0.4 Education level. Included in Gallup's background data. Level 1: Completed elementary education or less (up to 8 years of basic education). Level 2: Secondary - 3 year tertiary education and some education beyond secondary education (9-15 years of education). Level 3: Completed four years of education beyond high school and / or received a 4-year college degree.

5.I.1.0.5 Subjective self-assessment of math skills. *How well do the following statements describe you as a person? Please indicate your answer on a scale from 0 to 10. A 0 means "does not describe me at all" and a 10 means "describes me perfectly". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. I am good at math.*

5.I.1.0.6 Saved last year. Binary variable capturing whether the respondent saved any money in the previous year. Included in Gallup's background data.

5.I.1.0.7 Own business. Binary variable capturing whether the respondent is self-employed. Included in Gallup's background data.

5.I.1.0.8 Plan to start business. Binary variable capturing whether the respondent is planning to start their own business (only asked of those who are not self-employed). Included in Gallup's background data.

5.I.1.0.9 Smoking intensity. Variable capturing how frequently a respondent smokes (0=never, 1=occasionally, 2=frequently). Included in Gallup's background data.

5.I.1.0.10 Donated money. Binary variable capturing whether the respondent donated money in the previous month. Included in Gallup's background data.

5.I.1.0.11 Volunteered time. Binary variable capturing whether the respondent volunteered time to an organization in the previous month. Included in Gallup's background data.

5.I.1.0.12 Helped stranger. Binary variable capturing whether the respondent helped a stranger who needed help in the previous month. Included in Gallup's background data.

5.I.1.0.13 Sent help to individual. Binary variable capturing whether the respondent sent help (money or goods) to another individual in the previous year. Included in Gallup's background data.

5.I.1.0.14 Voiced opinion to official. Binary variable capturing whether the respondent voiced their opinion to a public official in the previous month. Included in Gallup's background data.

5.I.1.0.15 Donated money. Binary variable capturing whether the respondent has relatives or friends they can count on to help them whenever needed. Included in Gallup's background data.

5.I.1.0.16 In a relationship. Binary variable coded as zero if the respondents is single, separated, divorced, or widowed, and as 1 if respondent is married or has a domestic partner. Included in Gallup's background data.

5.I.2 Country-Level Variables

5.I.2.0.1 Distance to equator, longitude. Source: the CEPII geo database.

5.I.2.0.2 Land suitability for agriculture. Index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH, taken from Michalopoulos (2012).

5.I.2.0.3 Temperature. Average monthly temperature of a country in degree Celsius, 1961-1990, taken from Ashraf and Galor (2013). Data originally based on geospatial average monthly temperature data for this period reported by the G-ECON project (Nordhaus, 2006).

5.I.2.0.4 Precipitation. Average monthly precipitation of a country in mm per month, 1961-1990, taken from Ashraf and Galor (2013). Data originally based on geospatial average monthly precipitation data for this period reported by the G-ECON project (Nordhaus, 2006).

5.I.2.0.5 Predicted genetic diversity. Predicted genetic diversity of the contemporary population, adjusted for post-Columbian migration flows and genetic distance between ethnic groups. See Ashraf and Galor (2013).

5.I.2.0.6 GDP per capita. Average annual GDP per capita over the period 2003 – 2012, in 2005US$. Source: World Bank Development Indicators.

5.I.2.0.7 Democracy index. Index that quanties the extent of institutionalized democracy, as reported in the Polity IV dataset. Average from 2003 to 2012.

5.I.2.0.8 Percentage at risk of malaria. The percentage of population in regions of high malaria risk (as of 1994), multiplied by the proportion of national cases involving the fatal species of the malaria pathogen, P. falciparum. This variable was originally constructed by Gallup and Sachs (2000) and is part of Columbia University's Earth Institute data set on malaria. Data taken from Ashraf and Galor (2013).

5.I.2.0.9 Percentage in (sub-)tropical zones. Percentage of area within a country which forms part of each of the tropical or sub-tropical climatic zones. Data taken from John Luke Gallup, http://www.pdx.edu/econ/jlgallup/country-geodata.

5.I.2.0.10 Life expectancy. Average life expectancy at birth, average from 2003 to 2012, taken from World Bank Development Indicators.

5.I.2.0.11 Gini coefficient. Average from 2003 to 2012, taken from World Bank Development Indicators.

5.I.2.0.12 Redistribution (% of GDP). Government transfers as a fraction of national income. Average from 2003 to 2012, taken from World Bank Development Indicators.

5.I.2.0.13 Ethnic and religious fractionalization. Indices due to Alesina et al. (2003) capturing the probability that two randomly selected individuals from the same country will be from different ethnic (religious) groups.

5.I.2.0.14 Linguistic diversity. Index due to Fearon (2003) capturing the linguistic diversity within a given country, taking into account the structural similarity of languages using a language tree.

5.I.2.0.15 Labor protection index. Index capturing the rigidity of employment laws by Botero et al. (2004). Includes data on employment, collective relations, and social security laws and measures legal worker protection.

5.I.2.0.16 Homicide rate. Numbers of intentional homicides per 100,000 people. Average 2003–2012, taken from World Bank Development Indicators.

5.I.2.0.17 Religion shares. Source: Barro (2003).

6

The Ancient Origins of the Global Variation in Risk Preferences and Prosociality*

6.1 Introduction

Preferences over risk, the timing of rewards, and social interactions form the building blocks of a large class of models in both micro- and macroeconomics. Recently, empiricists have shown that these preferences vary substantially within populations and – in line with economic models – predict a plethora of individual-level economic decisions ranging from stock and labor market behavior over savings and schooling choices to volunteering, donating, and cooperating. However, do these preferences also vary across countries? And if so, what explains this variation, i.e., what are the ultimate determinants of heterogeneity in preferences? Using a novel, globally representative data set on key economic preferences, this paper seeks to provide an answer to these questions. Our key contribution is to show that the structure of mankind's ancient migration out of Africa and around the globe has had a persistent impact on the between-country distribution of preferences over risk and social interactions as of today. These findings add to the emerging literature on endogenous preferences (Fehr and Hoff, 2011) in highlighting very deep population-level historical events as a key driver in shaping preferences, and hence also contribute to understanding the ultimate sources of cross-country economic heterogeneity.

* We are grateful to Quamrul Ashraf, Marianne Bertrand, Klaus Desmet, Andreas Kleiner, Nathan Nunn, Uwe Sunde, Hans-Joachim Voth, and Romain Wacziarg for very helpful comments and discussions. Valuable comments were also received from seminar audiences at Bonn, Munich, the NBER Economics of Culture and Institutions Meeting 2014, SITE Stanford 2014, the North-American ESA conference 2014, and the Annual Congress of the EEA 2013. We thank Quamrul Ashraf, Oded Galor, and Ömer Özak for generous data sharing. Ammar Mahran provided oustanding research assistance. Armin Falk acknowledges financial support from the European Research Council through ERC # 209214. Becker, Enke, Falk: University of Bonn, Department of Economics, Adenauerallee 24-42, 53113 Bonn, Germany. anke.becker@uni-bonn.de, benjamin.enke@uni-bonn.de, armin.falk@uni-bonn.de

According to the widely accepted "Out of Africa hypothesis", starting around 50,000 years ago, early mankind migrated out of East Africa and continued to explore and populate our planet through a series of successive migratory steps. Each of these steps consisted of some sub-population breaking apart from the previous colony and moving on to found new settlements. This pattern implies that some contemporary population pairs have spent a longer time of human history apart from each other than others. As a result, the time elapsed since two groups shared common ancestors differs across today's population pairs. The key idea underlying our analysis is that these differential time frames of separation might have affected the cross-country distribution of preferences over risk, time, and social interactions. First, populations that have spent a long time apart from each other were exposed to differential historical experiences, which could affect risk, time, and social preferences (Callen et al., 2014; Rao, 2013; Kosse et al., 2014). Second, due to, e.g., random genetic drift, long periods of separation lead to different population-level genetic endowments, which might in turn shape attitudes.[1] We use a simple model to show that both of these channels imply the prediction that – on average – populations that have been separated for a longer time in the course of human history, should also exhibit more different preference profiles.

To investigate this hypothesis, we use a novel dataset on economic preferences across countries in combination with proxies for long-run human migration patterns and the resulting temporal distances. As part of the Global Preference Survey (GPS), we collected survey measures of risk, time, and social preferences (see Falk et al., 2015a). The sample of 80,000 people from 76 countries is constructed to provide representative population samples within each country and geographical representativeness in terms of countries covered. The survey items were selected and tested through a rigorous ex ante experimental validation procedure involving real monetary stakes. The elicitation followed a standardized protocol that was implemented through the professional infrastructure of the Gallup World Poll. These data allow the computation of nationally representative levels of risk aversion, patience, altruism, positive reciprocity, negative reciprocity, as well as trust, and hence facilitate the derivation of the absolute difference in a given trait within a country pair.

We combine these data with three classes of proxies for the temporal patterns of ancient population fissions, i.e., proxies for the length of time since two populations shared common ancestors. (i) First, we employ the F_{ST} and Nei genetic distances between populations, as originally measured by the population geneticists L. Cavalli-Sforza et al. (1994) and introduced into the economics literature by Spolaore and Wacziarg (2009). As population geneticists have long noted, whenever two populations split apart from each other in order to found separate settlements, their genetic distance increases over time due to random genetic drift. Thus, the genetic distance between two populations is a measure of *temporal distance* since separation. (ii) Second, we use measures of *predicted* migratory distance between contemporary populations, which were constructed by Ashraf and Galor (2013b) and Özak (2010), respectively. The predicted migratory distance variable of Ashraf and Galor (2013b) is based on a procedure which exploits

[1] Cesarini et al. (2008, 2009, 2012) use a series of twin studies to provide evidence for a genetic effect on risk, time, and social preferences.

information on the geographic patterns of early migratory movements and constitutes a proxy for the predicted length of separation of two populations. The human-mobility-index measure of Özak (2010), on the other hand, explicitly computes the walking time between two countries' capitals, taking into account topographic, climatic, and terrain conditions, as well as human biological abilities. (iii) Finally, we make use of the observation that linguistic trees closely follow the structure of separation of human populations and employ a measure of linguistic distance between two populations as explanatory variable. In sum, we use various independent sources of data which are known to reflect the length of separation of populations.

Our empirical analysis of the relationship between preferences and ancient migration patterns starts by investigating the raw relationship between the absolute difference in average preferences between two countries and genetic distance, which is known to be the theoretically most appealing proxy for the length of time since today's populations shared common ancestors. Results show that cross-country differences in risk aversion as well as altruism, positive reciprocity, and trust are all significantly increasing in our temporal distance proxy. In contrast, differences in patience and negative reciprocity are largely uncorrelated with temporal distance.

In a second step, we investigate to what extent the relationships between risk and social preferences and temporal distance are likely to be driven by contemporary environmental conditions, i.e., omitted variables. We establish that the relationships are robust to an extensive set of covariates, including controls for differences in the countries' demographic composition, their geographic position, prevailing climatic and agricultural conditions, institutions, and economic development. In all of the corresponding regressions, the point estimate is very stable, suggesting that unobserved heterogeneity is unlikely to drive our results, either (Altonji et al., 2005). Our results also hold when we (i) exclude observations from the tails of the genetic distance or preference distributions from the analysis, (ii) split the sample by level of economic development, (iii) restrict the sample to countries in the Old World, or (iv) make use of information on the precision of the genetic distance data.

We then proceed by investigating the robustness of our results by employing predicted migratory and linguistic distance as alternative explanatory variables. In both unconditional and conditional regressions, the results closely mirror those established with genetic distance, and hence highlight that our findings do not hinge on genetic distance as proxy for temporal distance.

In sum, several proxies for the temporal distance between populations are predictive of differences in risk and social preferences. In a final step, we provide evidence that these patterns indeed reflect the accumulation of preference changes over thousands of years subsequently to the original population breakups, rather than characteristics of the breakup process, such as selective migration by, e.g., the risk tolerant types. To this end, we investigate whether today's country-level preference profiles evolve monotonically along humans' migratory route out of East Africa, i.e., whether preferences are correlated with the length of the ancestors' migratory path. We find that the level of none of the preference traits exhibits a monotonic relationship with (predicted) migratory distance from East Africa as derived by Ashraf and Galor (2013b); instead, risk aversion, patience, and the prosocial traits are all significantly non-linearly related to migratory distance. These results suggest that the relationship between temporal distance and preference differences is not driven by features of the breakup process that persisted until today, but rather by the accumulation of idiosyncratic shocks over thousands of years. However, these results also lend themselves to an interpretation in terms of intrapopulation genetic diversity: whenever a sub-population split apart from its parental colony, those humans breaking new ground took with them only a fraction of the genetic diversity of the previous genetic pool, implying that the total diversity of the gene pool significantly decreases along human migratory routes out of East Africa. Thus, in essence, our results say that country-level preference profiles are non-linearly associated with genetic diversity.

A number of recent contributions argue that the (cultural) diversity caused by long-run migration thousands of years ago can have aggregate economic effects. Spolaore and Wacziarg (2009, 2011) find a strong relationship between genetic distance and income differences across countries and posit that lower cultural distance facilitates the diffusion of knowledge.[2] Ashraf and Galor (2013b) and Ashraf et al. (2014) establish a hump-shaped relationship between national income and genetic diversity and argue that the non-monotonicity reflects the trade-off between more innovation and lower cooperation that is associated with higher cultural diversity.[3] Our paper dovetails with these contributions as it shows that the genetic variables which proxy for migratory flows indeed capture variation in economically important traits.

This paper also forms part of an active recent literature on the historical, biological, and cultural origins of beliefs and preferences. Chen (2013) and Galor and Özak (2014) show that future-orientation is affected by a structural feature of languages and historical agricultural productivity. Tabellini (2008) and an earlier version of Guiso et al. (2009) relate interpersonal trust to linguistic features and the genetic distance between two populations. Nunn and Wantchekon (2011), Voigtländer and Voth (2012) and Alesina et al. (2013) establish the deep roots of trust, beliefs over the appropriate role of women in society, and anti-semitism, respectively. Desmet et al. (2009) and Spolaore and Wacziarg (2015) show that genetic and linguistic distance correlates with opinions and attitudes as expressed in the World Values Survey. However, perhaps given the previous lack of data, this paper is the first contribution to study the origins of cross-country variation in risk, time, and social preferences.

The remainder of this paper proceeds as follows. In Section 6.2, we develop our hypothesis on the relationship between the structure of migratory movements and preferences, while Section 6.3 presents the data. Section 4.4 discusses our result on the connection between preference differences and length of separation. Section 6.5 studies the relationship between preferences and genetic diversity, and Section 3.6 concludes.

[2] Gorodnichenko and Roland (2010), Spolaore and Wacziarg (2014), and Giuliano et al. (2014) analyze the relationships between genetic distance and individualism, conflict, and bilateral trade.

[3] Also see Arbatli et al. (2013) and Ashraf and Galor (2013a).

6.2 Preferences and the Great Human Expansion

According to the widely accepted theory of the origins and the dispersal of early humans, the single cradle of mankind is to be found in East or South Africa and can be dated back to roughly 100,000 years ago (see, e.g., Henn et al. (2012) for an overview). Starting from East Africa, a small sample of hunters and gatherers exited the African continent around 50,000-60,000 years ago and thereby started what is now also referred to as the "great human expansion". This expansion continued throughout Europe, Asia, Oceania, and the Americas, so that mankind eventually came to settle on all continents. A noteworthy feature of this very long-run process is that it occurred through a large number of discrete steps, each of which consisted of a sub-sample of the original population breaking apart and leaving the previous location to move on and found new settlements elsewhere (so-called serial founder effect). The main hypothesis underlying this paper is that the pattern of successive breakups and the resulting distribution of temporal distances across populations affected the distribution of economic preferences we observe around the globe today.

In particular, the series of migratory steps implied a frequent breakup of formerly united populations. After splitting apart, these sub-populations often settled geographically distant from each other, i.e., lived in separation. There are at least two channels through which the length of separation of two groups might have had an impact on between-group differences in preferences.[4]

First, if two populations have spent a long time apart from each other, they were subject to many differential historical experiences. Recent work highlights that economic preferences are malleable by idiosyncratic experiences or, more generally, by the composition of people's social environment (see, e.g., Callen et al. (2014) on risk preferences or Rao (2013) and Kosse et al. (2014) on prosocial attitudes). Thus, the differential historical experiences which have accumulated over thousands of years of separation might have given rise to differential preferences as of today.

Second, whenever two populations spend time apart from each other, they develop different population-level genetic pools due to, e.g., random genetic drift or location-specific selection pressures. Given that attitudes like risk aversion, trust, and altruism are transmitted across generations and that part of this transmission is genetic in nature (Cesarini et al., 2009; Dohmen et al., 2012), the different genetic endowments induced by long periods of separation could also generate differential preferences.

To formally illustrate both of these channels (historical experiences and genetic pools) using a simple example, suppose that there are three contemporary populations, A, B, and C. Suppose further that the historical migration tree is such that, in $t = 0$, A, B, and C formed one population, in $t = 1$, the union of B and C broke apart from A, and in $t = 2$ population C split away from B. Each population i has a scalar-representable preference endowment x_i^t, which we normalize to $x^0 = 0$. In each period, a given population is subject to a random shock to the preference endowment ϵ_i^t, which we assume to be independently and identically distributed according to $F(\cdot)$ across time and space, where

[4] It is conceivable that differences in preferences are correlated with temporal distance proxies because of the structure of the population breakups *as such*, rather than the temporal distances that were caused by the population breakups. Section 6.5 provides a discussion of this issue.

$F(\cdot)$ has mean zero. In other words, we assume that the preference endowment is subject to random drift, which could result from random historical experiences or changes in the genetic pool (either through random drift or location-specific selection pressures.)

Importantly, note that we make the assumption of random drift only to emphasize that *even* random drift would generate the theoretical prediction that longer periods of separation imply larger differences in preferences, in expectation. Clearly, if preferences evolved monotonically along the migratory path, then temporal distance ought to be predictive of preference differences – however, there is no biological principle according to which the evolution of a scalar-representable trait must follow a monotonic path. While there are reasons to believe that traits like risk aversion, time preference, or altruism are subject to local selection pressures, these selection pressures might operate in differerent directions along the migratory path as groups of humans and their descendants pass through many different environments. Thus, we show here that even the weakest possible assumption of random idiosyncratic preference changes gives rise to the hypothesis that longer separation leads to larger differences in preferences, on average. Specifically, in our empirical regression framework to be presented below, we compare, e.g., the contemporary absolute difference in preferences between populations A and C (long separation) with the absolute difference between B and C (short separation):

$$\begin{aligned} E[|x_A^2 - x_C^2| - |x_B^2 - x_C^2|] &= E[|\epsilon_A^1 + \epsilon_A^2 - \epsilon_{B,C}^1 - \epsilon_C^2| - |\epsilon_{B,C}^1 + \epsilon_B^2 - \epsilon_{B,C}^1 - \epsilon_C^2|] \\ &= E[|\epsilon_A^1 + \epsilon_A^2 - \epsilon_{B,C}^1 - \epsilon_C^2|] - E[|\epsilon_B^2 - \epsilon_C^2|] > 0 \end{aligned}$$

Thus, in expectation, longer separation should lead to larger absolute differences in preferences, implying a bilateral statement about between-country differences:

Hypothesis. *The absolute difference in (average) preferences between two countries increases in the length of separation of the respective populations in the course of human history.*

6.3 Data

6.3.1 Risk, Time, and Social Preferences Across Countries

Our data on risk, time, and social preferences are part of the Global Preference Survey (GPS), which constitutes a unique dataset on economic preferences from representative population samples around the globe. In many countries around the world, the Gallup World Poll regularly surveys representative population samples about social and economic issues. In 76 countries, we included as part of the regular 2012 questionnaire a set of survey items which were explicitly designed to measure a respondent's preferences (for details see Falk et al., 2015a).

Four noteworthy features characterize these data. First, the preference measures have been elicited in a comparable way using a standardized protocol across countries. Second, contrary to small- or medium-scale experimental work, we use preference measures that have been elicited from representative population samples in each country. This allows for inference on between-country differences in preferences, in contrast to existing cross-country comparisons of convenience (student) samples. The median sample size was 1,000 participants per country; in total, we collected preference measures for more than 80,000 participants worldwide. Respondents were selected through probability sampling and interviewed face-to-face or via telephone by professional interviewers. Third, the dataset also reflects geographical representativeness. The sample of 76 countries is not restricted to Western industrialized nations, but covers all continents and various development levels. Specifically, our sample includes 15 countries from the Americas, 24 from Europe, 22 from Asia and Pacific, as well as 14 nations in Africa, 11 of which are Sub-Saharan. The set of countries contained in the data covers about 90% of both the world population and global income. Fourth, the preference measures are based on experimentally validated survey items for eliciting preferences. In order to ensure behavioral relevance, the underlying survey items were designed, tested, and selected through an explicit ex-ante experimental validation procedure (Falk et al., 2015b). In this validation step, out of a large set of preference-related survey questions, those items were selected which jointly perform best in explaining observed behavior in standard financially incentivized experimental tasks to elicit preference parameters. In order to make these items cross-culturally applicable, (i) all items were translated back and forth by professionals, (ii) monetary values used in the survey were adjusted along the median household income for each country, and (iii) pretests were conducted in 21 countries of various cultural heritage to ensure comparability. The preference measures are derived as follows (see Appendix 5.A and Falk et al. (2015a) for details):[5]

6.3.1.0.1 Risk Taking. The set of survey items included two measures of the underlying risk preference – one qualitative subjective self-assessment and one quantitative measure. The subjective self-assessment directly asks for an individual's willingness to take risks: *"Generally speaking, are you a person who is willing to take risks, or are you not willing to do so? Please indicate your answer on a scale from 0 to 10, where a 0 means "not willing to take risks at all" and a 10 means "very willing to take risks". You can also use the values in between to indicate where you fall on the scale."*

The quantitative measure is derived from a series of five interdependent hypothetical binary lottery choices, a format commonly referred to as the "staircase procedure". In each of the five questions, participants had to decide between a 50-50 lottery to win x euros or nothing (which was the same in each question) and varying safe payments y. The questions were interdependent in the sense that the choice of a lottery resulted in an increase of the safe amount being offered in the next question, and conversely. For instance, in Germany, the fixed upside of the lottery x was 300 euros, and in the first question, the fixed payment was 160 euros. In case the respondent chose the lottery (the safe payment), the safe payment increased (decreased) to 240 (80) euros in the second question. In essence, by adjusting the fixed payment according to previous choices, the questions "zoom in" around the respondent's certainly equivalent and make efficient use of limited and costly survey time. This procedure yields one of 32 ordered outcomes. The subjective self-assessment and the outcome of the quantitative lottery staircase were then aggregated into a single index which describes an individual's degree of risk taking.

[5] The description of the survey items closely follows the one in Falk et al. (2015a).

6.3.1.0.2 Patience. The measure of patience is also derived from the combination of responses to two survey measures, one with a quantitative and one with a qualitative format. The quantitative survey measure consists of a series of five hypothetical binary choices between immediate and delayed financial rewards. In each of the five questions, participants had to decide between receiving a payment today or larger payments in 12 months. Conceptually similar to the elicitation of risk preferences, the questions were interdependent in the sense that the delayed payment was increased or decreased depending on previous choices. The qualitative measure of patience is given by the respondent's self-assessment regarding their willingness to wait on an 11-point Likert scale, asking "how willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?".

6.3.1.0.3 Prosociality: Altruism, Positive Reciprocity, and Trust. The GPS includes six survey items which map into three prosocial traits: altruism, positive reciprocity, and trust. While these behavioral traits are conceptually distinct, they share in common that they are commonly associated with "positive" social interactions.

Altruism was measured through a combination of one qualitative and one quantitative item, both of which are related to donation. The qualitative question asked people how willing they would be to give to good causes without expecting anything in return on an 11-point scale. The quantitative scenario depicted a situation in which the respondent unexpectedly received 1,000 euros and asked them to state how much of this amount they would donate.

People's propensity to act in a positively reciprocal way was also measured using one qualitative item and one question with a quantitative component. First, respondents were asked to provide a self-assessment about how willing they are to return a favor on an 11-point Likert scale. Second, participants were presented a choice scenario in which they were asked to imagine that they got lost in an unfamiliar area and that a stranger – when asked for directions – offered to take them to their destination. Participants were then asked which out of six presents (worth between 5 and 30 euros in 5 euros intervals) they would give to the stranger as a "thank you".

Finally, to measure trust, people were asked whether they assume that other people only have the best intentions (Likert scale, 0-10).

6.3.1.0.4 Negative Reciprocity. Negative reciprocity was elicited through three self-assess-ments. First, people were asked how willing they are to take revenge if they are treated very unjustly, even if doing so comes at a cost (0-10). The second and third item probed respondents about their willingness to punish someone for unfair behavior, either towards *themselves* or towards a *third person*.

6.3.2 Proxies for Ancient Migration Patterns

We use various separate but conceptually linked classes of variables to proxy for the length of time since two populations split apart: (i) Genetic distance, (ii) predicted migratory distance, and (iii) linguistic distance.

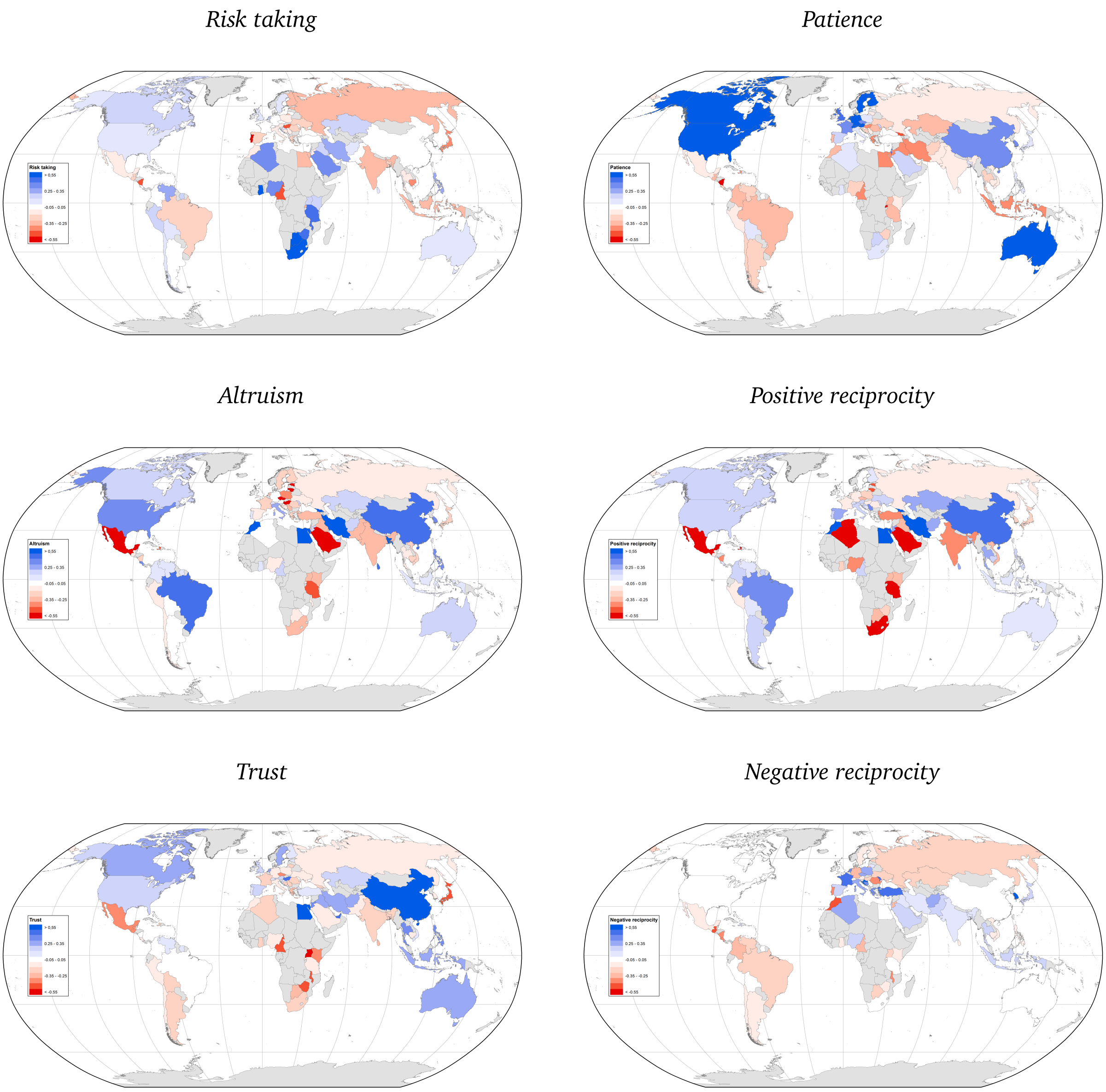


Figure 6.1. World maps of preferences. In each figure, white denotes the world average. Darker blue indicates higher values of a given trait, while darker red colors indicate lower values, all of which are measured in standard deviations from the world mean. Grey indicates missings.

6.3.2.0.1 Genetic Distance Between Countries. First, whenever populations break apart, they stop interbreeding, thereby preventing a mixture of the respective genetic pools. However, since every genetic pool is subject to random drift ("noise") or local selection pressures, geographical separation implies that over time the genetic distance between sub-populations gradually became (on average) larger. Thus, the genealogical relatedness between two populations reflects the length of time elapsed since these populations shared common ancestors. In fact, akin to a molecular clock, population geneticists have made use of this observation by constructing mathematical models to compute the timing of separation between groups. This makes clear that, at its very core, genetic distance constitutes not only a measure of genealogical relatedness, but also of *temporal distance* between two populations.

Technically, genetic distance constitutes an index of expected heterozygosity, which can be thought of as the probability that two randomly matched individuals will be genetically different from each other in terms of a pre-defined spectrum of genes. Indices of heterozygosity are derived using data on allelic frequencies, where an allele is a particular variant taken by a gene.[6] Intuitively, the relative frequency of alleles at a given locus can be compared across populations and the deviation in frequencies can then be averaged over loci. This is the approach pursued in the work of the population geneticists L. Cavalli-Sforza et al. (1994). The main dataset assembled by these researchers consists of data on 128 different alleles for 42 world populations. By aggregating differences in these allelic frequencies, the authors compute the F_{ST} genetic distance, which provides a comprehensive measure of genetic relatedness between any pair of 42 world populations. Using the same dataset, L. Cavalli-Sforza et al. (1994) also compute the so-called Nei distance for all population pairs. While this genetic distance measure has slightly different theoretical properties than F_{ST}, the two measures are highly correlated.

Since genetic distances are available only at the population rather than at the country level, Spolaore and Wacziarg (2009) matched the 42 populations in L. Cavalli-Sforza et al. (1994) to countries.[7] Thus, the genetic distance measures we use measure the expected genetic distance between two randomly drawn individuals, one from each country, according to the contemporary composition of the population. The key advantage of the genetic distance data relative to predicted measures of length of separation (see below) is that the measurement and imputation apply to *contemporary* populations. Thus, for example, the effects of smaller-scale migratory movements after the human exodus from Africa on the temporal distance between populations are by construction incorporated in these measures, while they are much more difficult to capture with theoretical temporal distance proxies.

[6] Such genetic measures are based on *neutral* genetic markers only, i.e., on genes which are not subject to selection pressure and only change due to random drift.

[7] To this end, the authors used ethnic composition data from Fearon (2003): the data by L. Cavalli-Sforza et al. (1994) contain information on the groups that were sampled to obtain genetic distance estimates, and these groups can be matched one-to-one to the ethnic groups that populate countries. Thus, the data from one group in L. Cavalli-Sforza et al. (1994) can be assigned to sub-populations in potentially multiple countries, so that, in principle, even the relatively small number of 42 populations is sufficient to compute genetic distances between more than 100 countries.

6.3.2.0.2 Predicted Migratory Distance Between Countries. Rather than physically *measure* the genetic composition of populations to investigate their kinship, one can also derive *predicted* migration measures (Ashraf and Galor, 2013b; Özak, 2010). Key idea behind using these variables is that populations that have lived far apart from each other (in terms of migratory, not necessarily geographic, distance), have usually spent a large portion of human history apart from each other. Notably, these data are independent of those on observed genetic distance and thus allow for an important out-of-sample robustness check.

First, the derivation of the predicted migratory distance variable of Ashraf and Galor (2013b) follows the methodology proposed in Ramachandran et al. (2005) by making use of today's knowledge of the migration patterns of our ancestors. Specifically, Ashraf and Galor (2013b) obtain an estimate of bilateral migratory distance by computing the shortest path between two countries' capitals. Given that until recently humans are not believed to have crossed large bodies of water, these hypothetical population movements are restricted to landmass as much as possible by requiring migrations to occur along five obligatory waypoints, one for each continent. By construction, these migratory distance estimates only pertain to the *native* populations of a given pair of countries. Thus, in contrast to the genetic distance measures, these distance estimates need to be adjusted to the extent that the contemporary populations in a country pair differ from the native ones. While this objective is difficult to achive for geographically scattered waves of temporally very distant events, adjustment for post-Columbian migration flows can be implemented using the "World Migration Matrix" of Putterman and Weil (2010), which describes the share of the year 2000 population in every country that has descended from people in different source countries as of the year 1500. To derive values of predicted migratory distance pertaining to the *contemporary* populations, we combine the dataset of Ashraf and Galor (2013b) with this migration matrix. Thus, the contemporary predicted migratory distance between two countries equals the weighted migratory distance between the contemporary populations.[8] Thus, this ancestry-adjusted predicted migratory distance between two countries can be thought of as the expected migratory distance between the ancestors of two randomly drawn individuals, one from each country. Further note that migratory distance and observed genetic distance tend to be highly correlated (Ramachandran et al., 2005). Ashraf and Galor (2013b) exploit this fact by linearly transforming migratory distance into a measure of predicted F_{ST} genetic distance. Our measure of predicted migratory distance might hence as well be interpreted as predicted genetic distance. Indeed, the correlation of our predicted (ancestry-adjusted) migratory distance measure with observed F_{ST} is $\rho = 0.54$.

---

[8] Formally, suppose there are N countries, each of which has one native population. Let $s_{1,i}$ be the share of the population in country 1 which is native to country i and denote by $d_{i,j}$ the migratory distance between the native populations of countries i and j. Then, the (weighted) predicted ancestry-adjusted migratory distance between countries 1 and 2 as of today is given by

$$\text{Predicted migratory distance}_{1,2} = \sum_{i=1}^{N}\sum_{j=1}^{N}(s_{1,i} \times s_{2,j} \times d_{i,j})$$

Second, as an additional independent measure of migratory distance, we use the so-called "human mobility index"-based migratory distance developed by Özak (2010). This measure is more sophisticated than the raw migratory distance using the five intermediate waypoints in that it measures the walking time along the optimal route between any two locations, taking into account the effects of temperature, relative humidity, and ruggedness, as well as human biological capabilities. Given that the procedure assumes travel by foot (as is appropriate if interest lies in migratory movements thousands of years ago), the data do not include islands, but assume that the Old World and the New World are connected through the Bering Strait, over which humans are believed to have entered the Americas. The original data contain the travel time between two countries' capitals, which we again adjust for post-Columbian migration flows using the ancestry-adjustment methodology outlined above. Thus, this variable measures the expected travel time between the ancestors of two randomly drawn individuals, one from each country.

There are strong ex ante reasons to suspect that genetic distance is a more powerful proxy for temporal distance than predicted migratory distance. First, the migratory distance measures are inherently coarse in nature. Second, as indicated above, migratory distance can only be adjusted for the post-1500 mass migratory movements that are captured in the "World Migration Matrix", but not for the more diverse migration waves earlier in history. Analyses using predicted migratory distance measures are hence more likely to suffer from attenuation bias.

6.3.2.0.3 Linguistic Distance Between Countries. Population geneticists and linguists have long noted the close correspondence between genetic distance and linguistic "trees", intuitively because population break-ups do not only produce diverging gene pools, but also differential languages. Hence, we employ the degree to which two countries' languages differ from each other as an additional proxy for the timing of separation. The construction of linguistic distances follows the methodology proposed by Fearon (2003). The Ethnologue project classifies all languages of the world into language families, subfamilies, sub-sub-families etc., which give rise to a language tree. In such a tree, the degree of relatedness between different languages can be quantified as the number of common nodes two languages share.[9] As in the case of predicted migratory distances, for each country pair, we calculate the weighted linguistic distance according to the population shares speaking a particular language in the respective countries today.

It is well-know in the population genetics and linguistics literatures that genetic distance appears to be a higher-quality measure of separation patterns. First, while languages generally maintain a certain structure over long periods of time, in some cases

[9] If two languages belong to different language families, the number of common nodes is 0. In contrast, if two languages are identical, the number of common nodes is 15. Following Fearon (2003), who argues that the marginal increase in the degree of linguistic relatedness is decreasing in the number of common nodes, we transformed these data according to

$$\text{Linguistic distance} \ = 1 - \sqrt{\frac{\#\ \text{Common nodes}}{15}}$$

to produce distance estimates between languages in the interval $[0, 1]$. We restricted the Ethnologue data to languages which make up at least 5% of the population in a given country.

they change or evolve very quickly, for example when the colonial powers brought Indo-European languages into Africa, or when Arabic was brought to Egypt during the Muslim conquest. Thus, in general, the slow-moving nature of aggregate genetic endowment makes genetic distance a more robust measure of ancient breakups of populations, also see the discussion in L. L. Cavalli-Sforza (1997). In addition, any quantitative measure of linguistic distance suffers from the fact that there is no natural metric on languages. While language trees are a useful tool to circumvent this problem, they remain coarse in nature, potentially introducing severe measurement error. Just as with migratory distance, we hence expect that genetic distance will be a more powerful explanatory variable than linguistic distance.

6.4 Preferences and Temporal Distance

6.4.1 Baseline Results

This section develops our main result on the relationship between differences in preferences between countries and the temporal distance between the respective populations. Since temporal distance is an inherently bilateral variable, this analysis will necessitate the use of a *dyadic* regression framework, which takes each possible pair of countries as unit of observation. Accordingly, we match each of the 76 countries with every other country into a total of 2,850 country pairs and, for each trait, compute the absolute difference in (average) preferences between the two countries.[10] We then relate our proxies for temporal distance to this absolute difference in preferences between the respective populations. Our regression equation is hence given by:

$$|\mathrm{pref}_i - \mathrm{pref}_j| = \alpha + \beta \times \text{temporal distance proxy}_{i,j} + \gamma_i \times d_i + \gamma_j \times d_j + \epsilon_{i,j}$$

where pref_i and pref_j represent some average trait in countries i and j, respectively, d_i and d_j country fixed effects, and $\epsilon_{i,j}$ a country pair specific disturbance term.

As is standard practice in dyadic analyses such as in gravity regressions of bilateral trade, every specification to be presented below will include country fixed effects d_i and d_j, i.e., a fixed effect for each of the two countries that appears in a country pair observation to take out any unobservables that are country-specific.[11] Heuristically speaking, with country fixed effects, the regressions do not relate, say, the raw difference in preferences between Sweden and Mexico to the respective raw genetic distance. Rather, the regression relates the difference in preferences between Sweden and Mexico *relative* to Sweden's and Mexico's average differences in preferences in all country pairs to their genetic distance, again relative to all other genetic distances involving these two countries. For instance, if Mexico had very large differences in preferences to all countries, then the fixed effects would ensure that these uniform large differences are treated as a Mexico-

---

[10] Since the analysis is not directional, each country pair is only used once, i.e., when country i is matched with country j, j cannot be matched with i, so that the bilateral dataset contains no redundant information. Due to a lack of data on genetic distance and / or migratory distance, the empirical analyses can only make use of sub-samples of the total set of 2,850 country pairs.

[11] Also see the working paper version of Spolaore and Wacziarg (2009).

specific effect, rather than attribute them to the bilateral relationships between Mexico and other countries. Thus, any country-specific factors are netted out of the analysis and the regression equation indeed estimates the bilateral effect of interest.[12]

Furthermore, regarding the noise term, because our empirical approach implies that each country will appear multiple times as part of the (in)dependent variable, we need to allow for clustering of the error terms at the country-level. We hence employ the two-way clustering strategy of Cameron et al. (2011), i.e., we cluster at the level of the first and of the second country of a given pair. This procedure allows for arbitrary correlations of the error terms within a group, i.e., within the group of country pairs which share the same first country or which share the same second country, respectively, see Appendix II of Spolaore and Wacziarg (2009).

Table 6.1 provides the results of OLS regressions of absolute differences in preferences on genetic distance. Throughout the paper, all regression coefficients (except for those of binary variables) are expressed in terms of standardized betas, i.e., both the dependent and the independent variables are normalized into z-scores and the dependent variable is then multiplied with 100, so that the coefficient can be interpreted as the percent change of a standard deviation in the dependent variable in response to a one standard deviation increase in the independent variable.

Columns (1) and (2) provide evidence that genetic distance is a significant predictor of differences in average risk attitudes. In quantitative terms, the standardized beta indicates that a one standard deviation increase in genetic distance is associated with an increase of roughly 20 percent of a standard deviation in differences in risk attitudes. In columns (3) through (8), we show that very similar results obtain for all of the prosocial traits, i.e., altruism, positive reciprocity, and trust. Given that these three traits are also positively correlated at the country-level, we keep the subsequent analysis consise by collapsing the three measures into a simple unweighted average that we refer to as "prosociality" and report robustness checks using each prososical trait separately in the Appendix.[13]

Finally, columns (9)-(10) and (11)-(12) present analogous analyses using patience and negative reciprocity as dependent variables. The resulting picture is very consistent: across specifications, the point estimates are positive, but small in magnitude, and statistically not significant. Thus, in sum, out of our six behavioral traits, four are robustly significantly related to temporal distance, while the other two are not. The remainder of this section will formally investigate the robustness of the relationship between temporal distance and preferences over risk and social interactions.

[12] The empirical results suggest that such country fixed effects indeed go a long way in addressing omitted variable concerns. For instance, in the analyses to be presented below, for patience and negative reciprocity we sometimes observe statistically significant *negative* coefficients on the temporal distance proxies, which we find very hard to interpret. These results entirely disappear with country fixed effects, providing evidence for the importance of including country fixed effects.

[13] The country-level correlations between the three measures range between 0.27 and 0.71, see Falk et al. (2015a). To derive the prosociality index, we computed a simple unweighted average of altruism, positive reciprocity, and trust at the individual level and collapsed this measure at the country-level.

Table 6.1. Preferences and temporal distance

	Dependent variable: Absolute difference in average...											
	Risk taking		Altruism		Pos. reciprocity		Trust		Patience		Neg. reciprocity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Fst genetic distance	0.19**		0.042**		0.15***		0.21***		0.045		0.012	
	(0.08)		(0.02)		(0.05)		(0.05)		(0.03)		(0.02)	
Nei genetic distance		0.17**		0.054**		0.17***		0.17***		0.046		0.012
		(0.07)		(0.03)		(0.05)		(0.05)		(0.04)		(0.02)
Observations	2701	2701	2701	2701	2701	2701	2701	2701	2701	2701	2701	2701
R^2	0.620	0.617	0.552	0.552	0.522	0.525	0.448	0.442	0.509	0.510	0.481	0.481

OLS estimates, twoway-clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.4.2 Conditional Regressions

The argument made in this paper is that the relationship between genetic distance on the one hand and risk as well as social preferences on the other hand reflects the impact of ancient migration patterns and the resulting distribution of temporal distances across populations, rather than contemporary differences in idiosyncratic country characteristics.

We hence proceed by investigating the robustness of the relationship between temporal distance and risk preferences as well as prosociality through conditional regressions. Since our dependent variable consists of absolute differences, all of our control variables will also be bilateral variables that reflect cross-country differences. In essence, in what follows, our augmented regression specification will be:

$$|pref_i - pref_j| = \alpha + \beta \times \text{genetic distance}_{i,j} + \gamma_i \times d_i + \gamma_j \times d_j + \eta \times g_{i,j} + \epsilon_{i,j}$$

where g_{ij} is a vector of bilateral measures between countries i and j (such as their geodesic distance or the absolute difference in per capita income). Details on the definitions and sources of all control variables can be found in Appendix 6.E.

We start our analysis by considering the case of risk preferences. To check that our coefficient of interest does not spuriously pick up the effect of demographic differences or differential population characteristics, column (2) of Table 6.2 adds to the baseline specification the absolute differences in average age, proportion of females, religious fractionalization, and the fraction of the population who are of European descent. This joint set of covariates reduces the point estimate of genetic distance by only about 10%, and the coefficient remains statistically significant.

A potential concern with our baseline specification is that it ignores differences in development and institutions across countries, in particular given that genetic distance is known to correlate with differences in national income (Spolaore and Wacziarg, 2009). Column (3) of Table 6.2 therefore introduces absolute differences in (log) GDP per capita, democracy, and a common legal origin dummy. The inclusion of this vector of controls has no further effect on the genetic distance coefficient.

Recall that our "world map" of risk preferences suggests the presence of geographic patterns in the distribution of risk attitudes. However, human migration patterns (and hence temporal distance proxies) are correlated with geographic and climatic variables. Thus, to ensure that effects stemming from variations in geography or climate are not attributed to genetic distance, we now condition on an exhaustive set of corresponding control variables. Column (4) introduces four distance metrics as additional controls into this regression. Our first geographical control variable consists of the geodesic distance (measuring the shortest distance between any two points on earth) between the most populated cities of the countries in a given pair. Relatedly, we introduce a dummy equal to one if two countries are contiguous. Finally, we also condition on the "distance" between two countries along the two major geographical axes, i.e., the difference in the distance to the equator and the longitudinal (east-west) distance. Again, the introduction of these variables has virtually no effect on the coefficient of genetic distance.

Table 6.2. Risk taking and genetic distance: Robustness

	Dependent variable: Absolute difference in average risk taking				
	(1)	(2)	(3)	(4)	(5)
Fst genetic distance	0.19**	0.17**	0.17**	0.17**	0.16**
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
Δ Average age		0.064*	0.10**	0.11**	0.11**
		(0.03)	(0.05)	(0.05)	(0.05)
Δ Proportion female		0.047	0.034	0.026	0.022
		(0.05)	(0.05)	(0.05)	(0.05)
Δ Religious fractionalization		0.024	0.026	0.021	0.021
		(0.03)	(0.03)	(0.03)	(0.03)
Δ % Of European descent		-0.025	-0.015	-0.0040	-0.0030
		(0.02)	(0.02)	(0.02)	(0.02)
Δ Democracy index			-0.032*	-0.038**	-0.041**
			(0.02)	(0.02)	(0.02)
Δ Log [GDP p/c PPP]			-0.075**	-0.067**	-0.066**
			(0.03)	(0.03)	(0.03)
Log [Geodesic distance]				0.087*	0.075*
				(0.04)	(0.04)
1 for contiguity				-0.046	-0.027
				(0.09)	(0.09)
Δ Distance to equator				-0.078***	-0.079**
				(0.03)	(0.04)
Δ Longitude				-0.091***	-0.084***
				(0.03)	(0.03)
Δ Land suitability for agriculture					0.026
					(0.03)
Δ Mean elevation					0.040
					(0.04)
Δ SD Elevation					-0.0087
					(0.03)
Δ Ave precipitation					0.025
					(0.05)
Δ Ave temperature					-0.0057
					(0.02)
Δ Log [Area]					0.013
					(0.02)
Colonial relationship dummies	No	No	Yes	Yes	Yes
Observations	2701	2628	2556	2556	2556
R^2	0.620	0.624	0.627	0.630	0.632

OLS estimates, twoway-clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Given that geographic distance as such does not seem to drive our result, we now control for more specific information about differences in the micro-geographic and climatic conditions between the countries in a pair. To this end, we make use of information on the agricultural productivity of land, different features of the terrain, and climatic factors. As column (4) shows, the inclusion of corresponding controls has no effect on the genetic distance point estimate. In sum, columns (4) and (5) suggest that the precise migration patterns of our ancestors, rather than simple shortest-distance calculations between contemporary populations, need to be taken into account to understand the cross-country variation in risk aversion.

Table 6.3 repeats the conditional regressions from Table 6.2 for the case of prosociality. Each column follows the same logic as the corresponding column in Table 6.2. As columns (1) through (5) show, the relationship between prosociality and genetic (temporal) distance is robust to this large and comprehensive vector of covariates. Appendix 6.B.1 shows that very similar results hold for each of the three prosocial traits separately, i.e., differences in altruism, positive reciprocity, and trust all exhibit significant conditional relationships with temporal distance.

In sum, conditioning on a large set of economic, institutional, geographic, climatic, and demographic variables, the relationships between genetic distance and differences in risk preferences and prosociality are highly significant. Furthermore, in both cases, the corresponding point estimate is very robust, suggesting that – in order for omitted variable bias to explain our results – unobservables would have to bias our results by much more than the very large and comprehensive set of covariates in our regressions (Altonji et al., 2005; Bellows and Miguel, 2009).[14]

Table 6.3. Prosociality and genetic distance: Robustness

	Dependent variable: Absolute difference in average prosociality				
	(1)	(2)	(3)	(4)	(5)
Fst genetic distance	0.22^{***}	0.23^{***}	0.23^{***}	0.24^{***}	0.24^{***}
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Population controls	No	Yes	Yes	Yes	Yes
Economic and institutional controls	No	No	Yes	Yes	Yes
Colonial relationship dummies	No	No	Yes	Yes	Yes
Distance controls	No	No	No	Yes	Yes
Geographic controls	No	No	No	No	Yes
Observations	2701	2628	2556	2556	2556
R^2	0.481	0.481	0.490	0.491	0.492

OLS estimates, twoway-clustered standard errors in parentheses. See Table 6.2 for a complete list of the control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[14] A potential concern is that genetic distance might simply pick up regional effects. Thus, we construct an extensive set of 28 continental dummies each equal to one if the two countries are from two given continents. For example, we have a dummy equal to one if one country is from Sub-Saharan Africa and the other one from North America. When we include this vector of fixed effects, we cannot condition on country fixed effects any longer because the resulting set of fixed effects would be too extensive to leave

6.4.3 Sub-Samples and Weighted Least Squares Estimates

Our sample of countries is highly heterogeneous economically. While our analyses already controlled for differences in per capita income, columns (1) and (4) of Table 6.4 restrict the sample to country pairs which are relatively similar in economic terms, i.e., which are below the median absolute difference in per capita income. In this restricted sample, genetic distance is significantly related to differences in both risk and prosocial preferences.

In many of the countries furthest from East Africa, the majority of the population is not indigenous. Our analysis addressed this aspect by employing observed genetic distance as main explanatory variable, which by construction pertains to contemporary populations. In addition, all analyses using predicted migratory distance to be presented below will make use of the ancestry-adjustment procedure to develop a meaningful representation of the temporal distance between contemporary populations. Still, to rule out that the mass migration post-1500 and its effect on temporal distances drives our results, columns (2) and (5) present the results of an additional robustness check in which we restrict the sample to countries in the Old World, i.e., we exclude Australia, the Americas, and the Caribbean. Reassuringly, the results are very similar to the baseline results.

Up to this point, all of our regressions were conducted using the raw F_{ST} genetic distance. However, since the data on allele frequencies are collected from different sample sizes across groups, the precision of the measurement varies across country pairs. Using bootstrap analysis, L. Cavalli-Sforza et al. (1994) compute standard errors of the F_{ST} genetic distance for each pair. Following the methodology proposed by Spolaore and Wacziarg (2009), we utilize this information by linearly downweighing each observation by its standard error (see Appendix 6.C). The results from the corresponding weighted least squares regression are presented in columns (3) and (6), respectively. The resulting standardized beta coefficients are highly significant and very similar to those from the unweighted regressions.

Finally, Appendix 6.B.3 presents an extensive set of robustness checks in which we restrict the sample by excluding observations from the left or right tail of the distributions of genetic distance and risk taking (prosociality). These analyses show that the relationships between differences in risk taking and prosociality on the one hand and genetic distance on the other hand is not driven by outliers.

6.4.4 Alternative Temporal Distance Proxies

So far, our analysis has made use of genetic distance as theoretically most appealing proxy for temporal distance. We now extend our analysis by employing predicted migratory distance, predicted HMI migratory distance, and linguistic distance as three additional (and conceptually slightly different) explanatory variables. Columns (1) through (6) of Table 6.5 describe the relationship between differences in risk preferences and temporal distance, while columns (7) through (12) analyze the effect of temporal distance on

meaningful variation to identify our coefficient of interest off. Appendix 6.B.2 presents the results, which are very similar to those using country fixed effects.

Table 6.4. Robustness: Sub-samples and weighted least squares

	Dependent variable: Abs. difference in average...					
	Risk taking			Prosociality		
	Δ GDP p/c < 50th pct	Old World	WLS	Δ GDP p/c < 50th pct	Old World	WLS
	(1)	(2)	(3)	(4)	(5)	(6)
Fst genetic distance	0.19**	0.23**	0.18**	0.20***	0.19***	0.23***
	(0.08)	(0.09)	(0.09)	(0.05)	(0.05)	(0.06)
Observations	2475	1653	2701	2475	1653	2701
R^2	0.619	0.602	0.613	0.488	0.442	0.488

OLS estimates, twoway-clustered standard errors in parentheses. In columns (1) and (4), the sample only includes country pairs whose difference in per capita income is below the median in our sample. In columns (2) and (5), the sample is restricted to countries in the Old World, i.e., we exclude Australia, the Americas, and the Caribbean. In columns (3) and (6), all pairwise observations are linearly downweighted by their standard error, see Appendix 6.C. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

differences in prosociality. For each dependent and explanatory variable, we present two specifications, one without controls (except for country fixed effects), and one including the full vector of controls from column (5) in Table 6.2.

Overall, the results provide evidence that the relationship between temporal distance and preferences extends beyond genetic distance: all proxy variables are positively related to differences in preferences, showing that our results do not hinge on genetic distance as proxy for temporal distance.[15] At the same time, consistent with increased measurement error in migratory and linguistic distance, the corresponding standardized beta coefficients are almost always smaller than the ones of genetic distance discussed above.[16]

[15] Appendix 6.B.1 shows that very similar results obtain when we use each of the three prosocial traits altruism, positive reciprocity, and trust separately in the conditional regressions.

[16] Unreported regressions show that when we use non-ancestry adjusted migratory distance measures in the regressions (as opposed to the ancestry adjusted variables used throughout this paper), the coefficients are either not statistically significant, or only marginally so, again suggesting that the precise migration patterns of our ancestors need to be taken into account to understand the cross-country variation in risk aversion and prosociality.

Table 6.5. Robustness: Alternative temporal distance proxies

	Dependent variable: Absolute difference in average...											
	Risk taking						Prosociality					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Migratory distance	0.16*	0.15					0.15**	0.22**				
	(0.09)	(0.10)					(0.06)	(0.09)				
HMI migratory distance			0.30***	0.35***					0.15**	0.20**		
			(0.12)	(0.13)					(0.06)	(0.09)		
Linguistic distance					0.089*	0.063*					0.072**	0.080**
					(0.05)	(0.04)					(0.03)	(0.04)
Population controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Economic and institutional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Colonial relationship dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Distance controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Geographic controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2775	2556	2211	2016	2850	2556	2775	2556	2211	2016	2850	2556
R^2	0.617	0.625	0.634	0.641	0.616	0.625	0.479	0.477	0.531	0.531	0.478	0.475

OLS estimates, twoway-clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.5 Preferences and the Length of the Migratory Path

The previous section established a bilateral statement: higher temporal distance is associated with larger differences in preferences. However, the economics literature that makes use of concepts from population genetics considers not only interpopulation temporal (genetic) distance, but also intrapopulation genetic diversity (Ashraf and Galor, 2013b). Whenever a sub-population split apart from its parental colony, those humans breaking new ground took with them only a fraction of the genetic diversity of the previous genetic pool, intuitively because they were usually small, and hence non-representative, samples. In consequence, through the sequence of successive fissions (serial founder effect), the total diversity of the gene pool significantly decreases along human migratory routes out of East Africa. Ashraf and Galor (2013b) make use of this observation by constructing a predicted measure of genetic diversity, which is essentially a linear transformation of migratory distance from East Africa. In light of the previous findings, the question emerges whether the *level* of a given preference is related to genetic diversity, i.e., migratory distance from Ethiopia. Understanding this relationship is interesting for two reasons.

First, Ashraf and Galor (2013b) provide evidence that national per capita income is hump-shaped in predicted genetic diversity. Thus, an analysis of the relationship between preferences and genetic diversity might produce insights into whether – or to what extent – preferences mediate the observed relationship between diversity and income.

Second, analyzing the relationship between the level of preferences and migratory distance from the cradle of mankind might also shed light on the mechanisms that drive the patterns reported in Section 4.4. The argument underlying this paper is that the relationship between temporal distance and preference differences reflects the accumulation of population-specific preference changes (through historical experiences and / or genetic drift) over thousands of years. However, it is also conceivable that this relationship is not driven by what happened *after* the population breakups, but rather by the structure of the breakups itself. In particular, it may be that the characteristics of the new founder population systematically differed from those of the parental colony, as would be the case if, e.g., only the least risk averse types split away. In such a scenario, preferences would change *monotonically* along the migratory path out of East Africa, hence mechanically producing the correlation between temporal distance and preference differences.[17] If true, this would still leave the main insight of the paper – that the structure and timing of population breakups in the very distant past have left in the footprint in the contemporary global distribution of preferences – intact. At the same time, the interpretation of this relationship would change slightly.

[17] Slightly more subtly, it is also possible that the correlation between temporal distance and preference differences is driven by a monotonic evolution of the *dispersion* of the preference pool along the migratory path out of Africa, akin to the serial founder effect in population genetics: if the dispersion of the preference pool decreased monotonically along the migratory path, the differences in preferences between later founder populations would mechanically be smaller than those between earlier ones because the respective parental preference pool has lower variation to begin with. To investigate this issue, Appendix 6.D.2 presents regressions which relate the standard deviation of the country-level preference pool to genetic diversity. Across preferences, the relationship between preference dispersion and genetic diversity is weak.

Because we only observe preferences today, we cannot evaluate whether such systematic population breakups actually took place. However, what is sufficient for our purposes is to investigate whether the results of such systematic breakups are still visible in the data today and hence potentially drive our result on the relationship between temporal distance and differences in preferences. To this end, we relate the level of the preference pool in a given country to the ancestry-adjusted length of the migratory path of the respective population from Ethiopia Ashraf and Galor (2013b):

$$\text{pref}_i = \alpha + \beta_1 \times \text{genetic diversity}_i + \beta_2 \times \text{genetic diversity sqr}_i + \gamma \times x_i + \epsilon_i$$

where pref_i is either the average trait in country i, x_i is a vector of covariates, and ϵ_i a disturbance term. Note that this regression does *not* constitute a special case of the bilateral migratory distance regressions discussed above, because here the dependent variable is the *level* of a given preference, rather than the absolute difference to East Africa, i.e., Ethiopia. Thus, the regressions estimated above do not imply any prediction on the sign or significance of β_1 and β_2.[18]

Table 6.6 provides an overview of the results. As explanatory variable we employ predicted genetic diversity as developed in Ashraf and Galor (2013b). Columns (1), (4), (7), and (10) show that none of our preference variables are significantly linearly related to genetic diversity. At the same time, as shown in columns (2), (5), and (8), risk taking, patience, and prosociality all exhibit significant non-linear relationships with genetic diversity, i.e., risk aversion, patience, and prosociality are all hump-shaped in genetic diversity.[19] Indeed, Appendix 6.D.1 shows that the non-linear associations between prosociality and genetic diversity also hold for altruism, positive reciprocity, and trust separately.[20] However, all of these significant non-linearities disappear with the inclusion of continent fixed effects (columns (3), (6), and (9)).[21]

These results show that – at least as of today – preferences do not evolve monotonically along the migratory path out of Africa. This pattern is indicative that the relationship between temporal distance and preference differences is indeed driven by what happened *after* the various population breakups, rather than selective breakup patterns as such.

Ashraf and Galor (2013b) and Ashraf et al. (2014) establish a significant hump-shaped relationship between per capita income and genetic diversity. The non-linear rela-

---

[18] A special case of the general bilateral regression framework estimated in Section 4.4 would be

$$|\text{pref}_i - \text{pref}_{Ethiopia}| = \alpha + \beta \times \text{genetic diversity}_i + \gamma \times |x_i - x_{Ethiopia}| + \epsilon_i$$

Since Ethiopia is not included in the Global Preference Survey, we cannot estimate this equation.

[19] The linear and squared genetic diversity coefficients are also jointly statistically significant.

[20] Again, this hump is not inconsistent with the results presented in Section 4.4 because here the dependent variable is not the absolute difference to Ethiopia.

[21] The fact that the non-linear patterns disappear with continent fixed effects is perhaps unsurprising given that a large fraction of the worldwide variation in genetic diversity is intercontinental rather than intracontinental in nature, which is a result of major population bottlenecks at key intercontinental waypoints during the "out of Africa" migration process. For instance, in our sample of countries, continent fixed effects alone explain 66% of the variation in genetic diversity. In addition, our sample size of 74 countries is substantially smaller than the one in, e.g., Ashraf and Galor (2013b), implying that we have less degrees of freedom to establish potential non-linear effects also within continents. Consistent with both of these arguments, the standard error on genetic diversity and its square ususally increases after the inclusion of continent fixed effects.

Table 6.6. Average preferences and genetic diversity

	Dependent variable: Average ...											
	Risk taking			Prosociality			Patience			Neg. reciprocity		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Genetic diversity	2.08*	-74.1**	52.9	-1.27	153.0***	76.4	-0.14	127.1**	73.5	1.47	31.5	-78.4*
	(1.15)	(33.86)	(60.18)	(1.68)	(56.60)	(64.27)	(1.08)	(57.46)	(56.00)	(1.03)	(35.86)	(42.51)
Genetic diversity sqr.		53.9**	-39.0		-109.3***	-54.8		-90.1**	-55.3		-21.3	57.5*
		(24.37)	(43.64)		(40.07)	(46.57)		(40.44)	(40.55)		(25.40)	(31.03)
Continent FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	74	74	74	74	74	74	74	74	74	74	74	74
R^2	0.038	0.074	0.309	0.009	0.107	0.279	0.000	0.067	0.317	0.023	0.029	0.191

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

tionships between risk, time, and social preferences and genetic diversity hence raise the question of whether preferences somehow mediate the observed reduced-form relationship between diversity and income. To address this question, Appendix 6.D.3 presents a set of regressions in which we relate per capita income to genetic diversity and its quare, and then successively introduce our preference variables as additional covariates. The results show that the inclusion of all preferences halves the point estimate of both the linear and the squared genetic diversity term, which is mostly driven by the strong correlation between patience and per capita income (Falk et al., 2015a). However, while the genetic diversity coefficients become smaller in magnitude, they remain statistically highly significant.

6.6 Conclusion

A growing body of empirical work highlights the importance of heterogeneity in risk, time, and social preferences for understanding a myriad of economic, social, and health behaviors. This paper takes a first step towards understanding the deep roots of variation in these preferences across entire populations. Our main contribution is to establish that a significant fraction of the substantial between-country heterogeneity in risk aversion, altruism, positive reciprocity, and trust has its historical origins in the structure and timing of very distant ancestral migration patterns, which highlights that if we aim to understand the ultimate roots of preference heterogeneity, we might have to consider events very far back in time.

In this respect, this paper did not attempt to shed light on the mechanisms through which temporal distance might drive preference differences. A priori, the observed patterns are consistent with both differential historical experiences and genetic drift. An important question for future research is which role socialization practices have played in these long-run processes.

An interesting question is why the relationship between temporal distance and preferences holds for risk aversion and all of the prosocial traits, but not for patience and negative reciprocity. A possible conjecture is that the roots of the cross-country variation in time preferences are more recent in nature, rather than being driven by (the accumulation of) very old historical events. Thus, an important avenue for future research is to identify the more proximate mechanisms that have shaped the distribution of time preference and negative reciprocity.

References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003): "Fractionalization." *Journal of Economic Growth*, 8 (2), 155–194. [306]

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013): "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics*, 128 (2), 469–530. [272]

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113 (1), 151–184. [271, 286]

Arbatli, Cemal Eren, Quamrul Ashraf, and Oded Galor (2013): "The Nature of Civil Conflict." *Working Paper*. [272]

Ashraf, Quamrul and Oded Galor (2013a): "Genetic Diversity and the Origins of Cultural Fragmentation." *American Economic Review: Papers & Proceedings*, 103 (3), 528–33. [272]

Ashraf, Quamrul and Oded Galor (2013b): "The Out of Africa Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review*, 103 (1), 1–46. [270–272, 279, 290, 291, 305, 306]

Ashraf, Quamrul, Oded Galor, and Marc P. Klemp (2014): "The Out of Africa Hypothesis of Comparative Development Reflected by Nighttime Light Intensity." *Working Paper*. [272, 291]

Bellows, John and Edward Miguel (2009): "War and Local Collective Action in Sierra Leone." *Journal of Public Economics*, 93 (11), 1144–1157. [286]

Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger (2014): "Violence and Risk Preference: Experimental Evidence from Afghanistan." *American Economic Review*, 104 (1), 123–148. [270, 273]

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011): "Robust Inference with Multiway Clustering." *Journal of Business & Economic Statistics*, 29 (2). [282]

Cavalli-Sforza, Luigi L. (1997): "Genes, Peoples, and Languages." *Proceedings of the National Academy of Sciences*, 94 (15), 7719–7724. [281]

Cavalli-Sforza, Luigi, Paolo Menozzi, and Alberto Piazza (1994): *The History and Geography of Human Genes*. Princeton university press. [270, 278, 287]

Cesarini, David, Christopher T. Dawes, James H. Fowler, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace (2008): "Heritability of Cooperative Behavior in the Trust Game." *Proceedings of the National Academy of Sciences*, 105 (10), 3721–3726. [270]

Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace (2009): "Genetic Variation in Preferences for Giving and Risk Taking." *Quarterly Journal of Economics*, 124 (2), 809–842. [270, 273]

Cesarini, David, Magnus Johannesson, Patrik K.E. Magnusson, and Björn Wallace (2012): "The Behavioral Genetics of Behavioral Anomalies." *Management Science*, 58 (1), 21–34. [270]

Chen, M. Keith (2013): "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets." *American Economic Review*, 103 (2), 690–731. [272]

Desmet, Klaus, Shlomo Weber, and Ignacio Ortuño-Ortín (2009): "Linguistic Diversity and Redistribution." *Journal of the European Economic Association*, 7 (6), 1291–1318. [272]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2012): "The Intergenerational Transmission of Risk and Trust Attitudes." *Review of Economic Studies*, 79 (2), 645–677. [273]

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2015a): "The Nature and Predictive Power of Preferences: Global Evidence." *Working Paper*. [270, 274, 275, 282, 293, 296]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde (2015b): "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Working Paper*. [275]

Fearon, James D (2003): "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, 8 (2), 195–222. [278, 280]

Fehr, Ernst and Karla Hoff (2011): "Introduction: Tastes, Castes and Culture: the Influence of Society on Preferences." *Economic Journal*, 121 (556), F396–F412. [269]

Galor, Oded and Ömer Özak (2014): "The Agricultural Origins of Time Preference." *Working Paper*. [272]

Giuliano, Paola, Antonio Spilimbergo, and Giovanni Tonon (2014): "Genetic Distance, Transportation Costs, and Trade." *Journal of Economic Geography*, 14 (1), 179–198. [272]

Gorodnichenko, Yuriy and Gerard Roland (2010): "Culture, Institutions and the Wealth of Nations." *Working Paper*. [272]

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2009): "Cultural Biases in Economic Exchange?" *Quarterly Journal of Economics*, 124 (3), 1095–1131. [272]

Henn, Brenna M., Luigi L. Cavalli-Sforza, and Marcus W. Feldman (2012): "The Great Human Expansion." *Proceedings of the National Academy of Sciences*, 109 (44), 17758–17764. [273]

Kosse, Fabian, Thomas Deckers, Hannah Schildberg-Hörisch, and Armin Falk (2014): "Formation of Human Prosociality: Causal Evidence on the Role of Sociocultural Environment." *Working Paper*. [270, 273]

Michalopoulos, Stelios (2012): "The Origins of Ethnolinguistic Diversity." *American Economic Review*, 102 (4), 1508–1539. [306]

Nordhaus, William D. (2006): "Geography and Macroeconomics: New Data and New Findings." *Proceedings of the National Academy of Sciences of the United States of America*, 103 (10), 3510–3517. [306]

Nunn, Nathan and Leonard Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review*, 101 (7), 3221–52. [272]

Özak, Ömer (2010): "The Voyage of Homo-Oeconomicus: Some Economic Measures of Distance." *Working Paper*. [270, 271, 279, 280, 305]

Putterman, Louis and David N. Weil (2010): "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality." *Quarterly Journal of Economics*, 125 (4), 1627–1682. [279, 306]

Ramachandran, Sohini, Omkar Deshpande, Charles C. Roseman, Noah A. Rosenberg, Marcus W. Feldman, and L. Luca Cavalli-Sforza (2005): "Support from the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa." *Proceedings of the National Academy of Sciences of the United States of America*, 102 (44), 15942–15947. [279]

Rao, Gautam (2013): "Familiarity Does Not Breed Contempt: Diversity, Discrimination and Generosity in Delhi Schools." *Working Paper*. [270, 273]

Spolaore, Enrico and Romain Wacziarg (2009): "The Diffusion of Development." *Quarterly Journal of Economics*, 124 (2), 469–529. [270, 272, 278, 281, 282, 284, 287, 301, 305]

Spolaore, Enrico and Romain Wacziarg (2011): "Long-Term Barriers to the International Diffusion of Innovations." In *NBER International Seminar on Macroeconomics 2011*. University of Chicago Press, 11–46. [272]

Spolaore, Enrico and Romain Wacziarg (2014): "War and Relatedness." *Working Paper*. [272]

Spolaore, Enrico and Romain Wacziarg (2015): "Ancestry, Language and Culture." *Working Paper*. [272]

Tabellini, Guido (2008): "Institutions and Culture." *Journal of the European Economic Association*, 6 (2-3), 255–294. [272]

Voigtländer, Nico and Hans-Joachim Voth (2012): "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany." *Quarterly Journal of Economics*, 127 (3), 1339–1392. [272]

Appendix 6.A Details on Global Preference Survey

See Falk et al. (2015a).

Appendix 6.B Additional Bilateral Regressions

6.B.1 Prosociality Variables Separately

Table 6.7. Prosociality and temporal distance: Robustness

	Dependent variable: Absolute difference in average...														
	Altruism					Positive reciprocity					Trust				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Fst genetic distance	0.042**	0.049**	0.049**	0.044	0.044	0.15***	0.15***	0.15***	0.17***	0.17***	0.21***	0.23***	0.21***	0.21***	0.21***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Population controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Economic and institutional controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Colonial relationship dummies	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Distance controls	No	No	No	Yes	Yes	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Geographic controls	No	No	No	No	Yes	No	No	No	No	Yes	No	No	No	No	Yes
Observations	2701	2628	2556	2556	2556	2701	2628	2556	2556	2556	2701	2628	2556	2556	2556
R^2	0.552	0.549	0.551	0.551	0.552	0.522	0.519	0.521	0.523	0.524	0.448	0.449	0.465	0.465	0.467

OLS estimates, twoway-clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.B.2 Continent Fixed Effects

Table 6.8. Preferences and temporal distance: Bilateral continent fixed effects

	Dependent variable: Absolute difference in average...			
	Risk taking	Altruism	Pos. reciprocity	Trust
	(1)	(2)	(3)	(4)
Fst genetic distance	0.17**	0.20***	0.33***	0.15*
	(0.07)	(0.07)	(0.08)	(0.09)
Population controls	Yes	Yes	Yes	Yes
Economic and institutional controls	Yes	Yes	Yes	Yes
Colonial relationship dummies	Yes	Yes	Yes	Yes
Distance controls	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes
Continent FE	Yes	Yes	Yes	Yes
Observations	2556	2556	2556	2556
R^2	0.318	0.105	0.147	0.207

OLS estimates, twoway-clustered standard errors in parentheses. The regressions do not include country fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.B.3 Excluding Tail Observations

Tables 6.9 and 6.10 present the results of regressions in which we restrict the sample of observations by excluding observations from the left or right tail of the distributions of genetic distance, risk taking, or prosociality. Specifically, the regressions either utilize observations below the 90th percentile or above the 10th percentile of the distribution of a given variable.

Note that location of any given country pair in the distribution of all bilateral variables may depend on whether country fixed effects are taken into account. For instance, if country A had very large genetic distances to all but one countries, and an average distance to country B, then restricting the sample by genetic distance would never exclude the $A-B$ observation. However, with country fixed effects, this may change, because (heuristically speaking) the fixed effects for country A take out the relatively large average genetic distance for country A, implying that the $A-B$ pair has a very small genetic distance in terms of residuals. Thus, after accounting for country fixed effects, this observation might get excluded based on the above sample restriction criteria. Thus, we apply our robustness exercises to both types of distributions, i.e., to the distributions of raw variables and the distributions of residuals, see the tablenotes for further details.

Table 6.9. Risk preferences and temporal distance: Excluding small and large values

	Dependent variable: Absolute difference in average risk taking							
	Δ Risk > 10th pct		Δ Risk < 90th pct		Genetic dist. > 10th pct		Genetic dist. < 90th pct	
	Raw	Residual	Raw	Residual	Raw	Residual	Raw	Residual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fst genetic distance	0.18**	0.071*	0.20***	0.20***	0.20**	0.16**	0.19**	0.19**
	(0.07)	(0.04)	(0.08)	(0.07)	(0.08)	(0.06)	(0.08)	(0.08)
Observations	2430	2430	2430	2430	2428	2430	2430	2430

OLS estimates, twoway-clustered standard errors in parentheses. In all columns, the sample is restricted to observations above or below a certain threshold, where the threshold is either computed with or without country fixed effects. For instance, in columns (1), the sample includes all observations whose absolute difference in risk taking is above the 90th percentile of the distribution of (raw) absolute differences in risk taking. In column (2), the sample includes all observations whose absolute difference in risk taking is above the 90th percentile of the distribution of residual absolute differences in risk taking after taking out country fixed effects. That is, we first regress absolute differences in risk taking on a vector of country fixed effects, compute the residual, and then restrict the sample based on the residuals. Likewise, in column (7), we restrict the sample to observations below the 90th percentile of the distribution of (raw) Fst genetic distances, while column (8) applies the 90th percentile to the distribution of genetic distances after accounting for country fixed effects, i.e., after regressing genetic distance on country fixed effects and computing residuals. All regressions include country fixed effects: the "raw" regressions are standard fixed effects regressions; the "residual" regressions are estimated by (i) partialing country fixed effects out of differences in risk taking and genetic distance (on the full sample), (ii) restricting the sample, (iii) regressing residuals on each other. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6.10. Prosociality and temporal distance: Excluding small and large values

	Dependent variable: Absolute difference in average prosociality							
	Δ Social > 10th pct		Δ Social < 90th pct		Genetic dist. > 10th pct		Genetic dist. < 90th pct	
	Raw	Residual	Raw	Residual	Raw	Residual	Raw	Residual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fst genetic distance	0.20***	0.16***	0.19***	0.17***	0.22***	0.21**	0.22***	0.20***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)	(0.10)	(0.05)	(0.05)
Observations	2430	2430	2430	2430	2428	2430	2430	2430

OLS estimates, twoway-clustered standard errors in parentheses. In all columns, the sample is restricted to observations above or below a certain threshold, where the threshold is either computed with or without country fixed effects. For instance, in columns (1), the sample includes all observations whose absolute difference prosociality is above the 90th percentile of the distribution of (raw) absolute differences in prosociality. In column (2), the sample includes all observations whose absolute difference in prosociality is above the 90th percentile of the distribution of residual absolute differences in prosociality after taking out country fixed effects. That is, we first regress absolute differences in prosociality on a vector of country fixed effects, compute the residual, and then restrict the sample based on the residuals. Likewise, in column (7), we restrict the sample to observations below the 90th percentile of the distribution of (raw) Fst genetic distances, while column (8) applies the 90th percentile to the distribution of genetic distances after accounting for country fixed effects, i.e., after regressing genetic distance on country fixed effects and computing residuals. All regressions include country fixed effects: the "raw" regressions are standard fixed effects regressions; the "residual" regressions are estimated by (i) partialing country fixed effects out of differences in prosociality and genetic distance (on the full sample), (ii) restricting the sample, (iii) regressing residuals on each other. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 6.C Details for WLS Regressions

Following Spolaore and Wacziarg (2009), the regression weights in Table 6.4 were computed as follows:

$$\text{Weight} = \frac{\text{Maximum standard error } + 1 - \text{standard error of observation}}{\text{Maximum standard error}}$$

Thus, the regression weights are between zero and one and linearly downweigh observations with a high standard error for genetic distance.

Appendix 6.D Additional Diversity Regressions

6.D.1 Prosocial Traits Separately

Table 6.11. Preferences and genetic diversity: Prosocial traits separately

	Dependent variable: Average ...								
	Altruism			Positive reciprocity			Trust		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Genetic diversity	-1.14	98.3**	91.2	-1.15	107.2**	25.8	-0.36	116.5***	43.6
	(1.31)	(42.84)	(57.65)	(1.50)	(52.75)	(65.35)	(1.23)	(42.37)	(40.67)
Genetic diversity sqr.		-70.4**	-65.9		-76.7**	-18.7		-82.8***	-30.5
		(30.46)	(41.69)		(37.31)	(47.64)		(30.29)	(29.54)
Continent FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	74	74	74	74	74	74	74	74	74
R^2	0.009	0.056	0.181	0.009	0.067	0.174	0.001	0.101	0.383

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6.12. Preferences and genetic diversity: Prosocial traits separately

	Dependent variable: SD in ...								
	Altruism			Positive reciprocity			Trust		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Genetic diversity	-0.92**	-11.6	-2.89	-0.30	-40.3*	-10.9	0.019	-17.4	19.2
	(0.43)	(15.33)	(17.54)	(0.61)	(23.91)	(25.57)	(0.37)	(13.97)	(12.74)
Genetic diversity sqr.		7.56	1.53		28.3*	7.56		12.4	-13.9
		(10.95)	(12.60)		(16.80)	(18.20)		(9.97)	(9.24)
Continent FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	74	74	74	74	74	74	74	74	74
R^2	0.055	0.061	0.194	0.006	0.073	0.245	0.000	0.020	0.276

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.D.2 Diversity and the Dispersion of the Preference Pool

Table 6.13. Preference variability and genetic diversity

	Dependent variable: Standard deviation in ...											
	Risk taking			Prosociality			Patience			Neg. reciprocity		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Genetic diversity	0.67*	-8.30	8.98	-0.56	-24.9	-8.04	-1.09*	39.2	-1.65	-0.44	-9.65	23.2
	(0.37)	(11.08)	(14.20)	(0.47)	(16.80)	(19.22)	(0.57)	(23.57)	(23.85)	(0.33)	(11.45)	(16.68)
Genetic diversity sqr.		6.36	-6.63		17.2	5.00		-28.6*	-0.25		6.53	-17.1
		(7.88)	(10.29)		(11.86)	(13.73)		(16.83)	(17.47)		(8.25)	(12.22)
Continent FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	74	74	74	74	74	74	74	74	74	74	74	74
R^2	0.044	0.050	0.102	0.030	0.071	0.177	0.038	0.075	0.258	0.014	0.019	0.140

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.D.3 National Income, Genetic Diversity, and Preferences

Table 6.14. Prosociality and temporal distance: Robustness

	Dependent variable: Log [GDP p/c]					
	(1)	(2)	(3)	(4)	(5)	(6)
Genetic diversity	762.9***	752.9***	779.2***	455.1***	725.0***	341.3**
	(195.93)	(196.46)	(201.82)	(149.47)	(178.84)	(150.80)
Genetic diversity sqr.	-542.2***	-534.9***	-553.9***	-323.9***	-516.6***	-242.2**
	(139.55)	(139.84)	(143.68)	(107.29)	(127.53)	(108.24)
Risk taking		-0.14				-1.22**
		(0.50)				(0.49)
Prosociality			-0.11			-0.12
			(0.41)			(0.38)
Patience				2.42***		2.61***
				(0.26)		(0.28)
Negative reciprocity					1.21**	0.57
					(0.49)	(0.51)
Observations	74	74	74	74	74	74
R^2	0.137	0.138	0.138	0.444	0.181	0.494

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 6.E Definitions and Data Sources of Main Variables

6.E.1 Explanatory Variables

6.E.1.0.1 Fst and Nei genetic distance. Genetic distance between contemporary populations, taken from Spolaore and Wacziarg (2009).

6.E.1.0.2 Linguistic distance. Weighted linguistic distance between contemporary populations. Derived from the Ethnologue project data, taking into account all languages which are spoken by at least 5% of the population in a given country.

6.E.1.0.3 Predicted migratory distance. Predicted migratory distance between two countries' capitals, along a land-restricted way through five intermediate waypoints (one on each continent). Taken from Ashraf and Galor (2013b).

6.E.1.0.4 HMI migratory distance. Walking time between two countries' capitals in years, taking into account topographic, climatic, and terrain conditions, as well as human biological abilities. Data from Özak (2010).

6.E.2 Covariates

6.E.2.0.1 Average age, proportion female. Computed from Gallup's sociodemographic background data.

6.E.2.0.2 Religious fractionalization. Indices due to Alesina et al. (2003) capturing the probability that two randomly selected individuals from the same country will be from different religious / linguistic groups.

6.E.2.0.3 Percentage of European descent. Constructed from the "World Migration Matrix" of Putterman and Weil (2010).

6.E.2.0.4 Contemporary national GDP per capita. Average annual GDP per capita over the period 2001 – 2010, in 2005US$. Source: World Bank Development Indicators.

6.E.2.0.5 Democracy index. Index that quanties the extent of institutionalized democracy, as reported in the Polity IV dataset. Average from 2001 to 2010.

6.E.2.0.6 Colonial relationship dummies. Taken from the CEPII Geodist database at http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=6.

6.E.2.0.7 Geodesic distance, contiguity, longitude, latitude, area Taken from CEPII GeoDist database. The longitudinal distance between two countries is computed as

$$\text{Longitudinal distance} \; = \; min\{|longitude_i - longitude_j|, 360 - |longitude_i| - |longitude_j|\}$$

6.E.2.0.8 Suitability for agriculture. Index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH, taken from Michalopoulos (2012).

6.E.2.0.9 Mean and standard deviation of elevation. Mean elevation in km above sea, taken from Ashraf and Galor (2013b). Data originally based on geospatial elevation data reported by the G-ECON project (Nordhaus, 2006).

6.E.2.0.10 Precipitation. Average monthly precipitation of a country in mm per month, 1961-1990, taken from Ashraf and Galor (2013b). Data originally based on geospatial average monthly precipitation data for this period reported by the G-ECON project (Nordhaus, 2006).

6.E.2.0.11 Temperature. Average monthly temperature of a country in degree Celsius, 1961-1990, taken from Ashraf and Galor (2013b). Data originally based on geospatial average monthly temperature data for this period reported by the G-ECON project (Nordhaus, 2006).

7

Patience and the Wealth of Nations*

7.1 Introduction

Time preference forms a key building block of all intertemporal choice theories. Accordingly, in recent years, economists have begun to measure individuals' patience and to empirically relate the resulting preference parameters to economically important behavior. Two insights derived from these efforts are that the elicited preference parameters correlate with a broad class of economic decisions and outcomes, often consistent with theories of intertemporal choice, and that time preferences exhibit pronounced heterogeneity across samples, but also even within fairly homogenous populations. As a consequence, patience has been stressed as a crucial non-cognitive skill in determining economic success, be it on the labor market or generally in life (see Borghans et al., 2008, for a review).

However, at a more aggregate level, time preference also constitutes a key primitive of dynamic micro-founded theories of comparative development. Given that any stock of production factors or knowledge necessarily arises from an accumulation process, a broad class of models posits that the time preferences of a country's representative agent are intimately linked to national income through the accumulation of human and physical capital as well as productivity improvements. At the same time, empirical evidence on the importance of global heterogeneity in time preferences for (individual or aggregate) accumulation processes and development is missing.

The present paper fills this gap by systematically studying the relationship between time preferences and future-oriented behaviors on a global scale, across individuals, subnational regions, and countries. The analysis makes use of a novel data set, the Global Preference Survey, which constitutes the first comparable data set on time preferences at the individual level from representative population samples for a large set of countries all over the world (Falk et al., 2015a). The data contain information on patience for more than 80,000 individuals from 76 countries. The sample is constructed to provide repre-

* For valuable comments and discussions we are grateful to Daron Acemoglu, Klaus Desmet, Oded Galor, Luigi Guiso, Robert Hall, Ömer Özak, Felipe Valencia, Nico Voigtländer, Joachim Voth, David Weil, David Yang, Fabrizio Zilibotti as well as seminar audiences at the 2015 NBER Political Economy Meeting, Columbia, IFN Stockholm, Konstanz, Stanford, and UCLA Anderson. Ammar Mahran provided outstanding research assistance.

sentative population samples within each country and geographical representativeness in terms of countries covered. The data contain information on two patience measures. One measure has a format similar to the standard procedure of eliciting time preferences in laboratory experiments, i.e., respondents were asked to make a series of hypothetical binary decisions between receiving monetary rewards today or in the future. The other measure elicits subjective evaluations of patience. These preference measures were selected and tested through a rigorous ex ante experimental validation procedure involving real monetary stakes, so that the survey items have a demonstrated ability to capture actual heterogeneity in intertemporal choice behavior as measured with financial incentives. To ensure comparability of preference measures across countries, the elicitation followed a standardized protocol that was implemented through the professional infrastructure of the Gallup World Poll. Monetary stakes involved comparable values in terms of purchasing power across countries, and the survey items were culturally neutral and translated using state-of-the-art procedures. Thus, the data provide an ideal basis for the first systematic analysis of the relationship between patience and future-oriented decisions on a global scale.

Our empirical analysis is based on the relationship between patience, accumulation processes, and income as predicted by a standard micro-founded development framework. For instance, in a Ramsey-Cass-Koopmans model, higher patience implies a higher propensity to save, a higher steady state level of physical capital and income, as well as faster growth along the convergence path towards the steady state.[1] Likewise, in the context of human capital theory, patience implies greater incentives to acquire education (G. S. Becker, 1962; Ben-Porath, 1967). In terms of residual productivity, endogenous growth theory suggests that higher patience raises the present value of R&D and thus research intensity and innovation (Romer, 1990; Aghion and Howitt, 1992).[2] It is also conceivable that higher patience leads to the design of better institutions. For example, if people face a tradeoff between creating an institutional environment suitable for sustained development and engaging in short-run rent extraction, time preferences will affect the design of these institutions. The global representativeness of our data allows us to present the first assessment of the consistency of the correlations predicted by this large and influential body of theory with the empirical facts.

Given that the dynamic neoclassical development paradigm derives its empirical content from microeconomic decisions, our empirical analysis starts by investigating the relationship between individual patience, human and physical capital, and income. We find that, in representative samples around the world, patient people have higher incomes, save more, have higher educational attainment, and report better health. These results hold within countries and subnational regions, and conditional on a comprehensive set of sociodemographic covariates (also see Falk et al., 2015a).

In light of the strong individual-level results, we proceed by considering the relationship between patience, accumulation processes, and income at a more aggregate level while keeping a subnational perspective. To this end, we exploit variation in average pa-

[1] Higher patience also implies faster growth in a human capital augmented model (Lucas, 1988).

[2] See also Acemoglu (2008) for a comprehensive overview of the role of time preferences for growth. The relation between income, income growth and patience is reinforced in a setting in which patience increases with the level of wealth, see, e.g., Strulik (2012).

tience, average educational attainment, and per capita income across subnational regions within countries, akin to the approach by Gennaioli et al. (2013). While the corresponding regressions investigate the correlates of patience at an aggregate level, as called for by dynamic development theories, they also allow us to keep many factors such as the overall institutional environment constant by including country fixed effects. Consistent with the individual-level results, we find robust evidence that, within countries, regions with more patient populations exhibit higher average educational attainment and higher per capita income.

Building on these subnational results, the main part of the paper analyzes the interplay between patience, the proximate determinants of development, and income at the country level. We first establish a strong raw correlation between patience and comparative development as measured by (log) per capita income. In a univariate regression, average patience explains about 40% of the between-country variation in income. This reduced-form relationship is robust across a wide range of regression specifications, which incorporate controls for many of the deep determinants previously identified in the empirical literature, such as geography, climate, the disease environment, or anthropological and cultural factors. The result also holds both within each continent separately, and when employing alternative definitions of development or national welfare. We conclude our reduced-form analysis by establishing a significant correlation between patience and economic growth. Across a large range of base years, patience explains a considerable fraction of the variation in growth rates both in the medium run (i.e., after World War II), and in the long run over the last 200 years.

Theory posits that patience affects development through accumulation processes, implying that patience should predict the stocks of the proximate determinants of development as well as the corresponding accumulation flows. Our analysis establishes coherent support for these predictions. For instance, the results reveal that patience explains large fractions of the cross-country variation in capital stocks, savings rates, educational attainment, education and health expenditure, research and development expenditure, innovative capacity, and institutional quality. These associations hold for alternative proxies, and are robust to the inclusion of a large and comprehensive vector of controls.

In sum, this paper establishes a coherent pattern of correlations linking patience to income and the accumulation of productive resources across individuals within regions, across regions within countries, and across countries. However, measuring time preferences across countries poses the difficulty that observed behavior might also reflect environmental conditions, raising the question of whether it is “true” or “revealed” patience that generates our findings. We remain largely agnostic about which particular component of the observed heterogeneity in the survey instruments generates the results, because, ultimately, the interest of this paper lies in understanding heterogeneity in future-oriented decisions both across and within countries, rather than in disentangling whether the sources of this heterogeneity are “innate”. Nonetheless, in the final part of the paper, we attempt to provide a first empirical assessment of the extent to which the country-level correlation between patience and per capita income is likely to be driven by “revealed”, rather than “true”, patience. First, we show that the correlation between patience and per capita income holds up when conditioning on variables that proxy for borrowing constraints, inflation and interest rates, the quality of the institutional environment, life expectancy, educational attainment, or cognitive skills. Second, we attempt to estimate an "actual" component of patience by purging individual-level patience of proxies for cognitive skills, education, life expectancy, or the subjectively perceived quality of the institutional environment. Even this cleaned patience variable, once aggregated up at the country level, correlates with per capita income, suggesting that at least part of the observed variation in patience across countries has deep roots, as suggested by recent work of Chen (2013) and Galor and Özak (2014).

This paper contributes to several strands of the literature. At the microeconomic level, a growing set of contributions measure individuals' patience in usually small and non-representative samples and empirically relate the resulting preference parameters to economically important field behaviors (DellaVigna and Paserman, 2005; Chabris et al., 2008; Tanaka et al., 2010; Castillo et al., 2011; Sutter et al., 2013; Meier and Sprenger, 2013; Golsteyn et al., 2014). Our work innovates on this set of contributions by systematically studying the relationship between individual patience and life outcomes in representative samples obtained for a large set of countries, and for a diverse set of outcomes spanning savings, education, health, and income.

At the macro level, our work is the first to empirically investigate the link between patience, aggregate accumulation processes, and national income. The empirical importance of human and physical capital as well as productivity has received considerable attention in the literature Solow (see, e.g., 1957), Glaeser et al. (2004), Caselli and Feyrer (2007), Erosa et al. (2010), and Manuelli and Seshadri (2014), but – presumably due to the lack of suitable data – evidence on the link between patience and the aggregate accumulation of these productive resources is missing. In a broader sense, our work also relates to the recent stream of papers which provide evidence for a set of "deep" determinants of development. These include geography, climate, and diseases (John Luke Gallup et al., 1999; Diamond, 2005; Olsson and Hibbs Jr, 2005; Alsan, 2015), colonial history (Acemoglu et al., 2001; Nunn, 2008), policies and institutions (e.g., Hall and Jones, 1999; La Porta et al., 1999; Acemoglu et al., 2002, 2005), anthropological factors (Alesina et al., 2003; Ashraf and Galor, 2013), cultural factors such as trust (Knack and Keefer, 1997; Guiso et al., 2009; Algan and Cahuc, 2010; Tabellini, 2010), diversity (Alesina et al., 2013b), religion (R. J. Barro and McCleary, 2003; Campante and Yanagizawa, forthcoming), or cultural distance to the technological frontier (Spolaore and Wacziarg, 2009).

The remainder of the paper proceeds as follows. The data and their sources are described in Section 6.3. Section 7.3 discusses the individual-level and Section 7.4 the regional-level results. Section 7.5 presents the reduced-form relationship between patience and aggregate development and the relation between patience and the proximate determinants. Section 7.6 discusses the role of actual patience and environmental conditions, while Section 3.6 offers a concluding discussion.

7.2 Data on Time Preferences Across Countries

7.2.1 Survey Procedure

Empirically relating comparative development to patience requires reliable and meaningful data on time preferences from representative population samples in a broad set of countries. Ideally, these data should reflect behaviorally relevant heterogeneity in time preference at the level of the individual. Our data on time preferences around the globe are part of the Global Preference Survey (GPS), a unique data set on economic preferences from representative population samples in 76 countries. In many countries around the world, the Gallup World Poll regularly surveys representative population samples about social and economic issues. In 76 countries, we included as part of the regular 2012 questionnaire a set of survey items which were explicitly designed to measure a respondent's time preferences, risk preferences, social preferences, and trust (for details see Falk et al., 2015a).

Four noteworthy features characterize these data. First, the preference measures have been elicited in a comparable way using a standardized protocol across countries. Second, we use preference measures that have been elicited from representative population samples in each country, in contrast to small- or medium-scale surveys or experiments, which use student or other convenience samples.[3] This allows for inferences about between-country differences in preferences. The median sample size was 1,000 participants per country; in total, we collected preference measures for more than 80,000 participants worldwide.[4] Respondents were selected through probability sampling and interviewed face-to-face or via telephone by professional interviewers.

A third important feature of the data is geographical representativeness in terms of the countries being covered. The sample of 76 countries is not restricted to Western industrialized nations, but covers all continents and various levels of development. Specifically, our sample includes 15 countries from the Americas, 24 from Europe, 22 from Asia and Pacific, as well as 14 countries in Africa, 11 of which are Sub-Saharan. This set of countries covers about 90% of the world population and of global income.

Fourth and finally, the preference measures are based on experimentally validated survey items for eliciting preferences. In order to ensure behavioral relevance of our measure of time preferences, the underlying survey items were designed, tested, and selected for the purpose of the GPS through a rigorous ex-ante experimental validation procedure (for details see Falk et al., 2015b). In this validation step, subjects participated in choice experiments that measured preferences using real money. They also answered large batteries of survey questions designed to elicit preferences. We then selected the survey items that were (jointly) the best predictors of actual behavior in the experiments, to form the survey module. In order to make these items cross-culturally applicable, (i) all items were translated back and forth by professionals, (ii) monetary values used in the survey were adjusted along the median household income for each country, and (iii) pretests were conducted in 22 countries of various cultural heritage to

[3] See, e.g., Wang (2011), Rieger et al. (forthcoming), Vieider et al. (2015), Vieider et al. (2014).

[4] Notable exceptions include China (2,574 obs.), Haiti (504 obs.), India (2,539 obs.), Iran (2,507 obs.), Russia (1,498 obs.), and Suriname (504 obs.).

ensure comparability. See Appendix 5.A and Falk et al. (2015a) for a description of the data set and the data collection procedure.

Our measure of patience is derived from the combination of responses to two survey measures, one with a quantitative and one with a qualitative format. These were the best predictors of behavior in experiments involving incentivized choices between earlier versus later rewards with a time delay of 12 months, thereby capturing annual time discounting. The quantitative survey measure consists of a series of five interdependent hypothetical binary choices between immediate and delayed financial rewards, a format commonly referred to as the "staircase" (or "unfolding brackets") procedure (Cornsweet, 1962). In each of the five questions, participants had to decide between receiving a payment today or larger payments in 12 months. The wording of the question was as follows:

> *Suppose you were given the choice between receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which one you would choose. Please assume there is no inflation, i.e., future prices are the same as today's prices. Please consider the following: Would you rather receive amount x today or y in 12 months?*

For example, in the German sample, in the first choice, the immediate payment x was 100 euros and the delayed payment y was 154 euros. The immediate payment x remained constant in all subsequent four questions, but the delayed payment y was increased or decreased depending on previous choices. To illustrate, suppose the respondent chose the immediate payment (the delayed payment) in the first decision. Then the delayed payment in the second decision was increased (decreased) to 185 (125) euros (see Appendix 5.A for an exposition of the entire sequence of binary decisions). In essence, by adjusting the delayed payment according to previous choices, the questions "zoom in" around the respondent's point of indifference between the smaller immediate and the larger delayed payment and make efficient use of limited and costly survey time. The sequence of questions has 32 possible ordered outcomes which partition the real line from 100 euros to 218 euros into roughly evenly spaced intervals. In the international survey, monetary amounts x and y were expressed in the respective local currency, scaled relative to median monthly household income in the given country.

The qualitative measure of patience is given by the respondents' self assessment regarding their willingness to wait on an 11-point Likert scale. The wording of the question was as follows:

> *We now ask for your willingness to act in a certain way. Please indicate your answer on a scale from 0 to 10, where 0 means you are "completely unwilling to do so" and a 10 means you are "very willing to do so". How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?*

Our patience measure is a linear combination of the quantitative and qualitative survey items, using the weights obtained from the experimental validation procedure.[5] As described in detail in Falk et al. (2015b), the survey items are strongly and significantly correlated with preference measures obtained from standard incentivized intertemporal choice experiments. The raw correlation between experimental choices and the two survey items are 0.53 and 0.39, respectively, and both items jointly explain more than 50% of the variation in the experimental choices.[6] Moreover, the measures predict experimental behavior out of sample. As established in Falk et al. (2015a), our patience measure also correlates with individual-level characteristics (such as cognitive skills) in a manner that is very similar to the correlations reported in data sets that use financially incentivized procedures. Arguably, the ex-ante validation of survey items constitutes a significant methodological advance over the often ad-hoc selection of questions for surveys based on introspective arguments of plausibility or relevance. Additionally, the quantitative staircase measure not only resembles standard experimental procedures of eliciting time preferences, but it is also relatively context neutral and precisely defined, arguably making it less prone to culture-dependent interpretations. This makes the patience measure particularly well-suited for a multinational study like the present one.

7.2.2 Summary Statistics

The analysis is based on individual-level patience measures that are standardized, i.e., we compute z-scores at the individual level. We then calculate a country's patience by averaging responses using the sampling weights provided by Gallup, see Appendix 5.A. Figure 7.1 depicts the resulting distribution of time preferences across countries, relative to the world's average individual level, colored in white. Darker red colors and darker blue colors indicate less and more patience, respectively, where differences are measured in terms of standard deviations from the world's average individual.[7] Visual inspection of the world map of time preferences already suggests the presence of noteable geographic and economic patterns. In particular, countries in North America and Western Europe appear considerably more patient than their South American or African counterparts. The map also illustrates the existence of considerable between-country differences in patience. The range of the country-averages is 1.7, implying that average patience varies by 1.7 of a standard deviation (in terms of the total individual-level variation).

---

[5] Specifically, responses to both items were standardized at the individual level and then aggregated using the following formula:

$$\text{Patience} \; = 0.7115185 \cdot \; \text{Staircase measure} \; + 0.2884815 \cdot \; \text{Qualitative measure.}$$

These weights are based on OLS estimates of a regression of observed behavior in financially incentivized laboratory experiments on the two survey measures. See Falk et al. (2015a) and Falk et al. (2015b) for details.

[6] The benchmark for this comparison is the explanatory power of a test-retest correlation between two incentivized elicitations.

[7] Appendix 5.A provides histograms of both average patience across countries and individual level patience.

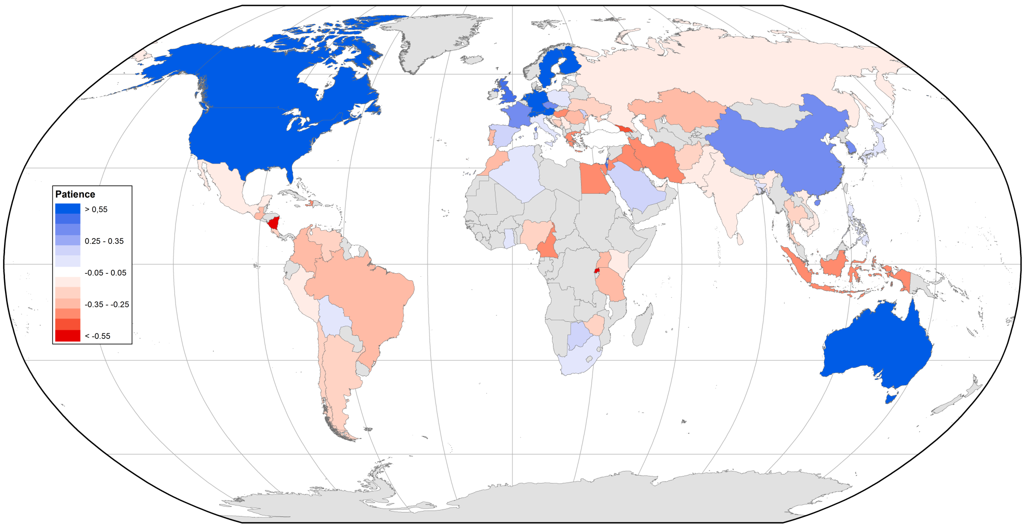


Figure 7.1. Distribution of Patience Across Countries

7.2.3 Further Variables of Interest

The GPS includes a broad range of individual-level variables. Apart from sociodemographic covariates and a self-reported proxy for cognitive skills regarding mathematical ability, the dataset also contains four variables which allow us to assess the relationship between patience, future-oriented choices, and income: household income, educational attainment as a three-step category, a binary indicator for whether the respondent saved in the previous year, and a binary variable measuring whether the respondent reports having health problems beyond what is normal given their age. The data also include a measure of trust that asks for respondents' self-assessment regarding the statement "I assume that people have only the best intentions (0-10)".

For our cross-regional analysis, we make use of the data set of Gennaioli et al. (2013) which includes regional per capita income as well as average years of education.

At the country level, the empirical analysis incorporates other variables from a variety of data sources, replicating measures that have been used in various contributions to the literature. In particular, the analysis makes use of macroeconomic variables that are taken mostly from the World Bank's World Development Indicators or the Penn World Tables. Whenever feasible, we use ten-year-averages (usually from 2003 to 2012) to smooth the data and eliminate variation due to measurement error or random fluctuations. Appendix 7.H contains information on the construction and sources of all variables used in the empirical analysis.

7.3 Individual Patience, Accumulation and Income

In line with the perspective taken by micro-founded theories of development, we begin our empirical analysis by an investigation of heterogeneity in patience and accumulation decisions at the individual level. The purpose of this analysis is twofold. First, the analysis explores the validity of the theoretical prediction that patience fosters the accumulation of resources and knowledge, in representative samples across economically and culturally highly heterogeneous samples. Second, the results provide information as to whether there are systematic patterns of patience affecting future-oriented decisions and, ultimately, income that cannot be explained by factors at the regional or national level, including institutions or the overall health environment.

The analysis is based on Gallup data on household income, savings, educational attainment, health status, and patience and exploits within-country variation in patience and the respective outcomes in terms of future-oriented decisions. Unfortunately, data for two of these outcome variables (health and savings) are only available for a subset of countries.

The results are presented in Table 7.1. For each dependent variable, we report the results of three OLS specifications, one without any covariates, one with country fixed effects, and one with regional fixed effects and additional individual-level covariates. Columns (1) through (3) report the results of OLS regressions of respondents' household income per capita on their patience. The estimates document that, both across and within countries, more patient people tend to earn significantly higher incomes. This pattern holds conditional on a comprehensive vector of individual-level covariates including age, gender, religion, cognitive skills, and three variables that proxy for the subjectively perceived quality of the institutional environment (these variables are collected and constructed by Gallup, see Appendix 7.H). At the same time, the coefficient of patience decreases substantially after the inclusion of country fixed effects; we will return to discussing this issue below.

Columns (3)-(9) consider the accumulation of physical and human capital, i.e., savings behavior and educational attainment, also see Falk et al. (2015a). Patient individuals save more and invest more into education, within countries and subnational regions, suggesting that accumulation decisions are affected by patience as predicted by micro-founded theories. Finally, we examine the relationship between time preference and health by relating a binary index of whether a respondent reports having health problems to patience. Again, consistent with the predictions of intertemporal choice theories, patient individuals exhibit better health. In sum, these results represent the first evidence for the relationship of patience with economic well-being as well as accumulation behavior at the global level, using comparable data elicited in representative samples, and holding the broader environment (in terms of regional heterogeneity) constant.

7.4 Patience and Development Across Regions

In light of the strong individual-level results, we proceed by incrementally aggregating the data to the macroeconomic level. The next step is to consider variation across subna-

Table 7.1. Individual patience, savings, human capital, and income

	Dependent variable:											
	Log [HH income p/c]			Saved last year			Education level			Health problems		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Patience	0.34***	0.056***	0.040***	0.050***	0.038***	0.032***	0.12***	0.069***	0.040***	-0.034***	-0.033***	-0.012***
	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Age			0.58***			-0.059			1.09***			0.024
			(0.20)			(0.32)			(0.29)			(0.09)
Age squared			-0.38			-0.056			-1.95***			0.81***
			(0.23)			(0.30)			(0.27)			(0.10)
1 if female			-0.086***			-0.0057			-0.019			0.022***
			(0.02)			(0.01)			(0.02)			(0.01)
Subj. math skills			0.035***			0.017***			0.047***			-0.0077***
			(0.00)			(0.00)			(0.00)			(0.00)
Subjective institutional quality			-0.00042*			0.00046			-0.00058***			-0.00014*
			(0.00)			(0.00)			(0.00)			(0.00)
Confidence in financial institutions			0.042***			0.052***			0.014			-0.028***
			(0.01)			(0.01)			(0.01)			(0.01)
Subjective law and order index			0.00058**			0.00012			-0.000048			-0.00054***
			(0.00)			(0.00)			(0.00)			(0.00)
Constant	7.92***	6.47***	5.88***	0.27***	0.17***	0.28***	1.87***	1.32***	0.97***	0.25***	0.27***	0.44***
	(0.13)	(0.00)	(0.09)	(0.03)	(0.00)	(0.08)	(0.03)	(0.00)	(0.08)	(0.01)	(0.00)	(0.03)
Country FE	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Regional FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Religion FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	79245	79245	46383	15260	15260	10438	79357	79357	46550	62727	62727	46428
R^2	0.051	0.608	0.638	0.011	0.073	0.137	0.030	0.205	0.308	0.006	0.029	0.153
Adjusted R^2	0.051	0.607	0.632	0.011	0.072	0.116	0.030	0.205	0.296	0.006	0.028	0.139

OLS estimates, standard errors (clustered at country level) in parentheses. The dependent variable in (4)-(6) is a binary indicator for whether the individual saved last year, in (5)-(6) it is educational attainment as a three-step category, and in (7)-(9) it is a binary variable capturing whether the respondent reports having health problems that are atypical given their age. Age is divided by 100. All results in columns (4)-(12) are robust to estimating (ordered) probit models. See Appendix 7.H for a detailed description of all dependent variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

tional regions. This is possible since the individual-level data contain regional identifiers (usually at the state or province level), which allows us to relate the average level of patience in a subnational region to the level of regional GDP per capita and the average years of education in a given region from data constructed by Gennaioli et al. (2013). In total, we were able to match 704 regions from 55 countries across the two data sets.[8]

While the regional level of analysis pertains to an aggregate view on accumulation processes and income, the corresponding regression analyses still have the important advantage of allowing us to account for unobserved heterogeneity at the country-level by using country fixed effects (Gennaioli et al., 2013). For example, potential concerns about the role of languages, institutions, and culture for survey responses are less relevant in within-country analyses. The benefits of considering regional data naturally come at the cost of losing representativeness, since the sampling scheme was constructed to achieve representativeness at the country level. Given a median sample size of 1,000 respondents per country, in some regions, we observe only a relatively small number of respondents. In consequence, average regional time preference is estimated less precisely for some regions. We explicitly take into account the differential precision with which patience is measured across regions by estimating weighted OLS regressions of regional (log) GDP per capita or average years of education on regional patience, in which each observation is weighted by the number of respondents in the respective region. This procedure ensures that regions with only a small number of respondents receive less weight in the estimations, as should be the case if more observations imply more precision.[9]

Table 7.2 reports estimates of these regressions of average education and average per capita income at the regional level on patience. For both dependent variables, we estimate a specification without country fixed effects, one with country fixed effects, and one with additional regional-level covariates (Gennaioli et al., 2013). The results mirror those established in the individual-level analysis: We find significant relationships between patience, per capita income, and human capital. These results hold conditional on country fixed effects (columns (2) and (5)), although again the size of the coefficient of patience is substantially reduced in the specifications with country fixed effects. At the same time, controlling for geographic and socio-cultural variables at the region level does not substantially alter the estimates (columns (3) and (6)). Nevertheless, even in these specifications, the coefficients of patience on regional income and education are substantially larger than the coefficients obtained at the individual level. We will return to this issue in Section 7.6 below.

In sum, heterogeneity in patience is predictive of the accumulation of productive resources as well as income differences within countries, on a global scale. While these results are in line with micro- and macroeconomic theories of intertemporal choice, a crucial question is whether similar patterns also hold across countries.

[8] See Appendix 7.E for an overview of the number of regions per country.

[9] An alternative route is to restrict the sample to regions for which the number of respondents exceeds a particular threshold. As Table 7.27 in Appendix 7.F shows, once the sample is restricted by eliminating regions with ten or less observations, the results are very similar to those reported in the main text.

Table 7.2. Regional patience, human capital, and income

	Dependent variable: Avg. years of education			Log [Regional GDP p/c]		
	(1)	(2)	(3)	(4)	(5)	(6)
Patience	3.34***	0.43**	0.32**	1.39***	0.21***	0.16**
	(0.55)	(0.17)	(0.14)	(0.23)	(0.07)	(0.07)
Temperature			-0.050***			-0.015
			(0.02)			(0.01)
Inverse distance to coast			1.36**			0.67
			(0.54)			(0.42)
Log [Oil production p/c]			-0.11			0.21***
			(0.13)			(0.06)
# Ethnic groups			-0.38**			-0.15**
			(0.15)			(0.06)
Log [Population density]			0.21**			0.078
			(0.10)			(0.05)
Constant	7.17***	7.37***	6.53***	8.74***	9.18***	8.66***
	(0.36)	(0.04)	(0.67)	(0.18)	(0.02)	(0.37)
Country FE	No	Yes	Yes	No	Yes	Yes
Observations	693	693	676	704	704	687
R^2	0.252	0.936	0.957	0.184	0.937	0.950
Adjusted R^2	0.251	0.931	0.953	0.183	0.932	0.946

Weighted least squares estimates, observations weighted by number of respondents in region. Standard errors (clustered at country level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7.5 Patience and Comparative Development: Cross-Country Evidence

Building on the within-country evidence presented above, the main part of the empirical analysis considers the relationship between country-level patience, aggregate accumulation processes, and development. We begin our corresponding investigation by providing evidence for a reduced-form relationship between patience and cross-country development, both in terms of contemporary levels of development and regarding income dynamics in terms of growth. We then study in detail the relationship between patience and the accumulation of physical capital, human capital, and productivity.

7.5.1 Patience and Contemporary Development

7.5.1.0.1 Baseline results. The left panel of Figure 7.2 provides a graphical illustration of the reduced-form relationship between comparative development and patience. The raw correlation between the log of GDP per capita (measured by averages over the period 2003-2012) and the patience measure is 0.63, implying that patience alone explains about 40% of the variation in log income per capita. To appreciate the quantitative magnitude of this relationship, note that the size of the standardized beta (63%) that corresponds to the regression line in the figure implies that an increase in patience by one standard deviation is associated with an increase of almost two-thirds of a standard deviation in (log) GDP, which is roughly equivalent to the income difference between Peru and the United States.[10]

To investigate the statistical significance and robustness of this relationship, we turn to multivariate regression analysis. Table 7.3 presents the reduced-form relationship between comparative development and patience accounting for different sets of covariates. Column (1) documents the existence of a significant unconditional relationship between (log) GDP per capita and patience as depicted in the left panel of Figure 7.2. This raw relationship is statistically highly significant with a t-statistic larger than 10. We proceed by investigating the robustness of the relation between patience and income to including various variables corresponding to deep-rooted factors that have been associated with development. Columns (2) through (4) successively add a large and comprehensive set of geographic and climatic covariates. Column (2) contains a set of continent fixed effects as well as a binary indicator for whether the country has ever been colonized in the past.[11] Column (3) contains additional controls for absolute latitude, longitude, the fraction of arable land, land suitability for agriculture, and the timing of the Neolithic transition. Column (4) adds average precipitation and temperature as well as the fractions of the population that live in the (sub-) tropics or in areas with the risk of contracting malaria. While the inclusion of this large vector of covariates reduces the the coefficient of patience by about 30%, the coefficient remains statistically significant and quantitatively large, with a standardized beta of 38%. Again, the coefficient estimate obtained with country-level data is substantially larger than the coefficient obtained with regional or individual data, suggesting reinforcing effects at the aggregate level through third factors or measurement error (Eric A. Hanushek et al., 1996).

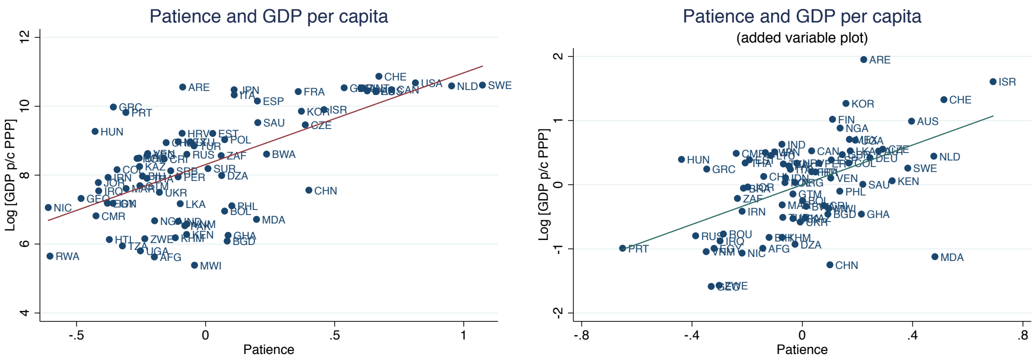


Figure 7.2. Patience and national income. The left panel depicts the raw correlation between log GDP per capita (purchasing-power parity) and patience ($\rho = 0.63$), while the right panel contains a plot conditional on the full set of baseline covariates in column (7) of Table 7.3 (partial $\rho = 0.52$, semi-partial $\rho = 0.24$).

[10] The standardized beta measures the change in the dependent variable in % of a standard deviation given a one standard deviation increase in the independent variable.

[11] Following the World Bank terminology, continents are defined as North America, Central and South America, Europe and Central Asia, East Asia and Pacific, South Asia, Middle East and North Africa, and South Africa.

In recent years, the literature on comparative development has argued that deep-rooted cross-country differences in the diversity of a population – such as genetic, ethnic, linguistic, or religious diversity – are partly responsible for differences in income (Alesina et al., 2003; Ashraf and Galor, 2013; Ashraf et al., 2014). In order to ensure that patience does not merely pick up deep-rooted cross-country differences in the diversity of a population, column (5) additionally controls for genetic diversity and its square, as well as for ethnic fractionalization. While the results for these variables by and large replicate those of the earlier literature, adding these covariates has little, if any, effect on the coefficient on patience.

Taken together, the results in columns (2) through (5) indicate that the relationship between patience and national income is unlikely to merely reflect the effect of other "deep" causes of comparative development that have received attention in the literature. The most extensive specification with patience and geographic, climatic, colonial, and diversity covariates explains more than 80% of the variation in per capita income, but patience continues to have strong explanatory power for cross-country income differences, with a partial R^2 associated with patience of 27%.[12]

Starting with Knack and Keefer (1997), it has been argued that social capital, measured in terms of interpersonal trust, is related to GDP. Column (6) presents the results of a regression of income on trust as explanatory variable. Consistent with previous findings in the literature based on the trust measure from the World Values Survey, we find that the GPS trust measure is significantly related to national income. Once we add patience and other controls in column (7), however, trust is no longer significant, while the relationship between patience and GDP remains strong. Virtually identical results obtain if we capture this cultural trait using the standard trust measure from the World Values Survey. The right-hand panel of Figure 7.2 illustrates the conditional relationship of column (7).[13] Additional robustness checks show that controlling for average risk aversion (measured similarly to patience, see Falk et al., 2015a), legal origin dummies, religious and linguistic fractionalization, major religion shares, the fraction of European descent, the genetic distance to the US, and other geographical variables, does not affect our main result.[14] Throughout this paper, we will employ the specification in column (7) as baseline set of controls.

7.5.1.0.2 Robustness: Patience and Income in Sub-Samples. Table 7.4 investigates the robustness of our main reduced-form relationship in various sub-samples. Columns (1) through (4) analyze whether the relationship between patience and income also prevails within each continent separately. In columns (5) through (8), we split the sample of countries by the level of development into OECD and non-OECD countries, and by historical legacy into former colonies and countries that have never been colonized. The relationship between income and patience is always positive and statistically significant, despite the rather small sample sizes.

---

[12] The semi-partial R^2 is 6%.

[13] Following Altonji et al. (2005), the results in Table 7.3 also show that the bias resulting from the omission of unobserved variables would have to be at least 1.5 times strong as the bias from excluding the large and comprehensive vector of covariates.

[14] See Table 7.22 in Appendix 7.F for details.

7.5.1.0.3 Robustness: Alternative Measures of Development. Next, we explore whether patience is also predictive of other measures of material and non-material well-being than GDP per capita. We employ three alternative measures of development. The first of these measures is GDP per worker, which is frequently used in growth empirics as a more useful measure of output than GDP per capita. Instead of confining comparative development to differences in income or consumption, we also employ two broader measures of well-being, namely (i) the United Nations Human Development Index, which combines GDP, years of schooling, and life expectancy, and (ii) subjective statements about well-being in terms of happiness. Table 7.5 reports the results of regressions of these dependent variables on our patience variable. Columns (1), (3) and (5) document that each of the alternative measures of economic development exhibits a strong pos-

Table 7.3. Patience and national income

	Dependent variable: Log [GDP p/c PPP]						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Patience	2.66***	2.06***	1.78***	1.61***	1.58***		1.55***
	(0.26)	(0.32)	(0.40)	(0.37)	(0.41)		(0.41)
1 if colonized		-0.30	-0.22	-0.44*	-0.31		-0.39*
		(0.32)	(0.34)	(0.24)	(0.24)		(0.23)
Distance to equator			0.021	0.017	-0.014		-0.012
			(0.02)	(0.02)	(0.02)		(0.02)
Longitude			-0.0027	0.0043	0.0055		0.0059
			(0.01)	(0.01)	(0.01)		(0.01)
Percentage of arable land			-0.027**	-0.016	-0.014		-0.015
			(0.01)	(0.01)	(0.01)		(0.01)
Land suitability for agriculture			0.81	0.12	0.12		-0.094
			(0.70)	(0.59)	(0.58)		(0.60)
Log [Timing neolithic revolution]			0.46	0.094	0.22		0.34
			(0.49)	(0.36)	(0.41)		(0.41)
Average precipitation				0.0072	0.0023		0.0019
				(0.00)	(0.00)		(0.00)
Average temperature				0.077**	0.051		0.057
				(0.03)	(0.03)		(0.03)
% living in (sub-)tropical zones				-1.53**	-1.31*		-1.16*
				(0.69)	(0.69)		(0.63)
% at risk of malaria				-1.46***	-1.45***		-1.49***
				(0.49)	(0.54)		(0.51)
Predicted genetic diversity					430.2**		451.4**
					(181.70)		(168.86)
Predicted genetic diversity sqr.					-308.7**		-324.0**
					(132.16)		(122.94)
Ethnic fractionalization					-0.89		-0.89
					(0.61)		(0.60)
Trust						1.58**	-0.57
						(0.68)	(0.51)
Constant	8.31***	9.30***	4.10	5.74*	-143.4**	8.33***	-151.8**
	(0.14)	(0.47)	(3.99)	(3.37)	(62.76)	(0.17)	(58.15)
Continent FE	No	Yes	Yes	Yes	Yes	No	Yes
Observations	76	76	74	74	74	76	74
R^2	0.397	0.691	0.730	0.819	0.845	0.079	0.850
Adjusted R^2	0.389	0.654	0.671	0.764	0.787	0.067	0.789

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7.4. Patience and national income in sub-samples

	Dependent variable: Log [GDP p/c PPP] in...							
	Africa & Middle East (1)	Europe & C. Asia (2)	SE Asia & Pacific (3)	Ameri-cas (4)	OECD (5)	Non-OECD (6)	Colo-nized (7)	Not colonized (8)
Patience	2.83***	1.82***	3.76***	2.42***	1.02***	1.43**	2.54***	2.23***
	(0.76)	(0.33)	(1.04)	(0.32)	(0.21)	(0.65)	(0.36)	(0.51)
Constant	7.84***	9.09***	7.40***	8.55***	9.75***	7.77***	8.10***	8.87***
	(0.34)	(0.19)	(0.33)	(0.20)	(0.15)	(0.20)	(0.16)	(0.30)
Observations	20	27	14	15	22	54	54	22
R^2	0.274	0.448	0.430	0.592	0.498	0.073	0.313	0.434
Adjusted R^2	0.234	0.426	0.383	0.560	0.473	0.055	0.300	0.406

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In the first column, the sample includes Africa and the Middle East, in the second column Europe and Central Asia, in the third South-East Asia and Pacific, in the fourth the Americas, in the fifth (sixth) all (non-) OECD members, and the seventh (eigth) all formerly colonized (never colonized) countries. The regional groups follow the World Bank definitions.

itive unconditional correlation with patience. The results in columns (2), (4), and (6) show that these positive relationships also hold conditional on the baseline set of control variables in column (7) of Table 7.3.

Table 7.5. Patience and alternative development measures

	Dependent variable:					
	Log [GDP per worker PPP]		Human Development Index		Subjective happiness	
	(1)	(2)	(3)	(4)	(5)	(6)
Patience	1.56***	0.64***	0.23***	0.13***	0.13***	0.21***
	(0.21)	(0.23)	(0.03)	(0.03)	(0.03)	(0.04)
Constant	9.87***	-66.1**	0.70***	-11.7**	0.72***	-11.1**
	(0.11)	(31.10)	(0.01)	(5.03)	(0.01)	(5.50)
Additional controls	No	Yes	No	Yes	No	Yes
Observations	71	69	76	74	76	74
R^2	0.305	0.896	0.335	0.882	0.140	0.737
Adjusted R^2	0.295	0.849	0.326	0.834	0.129	0.631

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

7.5.2 Patience and Growth

From a theoretical perspective, greater patience does not only imply greater steady state levels of income, but also faster growth along the balanced growth path (see, e.g., Lucas, 1988).[15] We therefore investigate the relationship between time preferences and growth rates over the past 200 years, and compute the (geometric) average annual growth rate in per capita GDP from different base years until 2010.[16]

[15] Patience also affects growth off the balanced growth path, including cases in which patience increases in income or wealth, see, e.g., Drugeon (1996), Das (2003), and Strulik (2012).

[16] For Afghanistan, Botswana, Nicaragua, and Rwanda, GDP in 2010 is not available in the Maddison data set on historical GDP. For these countries, we compute the annual growth rate until 2008 instead. None of the results change if we exclude these countries from the estimations.

Table 7.6. Patience and economic growth

	Dependent variable: Annual growth rate in GDP p/c (in %) since...											
	1820		1870		1925		1950			1975		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Patience	0.42***	0.41**	0.33***	0.47**	0.45*	0.88**	1.07***	1.16**	1.38***	1.09**	1.58***	1.75**
	(0.11)	(0.16)	(0.12)	(0.18)	(0.24)	(0.35)	(0.33)	(0.47)	(0.47)	(0.46)	(0.52)	(0.76)
Log [GDP p/c base year]		-0.64***		-0.44***		-0.91***		-0.73***	-1.16***		-0.87***	-1.32***
		(0.15)		(0.15)		(0.26)		(0.24)	(0.23)		(0.28)	(0.40)
Constant	1.34***	5.89***	1.61***	4.26***	1.90***	9.13***	2.06***	7.71***	-230.9***	1.56***	8.69***	-233.5*
	(0.05)	(1.00)	(0.06)	(1.00)	(0.12)	(2.07)	(0.14)	(1.92)	(74.01)	(0.20)	(1.98)	(134.25)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes
Additionals controls	No	No	No	No	No	No	No	No	Yes	No	No	Yes
Observations	31	31	42	42	33	33	63	63	63	66	66	65
R^2	0.299	0.632	0.119	0.479	0.091	0.496	0.121	0.516	0.751	0.065	0.501	0.679
Adjusted R^2	0.275	0.498	0.097	0.352	0.061	0.355	0.107	0.445	0.613	0.051	0.431	0.510

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Due to the lower number of observations, in columns (1)-(6) we only control for continent fixed effects. See column (7) of Table 7.3 for a complete list of the additional controls.

Table 7.6 presents the results of OLS regressions of annual growth rates computed for each of several base years on patience. The first of the respective columns shows the unconditional correlation between growth and patience, while the second column includes controls for log per capita income in the respective base year to capture convergence dynamics, and continent fixed effects. For base years after 1950, the larger number of observations enables us to present a third specification in which we additionally condition on the baseline control variables from column (7) in Table 7.3. Across base years, greater patience is significantly associated with higher growth rates. This pattern obtains for both very long-run growth rates and medium-run growth after World War II. In sum, patience correlates not only with contemporary development, but also with income growth over the last 200 years. Given this pattern, an immediate question is whether the relationship between patience and income levels was already present in the (distant) past. In Appendix 7.D, we investigate this issue by relating past development (both in pre- and post-industrial times) to today's patience. While these analyses naturally rest on the assumption that the distribution of patience across countries remained relatively stable over time, the corresponding results reveal strong relationships between patience and historical income per capita in the 19th and early 20th century as well as with economic development in 1500.

7.5.3 Patience and Factor Accumulation

The development accounting approach suggests that the reduced-form relationship between patience and development works through accumulation processes. In this section, we further investigate whether the data are consistent with patience affecting income through the "channels" of human and physical capital as well as (residual) factor productivity. Whenever feasible, in these analyses, we consider both stocks and flows as dependent variables, i.e., we analyze whether patience predicts the levels of production factors and productivity as well as the corresponding accumulation flows.

7.5.3.0.1 Physical Capital. We start by regressing the stock of physical capital on patience. Columns (1) and (2) of Table 7.7 present OLS estimates of the unconditional relationship and of the relationship conditional on the extensive set of baseline covariates from column (7) in Table 7.3. In line with a standard Ramsey-Cass-Koopmans model, the estimates reveal a significant positive relationship between patience and the stock of capital per capita. Patience alone explains about a third of the variation in capital.

The relationship between patience and physical capital accumulation can also be investigated by looking at flows. Columns (3) to (8) of Table 7.7 present the respective results for gross national savings rates, net adjusted national savings rates, as well as household savings rates as dependent variables. The definitions of the first two variables follow the World Bank terminology, according to which gross savings rates are given by gross national income net of consumption, plus net transfers, as a share of gross national income. Net adjusted savings rates correspond to gross savings net of depreciation, adding education expenditures and deducting estimates for the depletion of energy, minerals and forests, as well as damages from carbon dioxide emissions. Household savings rates are measured as household savings relative to household disposable income. These data are based on surveys and are only available for OECD countries. Throughout, the results reveal a significant positive relationship between patience and savings. In terms of size and statistical significance, the effect is largest for household savings and net adjusted savings, which includes education expenditures, and thus incorporates the accumulation of physical and human capital.[17] The finding that patience is related to household savings rates even within OECD countries is also noteworthy, given the similarity of this subset of countries in terms of economic development and other characteristics.[18]

7.5.3.0.2 Human Capital. As measures of human capital, we consider proxies for both the quantity and quality of schooling, as well as investments into education. Specifically, following the literature, we use average years of schooling as the baseline measure of education. However, since years of schooling does not account for quality differences across countries, we also use alternative measures. First, we take the human capital index provided by the Penn World Tables, which aims to provide a quality-adjusted index by combining years of schooling with returns to schooling. As a second alternative measure of quality-adjusted human capital, we employ a measure of cross-country differences in cognitive skills derived from educational achievement tests (Eric A Hanushek and Woessmann, 2012). Finally, we use education expenditure and public health expenditure as percentage of national income as a measure of the input into the human capital formation process.

Table 7.8 summarizes our corresponding findings. Columns (1) and (2) reveal a positive relation between patience and average years of schooling. The explained variation of roughly 40% indicates a strong unconditional relationship, which holds up when controlling for the baseline set of covariates.[19] Columns (3) through (10) present the analogous

Table 7.7. Patience, physical capital, and savings

	Dependent variable:							
	Log [Capital stock p/c]		Gross sav. (% GNI)		Net adj. sav. (% GNI)		HH savings rate	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Patience	2.03***	0.98***	6.66***	8.11**	7.70***	9.65***	6.21**	7.08**
	(0.28)	(0.31)	(2.29)	(4.01)	(2.23)	(2.96)	(2.70)	(3.13)
Constant	10.0***	-147.6***	22.1***	-1506.4**	10.2***	343.8	3.27**	2.86*
	(0.13)	(43.14)	(1.18)	(580.16)	(1.07)	(729.38)	(1.50)	(1.63)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes
Additional controls	No	Yes	No	Yes	No	Yes	No	No
Observations	71	69	73	71	68	68	21	21
R^2	0.327	0.861	0.058	0.474	0.102	0.537	0.231	0.272
Adjusted R^2	0.317	0.799	0.044	0.249	0.088	0.326	0.191	0.144

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Gross savings and net adjusted savings are national savings as % of GNI. Household savings as % of household disposable income. Household savings rates are only available for OECD countries. Due to the small number of observations, in column (8) the controls are restricted to continent dummies only. See column (7) of Table 7.3 for a complete list of the additional controls.

[17] Unreported results show that patience is also significantly correlated with net FDI outflows (as % of GDP).

[18] Figure 7.9 in the Appendix presents a graphical illustration of this result.

[19] Figure 7.10 in the Appendix graphically illustrates these regression results.

Table 7.8. Patience and human capital

	Dependent variable:									
	Years of schooling		Human capital index		Cognitive skills		Educ. exp. (% GNI)		Public health exp. (% GDP)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Patience	4.67***	3.53***	0.76***	0.42**	0.81***	0.35*	1.45***	1.48**	3.34***	1.88***
	(0.53)	(0.84)	(0.11)	(0.16)	(0.13)	(0.20)	(0.31)	(0.55)	(0.48)	(0.53)
Constant	5.40***	-85.0	2.62***	-30.8	4.39***	-11.9	4.28***	-68.9	4.11***	107.9
	(0.24)	(129.94)	(0.05)	(35.40)	(0.08)	(65.08)	(0.16)	(120.45)	(0.21)	(111.64)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	71	70	67	66	49	48	71	70	75	73
R^2	0.429	0.798	0.330	0.718	0.283	0.757	0.138	0.561	0.327	0.761
Adjusted R^2	0.421	0.710	0.319	0.583	0.268	0.561	0.125	0.369	0.318	0.663

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

results for the four alternative measures of human capital. We find a significant positive relationship of patience with all human capital proxies, both stocks and flows.

7.5.3.0.3 Patience, Productivity, and Institutions. The empirical literature has documented the importance of differences in residual factor productivity for comparative development. We complete the investigation of the role of patience for the proximate determinants by presenting evidence regarding the relationship between patience and different measures of factor productivity. In light of the literature that, at least since Hall and Jones (1999), has emphasized the role of institutions and social infrastructure for explaining cross-country productivity and income differences, we also consider institutions as a deeper determinant of productivity differences across countries.

We investigate the association between patience and productivity using a standard measure of total factor productivity (TFP) as well as three alternative measures that capture differences in productivity related to innovation, consistent with standard theories of endogenous growth. These measures are the share of GDP made up by R&D expenditures, the number of researchers in R&D (per 1,000 inhabitants), and the Global Innovation Index. This index is a summary statistic of innovative capacity, and hence productivity, that consists of over 80 qualitative and quantitative items, including measures of institutions, human capital and research, infrastructure, market sophistication, business sophistication, knowledge and technology outputs, and creative outputs.

Table 7.9 contains the respective estimation results. Using the standard TFP measure, we find a positive relation between patience and productivity (columns (1)-(2)). As shown in columns (3)-(6), we also find strong and significant associations with patience when using R&D expenditure or the number of researchers in R&D as dependent variable. Note that patience explains a substantially larger fraction of these R&D-related variables (roughly 60%) than it does for TFP. In columns (7) and (8), we employ the global innovation index as dependent variable. The relationship between patience and factor productivity measured in terms of this index is similarly strong as the one with R&D expenditure, and again remains significant when controlling for all baseline covariates.[20]

A broader interpretation of productivity differences refers to the quality of the institutional environment. In this respect, higher patience might lead to the design of higher-quality institutions if more patient decision-makers opt for creating institutions that support sustainable development rather than short-run rent extraction. We consider an index of democratic quality, an index of property rights as well as a social infrastructure index (Hall and Jones, 1999) as measures of institutional quality. Finally, as a proxy for the quality of the financial institutional environment, we employ Standard & Poor's long-term credit rating, which arguably captures the structural institutional reliability of a country in fulfilling its financial obligations. The regression results are presented in Table 7.10. For each institutional proxy, we again report estimation results for two different specifications, one with and one without controls. The estimates indicate a strong relationship between patience and institutions, confirming the hypothesis that patience is a

[20] Figures 7.11 and 7.12 in the Appendix show the raw and conditional correlations between patience and R&D expenditures, and between patience and the global innovation index, respectively.

Table 7.9. Patience, productivity and innovation

	Dependent variable:							
	TFP		R&D exp. (% GDP)		# Researchers in R&D		Innovation Index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Patience	0.36***	0.13	2.09***	1.78***	3.54***	2.45**	23.6***	18.1***
	(0.05)	(0.09)	(0.23)	(0.59)	(0.43)	(1.02)	(1.70)	(2.91)
Constant	0.62***	1.07	0.96***	1.28	1.54***	-29.8	39.1***	-114.4
	(0.03)	(14.63)	(0.08)	(47.10)	(0.16)	(94.18)	(0.82)	(462.69)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	60	59	64	63	62	61	72	71
R^2	0.338	0.724	0.574	0.716	0.526	0.738	0.619	0.825
Adjusted R^2	0.326	0.567	0.567	0.570	0.518	0.597	0.613	0.750

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Number of researchers in R&D are per 1,000 population. See column (7) of Table 7.3 for a complete list of the additional controls.

significant correlate of democracy, property rights, social infrastructure, and long-term credit ratings.[21]

In sum, patience predicts both, the stocks of and investments into physical capital, human capital, and productivity. In Appendix 7.F, we present extended specifications for all dependent variables in which we additionally condition on income per capita. Although controlling for GDP is likely to produce an underestimate of the relationship between patience and the respective proximate determinant, we report these specifications to illustrate that the consistent pattern linking time preferences to accumulation processes does not arise as a mere artefact of the correlations between national income on the one hand and proximate determinants on the other hand.

Table 7.10. Patience, institutions and social infrastructure

	Dependent variable:							
	Democracy		Property rights		Social infrastructure		S&P credit rating	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Patience	4.23***	2.95**	46.3***	35.0***	0.43***	0.17**	10.7***	9.04***
	(0.82)	(1.46)	(4.39)	(8.26)	(0.05)	(0.06)	(0.82)	(1.35)
Constant	6.63***	295.8	48.3***	-83.6	0.50***	11.9	14.5***	-271.9
	(0.37)	(227.41)	(1.95)	(1271.19)	(0.02)	(12.56)	(0.42)	(299.57)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	74	72	74	72	61	60	64	62
R^2	0.202	0.686	0.515	0.640	0.461	0.793	0.607	0.768
Adjusted R^2	0.191	0.554	0.508	0.489	0.451	0.678	0.601	0.647

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

7.6 Environmental Conditions and Revealed Patience

Measuring *any* preference parameter is burdened by the fact that responses in surveys or experiments might capture some additional factors besides the actual preference param-

[21] Figure 7.13 in the Appendix depicts the relationship between patience and property rights.

eter of interest used in theoretical work.[22] For the purposes of explaining the observed heterogeneity in accumulation decisions across individuals, regions, and countries, it is ultimately the observed variation in *revealed* patience that is the core object of interest because this is the variation that ultimately matters for behavior with real (monetary) consequences. Nonetheless, this section takes a more nuanced look at the data by investigating to which extent our main cross-country result on the correlation between patience and per capita income is likely to be driven be actual (or innate) as opposed to revealed patience.

7.6.1 Controlling for Country-Level Characteristics

7.6.1.0.1 Inflation and Interest Rates. If some respondents expect higher levels of inflation than others, or live in an environment with higher nominal interest rates, they might appear more impatient in their survey responses, even if they have the same time preference. Note, however, that the quantitative survey question explicitly asked people to imagine that there was zero inflation. Also, previous research has shown that differences in interest rates are unlikely to drive choices in experimental environments, see, e.g., Dohmen et al. (2010). Empirically, we check robustness to this concern by explicitly controlling for inflation (in terms of the consumer price index, or the GDP deflator) and deposit interest rates. We find that the reduced-form coefficient of patience remains quantitatively large and highly statistically significant after controlling for these factors, see Table 7.13 in Appendix 7.B.1.

7.6.1.0.2 Borrowing Constraints. Respondents might be more likely to opt for immediate payments in experimental choice situations if they face upward sloping income profiles and are borrowing constrained. To address this issue, we first establish that the correlation between patience and income is robust to controlling for covariates that capture different dimensions of the level of financial development or borrowing constraints of a given country. Specifically, we employ the ratio of external finance to GDP (Rajan and Zingales, 1998; Buera et al., 2011), as well as the (log) number of Automated Teller Machines (ATMs). In addition, we exploit the idea that borrowing constraints (if present) are likely to be less binding for relatively rich people. We hence employ the average patience of each country's top income quintile as explanatory variable. Again, the reduced-form relationship remains strong and significant, see Table 7.13 in Appendix 7.B.1.

7.6.1.0.3 Subjective Uncertainty. If respondents face subjective uncertainty in our quantitative decision task, people might seem more impatient than they really are. To check whether this drives the findings, we condition both on objective and subjective measures of the quality of the institutional environment as well as people's life expectancy. First, in column (1) of Table 7.11 we control for a property rights and a democracy index. Second, in column (2) we make use of the fact that Gallup's background data contain a series of questions which ask respondents to asses their confidence in their institutional environment. A first composite index incorporates people's confidence in the

[22] See the discussion on elicitation of attitudes by Dohmen et al. (2011) and Dohmen et al. (2014), or Dean and Sautmann (2014) for an analysis of the elicitation of time preferences.

national government, the legal system and courts, honesty of elections, and the military. An additional item asks for people's confidence in the country's financial institutions and banks, and thus arguably captures a dimension of financial uncertainty as it applies to our survey items. In column (3) we control for average life expectancy at birth. Results show that patience continues to be a strong correlate of national income, conditional on objective or subjective institutional quality, or life expectancy.

7.6.1.0.4 Education. Our survey requires respondents to think through abstract choice problems, which might be unfamiliar and cognitively challenging for some participants. This could induce people to decide based on heuristics, perhaps due to low education. Table 7.11 presents the results of OLS estimations of per capita income on patience conditional on several proxies for the average educational attainment in a given country. Column (4) conditions on average years of schooling, while columns (5) through (7) control for various school achievement measures derived from standardized test scores (Eric A Hanushek and Woessmann, 2012). Despite the large correlation of these measures with GDP, patience remains statistically significant.

7.6.1.0.5 Measurement Error. A related concern is that the quantitative time preference measure could suffer from measurement error arising from censoring. Specifically, about half of all individuals in the world sample choose the immediate payment in all five

Table 7.11. Robustness

	Dependent variable: Log [GDP p/c PPP]							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Patience	0.67**	1.47***	1.38***	1.29***	1.55***	1.67***	1.20***	2.06***
	(0.33)	(0.42)	(0.38)	(0.32)	(0.35)	(0.34)	(0.36)	(0.30)
Property rights	0.033***							
	(0.01)							
Democracy	0.020							
	(0.04)							
Confidence in financial institutions		-3.60***						
		(0.91)						
Subj. institutional quality		0.041**						
		(0.02)						
Life expectancy			0.12***					
			(0.03)					
Average years of schooling				0.25***				
				(0.06)				
Cognitive skills (math and science)					0.72*			
					(0.37)			
% of students reaching basic literacy						1.47		
						(0.90)		
Share of top-performing students							12.1***	
							(3.92)	
Hofstede long-term orientation								-0.00065
								(0.01)
Constant	5.29***	8.10***	-0.36	5.80***	5.82***	7.93***	8.71***	6.53***
	(0.29)	(0.69)	(2.08)	(0.25)	(1.72)	(0.77)	(0.29)	(0.51)
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	73	57	76	71	49	49	49	61
R^2	0.797	0.750	0.793	0.783	0.678	0.664	0.716	0.726
Adjusted R^2	0.768	0.708	0.768	0.755	0.614	0.597	0.659	0.683

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

questions of the decision sequence, so that the elicitation procedure only allows putting a lower bound on impatience. We check in various ways whether the results could hinge on particular assumptions about the level of impatience assigned to these censored individuals, see Appendix 7.B.1 for details. First, we directly manipulate the quantitative value assigned to censored individuals, and find that the results are robust to using a wide range of different values. Second, we sidestep the issue of bias in quantitative interpretation by taking an approach that avoids assigning a quantitative meaning: We collapse the patience measure into an (ordinal) binary indicator for whether an individual's patience exceeds some patience threshold or not, so that a country's average patience corresponds to the fraction of the population that exceeds a given patience level. Across different patience thresholds, countries with a higher proportion of impatient individuals have lower per capita income. Third, even when we drop all censored respondents from the analysis, average patience of the uncensored population still correlates strongly with national income. Finally, we make use of the idea that median patience is unaffected by censoring concerns, as long as the median individual in a country is not censored.[23] In sum, we conclude that measurement error arising from censoring is unlikely to drive the results.

7.6.1.0.6 Culture. It is conceivable that our survey measure is related to GDP not due to time preference, but rather because it is correlated with a broader cultural trait. Specifically, Hofstede (2001) developed a qualitative long-term orientation variable that is only available at the country-level and is occasionally used in the literature (e.g., Galor and Özak, 2014).[24] This long-term orientation variable is intended to capture various dimensions of a broad concept of the perception of time.[25] However, it lacks a tight association with intertemporal tradeoffs between utility flows at different points in time. Moreover, this measure is based on a composite of responses to qualitative items that were chosen based on an ad hoc procedure. In contrast, our patience measure is based on a rigorous experimental validation procedure, and captures tradeoffs between immediate and delayed monetary rewards, so that it appears more appropriate as measure of time preferences in economically relevant domains. While our patience measure is in fact correlated with the long-term orientation variable ($\rho = 0.35$, $p < 0.01$), it has substantially more predictive power for national income than the Hofstede measure, see column (8).

7.6.2 Eliminating Environmental Factors at the Individual Level

The results so far suggest that patience is a significant correlate of accumulation decisions and income, consistent with micro-founded theories of economic development. As shown in the previous section, the findings at the aggregate level cannot be easily ex-

[23] As we discuss in detail in Appendices 7.B.1 and 7.C, all other findings in the paper also stand up to robustness checks about censoring.

[24] As Hofstede (2001) notes, "Long Term Orientation stands for the fostering of virtues oriented towards future rewards, in particular perseverance and thrift. It's opposite pole, Short Term Orientation, stands for the fostering of virtues related to the past and present, in particular, respect for tradition, preservation of 'face' and fulfilling social obligations."

[25] Two of the underlying items ask respondents to assess the statements "We should honour our heroes from the past." or "Are you the same person at work (or at school if you're a student) and at home?", see Hofstede's Values Survey Module Manual at http://www.geerthofstede.nl/vsm-08.

plained by third factors at the country level. In this section, we pursue an alternative way of evaluating the importance of cognitive skills, education, mortality or institutional quality in generating our cross-country results, by directly accounting for individual-level variation in these aspects of the environment. In particular, we "clean" individual patience of the influence of these variables and then use employ the residual patience as proxy for people's actual patience.

The GPS contains detailed information on several relevant aspects of the environment, which we organize in four distinct groups. First, we condition patience on a proxy for cognitive skills in the GPS data set, in which respondents were asked to self-assess their own mathematical skills (on a scale from 0 to 10). This variable helps eliminating potential systematic differences in ability or cognitive skills that allow individuals to form better beliefs or make better predictions that allow them to make more patient choices. As second factor, we control for the respondents' educational attainment (as a 3-step category). While education has been treated as outcome so far, one might also argue that better educated individuals are more capable of making intertemporal decisions because they are better trained. Higher cognitive skills and better education together imply that individuals are more capable of computing the returns on delayed rewards, thus making them appear more patient.[26] Third, we account for the expected remaining years of life of each respondent, which are imputed at the country-cohort-gender level using data from the UN World Population Prospects, as well as the respondents' subjective perception of their health status. In other words, for each individual, we compute the average expected remaining years of life given their country, age, and gender. Individuals who face a shorter expected life and a higher risk of death might appear less patient because of their lower subjectively perceived time horizon. Controlling for individual life expectancy, we can rule out that time horizon effects interfere with measured patience. Finally, we condition patience on respondents' confidence in the quality of the institutional environment in terms of the subjectively perceived overall quality of institutions (including confidence in the national government, the legal system and courts, honesty of elections, and the military), a perception of law and order, of corruption, and of the respondents' confidence in financial institutions and banks.

Appendix 7.B.2 shows that each of these four variables explains between 1.1% and 3.1% of the observed variation in individual-level patience; jointly they explain 6.5%. Using the residuals from these regressions, we can construct a measure of "actual" individual patience that is purged of the potential influence of environmental factors at the individual level.

It turns out that the environmental component of patience (i.e., the predicted value from the respective regression) explains 42% of the variation in income per capita across countries, which is more than the raw measure of patience, which explains less than 40% (see Table 7.3). Hence, the purging procedure seems to be successful and consistent with the intuition that part of the effect of patience at the macro level is driven by unobserved heterogeneity that is systematically correlated with development as well as patience, such as institutional quality or life expectancy.

---

[26] We obtain similar results when not conditioning on education.

Then, the question remains whether the aggregated component of the patience measure that is not explained by these perceptions and factors at the country level (the purged measure of patience, i.e., the county-means of the residuals of the regression of patience on subjective perceptions of the environment) still have explanatory power for development differences.

The respective results are shown in Table 7.12. Regardless of the set of variables that are eliminated from the patience measure at the individual level, the relationship between per capita income and the purged patience variable is highly significant and positive. Quantitatively, the coefficient estimates are comparable to those obtained with the raw measure shown in Table 7.3, even though the purged measure explains less of the variation in incomes across countries (in particular when patience is purged of all variables jointly, which approximately halves the explained variance in terms of R^2). This documents that unobserved heterogeneity in terms of the subjective environment that might influence the patience revealed in the responses to the survey questions cannot explain the correlation between patience and income. This is strong evidence against a merely spurious effect, and in favor of the role of patience as implied by micro-founded theories of development.

Table 7.12. Patience and income: Purging patience of cognitive skills, life expectancy, health, education, and subjective institutional quality

	Dependent variable: Log [GDP p/c PPP]									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Patience purged of cognitive ability	2.66*** (0.29)	2.17*** (0.31)								
Patience purged of life expectancy and health			2.67*** (0.29)	2.15*** (0.30)						
Patience purged of education					2.47*** (0.33)	2.09*** (0.31)				
Patience purged of institutions							2.57*** (0.40)	2.13*** (0.39)		
Patience purged of all variables									2.36*** (0.49)	2.24*** (0.40)
Constant	8.30*** (0.14)	9.08*** (0.21)	8.26*** (0.14)	6.56*** (0.25)	8.23*** (0.15)	9.12*** (0.22)	8.15*** (0.17)	6.72*** (0.39)	8.09*** (0.18)	6.45*** (0.41)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	76	76	74	74	76	76	54	54	54	54
R^2	0.361	0.679	0.365	0.659	0.288	0.658	0.310	0.642	0.219	0.629
Adjusted R^2	0.353	0.646	0.357	0.629	0.279	0.623	0.297	0.596	0.204	0.582

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

7.7 Concluding Remarks

Using a unique novel data set, this paper has provided the first systematic investigation of the relationship between patience, accumulation behavior, and income on a global scale. Taken together, our findings could be interpreted as providing encouraging empirical support for a large body of theoretical work. After all, both micro- and macroeconomic theories of intertemporal choice highlight the crucial role of time preference in driving future-oriented behaviors and ultimately income. Given the strong and robust relationships between patience and accumulation processes across different levels of aggregation, and on a global scale, an interesting question concerns the origins of variation in time preferences, both at the individual and at the country level. Among the few candidate determinants that have been proposed are religion (Weber, 1930), cultural legacy as manifested in very old linguistic features (Chen, 2013) as well as historical agricultural productivity and crop yield (Galor and Özak, 2014). G. S. Becker and Mulligan (1997) and Doepke and Zilibotti (2008, 2013) emphasize the two-way links between patience and education as well as education and income. The results from our paper suggest that much insight is to be gained from further delving into the determinants of time preference.

References

Acemoglu, Daron (2008): *Introduction to Modern Economic Growth*. Princeton University Press. [308]

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91 (5), 1369–1401. [310]

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2002): "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics*, 1231–1294. [310]

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2005): "Institutions as a Fundamental Cause of Long-Run Growth." *Handbook of Economic Growth*, 1. Ed. by P. Aghion and S.N. Durlauf, 385–472. [310]

Aghion, Philippe and Peter Howitt (1992): "A Model of Growth through Creative Destruction." *Econometrica*, 60 (2), 323–351. [308]

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003): "Fractionalization." *Journal of Economic Growth*, 8 (2), 155–194. [310, 320, 363]

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013a): "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics*, 128 (2), 469–530. [349]

Alesina, Alberto, Johann Harnoss, and Hillel Rapoport (2013b): "Birthplace Diversity and Economic Prosperity." *NBER Working Paper*, 18699. [310]

Algan, Yann and Pierre Cahuc (2010): "Inherited Trust and Growth." *American Economic Review*, 2060–2092. [310]

Alsan, Marcella (2015): "The Effect of the TseTse Fly on African Development." *American Economic Review*, 105 (1), 382–410. [310]

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113 (1), 151–184. [320]

Ashraf, Quamrul and Oded Galor (2011): "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review*, 101 (5), 2003–2014. [350]

Ashraf, Quamrul and Oded Galor (2013): "The Out of Africa Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review*, 103 (1), 1–46. [310, 320, 350, 359, 362, 363]

Ashraf, Quamrul, Oded Galor, and Marc P. Klemp (2014): "The Out of Africa Hypothesis of Comparative Development Reflected by Nighttime Light Intensity." *Working Paper*. [320]

Barro, Robert (2003): *Religion Adherence Data*. http://scholar.harvard.edu/barro/publications/religion-adherence-data. [363]

Barro, Robert J. and Jong-Wha Lee (2012): "A New Data Set of Educational Attainment in the World, 1950–2010." *Journal of Development Economics*. [360]

Barro, Robert J. and Rachel M. McCleary (2003): "Religion and Economic Growth Across Countries." *American Sociological Review*, 68 (5), 760–781. [310]

Becker, Anke, Benjamin Enke, and Armin Falk (2015): "The Ancient Origins of Cross-Country Heterogeneity in Risk Preferences." *Working Paper*. [349]

Becker, Gary S. (1962): "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy*, 70 (5), 9–49. [308]

Becker, Gary S. and Casey B. Mulligan (1997): "The Endogenous Determination of Time Preference." *Quarterly Journal of Economics*, 729–758. [334]

Ben-Porath, Yoram (1967): "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, 75 (4), 352–365. [308]

Borghans, Lex, Angela Lee Duckworth, James J Heckman, and Bas Ter Weel (2008): "The economics and psychology of personality traits." *Journal of Human Resources*, 43 (4), 972–1059. [307]

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin (2011): "Finance and Development: A Tale of Two Sectors." *American Economic Review*, 101 (5), 1964–2002. [329, 339, 362]

Campante, Filipe R. and David Yanagizawa (forthcoming): "Does Religion Affect Economic Growth and Happiness?: Evidence from Ramadan." *Quarterly Journal of Economics*. [310]

Caselli, Francesco and James Feyrer (2007): "The Marginal Product of Capital." *Quarterly Journal of Economics*, 122 (2), 535–568. [310]

Castillo, Marco, Paul J. Ferraro, Jeffrey L. Jordan, and Ragan Petrie (2011): "The today and tomorrow of kids: Time preferences and educational outcomes of children." *Journal of Public Economics*, 95 (11), 1377–1385. [310]

Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace (2009): "Genetic Variation in Preferences for Giving and Risk Taking." *Quarterly Journal of Economics*, 124 (2), 809–842. [349]

Chabris, Christopher F., David Laibson, Carrie L. Morris, Jonathon P. Schuldt, and Dmitry Taubinsky (2008): "Individual Laboratory-Measured Discount Rates Predict Field Behavior." *Journal of Risk and Uncertainty*, 37 (2-3), 237–269. [310]

Chen, M. Keith (2013): "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets." *American Economic Review*, 103 (2), 690–731. [310, 334, 349]

Cornsweet, Tom N. (1962): "The Staircase-Method in Psychophysics." *American Journal of Psychology*, 75 (3), 485–491. [312]

Das, Mausumi (2003): "Optimal Growth with Decreasing Marginal Impatience." *Journal of Economic Dynamics and Control*, 27, 1881–1898. [322]

Dean, Mark and Anja Sautmann (2014): "Credit Constraints and the Measurement of Time Preferences." *Working Paper*. [329]

DellaVigna, Stefano and M Daniele Paserman (2005): "Job Search and Impatience." *Journal of Labor Economics*, 23 (3). [310]

Diamond, Jared M. (2005): *Guns, Germs, and Steel*. National Geographic. [310]

Doepke, Matthias and Fabrizio Zilibotti (2008): "Occupational Choice and the Spirit of Capitalism." *Quarterly Journal of Economics*, 123 (2), 747–793. [334]

Doepke, Matthias and Fabrizio Zilibotti (2013): "Culture, Entrepreneurship, and Growth." In *Handbook of Economic Growth*. Ed. by Philippe Aghion and Steven Durlauf. Vol. 2A. Elsevier, Amsterdam. [334]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, 100 (3), 1238–1260. [329, 339]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2012): "The Intergenerational Transmission of Risk and Trust Attitudes." *Review of Economic Studies*, 79 (2), 645–677. [349]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2014): "Interpreting Time-Horizon Effects in Inter-Temporal Choice." *Working Paper*. [329]

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011): "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association*, 9 (3), 522–550. [329]

Drugeon, Jean-Pierre (1996): "Impatience and Long-Run Growth." *Journal of Economic Dynamics and Control*, 20, 281–313. [322]

Erosa, Andres, Tatyana Koreshkova, and Diego Restuccia (2010): "How Important is Human Capital? A Quantitative Theory Assessment of World Income inequality." *Review of Economic Studies*, 77 (4), 1421–1449. [310]

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2015a): "The Nature and Predictive Power of Preferences: Global Evidence." *Working Paper*. [307, 308, 311–313, 315, 320, 339, 364]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde (2015b): "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Working Paper*. [311, 313]

Gallup, John L., Jeffrey D. Sachs, et al. (2000): "The Economic Burden of Malaria." *American Journal of Tropical Medicine and Hygiene*, 64 (1-2 Suppl), 85–96. [363]

Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger (1999): "Geography and Economic Development." *International Regional Science Review*, 22 (2), 179–232. [310, 363]

Galor, Oded and Ömer Özak (2014): "The Agricultural Origins of Time Preference." *Working Paper*. [310, 331, 334, 349]

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2013): "Human Capital and Regional Development." *Quarterly Journal of Economics*, 105, 164. [309, 314, 317, 364]

Glaeser, Edward L, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2004): "Do Institutions Cause Growth?" *Journal of Economic Growth*, 9 (3), 271–303. [310]

Golsteyn, Bart H.H., Hans Grönqvist, and Lena Lindahl (2014): "Adolescent Time Preferences Predict Lifetime Outcomes." *Economic Journal*. [310]

Grosjean, Pauline (forthcoming): "A History of Violence: The Culture of Honor as a Determinant of Homicide in the US South." *Journal of the European Economic Association*. [349]

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2009): "Cultural Biases in Economic Exchange?" *Quarterly Journal of Economics*, 124 (3), 1095–1131. [310]

Hall, Robert E. and Charles Jones (1999): "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 114 (1), 83–116. [310, 327, 361]

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor (1996): "Aggregation and the Estimated Effects of School Resources." *Review of Economics and Statistics*, 78 (4), 611–627. [319]

Hanushek, Eric A and Ludger Woessmann (2012): "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation." *Journal of Economic Growth*, 17 (4), 267–321. [325, 330, 341, 360]

Hofstede, Geert (2001): *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. 2nd ed. London: Sage. [331]

Knack, Stephen and Philip Keefer (1997): "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics*, 1251–1288. [310, 320]

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny (1999): "The Quality of Government." *Journal of Law, Economics, and Organization*, 15 (1), 222–279. [310, 363]

Lucas, Robert E. (1988): "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22 (1), 3–42. [308, 322]

Manuelli, Rodolfo E. and Ananth Seshadri (2014): "Human Capital and the Wealth of Nations." *American Economic Review*, 104 (9), 2736–2762. [310]

Meier, Stephan and Charles D. Sprenger (2013): "Discounting financial literacy: Time preferences and participation in financial education programs." *Journal of Economic Behavior & Organization*, 95, 159–174. [310]

Michalopoulos, Stelios (2012): "The Origins of Ethnolinguistic Diversity." *American Economic Review*, 102 (4), 1508–1539. [362]

Nordhaus, William D. (2006): "Geography and Macroeconomics: New Data and New Findings." *Proceedings of the National Academy of Sciences of the United States of America*, 103 (10), 3510–3517. [362, 363]

Nunn, Nathan (2008): "The Long-Term Effects of Africa's Slave Trades." *Quarterly Journal of Economics*, 123 (1), 139–176. [310]

Nunn, Nathan and Leonard Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review*, 101 (7), 3221–52. [349]

Olsson, Ola and Douglas A. Hibbs Jr (2005): "Biogeography and Long-Run Economic Development." *European Economic Review*, 49 (4), 909–938. [310]

Putterman, Louis and David N. Weil (2010): "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality." *Quarterly Journal of Economics*, 125 (4), 1627–1682. [350, 359, 364]

Rajan, Raghuram G. and Luigi Zingales (1998): "Financial Dependence and Growth." *American Economic Review*, 559–586. [329, 339]

Rieger, Marc-Oliver, Mei Wang, and Thorsten Hens (forthcoming): "Risk Preferences Around the World." *Management Science*. [311]

Romer, Paul M. (1990): "Endogenous Technological Change." *Journal of Political Economy*, 98 (5), S71–S102. [308]

Solow, Robert M. (1957): "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, 39 (3), 312–320. [310]

Spolaore, Enrico and Romain Wacziarg (2009): "The Diffusion of Development." *Quarterly Journal of Economics*, 124 (2), 469–529. [310, 364]

Strulik, Holger (2012): "Patience and Prosperity." *Journal of Economic Theory*, 147, 336–352. [308, 322]

Sutter, Matthias, Martin G Kocher, Daniela Glätzle-Rützler, and Stefan T Trautmann (2013): "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *The American Economic Review*, 103 (1), 510–531. [310]

Tabellini, Guido (2010): "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association*, 8 (4), 677–716. [310]

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen (2010): "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review*, 100 (1), 557–571. [310]

Vieider, Ferdinand M., Thorsten Chmura, Tyler Fisher, Takao Kusakawa, Peter Martinsson, Frauke Mattison-Thompson, and Adewara Sunday (2015): "Within- versus Between-Country Differences in Risk Attitudes: Implications for Cultural Comparisons." *Theory and Decision*, 78 (2), 209–218. [311]

Vieider, Ferdinand M., Mathieu Lefebvre, Ranoua Bouchouicha, Thorsten Chmura, Rustamdjan Hakimov, Michal Krawczyk, and Peter Martinsson (2014): "Common Components of Risk and Uncertainty Attitudes Across Contexts and Domains: Evidence from 30 Countries." *Journal of the European Economic Association*, 13 (3). [311]

Voigtländer, Nico and Hans-Joachim Voth (2012): "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany." *Quarterly Journal of Economics*, 127 (3), 1339–1392. [349]

Wang, Stephanie W. (2011): "Incentive Effects: The Case of Belief Elicitation from Individuals in Groups." *Economics Letters*, 111 (1), 30–33. [311]

Weber, Max (1930): *The Protestant Ethic and the "Spirit" of Capitalism*. New York: Charles Scribner's Sons. [334]

Appendix 7.A Details on Data Collection and Patience Measure

See Falk et al. (2015a).

Appendix 7.B Environmental Conditions and Revealed Patience

7.B.1 Controlling for Country-Level Observables: Details

7.B.1.0.1 Financial Environment. Respondents who expect higher levels of inflation might appear more impatient in our quantitative choice task as they require a compensation for inflation. Similarly, high market interest rates could induce people to behave as if they were impatient because they might try to "arbitrage" between the local credit market and the hypothetical interest rates implied in the quantitative survey measure. Cross-country differences in inflation and interest rates could bias our results on the relationship between comparative development and patience. It is important to recall, however, that our survey question explicitly asked people to imagine a zero inflation environment. Likewise, previous research has found that differences in interest rates are unlikely to drive choices in small-stakes experimental environments (e.g. Dohmen et al., 2010). To empirically address this issue, we explicitly control for inflation (in terms of the consumer price index, or the GDP deflator) and deposit interest rates. Columns (1) through (4) of Table 7.13 present the corresponding results. As one would expect, both inflation and high interest rates are negatively correlated with GDP. The coefficient of patience, however, remains quantitatively large and highly statistically significant when conditioning on these variables. In addition, the coefficient estimate is only slightly smaller in size.

7.B.1.0.2 Borrowing Constraints. Ceteris paribus, respondents who face upward-sloping income profiles and are borrowing-constrained might be more likely to opt for immediate payments in experimental choice situations not because of intrinsic preferences, but rather because of a current cash shortfall. Since participants in rather poor countries seem more likely to face such constraints, responses in our survey could make such populations appear less patient than they actually are, and hence drive the relationship between patience and development. Note, however, that all monetary values in the elicitation of the quantitative patience measure were adjusted in terms of purchasing power parity to be approximately comparable across countries, hence minimizing problems of income differences. Empirically, we approach the issue of borrowing constraints and financial development from two separate angles. On the one hand, we complement our baseline specification by including two additional covariates, which capture different dimensions of the level of financial development of a given country. In particular, we make use of the commonly employed ratio of external finance to GDP, where external finance is defined as the sum of private credit, private bond market capitalization, and stock market capitalization (Rajan and Zingales, 1998; Buera et al., 2011). In addition, we use the (log) number of Automated Teller Machines (ATMs), which arguably captures elements of the accessibility of cash for private households, as a measure of financial development. Columns (5)-(7) of Table 7.13 present the corresponding regression results, which provide reassuring evidence that the relationship between comparative develop-

Table 7.13. Patience and income: Robustness against inflation, interest rates, and borrowing constraints

	Dependent variable: Log [GDP p/c PPP]							
	The specifications address concerns regarding...							
	Inflation and interest rates				Borrowing constraints			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Patience	1.15***	1.09***	0.95*	0.79*	1.11***	1.02**		
	(0.39)	(0.37)	(0.48)	(0.43)	(0.34)	(0.49)		
Patience of top income quintile							2.28***	1.37***
							(0.18)	(0.27)
Log [CPI]	-0.38***							
	(0.13)							
Log [GDP deflator]		-0.79***		-0.59***				
		(0.16)		(0.17)				
Log [Deposit interest rate]			-0.69***	-0.35				
			(0.16)	(0.21)				
Log [# ATMs]					0.56***			
					(0.11)			
External finance as % of GDP						0.25		
						(0.18)		
Constant	-161.5***	-221.6***	-193.5***	-234.6***	-41.8	-178.4***	8.01***	-136.3**
	(56.38)	(56.79)	(59.31)	(56.93)	(52.16)	(48.69)	(0.14)	(52.64)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Observations	73	73	63	62	74	55	76	74
R^2	0.884	0.895	0.884	0.898	0.907	0.910	0.453	0.860
Adjusted R^2	0.833	0.848	0.821	0.836	0.867	0.848	0.446	0.804

OLS estimates, robust standard errors in parentheses. Log CPI, GDP deflator, and deposit interest rate are calculated as log (1 + x), where x is the respective variable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

ment and our patience measure is largely unaffected by the level of a country's financial development. To reiterate this point from a different angle, we make use of the idea that borrowing constraints (if present) are likely to be less binding for relatively rich people. Thus, rather than computing simple country averages of our patience measure across all respondents, we compute the average patience of each country's top income quintile only and use this measure instead of the population average patience measure. As shown in columns (6) and (7), both in unconditional and conditional regressions the relationship between GDP and patience is very similar to the baseline results using all respondents, which again suggests that borrowing constraints on the part of respondents are unlikely to be a main driver of our results.

7.B.1.0.3 Context and Cross-Cultural Differences. Conducting surveys in culturally heterogeneous samples poses the difficulty that respondents might interpret survey items in different ways. This problem appears particularly severe in the case of qualitative and context-specific items. Recall, however, that our quantitative question format provided a specific, neutral, and quantitative choice context for respondents, hence alleviating the need to construe alternative choice scenarios. Given that this quantitative item is arguably less prone to culture-dependent interpretations, we conduct a further robustness check in which we show that similar results obtain if we only employ the quantitative measure. The latter measure is also more in line with how economists define and measure time preferences, i.e., choices over monetary rewards at different points in time. Columns (1) and (2) of Table 7.14 show that using only the quantitative measure in fact strengthens the results; in contrast, the qualitative item alone is only weakly correlated

Table 7.14. Patience and income: Decision heuristics

	Dependent variable: Log [GDP p/c PPP]					
	OLS		WLS		WLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Staircase patience	0.27***	0.18***	0.24***	0.17***	0.23***	0.13***
	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.04)
Constant	6.25***	-178.6***	6.79***	-138.8**	6.98***	-148.4
	(0.25)	(59.06)	(0.27)	(60.58)	(0.29)	(129.64)
Additional controls	No	Yes	No	Yes	No	Yes
Observations	76	74	71	70	49	48
R^2	0.465	0.865	0.535	0.900	0.537	0.867
Adjusted R^2	0.457	0.811	0.528	0.856	0.527	0.760

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls. In columns (3) and (4), each observation is weighted by average years of schooling, and in columns (5) and (6) by average cognitive skills.

with national income, suggesting that culture-dependent interpretations might indeed be an issue regarding this particular qualitative self-assessment. In this respect, it should also be noted that the qualitative item is also a weaker predictor of financially incentivized behavior in the validation experiments as compared to the quantitative measure (though it is significantly related to experimental choices).

7.B.1.0.4 Cognitive Limitations and Heuristics. An issue that specifically concerns the quantitative patience measure is the possibility of measurement error, due to the use of heuristics tied to limited cognitive resources. For example, since our five-step procedure forces respondents into five fairly similar decision problems, respondents might adopt simple rules such as "always choose money today" in order to reduce cognitive burden. Indeed, in our data, 55 % of respondents do always choose the immediate payment. Notably, to bias our results, the prevalence of decision shortcuts would have to be systematic at the *country* level, rather than at the individual level, presumably due to differing levels of education. As a first robustness check, we repeat our baseline specifications, but now weight each country by its average years of schooling, to reflect potentially less accurate measures of time preference in general for populations with less education. Columns (3) and (4) of Table 7.14 show that this procedure yields results similar to the baseline results. In Columns (5) and (6) we try an alternative approach, weighting countries by average cognitive skills as proxied by performance on standardized achievement tests (see Eric A Hanushek and Woessmann (2012) for a discussion of the cognitive skills measure), and find similar results. In unreported regressions we also verify that the results on development and the regional and individual levels also go through when we weight responses by appropriate measures of educational attainment, or cognitive skills. Taken together, we find little indication that variation in the use of heuristics across countries or individuals could drive our results.

7.B.1.0.5 Censoring of the Quantitative Patience Measure. A related concern is that the quantitative patience measure might suffer from measurement error due to censoring. For individuals who always choose the immediate payment in all choices, up to the maximum possible delayed payment, the staircase procedure gives only a lower bound for the level of impatience (see Figure 5.5). This could potentially bias our results. In our robustness checks we distinguish between two cases: (1) A narrower case in which censored individuals might happen to have approximately the same true patience value, but we overstate this patience value (relative to uncensored observations) by assigning an upper bound level of 1; (2) a broader case in which individuals in the censored range might have substantially different true patience values, so that there is an additional bias from ignoring heterogeneity when we assign them an identical (upper bound) patience value.[27] Both cases would introduce bias at the individual level, and at higher levels of aggregation, because we assign a quantitative meaning to the difference in patience between the censored and uncensored observations, even though we do not observe the true switching point of censored individuals.

Starting with case (1), we test whether the results are robust to assigning alternative quantitative values of patience to the group of censored individuals. In columns (1)-(6) of Table 7.15, we impute arbitrary values to censored individuals, ranging from -2 to -50. The resulting estimates show that regardless of which value of patience one assumes for the censored individuals, our result about average patience and national income still holds.[28] This suggests that in case (1), a particular quantitative interpretation of the censoring value does not drive the findings. Notably, moving the censoring value to minus infinity corresponds to collapsing the patience measure to a binary indicator distinguishing between censored and non-censored observations. We check robustness to this specification below.

Turning to case (2), in which there is an additional bias due to unobserved heterogeneity, we first address what is fundamentally a problem about quantitative interpretation by taking an approach that avoids a quantitative interpretation altogether. To this end, we collapse the quantitative patience measure into a binary indicator for whether an individual's patience exceeds some patience threshold or not, so that a country's average patience is given by the fraction of the population that exceeds a given patience level. For instance, in columns (7) and (8) in Table 7.15, we binarize the data according to whether a given individual was censored or not. The results show that, despite the ordinal interpretation of the data, countries with a higher proportion of censored respondents have lower income per capita. Similarly, in columns (9) through (14), we introduce binary individual-level indicators for different patience cutoff levels (recall that the staircase variable is coded to be between 1 and 32, so that higher cutoffs at the margin discriminate between increasingly patient people). The results are robust to these different choices of cut-offs. This suggests that the conclusions are robust to a range of different qualitative definitions of patience, and do not hinge on a strict quantitative interpretation of variation in the patience measure.

---

[27] This corresponds to the notion of "expansion bias" arising from a censored regressor. If the true relationship between patience and GDP is linear, the "piling" up of observations at the (left-) censoring boundary for patience leads to an inflated (in absolute value) OLS coefficient on patience.

[28] Note that the manipulation affects some country averages more than others, due to varying proportions of censored individuals across countries, and thus the coefficient on patience need not change monotonically across the columns.

A second approach to addressing case (2) is to minimize the influence of censored observations on the analysis. In columns (15) and (16), we report results in which all censored individuals (equivalently, those using a heuristic to always choose immediate payments) are excluded from the calculations of country averages. The results show that countries with higher GDP have more patient uncensored populations.[29] In columns (17) and (18) we take a different approach, switching to median levels of patience because, unlike country averages, median values are unaffected by the presence of censored individuals in the population, as long as the median individual in a country is not censored. We then exclude entirely those countries for whom the median individual is censored. The regressions with the remaining set of countries indicate a strong and robust relationship between GDP and median level of patience.[30] Here, censoring bias is excluded, and the quantitative patience measure explains more than 50% of the variation in national income (compared to 42% when we include all median-censored countries). This indicates that the relationship between patience and GDP is not driven by censoring, and interestingly, shows that patience can explain differences in national income even among relatively rich nations in which the median individual is not censored. In addition, if anything, the higher explained variance within the group of uncensored countries suggests that the missing variation in the left tail of the country-level distribution prevents an even stronger relationship between patience and national income in terms of variation explained.

Other results in the paper also pass robustness checks for censoring. For example, the finding that even binary variants of the quantitative measure predict outcomes is neither confined to analyses with GDP as dependent variable nor to country-level analyses as a whole. Section 7.C illustrates, repeating all of our main country-, regional-, and individual-level analyses with a binary version of the quantitative measure in which a value of zero is assigned to censored individuals and of one to non-censored respondents. The results also hold for all measures of development at the country level, and at the individual level, if we perform similar analyses excluding all censored individuals, although the regional-level results weaken with this approach. Finally, note that to the extent that censoring is a manifestation of decision heuristics, these robustness checks provide further evidence that heterogeneity in cognitive skills does not drive the results.

[29] Notably, censoring is less prevalent in wealthier countries; this is consistent with a distribution of patience values that is shifted to the right for wealthy countries, and suggests censored populations are likely more patient in rich countries as well.

[30] We also find similar results if we regress GDP on patience of the median individual, and leave countries with a censored median in the analysis.

Table 7.15. Patience and national income: Robustness against censoring

	Dependent variable: Log [GDP p/c PPP]																	
	Independent variable is staircase patience in the following variations:																	
	Patience at lower censoring point set to...						Fraction of respondents who are more patient than outcome node...								Exclude all censored individuals		Exclude all median-censored countries	
	-2		-20		-50		1		8		16		24					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Staircase patience	2.71***	1.73***	2.67***	1.69***	2.69***	1.68***	5.57***	3.38***	6.42***	3.87***	7.73***	4.88***	9.93***	6.11***	0.18***	0.11**		
	(0.22)	(0.35)	(0.21)	(0.33)	(0.21)	(0.32)	(0.48)	(0.67)	(0.57)	(0.83)	(0.72)	(1.09)	(1.10)	(1.51)	(0.05)	(0.05)		
Staircase patience (median)																	0.14***	0.14***
																	(0.03)	(0.04)
Constant	8.32***	-172.1***	8.32***	-153.9***	8.32***	-143.1**	5.90***	-129.9**	6.44***	-162.3***	6.84***	-186.8***	7.07***	-187.6***	5.56***	-178.1**	8.30***	8.68***
	(0.13)	(58.03)	(0.13)	(56.15)	(0.13)	(55.63)	(0.27)	(55.56)	(0.24)	(56.71)	(0.22)	(61.99)	(0.21)	(66.16)	(0.77)	(78.81)	(0.39)	(0.58)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	No
Observations	76	74	76	74	76	74	76	74	76	74	76	74	76	74	76	74	25	25
R^2	0.473	0.866	0.484	0.864	0.482	0.862	0.470	0.856	0.434	0.860	0.422	0.861	0.383	0.849	0.121	0.819	0.522	0.543
Adjusted R^2	0.466	0.811	0.477	0.809	0.475	0.806	0.463	0.799	0.426	0.804	0.415	0.804	0.375	0.788	0.109	0.745	0.502	0.423

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls. In columns (1)-(6), the independent variable consists of the quantitative patience measure, where varying values are assumed for the lower censoring point, i.e., for those respondents who always opt for the payment today (these variables are standardized at the individual level). In columns (7)-(14), the independent variable consists of the fraction of respondents in a given country who are more patient than a given outcome node, see Figure 5.5. In columns (15)-(16), the staircase patience variable is computed is average staircase outcomes of those respondents who are not censored, i.e., all censored individuals are excluded from the sample. In columns (17) and (18), we the independent variable is the median quantitative patience measure and we exclude all countries from the analysis in which the median is censored.

7.B.2 Purging Individual Patience

Table 7.16 presents the results of the first stage of purging individual patience of cognitive skills, education, life expectancy, health, and various dimensions of subjective institutional quality. In column (1), we regress all individual-level observations of patience on people's self-assessment of their math skills (as 11 distinct categories). In column (2), we use each individual's educational attainment (as a 3-step category). In column (3), we employ the remaining years of life of each individual as a continuous variable. This measure is imputed separately for each cohort and gender in a given country. In addition, column (3) includes a continuous subjective physical health index. This index is included in Gallup's background data and aggregates responses to five items, e.g., "Do you have any health problems that prevent you from doing any of the things people your age normally can do?", see Appendix 7.H for details. In column (4), we use four variables which proxy for the respondent's overall assessment of the quality of their institutional environment. The first index concerns the overall quality of institutions (including confidence in the national government, the legal system and courts, honesty of elections, and the military); the second index is about law and order (asking, e.g., whether people recently had money or porperty stolen), and the third continuous index about corruption (asking whether corruption is widespread in business and the government). Finally, this specification also includes a binary variable measuring respondendent's confidence in financial institutions and banks. Column (5) combines all previous specifications. In sum, our proxies for individual-level characteristics cognitive skills, education, mortality, and subjective institutional quality jointly explain 6.3% of the global variation in time preferences. We proceed by generating the residuals of these regressions and employing these as a "cleaned" patience variable for further analysis, see Section 7.6.2.

Table 7.16. Purging individual patience: First stage

	Dependent variable: Patience				
	(1)	(2)	(3)	(4)	(5)
Constant	-0.20***	-0.21***	-0.36***	-0.037**	-0.53***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Cognitive skills	Yes	No	No	No	Yes
Education	No	Yes	No	No	Yes
Life expectancy and health	No	No	Yes	No	Yes
Subj. institutional quality	No	No	No	Yes	Yes
Observations	78755	79357	77414	50482	49603
R^2	0.019	0.031	0.011	0.021	0.065
Adjusted R^2	0.019	0.031	0.011	0.021	0.064

OLS estimates.

Appendix 7.C Main Results Based on a Binary Version of the Quantitative Patience Measure

This section provides a robustness check for all main results in the paper using a binary version of the quantitative patience measure as explanatory variable. Specifically, we assign a value of zero (one) to all (non-) censored individuals, so that the country-level patience measure consists of the fraction of respondents who are not censored. Figure 7.3 illustrates the distribution of this patience variable across countries. Even using this much coarser measure of patience do all of our results on the relationship between patience, national income, growth rates, and the proximate determinants hold, see Tables 7.17 and 7.18.[31] In fact, as Figure 7.4 illustrates, this binarized patience measure confers an even stronger relationship with national income than our baseline measure because it also captures meaningful variation in patience within the group of fairly impatient countries (which are often "almost censored" using the average baseline staircase measure).

As Tables 7.19 and 7.20 show, the subnational results are also robust to employing this binarized patience measure.

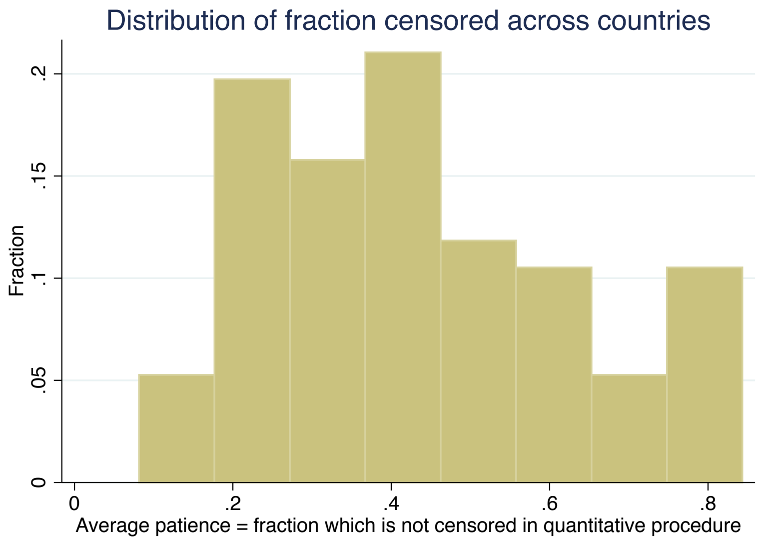


Figure 7.3. Distribution of average binarized patience variable across countries. Each individual is assigned a value of one if they are not censored and zero otherwise.

[31] Results for all other dependent variables (i.e., other proxies for the proximate determinants) closely mirror those established in the main text and are available upon request.

Table 7.17. Replicate main country-level results with binarized quantitative patience measure (1/2)

	Dependent variable:													
	Log [current GDP p/c]		Log [GDP p/c 1925]		Log [GDP p/c 1870]		Log [GDP p/c 1820]		Growth since 1975		Growth since 1925		Growth since 1820	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Staircase patience (fraction not censored)	5.57***	3.38***	1.84***	1.67***	1.75***	1.25**	1.03***	-0.10	1.45*	2.37*	1.09**	2.46***	0.71***	0.57
	(0.48)	(0.67)	(0.33)	(0.48)	(0.32)	(0.51)	(0.32)	(0.31)	(0.85)	(1.29)	(0.52)	(0.70)	(0.20)	(0.34)
Log [GDP p/c base year]										-0.80**		-1.04***		-0.63***
										(0.34)		(0.22)		(0.15)
Constant	5.90***	-129.9**	6.83***	7.31***	6.07***	6.01***	6.09***	6.67***	0.95**	7.16***	1.44***	9.02***	1.06***	4.86***
	(0.27)	(55.56)	(0.16)	(0.39)	(0.13)	(0.29)	(0.13)	(0.17)	(0.42)	(2.06)	(0.28)	(1.59)	(0.09)	(0.96)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Additional controls	No	Yes	No	No	No	No	No	No	No	No	No	No	No	No
Observations	76	74	33	33	42	42	31	31	66	66	33	33	31	31
R^2	0.470	0.856	0.400	0.642	0.369	0.612	0.240	0.704	0.030	0.472	0.125	0.582	0.195	0.584
Adjusted R^2	0.463	0.799	0.380	0.559	0.353	0.532	0.214	0.614	0.015	0.397	0.097	0.465	0.167	0.432

OLS estimates, robust standard errors in parentheses. The patience variable is the fraction of respondents who are not left-censored in the quantitative patience procedure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

Table 7.18. Replicate main country-level results with binarized quantitative patience measure (2/2)

	Dependent variable:													
	Average years of schooling		Education expenditure		Log [Capital stock p/c]		Net adjusted savings		Total factor productivity		R&D expenditure		Property rights	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Staircase patience (fraction not censored)	9.79***	7.30***	3.06***	3.42***	4.52***	2.35***	12.3***	20.5***	0.72***	0.29	4.02***	3.35**	85.7***	71.9***
	(0.90)	(1.70)	(0.63)	(1.03)	(0.53)	(0.67)	(4.33)	(6.68)	(0.12)	(0.24)	(0.51)	(1.29)	(8.56)	(15.08)
Constant	1.14**	-32.2	2.96***	-50.3	8.07***	-126.5***	4.94**	496.9	0.30***	2.07	-0.81***	27.9	11.4***	461.9
	(0.45)	(117.14)	(0.33)	(116.76)	(0.31)	(38.51)	(2.34)	(708.39)	(0.07)	(14.28)	(0.20)	(47.20)	(4.26)	(1293.34)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	71	70	71	70	71	69	68	68	60	59	64	63	74	72
R^2	0.509	0.801	0.170	0.577	0.440	0.871	0.070	0.536	0.350	0.724	0.534	0.674	0.484	0.641
Adjusted R^2	0.502	0.713	0.158	0.392	0.432	0.813	0.056	0.324	0.339	0.567	0.527	0.507	0.477	0.491

OLS estimates, robust standard errors in parentheses. The patience variable is the fraction of respondents who are not left-censored in the quantitative patience procedure. Education and R&D expenditure are % of GDP, while net adjusted savings are % of GNI. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

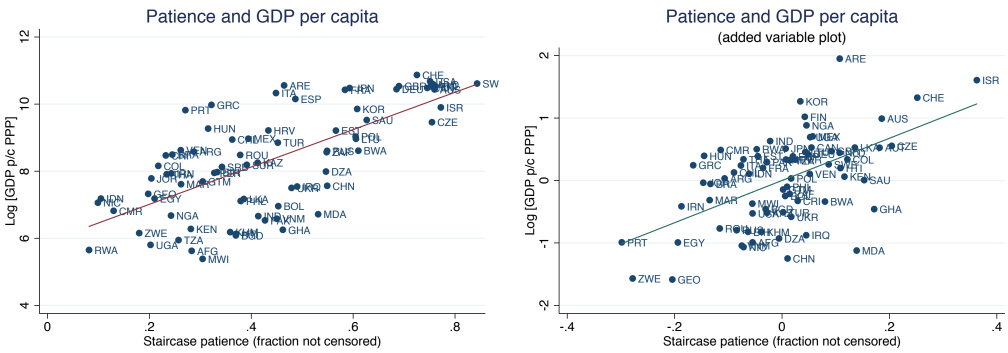


Figure 7.4. Patience and national income. The left panel depicts the raw correlation between log GDP per capita (purchasing-power parity) and patience. Patience is the fraction of respondents in a given country who never switch to preferring the delayed payment in the staircase elicitation procedure, i.e., who are left-censored. The right panel contains a plot conditional on the full set of baseline covariates in column (7) of Table 7.3.

Table 7.19. Main regional-level results replicated with binarized quantitative patience measure

	Dependent variable:					
	Log [Regional GDP p/c]			Avg. years of education		
	(1)	(2)	(3)	(4)	(5)	(6)
Staircase patience (fraction not censored)	3.36***	0.50***	0.58***	7.76***	0.64	0.85**
	(0.60)	(0.18)	(0.15)	(1.01)	(0.52)	(0.39)
Temperature			-0.013			-0.048***
			(0.01)			(0.02)
Inverse distance to coast			0.69			1.40**
			(0.43)			(0.54)
Log [Oil production p/c]			0.21***			-0.11
			(0.06)			(0.13)
# Ethnic groups			-0.14**			-0.37***
			(0.05)			(0.13)
Log [Population density]			0.083*			0.22**
			(0.04)			(0.09)
Constant	7.31***	8.99***	8.38***	3.84***	7.09***	6.09***
	(0.37)	(0.05)	(0.36)	(0.55)	(0.15)	(0.49)
Country FE	No	Yes	Yes	No	Yes	Yes
Observations	704	704	687	693	693	676
R^2	0.274	0.938	0.952	0.345	0.936	0.957
Adjusted R^2	0.273	0.932	0.947	0.344	0.930	0.954

Weighted least squares estimates, observations weighted by number of respondents in region. Standard errors (clustered at country level) in parentheses. The patience variable is the fraction of respondents who are not left-censored in the quantitative patience procedure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7.20. Main individual-level results replicated with binarized quantitative patience measure

	Dependent variable:							
	Log [HH income p/c]		Saved last year		Education level		Health problems	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Staircase patience (1 if not censored)	0.092*** (0.01)	0.076*** (0.01)	0.066*** (0.02)	0.047** (0.02)	0.089*** (0.01)	0.052*** (0.01)	-0.038*** (0.01)	-0.014*** (0.00)
Age		0.67*** (0.17)		-0.070 (0.31)		1.07*** (0.26)		0.15* (0.09)
Age squared		-0.42** (0.20)		-0.050 (0.30)		-1.89*** (0.25)		0.66*** (0.10)
1 if female		-0.072*** (0.02)		-0.0062 (0.01)		-0.027** (0.01)		0.027*** (0.00)
Subj. math skills		0.037*** (0.00)		0.018*** (0.00)		0.047*** (0.00)		-0.0081*** (0.00)
Confidence in financial institutions				0.067*** (0.01)				
Constant	6.43*** (0.00)	5.89*** (0.07)	0.15*** (0.00)	0.26*** (0.07)	1.28*** (0.00)	0.90*** (0.06)	0.29*** (0.00)	0.43*** (0.03)
Country FE	Yes	No	Yes	No	Yes	No	Yes	No
Regional FE	No	Yes	No	Yes	No	Yes	No	Yes
Religion FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	79245	68632	15260	10438	79357	68855	62727	59043
R^2	0.607	0.656	0.071	0.134	0.200	0.326	0.025	0.152
Adjusted R^2	0.607	0.650	0.071	0.113	0.199	0.316	0.025	0.139

Columns (1)-(4) are OLS and columns (5)-(8) ordered probit estimates. Standard errors (clustered at country level) in parentheses. The dependent variable in (5)-(8) is educational attainment as a three-step category. In columns (5)-(8), the R^2 is a Pseudo-R^2. The patience variable is a dummy equal to one if the respondent was not censored in the staircase procedure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 7.D Patience and Historical Income

Accumulating evidence indicates that preferences are transmitted across generations (see, e.g., Cesarini et al., 2009; Dohmen et al., 2012), and that differences in future-orientation or time preferences may have deep cultural or environmental roots in the distant past (Chen, 2013; Galor and Özak, 2014).[32] If the relative distribution of patience (but not necessarily the absolute levels) across countries originates from agro-climatological conditions, becomes manifest in linguistic patterns, and exhibits substantial persistence over time, then the patience patterns found in contemporaneous populations should be related to not just contemporary, but also historical development. To test this hypothesis, we repeat the analysis with measures of historical income before and after the Industrial Revolution.

Columns (1)-(6) of Table 7.21 present the results of OLS regressions, in which we relate patience to (log) per capita income in 1925, 1870, and 1820, respectively.[33] Throughout, the results reveal positive and significant relationships, which hold up conditional on continent fixed effects.

In order to investigate whether such a relationship was already present in pre-industrial times, i.e., around 1500, we follow the literature and use (log) population

[32] Additional evidence for the persistence of preferences and cultural values in general comes from the work of, e.g., Nunn and Wantchekon (2011), Voigtländer and Voth (2012), Alesina et al. (2013a), A. Becker et al. (2015), and Grosjean (forthcoming).

[33] The choice of these years is due to data availability constraints in the Maddison data set.

density as proxy for economic development in the Malthusian epoch (Ashraf and Galor, 2011). We account for the compositional changes in the population since 1500 due to migration flows and compute an ancestry-adjusted measure of population density by adjusting the historical population density figures by post-Columbian migration flows using the world migration matrix of Putterman and Weil (2010).[34] In essence, we relate patience of today's population to the weighted average of population density that prevailed in the country of residence of their ancestors in 1500. For example, we relate the average patience of the contemporary US population to a weighted average of the past population density of the "source countries" of US immigrants such as the UK, China, or Angola.[35]

Columns (7) and (8) of Table 7.21 report the corresponding results. Consistent with the findings for contemporary income, patience exhibits a significant unconditional correlation with past population density. The inclusion of control variables leads to an even stronger correlation between patience and past development.[36] Columns (9)-(12) report the results from complementary regressions with non-adjusted population density in 1500 as dependent variable and excluding countries with particularly high migratory inflows, or countries from the New World, from the analysis.[37] These robustness checks deliver qualitatively and quantitatively similar results.

Appendix 7.E Details for Regional-Level Analysis

Our regional-level data contain 704 regions (typically states or provinces) from the following countries: Argentina (16), Australia (8), Austria (9), Bolivia (8), Brazil (24), Cambodia (14), Cameroon (10), Canada (10), Chile (12), China (23), Colombia (23), Czech Republic (7), Egypt (3), Germany (16), Finland (4), France (22), Georgia (10), Ghana (10), Great Britain (12), Greece (13), Hungary (7), India (24), Indonesia (17), Iran (30), Israel (6), Italy (17), Jordan (6), Kazakhstan (6), Kenya (8), Lithuania (10), Macedonia (3), Malawi (3), Mexico (28), Morocco (13), Nigeria (6), Nicaragua (17), Netherlands (12), Pakistan (4), Poland (16), Portugal (7), Romania (8), Russia (27), Serbia (2), Spain (19), Sri Lanka (9), Sweden (8), Tanzania (20), Thailand (5), Turkey (4), Uganda (4), Ukraine (27), United Arab Emirates (7), USA (51), South Africa (9), Zimbabwe (10)

---

[34] This procedure of computing ancestry-adjusted values is analogous to the standard procedure in the literature, see, e.g., Ashraf and Galor (2011, 2013).

[35] Notice that the reverse, a computation of the distribution of patience for historical populations, corrected for post-1500 migration flows, is not possible due to the missing information on the historical ancestry of the survey respondents.

[36] Figure 7.7 in Appendix 7.G visualizes these two results.

[37] These countries are Argentina, Australia, Brazil, Canada, and the United States. Columns (9) and (10) also exclude Serbia, for which no data on (non-adjusted) historical population density are available.

Table 7.21. Patience and historical development

	Dependent variable:											
	Log [GDP p/c] in...						Log [Population density in 1500]					
	1925		1870		1820		Full sample (ancestry-adjusted)		Low migration sample (non-adjusted)		Old World (non-adjusted)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Patience	0.91^{***}	0.74^{***}	0.84^{***}	0.63^{**}	0.54^{***}	-0.018	0.65^{*}	0.79^{**}	0.95^{*}	1.05^{***}	0.69	0.96^{**}
	(0.16)	(0.22)	(0.16)	(0.24)	(0.15)	(0.19)	(0.36)	(0.34)	(0.49)	(0.35)	(0.50)	(0.37)
Constant	7.58^{***}	7.99^{***}	6.81^{***}	6.58^{***}	6.51^{***}	6.62^{***}	1.83^{***}	-13.5^{***}	1.43^{***}	-10.7^{***}	1.60^{***}	-8.03^{*}
	(0.09)	(0.21)	(0.07)	(0.18)	(0.06)	(0.01)	(0.13)	(5.02)	(0.16)	(3.97)	(0.17)	(4.12)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Additional controls	No	No	No	No	No	No	No	Yes	No	Yes	No	Yes
Observations	33	33	42	42	31	31	75	74	69	69	59	59
R^2	0.421	0.628	0.390	0.625	0.293	0.703	0.048	0.598	0.061	0.772	0.035	0.748
Adjusted R^2	0.402	0.542	0.375	0.547	0.268	0.612	0.035	0.476	0.047	0.702	0.018	0.660

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In columns (1)-(6), the dependent variable is historical national income per capita. Due to the small number of obervations, we only control for continent fixed effects in these columns. In columns (7) and (8), the dependent variable is ancestry-adjusted population density in 1500. In columns (9) and (10), it is non-adjusted population density, and the sample excludes Argentina, Australia, Brazil, Canada, Serbia, and the United States, see footnote 37. In columns (11)-(12), we exclude the New World. See the text for details on the construction of the ancestry-adjusted variable. See column (7) of Table 7.3 for a complete list of the additional controls. In this table, the control vector excludes the colonization dummy, genetic diversity and its square, and ethnic fractionalization.

Appendix 7.F Additional Tables

Table 7.22. Patience and national income: Additional control variables

	Dependent variable: Log [GDP p/c PPP]					
	(1)	(2)	(3)	(4)	(5)	(6)
Patience	2.02***	1.87***	1.85***	1.40**	1.36**	1.47**
	(0.47)	(0.46)	(0.50)	(0.53)	(0.58)	(0.58)
Will. to take risks	-1.07**	-0.78*	-0.64	-1.05**	-0.84	-0.95*
	(0.41)	(0.42)	(0.43)	(0.49)	(0.54)	(0.55)
Mean elevation		-1.04*	-1.69***	-0.92	-1.08	-1.05
		(0.60)	(0.56)	(0.57)	(0.67)	(0.78)
Standard deviation of elevation		-0.64	-0.14	-0.23	-0.019	-0.016
		(0.56)	(0.52)	(0.40)	(0.45)	(0.48)
Terrain roughness		3.36**	3.06**	0.68	1.10	0.94
		(1.27)	(1.23)	(1.51)	(2.19)	(2.32)
Mean distance to nearest waterway		-0.62**	-0.87***	-1.01***	-0.83**	-0.84**
		(0.29)	(0.31)	(0.33)	(0.34)	(0.39)
1 if landlocked		0.38	0.61	0.59	0.43	0.45
		(0.36)	(0.37)	(0.44)	(0.40)	(0.45)
Log [Area]		0.099	0.14	0.15	0.10	0.11
		(0.11)	(0.12)	(0.12)	(0.12)	(0.13)
Linguistic fractionalization			0.058	0.35	-0.027	0.13
			(0.55)	(0.51)	(0.57)	(0.60)
Religious fractionalization			-0.60	-1.14**	-1.13**	-0.77
			(0.46)	(0.44)	(0.55)	(0.61)
% of European descent						0.15
						(0.74)
Genetic distance to the U.S. (weighted)						0.030
						(0.06)
Constant	-189.4***	-160.8**	-176.5**	-223.0***	-248.4***	-254.3***
	(62.84)	(68.36)	(69.44)	(63.75)	(68.73)	(69.99)
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin FE	No	No	No	Yes	Yes	Yes
Major religion shares	No	No	No	No	Yes	Yes
Observations	74	74	72	72	72	71
R^2	0.866	0.895	0.904	0.929	0.948	0.950
Adjusted R^2	0.809	0.829	0.834	0.863	0.881	0.875

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Major religion shares include the share of Protestants, Catholics, Muslims, Buddhists, Hinduists, and Atheists. See column (7) of Table 7.3 for a complete list of the additional controls.

Table 7.23. Patience, Physical Capital, and Savings: Conditioning on per capita income

	Dependent variable:															
	Log [Capital stock p/c]				Gross savings (% of GNI)				Net adjusted savings (% of GNI)				Household savings rate			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Patience	2.03***	0.98***	-0.17	0.038	6.66***	8.11**	8.11**	8.33*	7.70***	9.65***	13.1***	13.9***	6.21**	7.08**	8.13**	8.82**
	(0.28)	(0.31)	(0.15)	(0.15)	(2.29)	(4.01)	(3.80)	(4.69)	(2.23)	(2.96)	(3.47)	(3.21)	(2.70)	(3.13)	(3.19)	(3.46)
Log [GDP p/c PPP]			0.84***	0.71***			-0.54	-0.17			-2.03**	-3.36**			-2.17	-2.01
			(0.05)	(0.09)			(1.16)	(2.61)			(0.91)	(1.51)			(2.34)	(2.55)
Constant	10.0***	-147.6***	3.03***	-76.4**	22.1***	-1506.4**	26.6**	-1524.8**	10.2***	343.8	27.1***	35.2	3.27**	2.86*	24.7	22.8
	(0.13)	(43.14)	(0.41)	(35.71)	(1.18)	(580.16)	(10.23)	(679.89)	(1.07)	(729.38)	(8.01)	(790.61)	(1.50)	(1.63)	(22.26)	(25.67)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	No	No	No
Observations	71	69	71	69	73	71	73	71	68	68	68	68	21	21	21	21
R^2	0.327	0.861	0.893	0.948	0.058	0.474	0.062	0.474	0.102	0.537	0.173	0.578	0.231	0.272	0.255	0.293
Adjusted R^2	0.317	0.799	0.890	0.924	0.044	0.249	0.035	0.233	0.088	0.326	0.147	0.372	0.191	0.144	0.173	0.116

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

Table 7.24. Patience and Human Capital Accumulation: Conditioning on per capita income

	Dependent variable:															
	Average years of schooling				Human capital index				Cognitive skills				Education expenditure (% of GNI)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Patience	4.67***	3.53***	1.74***	2.14**	0.76***	0.42**	0.22*	0.24	0.81***	0.35*	0.31*	0.031	1.45***	1.48**	0.79	1.38*
	(0.53)	(0.84)	(0.64)	(0.97)	(0.11)	(0.16)	(0.12)	(0.17)	(0.13)	(0.20)	(0.18)	(0.18)	(0.31)	(0.55)	(0.55)	(0.76)
Log [GDP p/c PPP]			1.08***	0.85**			0.21***	0.16**			0.23***	0.27**			0.24	0.081
			(0.15)	(0.37)			(0.03)	(0.08)			(0.06)	(0.10)			(0.16)	(0.34)
Constant	5.40***	-85.0	-3.64***	29.2	2.62***	-30.8	0.90***	-20.1	4.39***	-11.9	2.40***	21.1	4.28***	-68.9	2.25	-61.1
	(0.24)	(129.94)	(1.17)	(117.52)	(0.05)	(35.40)	(0.26)	(35.54)	(0.08)	(65.08)	(0.53)	(69.85)	(0.16)	(120.45)	(1.38)	(124.25)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	71	70	71	70	67	66	67	66	49	48	49	48	71	70	71	70
R^2	0.429	0.798	0.673	0.827	0.330	0.718	0.579	0.746	0.283	0.757	0.436	0.810	0.138	0.561	0.177	0.562
Adjusted R^2	0.421	0.710	0.663	0.745	0.319	0.583	0.566	0.616	0.268	0.561	0.412	0.643	0.125	0.369	0.153	0.357

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

Table 7.25. Patience and Total Factor Productivity: Conditioning on per capita income

	Dependent variable:															
	Total factor productivity				R&D expenditure (% of GDP)				# of researchers in R&D (per 1,000)				Global Innovation Index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Patience	0.36***	0.13	0.043	0.054	2.09***	1.78***	1.44***	1.06**	3.54***	2.45**	1.77***	0.97	23.6***	18.1***	12.5***	10.1***
	(0.05)	(0.09)	(0.06)	(0.08)	(0.23)	(0.59)	(0.24)	(0.46)	(0.43)	(1.02)	(0.41)	(0.74)	(1.70)	(2.91)	(1.74)	(2.43)
Log [GDP p/c PPP]			0.13***	0.074			0.27***	0.51***			0.68***	1.21***			4.30***	5.46***
			(0.03)	(0.06)			(0.07)	(0.13)			(0.09)	(0.25)			(0.44)	(0.77)
Constant	0.62***	1.07	-0.49*	8.45	0.96***	1.28	-1.36**	78.8*	1.54***	-29.8	-4.18***	99.6	39.1***	-114.4	3.15	672.2**
	(0.03)	(14.63)	(0.25)	(14.52)	(0.08)	(47.10)	(0.54)	(42.16)	(0.16)	(94.18)	(0.73)	(75.06)	(0.82)	(462.69)	(3.82)	(309.92)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	60	59	60	59	64	63	64	63	62	61	62	61	72	71	72	71
R^2	0.338	0.724	0.621	0.747	0.574	0.716	0.657	0.796	0.526	0.738	0.711	0.860	0.619	0.825	0.834	0.913
Adjusted R^2	0.326	0.567	0.608	0.592	0.567	0.570	0.646	0.684	0.518	0.597	0.701	0.779	0.613	0.750	0.830	0.873

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

Table 7.26. Patience and Institutions: Conditioning on per capita income

	Dependent variable:															
	Democracy				Property rights				Social infrastructure				S&P credit rating			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Patience	4.23***	2.95**	1.52	1.52	46.3***	35.0***	24.8***	13.6	0.43***	0.17**	0.13**	-0.069	10.7***	9.04***	7.44***	5.86***
	(0.82)	(1.46)	(1.27)	(1.54)	(4.39)	(8.26)	(5.86)	(8.62)	(0.05)	(0.06)	(0.06)	(0.08)	(0.82)	(1.35)	(1.27)	(1.77)
Log [GDP p/c PPP]			1.03***	0.97*			8.16***	14.6***			0.11***	0.17***			1.31***	2.04***
			(0.34)	(0.54)			(1.50)	(3.11)			(0.01)	(0.04)			(0.30)	(0.66)
Constant	6.63***	295.8	-1.97	423.7*	48.3***	-83.6	-19.7	1933.6	0.50***	11.9	-0.42***	43.0***	14.5***	-271.9	3.33	55.9
	(0.37)	(227.41)	(2.81)	(242.87)	(1.95)	(1271.19)	(12.51)	(1335.14)	(0.02)	(12.56)	(0.10)	(14.93)	(0.42)	(299.57)	(2.48)	(322.31)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	74	72	74	72	74	72	74	72	61	60	61	60	64	62	64	62
R^2	0.202	0.686	0.326	0.713	0.515	0.640	0.683	0.773	0.461	0.793	0.729	0.882	0.607	0.768	0.688	0.810
Adjusted R^2	0.191	0.554	0.308	0.584	0.508	0.489	0.674	0.671	0.451	0.678	0.720	0.812	0.601	0.647	0.678	0.703

OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See column (7) of Table 7.3 for a complete list of the additional controls.

Table 7.27. Regional patience, human capital, and income: Robustness

	Dependent variable:															
	Log [Regional GDP p/c]								Average years of education							
	Sub-samples: Number of observations larger than...															
	N > 0		N > 10		N > 20		N > 50		N > 0		N > 10		N > 20		N > 50	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Patience	0.085	0.072	0.17^{***}	0.16^{***}	0.18^{***}	0.16^{***}	0.23^{**}	0.21^{**}	0.27^{**}	0.26^{**}	0.46^{***}	0.43^{***}	0.42^{**}	0.37^{**}	0.46	0.42^{*}
	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.09)	(0.08)	(0.12)	(0.13)	(0.14)	(0.14)	(0.18)	(0.15)	(0.31)	(0.21)
Temperature		-0.019^{**}		-0.025^{**}		-0.020^{*}		-0.0054		-0.051^{***}		-0.054^{***}		-0.057^{***}		-0.052^{***}
		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)		(0.01)		(0.02)		(0.02)
Inverse distance to coast		0.44^{**}		0.38		0.48		0.63		0.64		0.78		1.04^{*}		1.17^{*}
		(0.21)		(0.25)		(0.31)		(0.40)		(0.50)		(0.57)		(0.58)		(0.62)
Log [Oil production p/c]		0.24^{***}		0.29^{***}		0.30^{***}		0.093		0.053		0.050		0.042		-0.074
		(0.06)		(0.07)		(0.11)		(0.08)		(0.04)		(0.05)		(0.08)		(0.12)
# Ethnic groups		-0.13^{*}		-0.12^{*}		-0.11^{*}		-0.20^{***}		-0.32^{*}		-0.26^{**}		-0.32^{**}		-0.47^{**}
		(0.07)		(0.06)		(0.07)		(0.07)		(0.16)		(0.13)		(0.15)		(0.18)
Log [Population density]		0.050		0.071^{**}		0.094^{**}		0.080		0.12		0.19^{***}		0.23^{***}		0.24^{**}
		(0.04)		(0.03)		(0.04)		(0.05)		(0.08)		(0.06)		(0.07)		(0.11)
Constant	8.94^{***}	8.84^{***}	9.05^{***}	8.98^{***}	8.97^{***}	8.66^{***}	9.56^{***}	8.81^{***}	7.02^{***}	7.28^{***}	7.16^{***}	7.02^{***}	7.17^{***}	6.73^{***}	8.08^{***}	6.94^{***}
	(0.02)	(0.24)	(0.02)	(0.27)	(0.02)	(0.37)	(0.01)	(0.48)	(0.04)	(0.55)	(0.05)	(0.54)	(0.05)	(0.61)	(0.05)	(1.00)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	704	687	657	640	544	527	361	344	693	676	646	629	534	517	355	338
R^2	0.925	0.936	0.928	0.941	0.933	0.949	0.943	0.958	0.936	0.947	0.937	0.951	0.938	0.956	0.935	0.960
Adjusted R^2	0.919	0.931	0.921	0.935	0.926	0.942	0.933	0.950	0.930	0.942	0.931	0.946	0.931	0.951	0.923	0.952

OLS estimates, standard errors (clustered at country level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 7.G Additional Figures

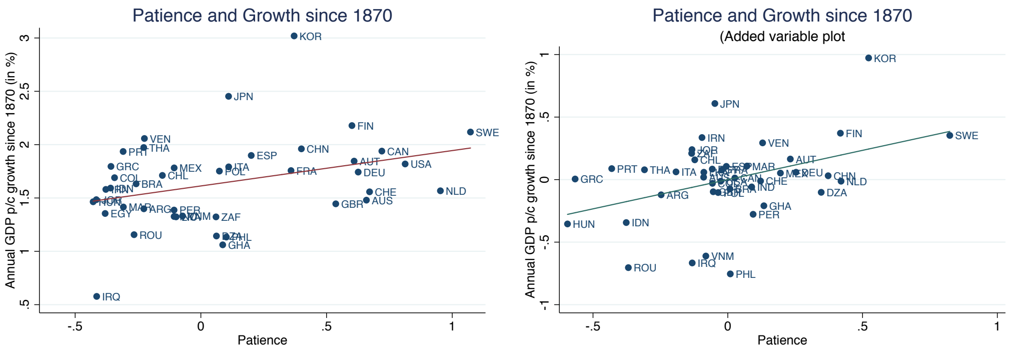


Figure 7.5. Patience and long-run growth. The left panel depicts the raw correlation between annual growth rates in GDP per capita (in %) since 1870 and patience, while the right panel contains a plot conditional on continent fixed effects.

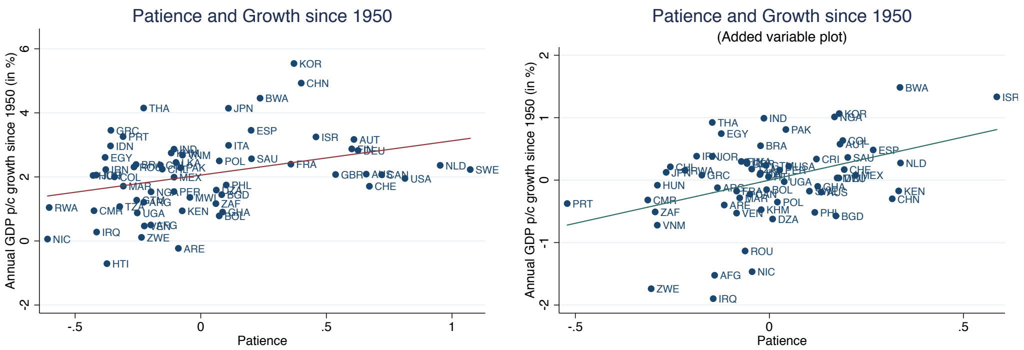


Figure 7.6. Patience and medium-run growth. The left panel depicts the raw correlation between annual growth rates in GDP per capita (in %) since 1950 and patience, while the right panel contains a plot conditional on the full set of baseline covariates in column (7) of Table 7.3.

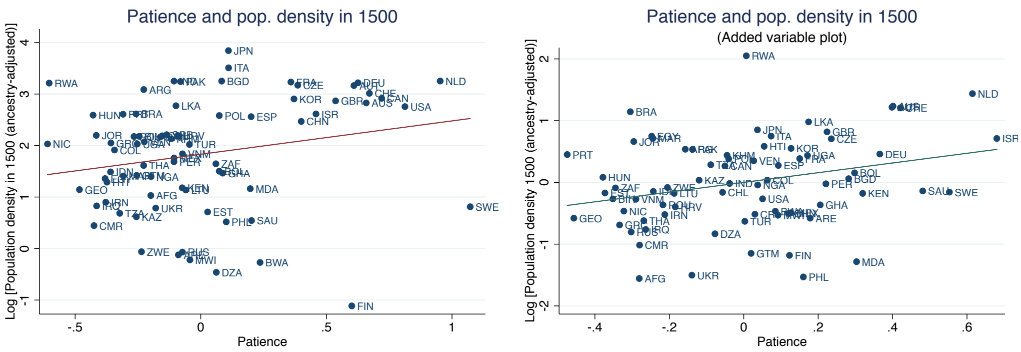


Figure 7.7. Patience and ancestry-adjusted population density in 1500. The left panel depicts the raw correlation between ancestry-adjusted population density in 1500 and patience, while the right panel contains a plot conditional on the full set of baseline covariates in column (2) of Table 7.21.

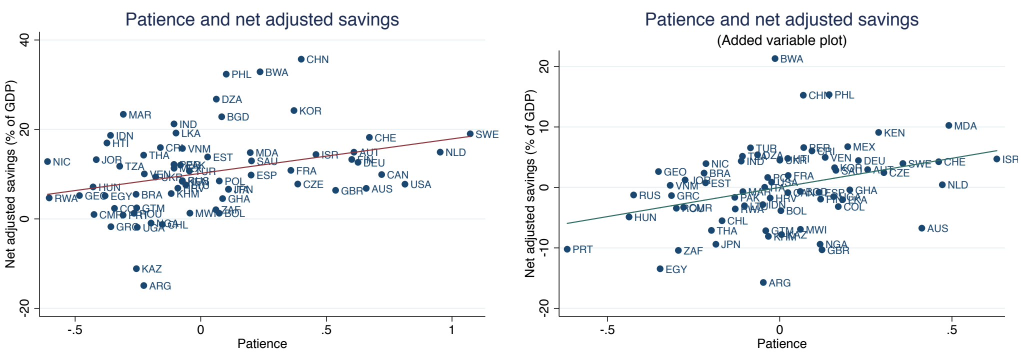


Figure 7.8. Patience and Savings. The left panel depicts the raw correlation between net adjusted savings (% of GDP) and patience, while the right panel contains a plot conditional on the full set of baseline covariates in column (7) of Table 7.3.

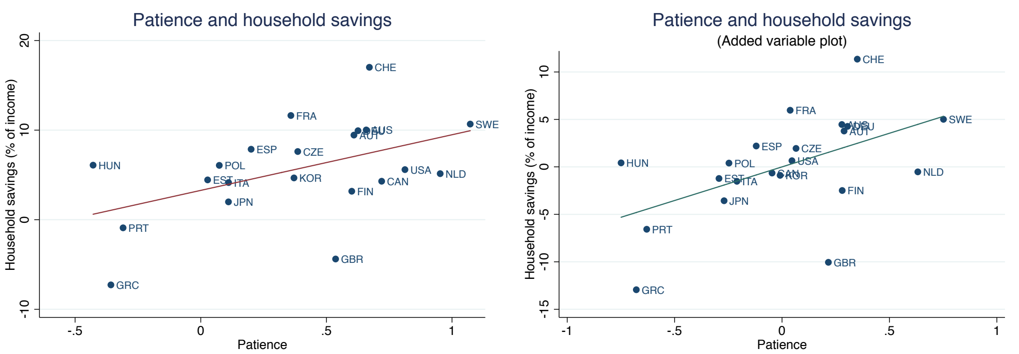


Figure 7.9. Patience and Household Savings. The left panel depicts the raw correlation between household savings (% of disposable income) and patience, while the right panel contains a plot conditional on the full set of baseline covariates in column (7) of Table 7.3.

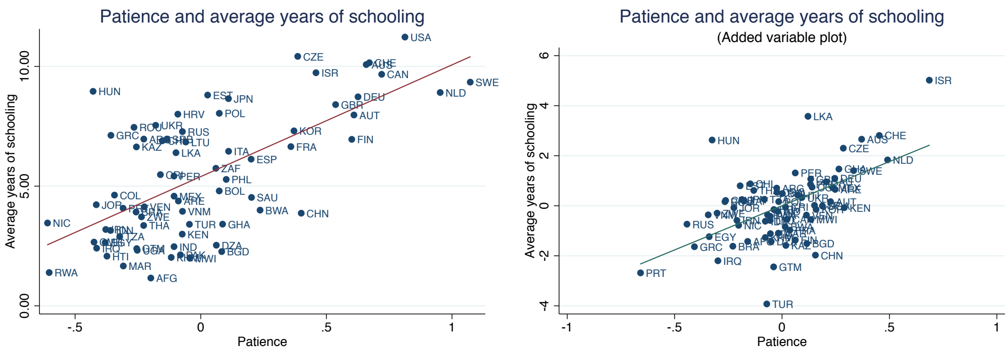


Figure 7.10. Patience and Average Years of Schooling. The left panel depicts the raw correlation between average years of schooling and patience, while the right panel contains a plot conditional on the full set of baseline covariates in column (7) of Table 7.3.

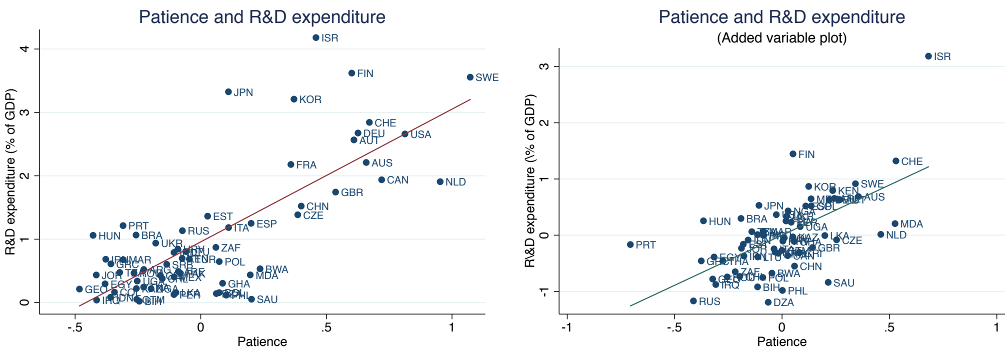


Figure 7.11. Patience and innovation. The left panel depicts the conditional correlation between R&D expenditure (as % of GDP) and patience, while the right panel contains a conditional plot of the relationship between the global innovation index and patience. Both plots are conditional on the full set of baseline covariates in column (7) of Table 7.3.

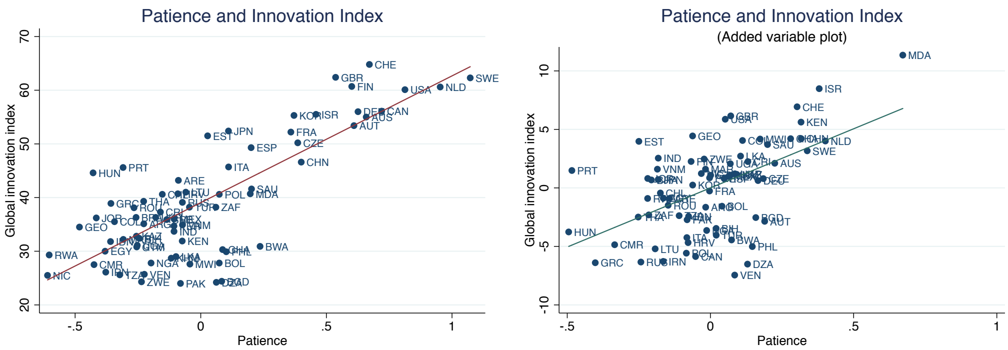


Figure 7.12. Patience and innovation. The left panel depicts the raw correlation between R&D expenditure (as % of GDP) and patience, while the right panel plots the raw correlation between the Global Innovation Index and patience

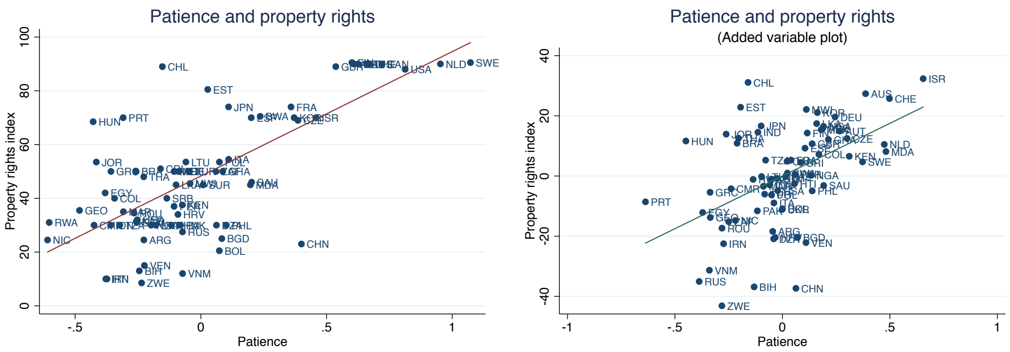


Figure 7.13. Patience and property rights. The left panel depicts the raw correlation between the property rights index and patience, while the right panel contains a plot conditional on the full set of baseline covariates in column (7) of Table 7.3.

Appendix 7.H Description and Sources of Main Variables

7.H.1 Country-Level Variables

7.H.1.1 Outcome Variables

7.H.1.1.1 Contemporary national GDP per capita. Average annual GDP per capita over the period 2003 – 2012, in 2005US$. Source: World Bank Development Indicators.

7.H.1.1.2 National GDP per worker. GDP per worker, 1990US$. Source: World Bank Development Indicators, average 2003 – 2012.

7.H.1.1.3 Human Development Index. The Human Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living. The HDI is the geometric mean of normalized indices for each of the three dimensions. Average 2000-2010, taken from UNDP.

7.H.1.1.4 Average subjective happiness. In Gallups' World Poll, respondents are asked to evaluate the current state of their lives, using the image of a ladder, with the best possible life for them as a 10 and the worst possible life as a 0. Source: the World Happiness Report 2013, at http://unsdsn.org/resources/publications/world-happiness-report-2013/.

7.H.1.1.5 Historical Income Data and Growth rates in GDP per capita. Source: the Maddison project.

7.H.1.1.6 Population density in 1500. Persons per square km, original data taken from Ashraf and Galor (2013). The ancestry-adjusted population density measure is computed by multiplying the contemporary population shares (as obtained from Putterman and Weil (2010)) with the historical population density of the respective population's ancestor countries.

7.H.1.1.7 Average years of schooling. The mean over the 2000-2010 time period, of the 5-yearly figure, reported by R. J. Barro and Lee (2012), on average years of schooling amongst the population aged 25 and over.

7.H.1.1.8 Human capital index. Human capital index provided by the Penn World Tables, which aims to provide a quality-adjusted index of human capital by combining years of schooling with returns to schooling. The index is defined as $e^{f(s)}$, where $f(s) = 0.134s$ if $s \leq 4$, $f(s) = 0.134s \times 4 + 0.101(s-4)$ if $4 < s \leq 8$ and $f(s) = 0.134 \times 4 + 0.101 \times 4 + 0.068(s-8)$ if $s > 8$, where s = years of schooling.

7.H.1.1.9 Cognitive skills. Measure of cognitive skills derived from a series of standardized tests in math, science, and reading across countries, see Eric A Hanushek and Woessmann (2012).

7.H.1.1.10 Education expenditure. Current operating expenditures in education, including wages and salaries and excluding capital investments in buildings and equipment. Source: World Bank Development Indicators, average 2003 – 2012.

7.H.1.1.11 Capital stock. Capital stock at constant 2005 national prices (in mil. 2005US$), average from 2002 to 2011. Data taken from the Penn World Tables.

7.H.1.1.12 National savings. Gross savings are calculated as gross national income less total consumption, plus net transfers. Net national savings are equal to gross national savings less the value of consumption of fixed capital. Adjusted net savings are equal to net national savings plus education expenditure and minus energy depletion, mineral depletion, net forest depletion, and carbon dioxide and particulate emissions damage. Source: World Bank Development Indicators, average 2003 – 2012.

7.H.1.1.13 Household savings rate. The household saving rate is calculated as the ratio of household saving to household disposable income (plus the change in net equity of households in pension funds). Source: the OECD statistics database. We use the most recent available data point (either projected or realized), working backwards from 2010.

7.H.1.1.14 Total factor productivity. TFP level at current PPPs (USA=1), average from 2002 to 2011. Source: the Penn World Tables.

7.H.1.1.15 R&D expenditure. Expenditures for research and development are current and capital expenditures (both public and private) on creative work undertaken systematically to increase knowledge, including knowledge of humanity, culture, and society, and the use of knowledge for new applications. R&D covers basic research, applied research, and experimental development. Source: World Bank Development Indicators, average 2003 – 2012.

7.H.1.1.16 Number of researchers in R&D. Researchers in R&D are professionals engaged in the conception or creation of new knowledge, products, processes, methods, or systems and in the management of the projects concerned. Average from 2003 to 2012, taken from the World Bank Development Indicators.

7.H.1.1.17 Global innovation index. This index is a summary statistic of innovative capacity that consists of over 80 qualitative and quantitative items, including measures of institutions, human capital and research, infrastructure, market sophistication, business sophistication, knowledge and technology outputs, and creative outputs. Data from 2014, taken from https://www.globalinnovationindex.org/content.aspx?page=data-analysis.

7.H.1.1.18 Democracy index. Index that quanties the extent of institutionalized democracy, as reported in the Polity IV dataset. Average from 2003 to 2012.

7.H.1.1.19 Social infrastructure index. Index due to Hall and Jones (1999) which measures the wedge between the private return to productive activities and the social return to such activities. This index is derived from two separate indices. First, an index of government antidiversion is policies created from data assembled by Political Risk Services and covers the categories law and order, bureaucratic quality, corruption, risk of expropriation, and government repudiation of contracts. The second element of the index captures the extent to which a country is open to international trade.

7.H.1.1.20 Property rights. This factor scores the degree to which a country's laws protect private property rights and the degree to which its government enforces those laws. It also accounts for the possibility that private property will be expropriated. In addition, it analyzes the independence of the judiciary, the existence of corruption within the judiciary, and the ability of individuals and businesses to enforce contracts. Average 2003-2012, taken from the Quality of Government dataset, http://www.qogdata.pol.gu.se/codebook/codebook_basic_30aug13.pdf.

7.H.1.1.21 Standard & Poor's long-term credit rating. Captures a country's likelihood of payment-capacity and willingness to meet its financial commitments, the nature of and provisions of the underlying debt, as well as the protection in case of bankruptcy. Source: http://www.standardandpoors.com/ratings/sovereigns/ratings-list/en/us/?subSectorCode=39 on 9 October 2014.

7.H.1.2 Covariates

7.H.1.2.1 Consumer price index. Average 2003-2012, taken from the World Bank Development Indicators.

7.H.1.2.2 GDP deflator. Average 2003-2012, taken from the World Bank Development Indicators.

7.H.1.2.3 Ratio of external finance and GDP. External finance is defined as the sum of private credit, private bond market capitalization, and stock market capitalization. Source: Buera et al. (2011).

7.H.1.2.4 Number of automated telling machines. Average 2003-2012. Source: World Bank Development Indicators.

7.H.1.2.5 Long-term orientation. Hofstede defines this concept by noting that every society has to maintain some links with its own past while dealing with the challenges of the present and the future. Societies prioritize these two existential goals differently. Societies who score low on this dimension, for example, prefer to maintain time-honoured traditions and norms while viewing societal change with suspicion. Those with a culture which scores high, on the other hand, take a more pragmatic approach: they encourage thrift and efforts in modern education as a way to prepare for the future. Source: http://geerthofstede.eu/research--vsm, retrieved on March 25, 2015.

7.H.1.2.6 Colonization dummy. Dummy equal to one if the respective country had at least one colonizer over a long period of time and with substantial participation in governance. Source: the CEPII geo database.

7.H.1.2.7 Area, distance to equator, longitude, landlocked dummy. Source: the CEPII geo database.

7.H.1.2.8 Mean and standard deviation of elevation. Elevation in km above sea level, taken from Ashraf and Galor (2013). Data originally based on geospatial elevation data reported by the G-ECON project (Nordhaus, 2006).

7.H.1.2.9 Percentage of arable land. Fraction of land within a country which is arable, taken from the World Bank Development Indicators.

7.H.1.2.10 Land suitability for agriculture. Index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH, taken from Michalopoulos (2012).

7.H.1.2.11 Neolithic revolution timing. The number of thousand years elapsed, until the year 2000, since the majority of the population residing within a country's modern national borders began practicing sedentary agriculture as the primary mode of subsistence. The measure is weighted within each country, where the weight represents the fraction of the year 2000 population (of the country for which the measure is being computed) that can trace its ancestral origins to the given country in the year 1500. Measure taken from Ashraf and Galor (2013).

7.H.1.2.12 Precipitation. Average monthly precipitation of a country in mm per month, 1961-1990, taken from Ashraf and Galor (2013). Data originally based on geospatial average monthly precipitation data for this period reported by the G-ECON project (Nordhaus, 2006).

7.H.1.2.13 Temperature. Average monthly temperature of a country in degree Celsius, 1961-1990, taken from Ashraf and Galor (2013). Data originally based on geospatial average monthly temperature data for this period reported by the G-ECON project (Nordhaus, 2006).

7.H.1.2.14 Percentage in (sub-)tropical zones. Percentage of area within a country which forms part of each of the tropical or sub-tropical climatic zones. Data taken from John Luke Gallup, http://www.pdx.edu/econ/jlgallup/country-geodata.

7.H.1.2.15 Percentage at risk of malaria. The percentage of population in regions of high malaria risk (as of 1994), multiplied by the proportion of national cases involving the fatal species of the malaria pathogen, P. falciparum. This variable was originally constructed by John L. Gallup and Sachs (2000) and is part of Columbia University's Earth Institute data set on malaria. Data taken from Ashraf and Galor (2013).

7.H.1.2.16 Predicted genetic diversity. Predicted genetic diversity of the contemporary population, adjusted for post-Columbian migration flows and genetic distance between ethnic groups. See Ashraf and Galor (2013).

7.H.1.2.17 Ethnic, linguistic, and religious fractionalization. Indices due to Alesina et al. (2003) capturing the probability that two randomly selected individuals from the same country will be from different ethnic (religious) groups.

7.H.1.2.18 Terrain roughness. Degree of terrain roughness of a country, taken from Ashraf and Galor (2013). Data originally based on geospatial undulation data reported by the G-ECON project (Nordhaus, 2006).

7.H.1.2.19 Distance to nearest waterway. The distance, in thousands of km, from a GIS grid cell to the nearest ice-free coastline or sea-navigable river, averaged across the grid cells of a country. Source: Ashraf and Galor (2013), originally constructed by John Luke Gallup et al. (1999).

7.H.1.2.20 Legal origins. Origin of legal system: UK, French, German, Scandinavian, Soviet. Source: La Porta et al. (1999).

7.H.1.2.21 Major religion shares. Source: R. Barro (2003).

7.H.1.2.22 Fraction of European descent. Fraction of the population which is of European dexcent. Constructed from the "World Migration Matrix" of Putterman and Weil (2010).

7.H.1.2.23 Genetic distance to the United States. Fst genetic distance of a country's contemporary population to the population of the United States. Source: Spolaore and Wacziarg (2009).

7.H.1.2.24 Trust. Part of Global Preference Survey. Elicited through respondents' self-assessment regarding the following statement on an 11 point scale: "I assume that people have only the best intentions."

7.H.1.2.25 Risk preferences. Risk preferences were measured in the Global Preference Survey using two survey items. First, respondents went through a quantitative five-step staircase procedure:
Please imagine the following situation. You can choose between a sure payment of a particular amount of money. or a draw. where you would have an equal chance of getting 300 Euro or getting nothing. We will present to you five different situations. What would you prefer: a draw with a 50 percent chance of receiving 300 Euro, and the same 50 percent chance of receiving nothing, or the amount of 160 Euro as a sure payment? See Falk et al. (2015a) for an exposition of the entire sequence of survey items.
In addition, respondents provided a self-assessment:
Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.
These items were combined on standardized data using the following formula:

$$\text{Will. to take risks} = 0.4729985 \times \text{Staircase risk} + 0.5270015 \times \text{Qualitative item}$$

7.H.2 Regional-Level Data

Except for the patience measures and a region's size (area), all regional-level data are taken from Gennaioli et al. (2013). The area data were collected by research assistants from various sources on the web.

7.H.3 Individual-Level Data

7.H.3.0.1 Household income per capita. Included in Gallup's background data. To calculate income, respondents are asked to report their household income in local currency. Those respondents who have difficulty answering the question are presented a set of ranges in local currency and are asked which group they fall into. Income variables are created by converting local currency to International Dollars (ID) using purchasing power parity (PPP) ratios. Log household income is computed as log (1+ household income).

7.H.3.0.2 Education level. Included in Gallup's background data. Level 1: Completed elementary education or less (up to 8 years of basic education). Level 2: Secondary - 3 year tertiary education and some education beyond secondary education (9-15 years of education). Level 3: Completed four years of education beyond high school and / or received a 4-year college degree.

7.H.3.0.3 Subjective self-assessment of math skills. Included in Gallup's background data. *How well do the following statements describe you as a person? Please indicate your answer on a scale from 0 to 10. A 0 means "does not describe me at all" and a 10 means "describes me perfectly". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. I am good at math.*

7.H.3.0.4 Saved last year. Binary variable capturing whether the respondent saved any money in the previous year. Included in Gallup's background data.

7.H.3.0.5 Health problems. Included in Gallup's background data. Binary response to the question "'Do you have any health problems that prevent you from doing any of the things people your age normally can do?".

7.H.3.0.6 Confidence in financial institutions. Included in Gallup's background data. Binary response to the question "In this country, do you have confidence in each of the following, or not? How about financial institutions or banks?"

7.H.3.0.7 Subjective institutional quality. Included in Gallup's background data. This variable consists of a perceived institutional quality index as it is provided by Gallup. This index combines binary questions (yes / no) about whether people have confidence in the military, the judicial system and courts, the national government, and the honesty of elections. The index is then constructed by averaging the yes / no answers across items.

7.H.3.0.8 Subjective law and order index. Included in Gallup's Background data. Derived from responses to three questions: "In the city or area where you live, do you have confidence in the local police force?"; "Do you feel safe walking alone at night in the city or area where you live?"; "Within the last 12 months, have you had money or property stolen from you or another household member?".

7.H.3.0.9 Subjective physical health index. Included in Gallup's Background data. Derived from responses to five questions: "Do you have any health problems that prevent you from doing any of the things people your age normally can do?"; "Now, please think about yesterday, from the morning until the end of the day. Think about where you were, what you were doing, who you were with, and how you felt. Did you feel well-rested yesterday?"; "Did you experience the following feelings during a lot of the day yesterday? How about physical pain?"; "Did you experience the following feelings during a lot of the day yesterday? How about worry?"; "Did you experience the following feelings during a lot of the day yesterday? How about sadness?".

7.H.3.0.10 Subjective corruption index. Included in Gallup's Background data. Derived from two questions: "Is corruption widespread within businesses located in (country), or not?"; "Is corruption widespread throughout the government in (country), or not?".

7.H.3.0.11 Remaining years of life. For each individual, we impute the remaining years of life at the country-age-gender level using data from the UN World Population Prospects, see http://esa.un.org/wpp/.