Correlation Neglect in Belief Formation*

Benjamin Enke Florian Zimmermann

October 21, 2015

Abstract

Many information structures generate correlated rather than mutually independent signals, the news media being a prime example. This paper provides clean experimental evidence that many people neglect the resulting double-counting problem in the updating process, so that beliefs are excessively sensitive to well-connected information sources and follow an overshooting pattern. 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. These findings lend support to recent models of boundedly rational social learning and are reminiscent of popular narratives according to which excessive confidence swings may be driven by the ubiquitous "telling and re-telling of stories". 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.

JEL classification: C91, D03, D83, D84, D40.

Keywords: Beliefs; Correlation Neglect; Overshooting; Markets; Attention.

---

*For valuable comments and discussions we are grateful to Steffen Altmann, Sandro Ambuehl, Anke Becker, Stefano DellaVigna, 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, INSEAD, Maastricht, UPenn Wharton, Pompeu Fabra, 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. Enke: University of Bonn, Department of Economics, Adenauerallee 24-42, 53113 Bonn, Germany; benjamin.enke@uni-bonn.de. Zimmermann: University of Zurich, Department of Economics, Bluemlisalpstr. 10, 8006 Zurich, Switzerland; florian.zimmermann@econ.uzh.ch.

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. 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

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

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.

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 (see, e.g., Charness et al., 2010; Andreoni and Mylovanov, 2012; 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]

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

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

[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

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 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 3 and 4 investigate the validity of face value bias and the mechanisms underlying correlation neglect, respectively. Section 5 concludes.

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.

2.1 Experimental Design

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.

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.

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 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".

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

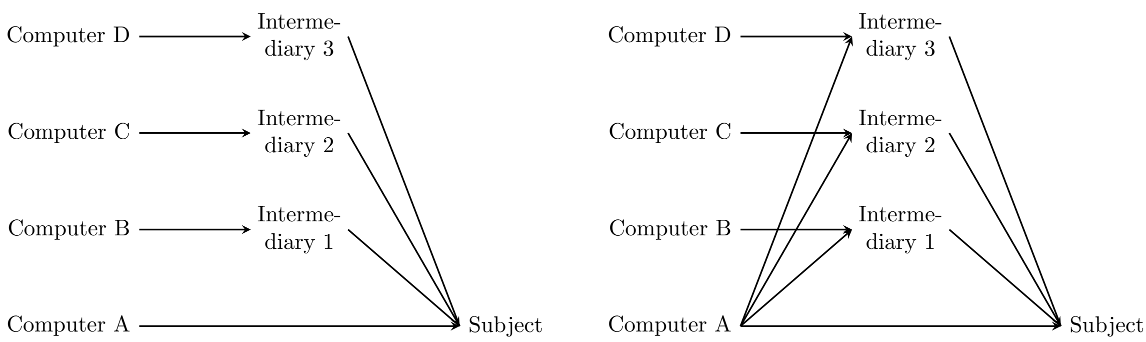

Figure 1: Uncorrelated (left panel) and correlated (right panel) information structure

[6] Section 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.

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.

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 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 for details). Throughout, we employ the term "belief" to denote the mean

[8]This holds provided that the treatment did not affect prior beliefs. As we show in Section 2.4.1, our results are robust to explicitly inducing equal priors across treatments.

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

Table 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.

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 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]

[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 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

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.

2.1.2 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.

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 E.7 for details). In any case, since budgets and exchange rates were identical across treatments, this procedure cannot explain potential treatment differences.

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.

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 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 $h \in \{1, \ldots, 4\}$. 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

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

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

[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.

$$b_{CN} = \frac{s_1 + \sum_{h=1}^{3} \hat{s}_h}{4} = \bar{s} + \frac{3}{8}\chi(s_1 - \bar{s}_{-1}) \tag{2}$$

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.

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 E.1, under heterogeneity the magnitude of a potential price distortion will depend on the naïveté of the marginal traders.[15]

[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.

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

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 ($\approx$ 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 ($\approx$ USD 25) on average. In all treatments, payments included a 6 euros show-up fee.

[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.

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.

2.4.1 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 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.

Table 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* treatment.[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 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

[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 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.

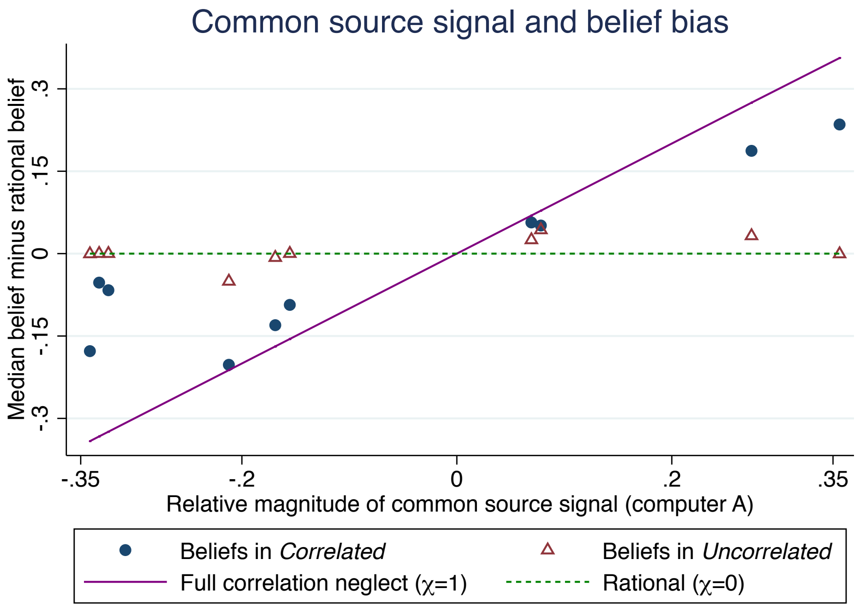

Figure 2: Beliefs in the *Correlated* and *Uncorrelated* treatments plotted against the relative magnitude of the signal of computer A. The logic of the figure is that if the signal of computer A is relatively high (low) compared to the other signals, correlation neglect predicts that beliefs should 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, 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).

Table 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 1 for details of the computation of the rational and the correlation neglect benchmarks. Note that subjects faced the ten tasks in randomized order.

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).

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 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, 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:

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

The left panel of Figure 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 visual inspection suggests, comparing median normalized beliefs across treatments also reveals a pronounced treatment difference (p-value < 0.0001, Wilcoxon ranksum test). Appendix 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]

[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.

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

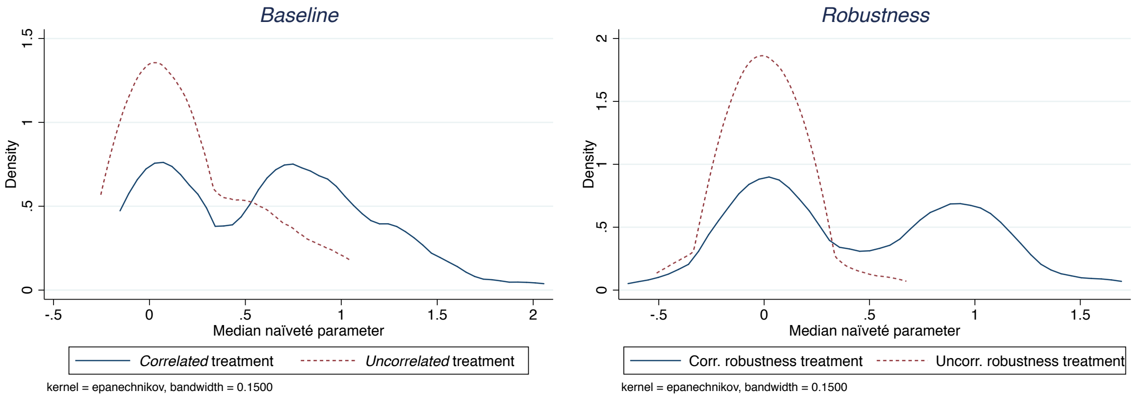

Figure 3: Kernel density estimates of median naïveté parameters. The left panel depicts the distribution of naïveté in the baseline treatments, and the right panel in the robustness treatments.

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 C.5, we pursue a different approach by structurally estimating the belief formation rule proposed in Section 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 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 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 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 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).

2.4.2 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.*

[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 (3), also see Appendix 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 E.2 and E.3.

Table 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.

Table 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, 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 E.2 and 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.

(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 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 E.4 formally confirms that the bias reflected in market prices does not become smaller over the course of the ten trading

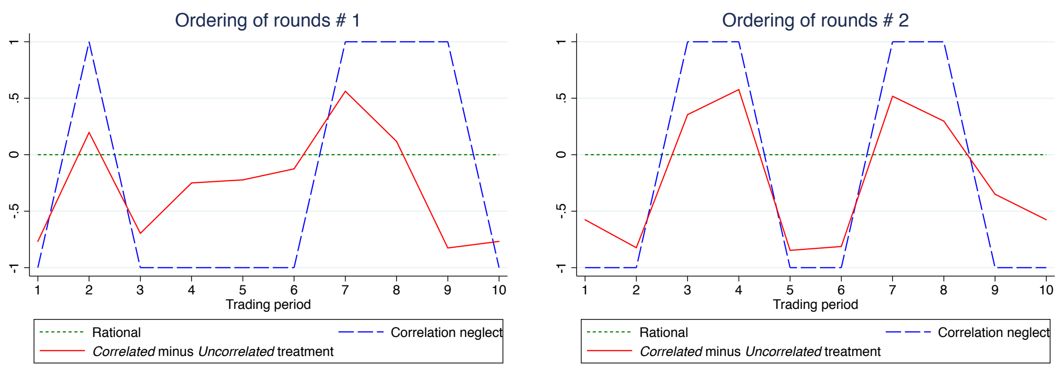

Figure 4: Difference between median normalized market prices in the *Correlated* and *Uncorrelated* treatments across trading rounds for the two randomized orders of rounds

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

periods. Appendix 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 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.

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 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 (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 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

[26]To this end, as we detail in Appendix 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.

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)).

Table 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***	0.65***						
	(0.12)	(0.14)						
Individual median naïveté (χ)			-1.53***	-1.30***	0.64***	0.26*	-0.12**	-0.11**
			(0.17)	(0.19)	(0.12)	(0.14)	(0.05)	(0.05)
Constant	0.02	0.68	2.85***	2.37**	1.43***	4.04***	10.1***	10.2***
	(0.11)	(1.41)	(0.12)	(0.94)	(0.14)	(1.31)	(0.03)	(0.28)
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 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 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 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*

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

condition, and (iii) the correlation neglect and face value predictions do not coincide. Specifically, as depicted in Figure 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 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

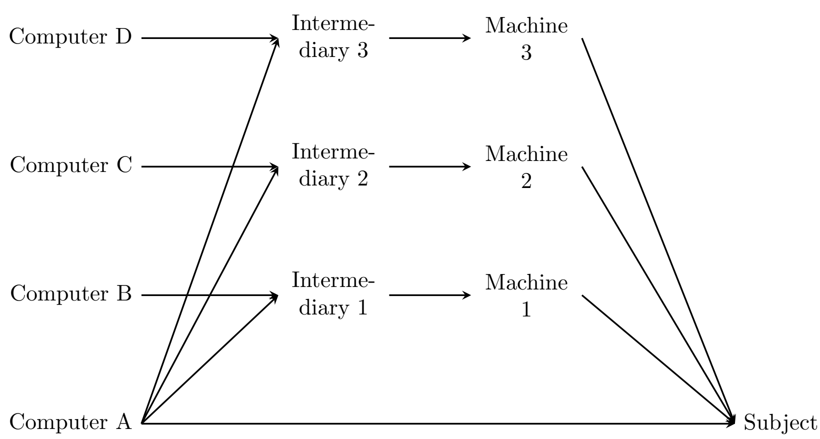

Figure 5: Treatment *Face value*. The machines add X to the reports of the intermediaries.

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 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 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 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.

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.

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 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.

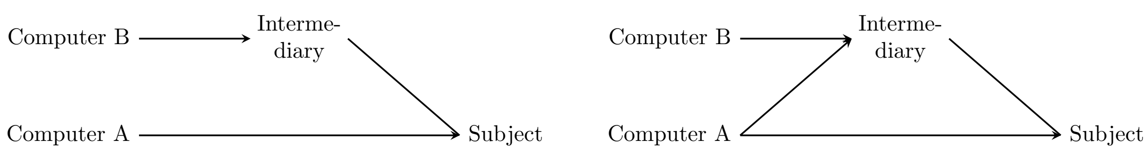

Figure 6: Simple uncorrelated (left panel) and correlated (right panel) information structure

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) 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 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

Table 5: Correlation neglect and cognitive skills

	Dependent variable: Median naiveté			
	(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)
Constant	1.50***	1.20***	2.20***	2.20***
	(0.40)	(0.22)	(0.48)	(0.73)
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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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

[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.

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

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

[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.

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

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* –

[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.

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 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 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 F.3 visualizes, very similar results obtain in *Alternating*. Appendix 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 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 $\chi = 0.09$ (*Intermediaries*) 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

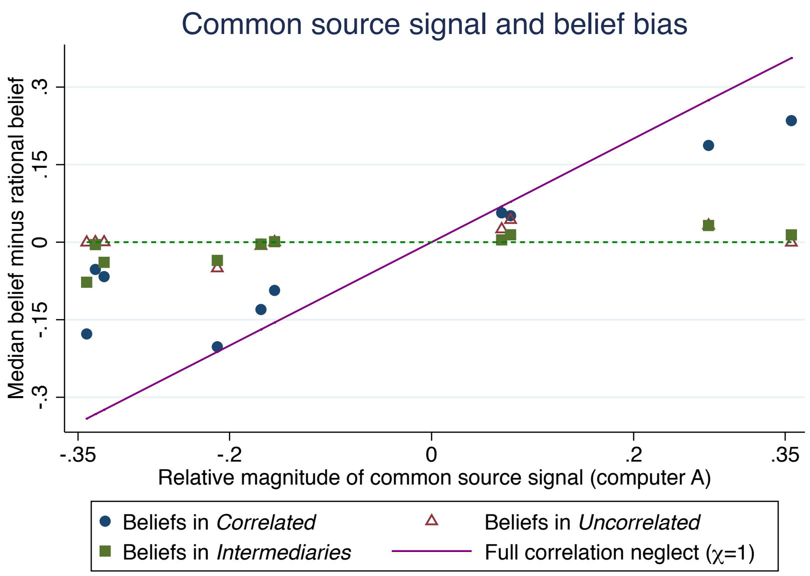

Figure 7: Beliefs in the *Correlated*, *Uncorrelated* and *Intermediaries* treatments plotted against the relative magnitude of the signal of computer A. See the notes of Figure 2 for details on the construction of this figure.

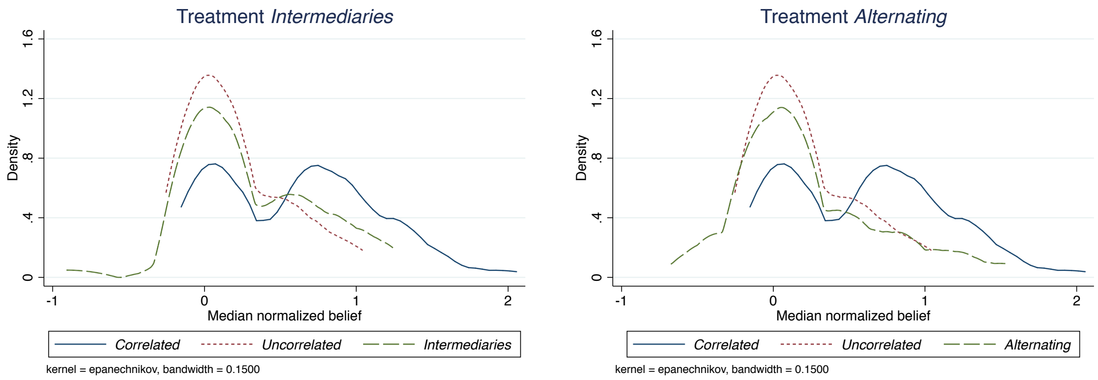

Figure 8: Kernel density estimates of median normalized beliefs in the *Intermediaries* treatment (median of ten tasks) and the *Alternating* treatment (median of five tasks), each compared with median beliefs in the baseline *Correlated* and *Uncorrelated* treatments

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 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 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 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.

[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 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 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]

[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.

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 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, 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, *Animal spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism*, Princeton University Press, 2009.

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

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

Bohren, Aislinn, "Informational Herding with Model Misspecification," *Working Paper*, 2013.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience and Consumer Choice," *Journal of Political Economy*, 2013, *121*, 803–843.

_ , _ , and _ , "Memory, Attention and Choice," *Working Paper*, 2015.

_ , _ , and _ , "Stereotypes," *Working Paper*, 2015.

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

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

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

_ and Hsiu-Ting Yu, "Aggregation of Opinion Based on Correlated Cues and Advisors," *Journal of Behavioral Decision Making*, 2007, *20*, 153–177.

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

Caplin, Andrew and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, forthcoming.

— , — , **and Daniel Martin**, "Search and Satisficing," *American Economic Review*, 2011, *101* (7), 2899–2922.

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

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

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

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

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

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

Eyster, Erik and Georg Weizsäcker, "Correlation Neglect in Financial Decision-Making," *Working Paper*, 2011.

— **and Matthew Rabin**, "Cursed Equilibrium," *Econometrica*, 2005, *73* (5), 1623–1672.

— **and** — , "Naïve Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, 2010, *2* (4), 221–243.

— **and** — , "Extensive Imitation is Irrational and Harmful," *Quarterly Journal of Economics*, 2014, *129* (4), 1861–1898.

— , — , **and Dimitri Vayanos**, "Financial Markets Where Traders Neglect the Informational Content of Prices," *Working Paper*, 2013.

Fiedler, Klaus and Peter Juslin, *Information Sampling and Adaptive Cognition*, Cambridge University Press, 2006.

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

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

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

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

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

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

Golub, Benjamin and Matthew O. Jackson, "Naïve learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2010, *2* (1), 112–149.

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

Grimm, Veronika and Friederike Mengel, "An Experiment on Belief Formation in Networks," *Working Paper*, 2014.

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

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

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

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

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

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

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

Levy, Gilat and Ronny Razin, "Correlation Neglect, Voting behaviour and Polarization," *American Economic Review*, forthcoming.

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

_ , "An Experimental Examination of Subjective Forecast Combination," *International Journal of Forecasting*, 1996, *12*, 223–233.

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

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

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

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

_ , "Incorporating Limited Rationality into Economics," *Journal of Economic Literature*, 2013, *51* (2), 528–43.

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

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

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

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

Shiller, Robert J., *Irrational Exuberance*, Princeton University Press, 2000.

Simon, Herbert A., "Rational Choice and the Structure of the Environment," *Psychological Review*, 1956, *63*, 129–138.

Spiegler, Ran, "Bayesian Networks and Boundedly Rational Expectations," *Working Paper*, 2015.

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

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

ONLINE APPENDIX

A Overview of Treatments

Table 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

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.

C Additional Analyses for Individual Baseline Treatments

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 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 9 and 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 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.

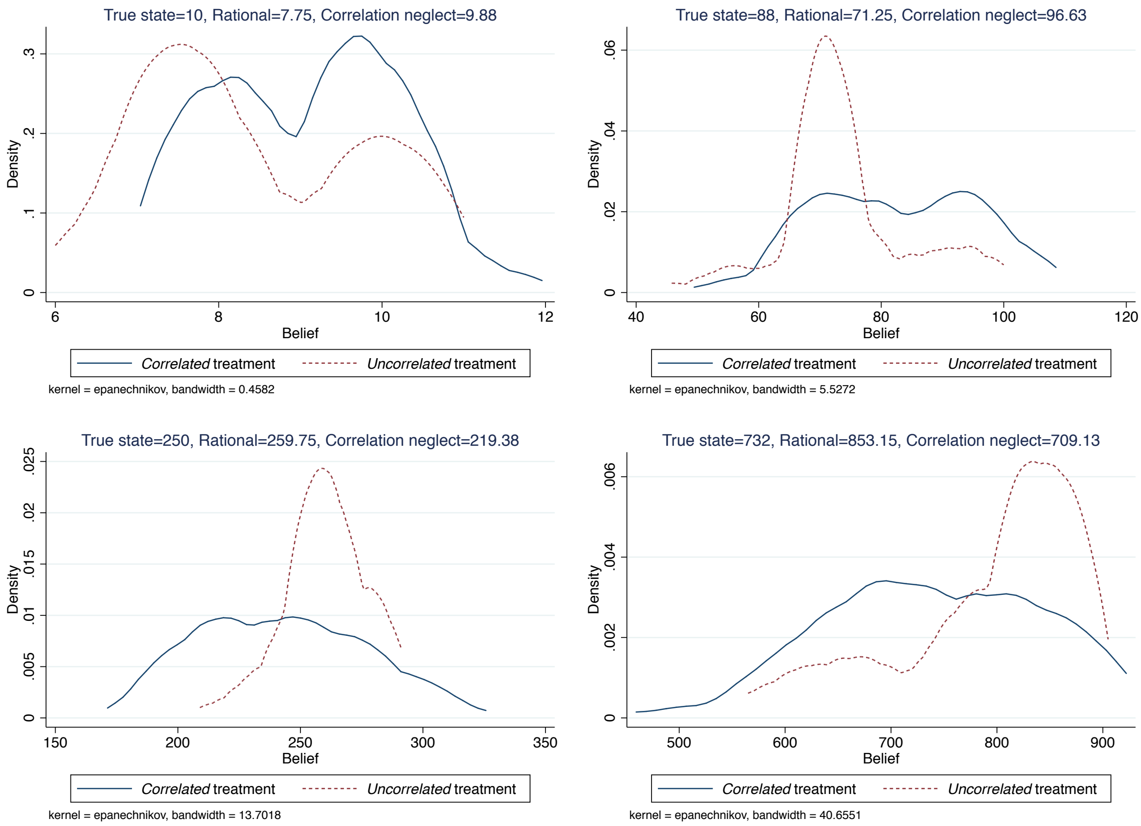

Figure 9: Kernel density estimates of beliefs in individual belief formation treatments (1/2)

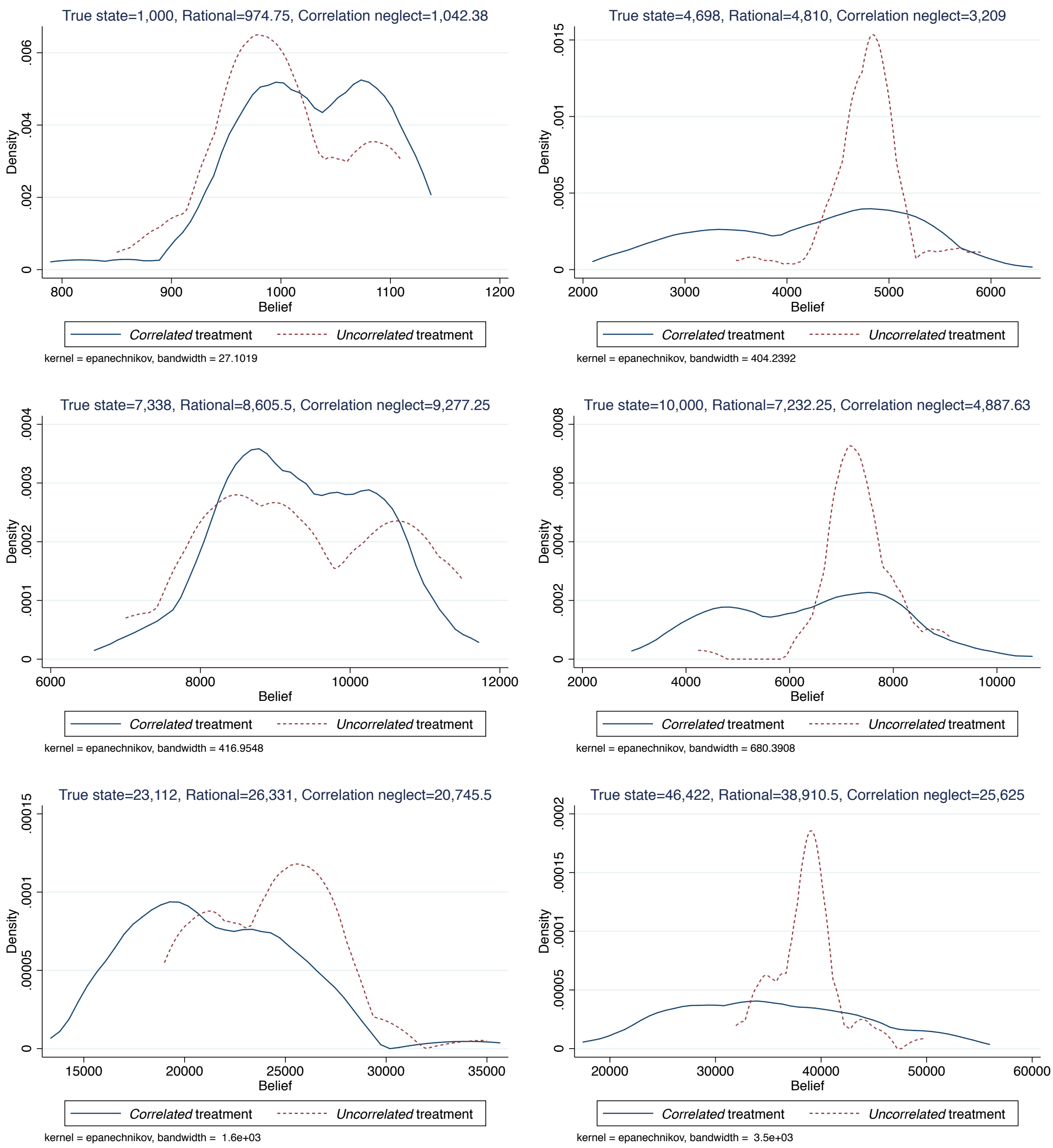

Figure 10: Kernel density estimates of beliefs in individual belief formation treatments (2/2)

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 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).

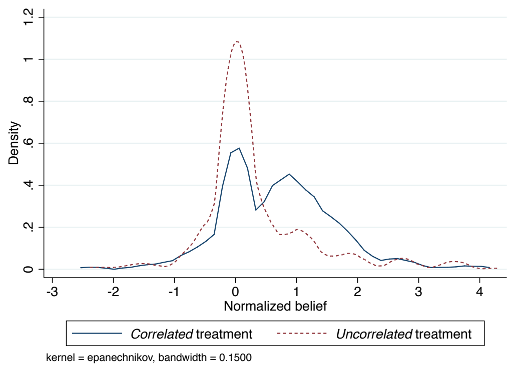

Figure 11: Kernel density estimates of all normalized individual beliefs

To statistically confirm this visual impression, columns (1) and (2) of Table 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 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). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 re-compute 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 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.

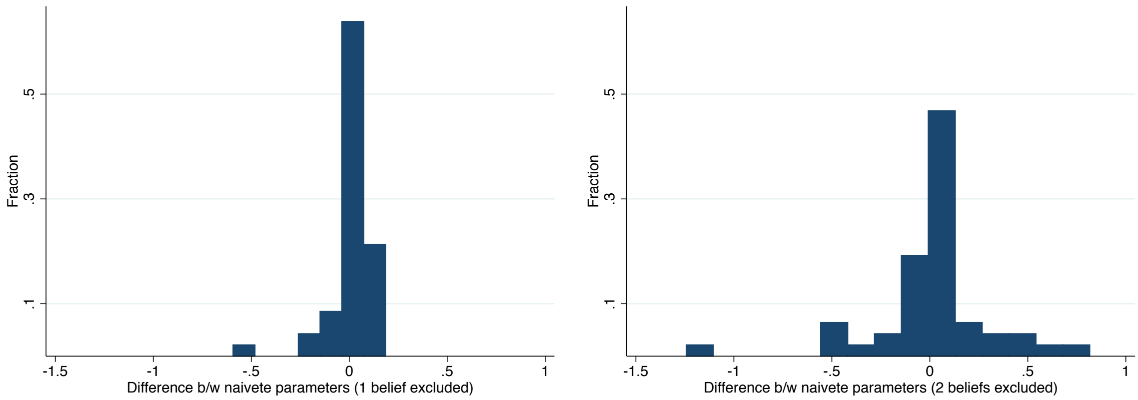

Figure 12: Histograms of the difference between original naïveté parameters and modified naïveté parameters when excluding the one or two beliefs that are closest to the original (implied) naïveté parameter

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

Column (1) of Table 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 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 (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

$$L_i(\chi, \sigma, w) = \sum_{k=1}^{K} w_k \prod_{j=1}^{10} P(\tilde{b}_i^j | \chi_k, \sigma_k) \tag{4}$$

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 11 presents the key results from these estimations. The table reports the estimated parameters of our belief formation model for three different specifications,

Table 11: Results of finite mixture model

		Model parameters			*Goodness of fit*		
Model	Type	χ	σ	w (%)	LL	AIC	BIC
$K = 1$	$k = 1$	0.68 (0.07)	0.91 (0.05)	100	−607	1219	1222
$K = 2$	$k = 1$	0.05 (0.02)	0.26 (0.03)	19.1 (5.8)	−531	1073	1082
	$k = 2$	0.83 (0.06)	0.95 (0.06)	80.9 (5.8)			
$K = 3$	$k = 1$	0.05 (0.02)	0.26 (0.03)	19.1 (5.8)	−512	1041	1056
	$k = 2$	0.74 (0.19)	1.08 (0.07)	55.7 (15.8)			
	$k = 3$	1.02 (0.27)	0.52 (0.18)	25.2 (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. (3)). This resulted in the exclusion of 4 (out of 462) observations.

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.

D Details for Individual Robustness Treatments

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 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 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.

D.2 Results

Table 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).

Table 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 12 for details of the computation of the rational and the correlation neglect benchmarks. Note that subjects faced the ten rounds in randomized order.

E Details, Hypotheses, and Robustness Checks for Market Treatments

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.

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 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 generality we assume

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

[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.

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 $xd_j \in xd_j(p), \forall j$.*

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.

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

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

   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

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

   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

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

   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 $p = \bar{s}$.

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.

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.

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).

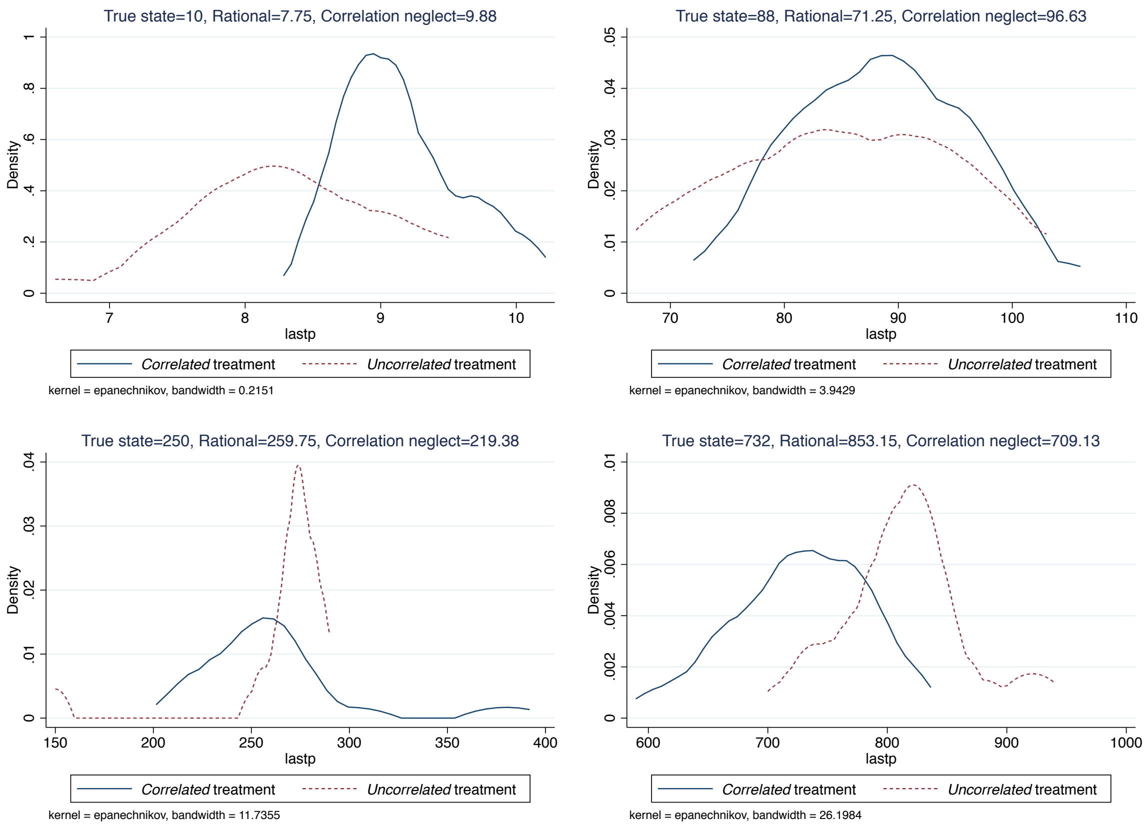

Figure 13: Kernel density estimates of market prices (1/2)

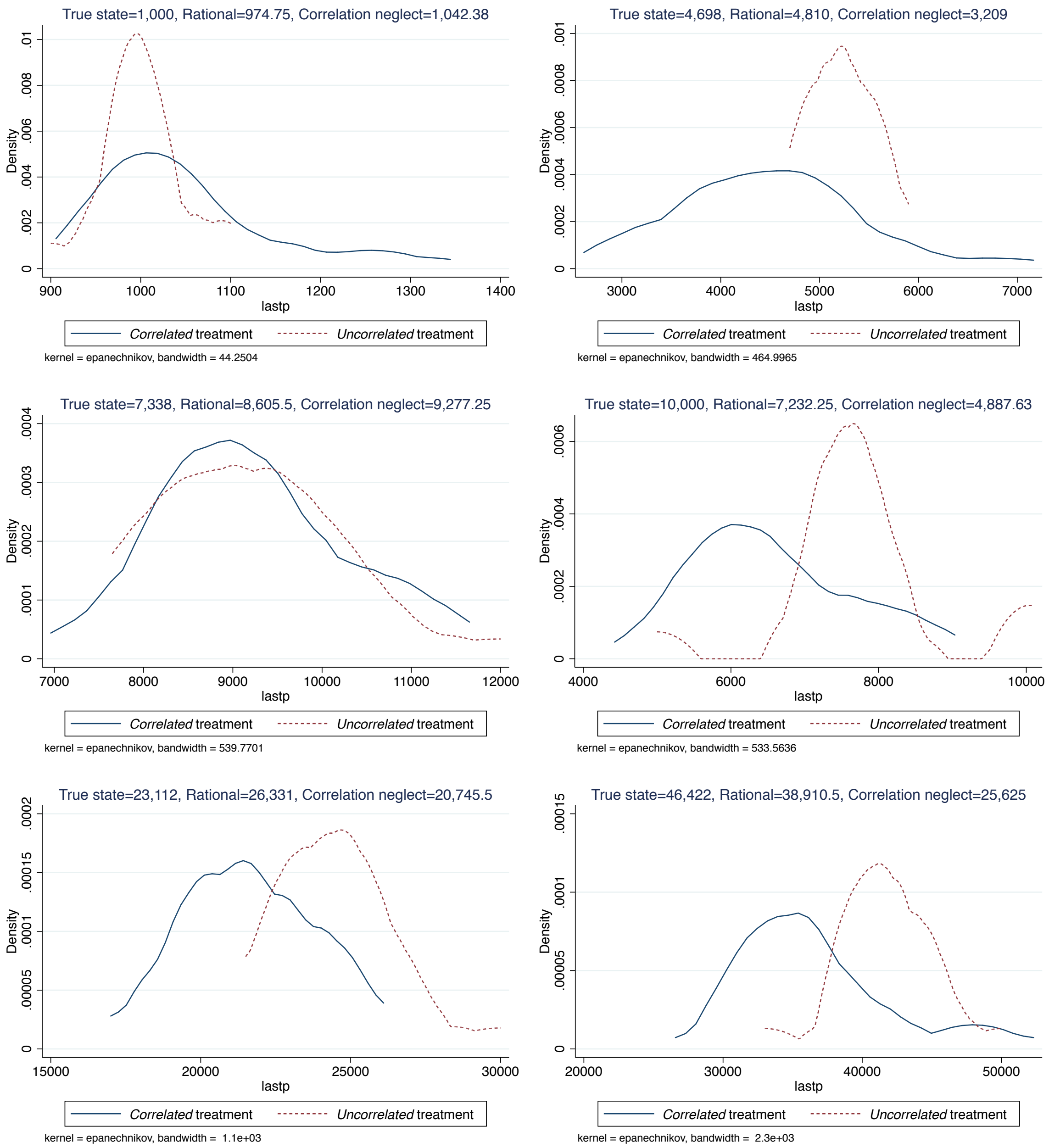

Figure 14: Kernel density estimates of market prices (2/2)

Next, we show that the strong treatment difference in price levels is not driven by our definition of the market price. Table 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 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).

Table 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.

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 (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 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 decision making treatments. Specifically, for each market group we use

Table 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the median normalized market price over the ten rounds to plot the distribution of market prices across treatments.

Figure 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 16 presents the p-value of ranksum tests for the equality of the aggregated market price between treatments.

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

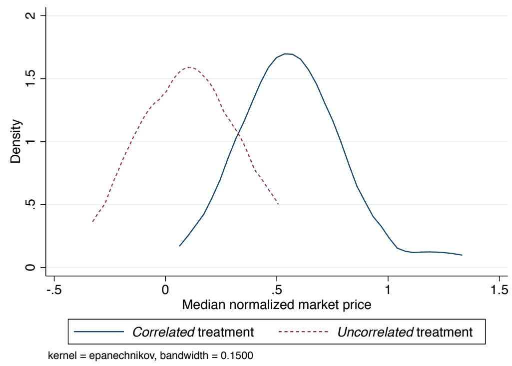

Figure 15: Kernel density estimates of median market prices

Table 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

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 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.

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

Table 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	-0.0091	-0.024	-0.0069
	(0.03)	(0.02)	(0.03)	(0.02)
Constant	0.71***	0.73***	0.57***	0.48**
	(0.19)	(0.17)	(0.16)	(0.17)
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 $|p_i^j| > 10$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E.5 Time Trend of Beliefs in Market Experiments

Table 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.

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 all buyers offering to buy the asset at, say, 20, this

Table 18: Time trend of normalized beliefs in the *Correlated market* treatment

	Dependent variable: Normalized belief				
	(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)
Constant	0.64***	0.67***	0.80***	1.24***	1.34***
	(0.08)	(0.07)	(0.06)	(0.06)	(0.09)
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 $b_i^j < 2$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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

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

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

E.7 Endowments and Exchange Rates in Market Treatments

Table 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.

F Treatments to Investigate the Mechanisms Underlying the Bias

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 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 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 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 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.

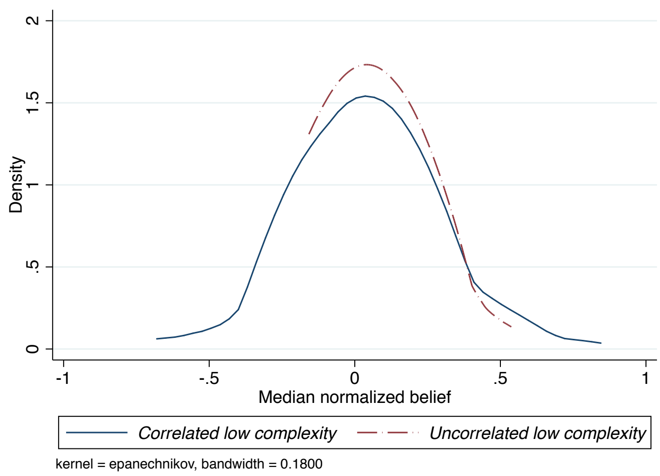

Figure 16: Kernel density estimates of beliefs in the reduced complexity treatments

Table 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F.2 Treatment *Math*

Figure 17 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 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 18.

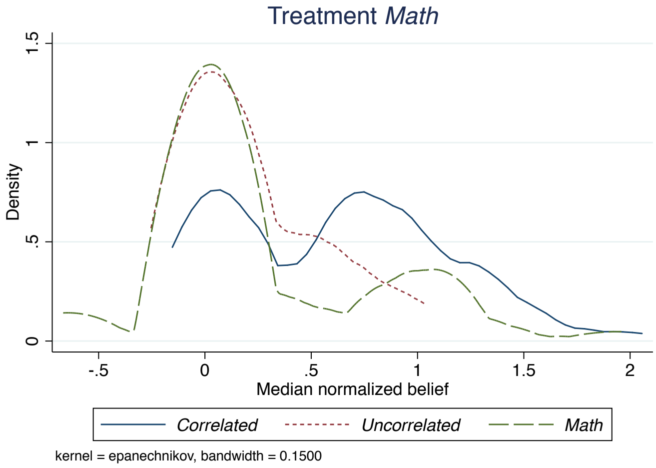

Figure 17: Kernel density estimates of beliefs in the *Math* treatment

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.

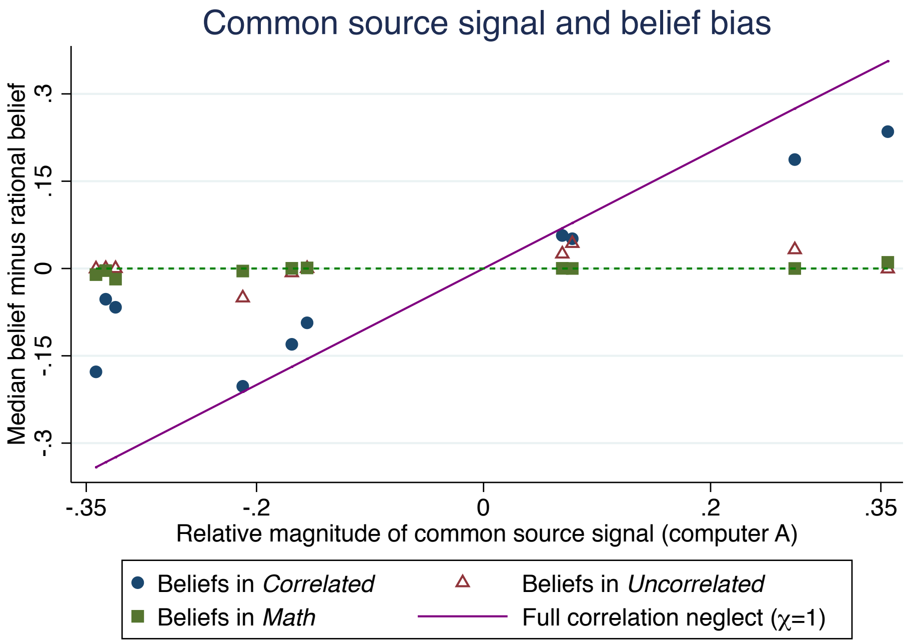

Figure 18: Beliefs in the *Correlated*, *Uncorrelated* and *Math* treatments plotted against the relative magnitude of the signal of computer A. See the notes of Figure 2 for details on the construction of this figure.

Table 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.

Table 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.

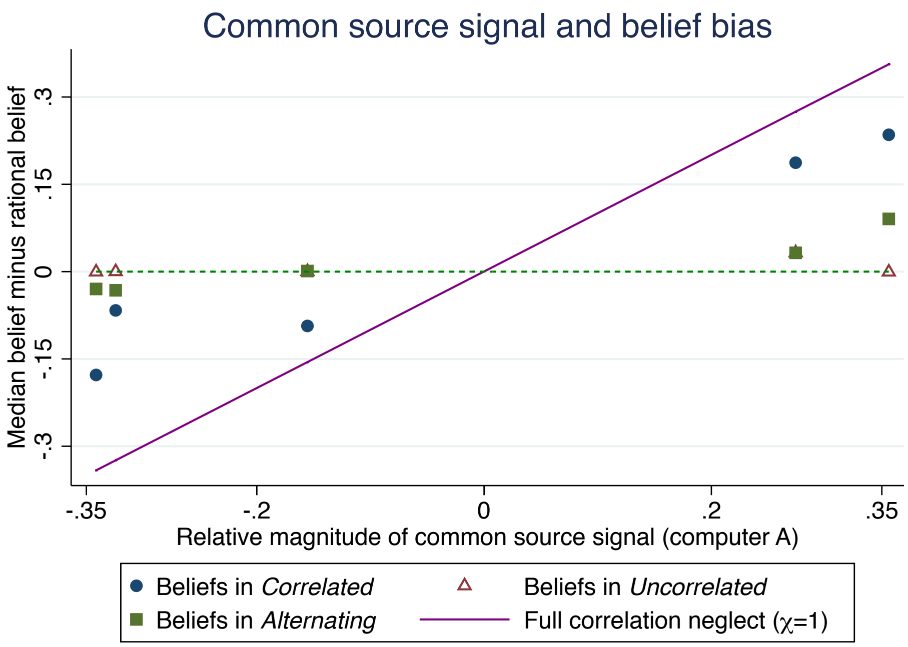

Figure 19: Beliefs in the *Correlated*, *Uncorrelated* and *Alternating* treatments plotted against the relative magnitude of the signal of computer A. See the notes of Figure 2 for details on the construction of this figure.

Table 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.

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 20 provides kernel density estimates of the median naïveté parameters (see equation (3)) in the baseline and high-stakes conditions, which suggest that beliefs in these treatments are almost indistinguishable from each other. As Figure 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.

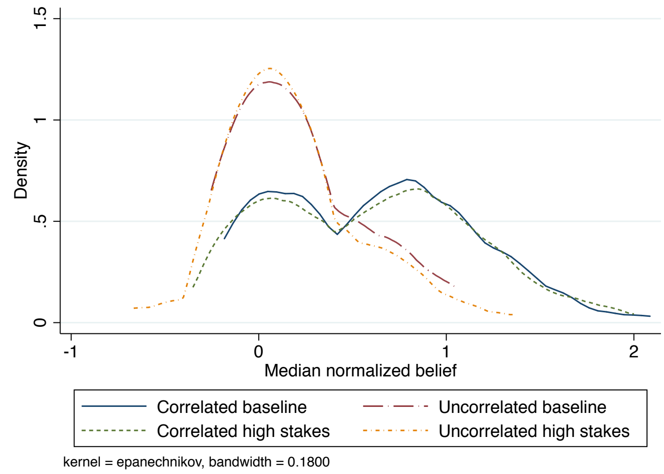

Figure 20: Kernel density estimates of beliefs in the baseline and high stakes conditions

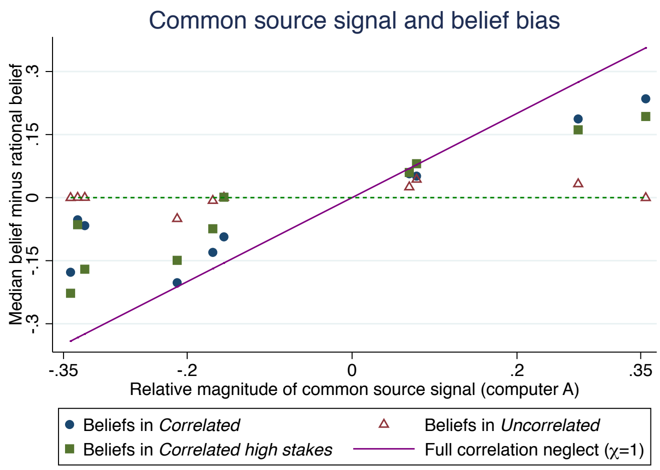

Figure 21: Beliefs in the baseline *Correlated* and *Uncorrelated* as well as the *Correlated high stakes* treatments plotted against the relative magnitude of the signal of computer A. See the notes of Figure 2 for details on the construction of this figure.

F.5 *Multiply* Treatment

In treatment *Multiply*, the intermediaries 1-3 each received one signal and multiplied it by *1.5.* Table 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 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 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 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.

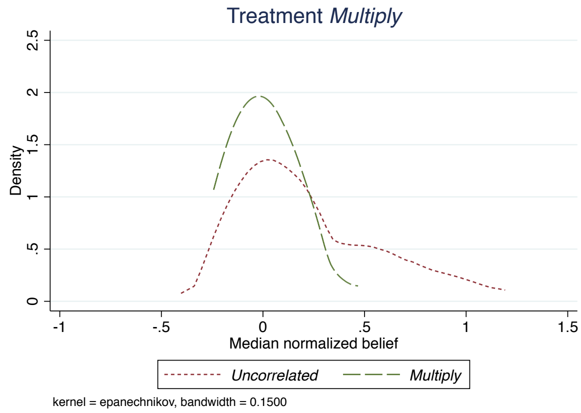

Figure 22: Kernel density estimates of median naïveté parameters in *Multiply* and the *Uncorrelated* treatment. Naïveté parameters are computed akin to the procedure in eq. (3)

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 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 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

[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.

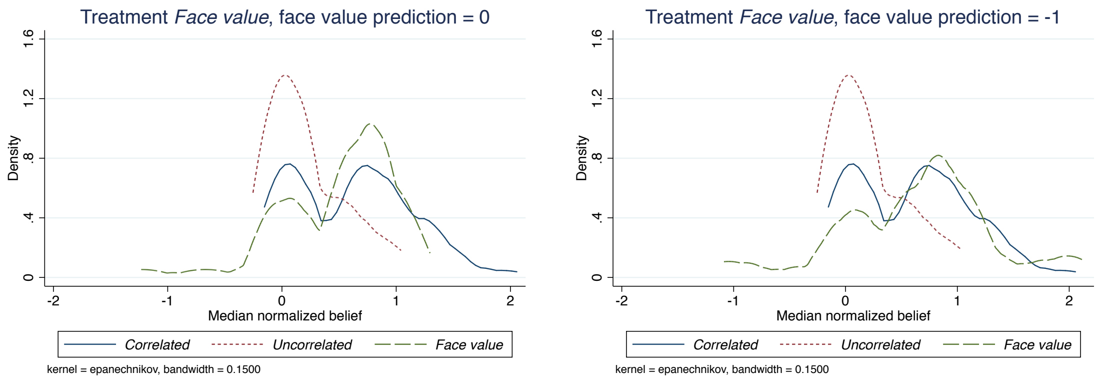

Figure 23: Kernel density estimates of median normalized beliefs in the *Face value* treatment, compared with those from the baseline *Correlated* and *Uncorrelated* conditions. The left panel illustrates the five tasks in which the face value belief equals the rational belief, while the right panel 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.

a prediction different from correlation neglect, we identify significant evidence for people's neglect of correlations.

Table 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 *Correlated* 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 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 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 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 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.

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

Table 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 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 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.

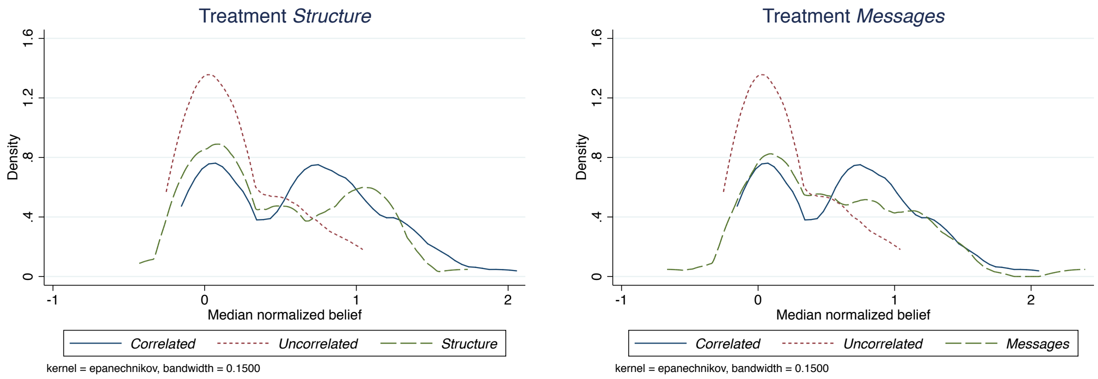

Figure 24: Kernel density estimates of median normalized beliefs in the *Structure* and the *Messages* treatments, each compared with median normalized beliefs in the baseline *Correlated* and *Uncorrelated* treatments

Table 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.

G Correlation Neglect in Newspaper Articles

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.

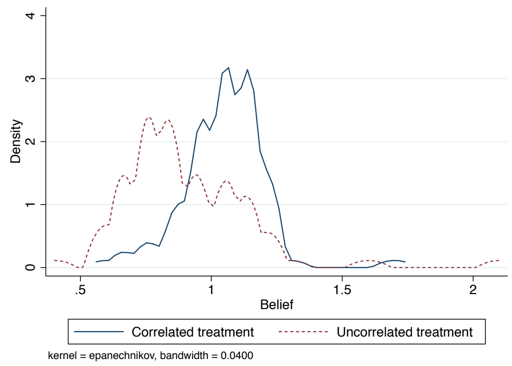

Figure 25: Kernel density estimates of beliefs in the two main newspaper treatments

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.

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.

[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."

G.4 Newspaper Articles and Instructions

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. 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): ...

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.