Review of Economic Studies (2019) **86**, 313–332 doi:10.1093/restud/rdx081

Advance access publication 22 December 2017

Correlation Neglect in Belief Formation

BENJAMIN ENKE
Harvard University, Department of Economics, and NBER
and
FLORIAN ZIMMERMANN
University of Bonn and briq

First version received October 2015*; Editorial decision August* 2017*; Accepted December* 2017 (*Eds.*)

Many information structures generate correlated rather than mutually independent signals, the news media being a prime example. This article provides experimental evidence that many people neglect the resulting double-counting problem in the updating process. In consequence, beliefs are too sensitive to the ubiquitous "telling and re-telling of stories" and exhibit excessive swings. We identify substantial and systematic heterogeneity in the presence of the bias and investigate the underlying mechanisms. The evidence points to the paramount importance of complexity in combination with people's problems in identifying and thinking through the correlation. Even though most participants in principle have the computational skills that are necessary to develop rational beliefs, many approach the problem in a wrong way when the environment is moderately complex. Thus, experimentally nudging people's focus towards the correlation and the underlying independent signals has large effects on beliefs.

Key words: Beliefs, Correlation neglect, Bounded rationality, Complexity, Attention

JEL Codes: C91, D03, D83, D84, D40

1. INTRODUCTION

A pervasive feature of information structures is that decision makers are exposed to correlated signals. For example, various news media 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 (Akerlof and Shiller, 2009), which implies 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 two contributions. First, we provide clean evidence that in a relatively simple and completely transparent setting people neglect correlations in information sources when forming beliefs, albeit

The editor in charge of this paper was Botond Köszegi.

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. Secondly, we develop a series of treatment variations to uncover that people do in principle possess the mathematical and computational skills that are necessary to process correlated information in our setting. However, when the informational environment is sufficiently complex, many people exhibit conceptual problems in identifying and thinking through the correlation in the first place. As a consequence, exogenously shifting subjects' focus towards the correlation and the underlying independent signals has large effects on beliefs.

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. The entire signal-generating process is computerized, and subjects know the precise process generating the data. In this setup, computers A through D generate four unbiased i.i.d signals about the state of the world. In the *Uncorrelated* condition, subjects observe these four independent signals. In the *Correlated* treatment, participants also receive four messages, which consist of 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, the signal of the common source A is partially recurring in multiple messages, implying a potential double-counting problem. Viewed through the lense of our motivating examples, this setup could reflect a news reader who has access to different news sources, all of which partially rely on the same press agency. Similarly, the setup mirrors a network context in which an individual communicates with various friends, all of which have previously communicated with a mutually shared acquaintance.

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. Despite extensive instructions and control questions, our results indicate that, on average, subjects treat correlated information partially as independent and hence double-count the signal of the common source A. Thus, while beliefs remain statistically unbiased *ex ante*, they are highly sensitive to the well-connected information source and exhibit excessive swings, an effect that is sizeable, 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 roughly bimodal distribution: most people are either fully sophisticated or very naïve about the correlation, which points to the presence of two fundamentally different belief formation types. In particular, those subjects that do not successfully process correlations form beliefs by following a particular simple heuristic of averaging the correlated messages. The strong type heterogeneity is significantly associated with cognitive skills. At the same time, the relationship between subjects' response times (a commonly used proxy for cognitive effort; Rubinstein, 2007, 2016) and beliefs is weak at best, both within the main treatment condition and when we exogenously increase response times through a moderate increase in financial incentives.

The second part of the article investigates the mechanisms underlying the observed neglect of correlations. We start our corresponding quest by examining the role of complexity, which has previously been shown to affect updating mistakes and thus serves as a natural starting point for our analysis (*e.g.* Charness and Levin, 2009). We exogenously manipulate the complexity of the

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

updating problem by reducing the number of signals and resulting messages, so that subjects only need to process two pieces of information. In this low complexity version of our experiments, correlation neglect essentially disappears.

While this finding highlights that correlation neglect is not universal, but rather a function of the environment, it leaves open the precise mechanism through which complexity generates neglect. To make progress, we conceptualize belief formation as three steps of reasoning, all of which are potentially affected by our complexity manipulation: first, people need to notice the double-counting problem inherent in our experimental environment, *i.e*. they need to realize that taking the correlated messages at face value is suboptimal; secondly, subjects need to understand that this double-counting problem can be overcome by backing out the underlying independent signals; thirdly, they need to be willing and able to execute the mathematical computations that are necessary to develop unbiased beliefs. Crucially, in this framework, the first two steps refer to *conceptual* problems that people might have in processing correlations, while the last step is about *mathematical* or *computational* problems.

Given that conceptual and mathematical limitations likely have different implications for both policy and potential formalizations of correlation neglect, we develop two treatment variations to separate these two broad mechanisms. First, building on the low complexity environment, we elucidate the role of the size of the information structure. We design a treatment in which we fix the mathematical steps that are required to solve the problem, but manipulate how many messages subjects observe based on the independent signals. The corresponding results establish that a "larger" information structure causes significantly more correlation neglect even when the required mathematical operations are unaffected. Thus, the complexity of information structures seems to affect belief updating (also) through its effect on people's ability to notice and think through the correlation, which rationalizes the observed difference in correlation neglect between the baseline and low complexity experiments.

To lend further credence to the idea that subjects struggle predominantly with the conceptual difficulty of detecting and thinking through the correlation, and to provide evidence on how subjects could be debiased, we design an additional treatment variation. Here, the experimental procedures exogenously draw people's focus towards the mechanics that generate the correlation, but again hold fixed the mathematical steps that are required to be rational. In a within-subjects treatment, participants are confronted with both the correlated and the uncorrelated information structure from the baseline treatments, which is meant to induce subjects to focus on the key difference between the two environments. The results show that the vast majority of subjects states rational beliefs in this condition. Thus, taken together, two conceptually distinct treatment variations show that correlation neglect can be meaningfully affected by features of the environment that are independent of purely math-based explanations. This set of results points to the importance of people's problems in noticing and thinking through the correlation in the first place, and speaks against the notion that people cannot (or do not want to) engage in the calculations that are necessary to process relatively simple correlated messages.

This article contributes to the literature on boundedly rational belief formation by identifying an error in statistical reasoning that is associated with a pervasive feature of real information structures such as the news media (see *e.g*. Charness and Levin, 2009; Charness *et al*., 2010; Benjamin *et al*., 2013; Esponda and Vespa, 2014; Hanna *et al*., 2014; Ngangoue and Weizsäcker, 2015; Jin *et al*., 2016, for recent documentations of bounded rationality in other contexts). Conceptually, our article moves beyond existing work by studying in detail the roles of complexity and focus for biased statistical reasoning.[2]

2. Brocas *et al*. (2014) highlight the relevance of attention in strategic settings. Gennaioli and Shleifer (2010), Bordalo *et al*. (2016), and Schwartzstein (2014) provide models of attention-driven updating errors.

Our experiments can be interpreted in terms of learning in networks. Eyster and Rabin (2014) develop a model to show that rationality often requires people to anti-imitate others 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. We establish correlation neglect (and the resulting excessive "imitation") in a setup in which the signal-generating process is known and simple.[3] Our findings hence support the assumptions underlying recent theories of inferential naïveté in social interactions (*e.g.* DeMarzo *et al*., 2003; Eyster and Rabin, 2010; Golub and Jackson, 2010; Bohren, 2016) as well as bounded rationality models in political economy (Levy and Razin, 2015; Ortoleva and Snowberg, 2015).[4]

Relatedly, Shiller (2000) and Akerlof and Shiller (2009) have argued that "exuberant" public opinions or "panics", driven by the multiple occurrence of similar stories in social networks, may be a driver of aggregate distortions. In Online Appendix F, we report on experiments along these lines and show that, in an experimental asset market, the incidence of correlated (and hence partially recurring) news leads to pronounced and predictable price distortions.

The remainder of the article is organized as follows. In the next section, we present our baseline experiments. Section 3 investigates the mechanisms underlying correlation neglect. Section 4 discusses extensions of our experiments and concludes.

2. EVIDENCE FOR CORRELATION NEGLECT

2.1. *Experimental design*

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 of these features.

Subjects were asked to estimate an *ex ante* unknown continuous state of the world μ and were paid for accuracy. The task was framed as guessing how many items are contained in an imaginary container. In order to keep the experiment as simple as possible, we refrained from inducing prior beliefs.[5] 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 i.i.d signals about μ, which were identical across treatments and subjects. Technically, this was implemented by random draws from a truncated discretized normal distribution with mean μ and standard deviation $\sigma = \mu/2$.[6] In the *Uncorrelated* treatment (left panel), the intermediaries 1–3, who are fictitious computers themselves, observed the signals of computers B through D, respectively, and simply transmitted these signals to

3. 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. Section 4 relates our findings to the experimental literature on learning in networks. Spiegler (2016) uses Bayesian networks to provide a formal framework for boundedly rational belief formation.

5. Online Appendix D shows that inducing prior beliefs does not affect our findings.

6. Truncation was at $\mu \pm 2\sigma = \mu \pm \mu$ in order to avoid negative signals.

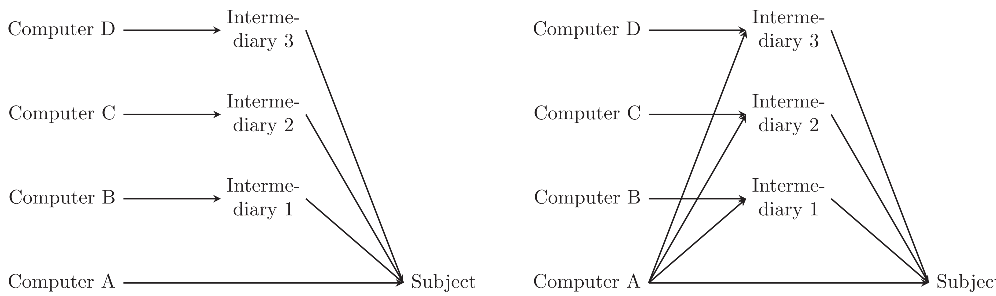

FIGURE 1

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

Notes: In the left panel, the intermediaries directly transmit the signal they observe from the computer they are connected to. In the right panel, the intermediaries take the average of the signals of the two computers they are connected with, and transmit this average to the subjects.

the subject. Thus, subjects received information from computer A as well as from the three intermediaries. For example, in one experimental 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 1–3 observed both the signal of computer A and of computers B–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 article, we will also refer to computer A’s signal as common source signal. Continuing the example from above, each of the three intermediaries took the average of 12 and the corresponding signal of the other computer it communicated with. Thus, computer A reported 12, intermediary 1 reported 10.5, intermediary 2 reported 11, and intermediary 3 reported 6. In the terminology of Eyster and Rabin (2014), this information structure constitutes a “shield”. Here, people need to “anti-imitate” because they predominantly see messages larger than 9, while the majority of signals and the rational belief are smaller than 9. 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, including the rational benchmark. Thus, comparing beliefs across *Correlated* and *Uncorrelated* allows us to identify correlation neglect. 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 that 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 Online Appendix B. At the end of the experiment, subjects were paid according to the precision of their belief in one randomly selected

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	1,110	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

Notes: 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.

task using a quadratic scoring rule (Selten, 1998).[7] 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 messages subjects received in the *Correlated* treatment (see Section 2.2 for details). Throughout, we employ the term "belief" to denote the mean of the belief distribution.

Subjects received extensive written instructions that explained the details of the task and the incentive structure.[8] In particular, the signals of the four computers, how these signals mapped 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 exemplary 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.[9]

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

8. See the Online Supplementary Material 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/.

9. 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 thirteen response options across four questions (*i.e.* $2 \times 3 \times 4 \times 4 = 96$ combinations of responses), making trial and error *extremely* cumbersome. In addition, the BonnEconLab has a control room in which the decision screens of all subjects can be monitored. From this monitoring, no attempts to solve the control questions by random guessing were detectable. Furthermore, whenever a subject appeared to have trouble solving the control questions, an experimenter approached the

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.

Apart from the two baseline *Correlated* and *Uncorrelated* treatments, we implemented a number of straightforward extensions and robustness checks. (i) First, we ran high stakes versions of the two baseline treatments. These experiments featured the same procedures, except that the stake size was tripled.[10] (ii) Secondly, we re-ran the *Correlated* treatment using a slightly different procedure: in treatment *Reading time*, subjects were free to start the control questions and the experiment at any point in time after we had distributed the paper-based instructions, which allows us to measure the time subjects take to read and engage with the instructions before completing the tasks. (iii) Thirdly, we conducted two robustness treatments in which we slightly altered certain aspects of the design, including inducing a prior belief (see Online Appendix D for details). Online Appendix A presents an overview of all treatments that are part of this study, including extensions and further robustness checks.

2.2. *Hypothesis*

In the information structure described above, the computers generated four i.i.d signals of the form $s_h \sim \mathcal{N}(\mu,(\mu/2)^2)$ (truncated at $(0,2\mu)$) for $h \in \{1,\dots,4\}$. In the *Correlated* condition, subjects observed messages s_1 and $\tilde{s}_h = (s_1+s_h)/2$ for $h \in \{2,3,4\}$. When prompted to estimate μ, a rational decision maker would extract the underlying independent signals from the messages $\tilde{s}_h$ and compute the mean rational belief as $b_B = \sum_{h=1}^{4} s_h/4$, which by design also equals the rational belief in *Uncorrelated*.[11]

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

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

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 $\bar{s}$ plus a belief bias component which depends on the degree of naïveté χ and the magnitude of the common source signal relative to the other signals.

subject, clarified open questions, and (very rarely) excluded the subject if they did not show an adequate understanding of the task.

10. Variable earnings in euros were given by $\pi = max[0, 30 - 480 \times (\text{Belief / True state} - 1)^2]$.

11. 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 behaviour in the *Uncorrelated* treatment, where subjects tended to merely take the average of the four signals. Finally, and most importantly, regardless of the precise definition of the rational benchmark, beliefs should be identical across treatments.

Hypothesis. *Assuming that $\chi > 0$, beliefs in the Correlated treatment exhibit an overshooting pattern. 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.*

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. At the same time, note from equation (2) 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.

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 234 subjects participated in the individual belief formation treatments, 94 in the baseline, 94 in the high stakes, and 46 in the *Reading time* treatments. Within the baseline and high stake treatments, the *Correlated* and *Uncorrelated* condition were randomized within session. Sessions lasted about 1.5 hours and average earnings equalled €11.60 in the baseline treatments (≈ USD 15 at the time) and €21.90 in the high stakes treatments (≈ USD 28). In all treatments, payments included a €6 show-up fee.

2.4. *Results*

2.4.1. Beliefs across treatments.

Result 1. *In all but one belief formation task, beliefs differ significantly between treatments in the direction predicted by correlation neglect. This pattern is unaffected by the tripling of the stake size.*

As we will establish formally below, beliefs are strikingly similar between the baseline and the high stakes treatments. Given the otherwise identical procedures, we hence pool the data across stake size conditions in all analyses unless noted otherwise.

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. The bias is very stable across tasks and does not seem to depend on the magnitude of the true state.[12] As a consequence of these biased beliefs, subjects in the baseline condition earned roughly €2.70 less than those in the *Uncorrelated* group, which amounts to almost 50% of subjects' average variable earnings. The earnings difference is significant ($p = 0.0025$, Wilcoxon ranksum test). In the high stakes treatments, the earnings difference is €5.40 ($p = 0.0887$).

Our experiments provide no feedback and hence little scope for learning. Indeed, in the data, subjects do not seem to learn to deal with correlations over time (see Online Appendix C.4). It is doubtful for at least two reasons that subjects would learn within the course of ten experimental periods even in the presence of feedback. First—given the small sample of four signals—occasionally the "naïve" (correlation neglect) belief is closer to the true state than the rational

12. Our two robustness treatments replicate these findings, see Online Appendix D for details.

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	9.1	0.0002
88	71.25	96.63	71.25	87.5	0.0001
250	259.75	219.38	260	250	0.0028
732	853.15	709.13	850	752	0.0018
1,000	974.75	1,042.38	999	1,030	0.0165
4,698	4,810	3,209	4,810	4,505	0.0001
7,338	8,604.5	9,277.25	9,000	9,152.5	0.8317
10,000	7,232.25	4,887.63	7,232	6,200	0.0001
23,112	26,331	20,745.5	25,000	21,506	0.0001
46,422	38,910.5	25,625	38,885.5	30,277	0.0014

Notes: This table presents an overview of beliefs in the *Uncorrelated* and *Correlated* treatments across the ten estimation tasks. The p-values refer to a Wilcoxon ranksum test between beliefs in the *Correlated* and *Uncorrelated* conditions. The data are pooled across the high stakes and baseline treatments. For reference, we also provide the benchmarks of rational and fully naïve beliefs. 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.

belief (compare Table 1), creating a rather coarse environment for learning. In addition, recall that the correlation neglect belief is statistically unbiased. Secondly, we can actually derive first insights into whether and how people learn over time in the presence of feedback from market experiments that are reported upon in Section 4. These market trading experiments build on our individual decision-making design and feature the same information structure. Here, subjects were provided with extensive feedback after each period, including the true state of the world and losses and profits from trading activities, yet the data reveal little, if any, learning over time.

2.4.2. Heterogeneity. Thus far, we have established a significant amount of correlation neglect *on average*. These average patterns may mask a substantial amount of heterogeneity. To investigate this, we develop a measure of an individual's belief type. Specifically, our experimental design in combination with the simple model of belief formation introduced in Section 2.2 allows us to estimate individual's naïveté χ. For each belief, we compute the naïveté parameter χ in equation 2. The median of those naïveté values then serves as estimator for the subject-level naïveté parameter:

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

Figure 2 provides kernel density estimates of the distribution of these naïveté parameters for both the *Correlated* and the *Uncorrelated* treatment, pooled across stake size conditions. The plots reveal that in the *Uncorrelated* treatment the vast majority of subjects behaves approximately 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 parameter $\chi = 1$, respectively, which suggests the presence of different updating types. In particular, those subjects that do not succesfully process correlations form beliefs by following a particular simple heuristic of essentially full neglect.[13] Finally, beliefs in *Correlated* exhibit a much larger heterogeneity than those in *Uncorrelated*: in eight out of the ten belief formation tasks, the within-treatment belief variance is statistically significantly higher in the *Correlated* condition ($p < 0.05$).

13. Online Appendix C.3 analyses the stability of the individual-level naïveté parameters across tasks.

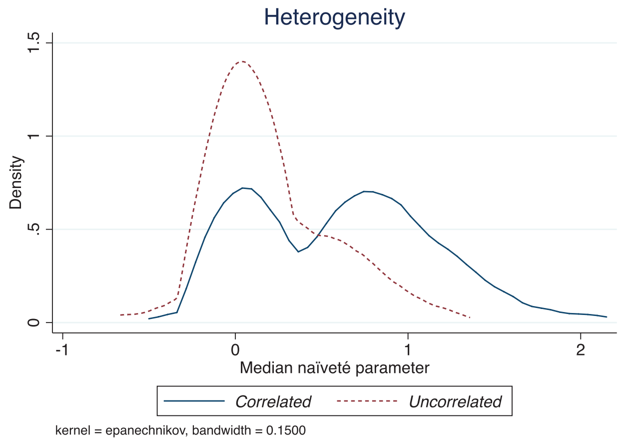

FIGURE 2

Kernel density estimates of median naïveté parameters.

Note: The two kernels depict the distributions of naïveté in the *Correlated* and *Uncorrelated* conditions, pooled across the high stakes and baseline treatments.

Our procedure of computing an individual's type only makes use of the first moment of the distribution of each subject's beliefs, and hence ignores the variability in beliefs. In Online Appendix C.5, we pursue a different approach by estimating the belief formation rule proposed in Section 2.2 through a finite mixture model. The picture resulting from these estimations is very similar to what can be inferred from Figure 2.

2.4.3. Cognitive ability and cognitive effort. Before we develop a structured analysis of mechanisms in the next section, we conclude the baseline analysis by studying the roles of cognitive ability and cognitive effort in generating correlation neglect. Columns (1) and (2) of Table 3 provide evidence for the treatment difference in beliefs, again pooling across the baseline and high stakes conditions. Here, the dependent variable is the full set of ten beliefs, expressed in terms of χ (equation 2).

The remainder of the table restricts attention to the *Correlated* treatments. Columns (3) and (6) show that correlation neglect is significantly associated with cognitive skills, as derived from participants' high school GPA and their score on a post-experimental Raven IQ test. Columns (4) and (6) through (10) investigate the relationship between correlation neglect and subjects' response times, which are commonly used as proxy for cognitive effort (Rubinstein, 2007, 2016). Indeed, it is conceivable that subjects face prohibitively high cognitive costs in developing or executing a solution strategy (Caplin *et al.*, 2011; Gabaix, 2014) and hence opt for a simplifying heuristic, perhaps akin to Kahneman's (2011) system 1 versus 2 terminology.[14] The standard approach in the literature is to record the time subjects take to complete an experimental task, *after they have read and contemplated about the experimental instructions*. However, it is *ex ante* unclear at which point during the experiment subjects develop their solution strategy. As explained in Section 2.1, we implemented a follow-up treatment called *Reading Time* to explicitly

14. In addition, subjects might exhibit self-serving biases regarding their effort level: in principle, subjects might only be able to rationalize to themselves not to exert effort to develop rational beliefs if the underlying problem is sufficiently complex. For recent evidence on such "wiggling" behaviour see, *e.g.* Dana *et al.* (2007), Haisley and Weber (2010), and Exley (2015).

TABLE 3
Heterogeneity in correlation neglect

	Dependent variable:									
	Naiveté χ								*Response time*	
	Full sample		*Correlated* treatments							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0 if *Uncorrelated*, 1 if *Correlated*	0.38***	0.35***								
	(0.06)	(0.06)								
Cognitive skills		−0.15***	−0.27***			−0.27***		0.20**		0.24*
		(0.03)	(0.05)			(0.06)		(0.10)		(0.14)
Response time (in minutes)				−0.15***		−0.13***		−0.10*		
				(0.04)		(0.04)		(0.05)		
0 if *Baseline*, 1 if *High stakes*					−0.029	−0.0093			0.35*	0.42**
					(0.10)	(0.09)			(0.20)	(0.19)
Reading time (in minutes)							−0.017	−0.0026		
							(0.03)	(0.03)		
Constant	0.24***	0.31	0.66***	0.85***	0.63***	0.58*	0.50	−0.17	1.43***	0.093
	(0.03)	(0.20)	(0.05)	(0.07)	(0.07)	(0.35)	(0.33)	(0.92)	(0.13)	(1.10)
Additional controls	No	Yes	No	No	No	Yes	No	Yes	No	Yes
Observations	1799	1785	889	889	889	875	386	382	889	875
R^2	0.06	0.13	0.06	0.05	0.00	0.15	0.00	0.14	0.02	0.15

Notes: OLS estimates, robust standard errors (clustered at subject level) in parentheses. The table analyses the determinants and correlates of subjects' naïveté as implied in each of the ten beliefs. In columns (1)–(2), observations include all subjects from the *Correlated* and *Uncorrelated* treatments, both baseline and high stakes. In columns (3)–(6) as well as (9)–(10), the sample includes all subjects from the baseline and high stakes *Correlated* conditions. Columns (7) and (8) analyse treatment *Reading time*. Additional controls include age, gender, monthly income, marital status fixed effects, and task fixed effects. Cognitive skills are the z-score of the unweighted average of the z-scores of high school GPA and a Raven test score. All regressions exclude extreme outliers with $|\chi_i^j| > 3$, but all results are robust to including these observations when employing median regressions. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

measure not only standard response times, but also the time subjects take to read the instructions. In sum, we have access to a standard measure of response time for the baseline and high stakes *Correlated* and the *Reading Time* treatments, as well as to the time subjects took to read the instructions in *Reading Time*.

Turning to the analysis, columns (4)–(8) document that both of our proxies for cognitive effort (reading and response times) are only weakly related to correlation neglect. First, in the baseline treatments (analysed in columns (4) and (6)), a longer time spent on the problem is significantly correlated with less correlation neglect, but the quantitative magnitude of this relationship is small: interpreted causally, the OLS coefficients suggest that response times would have to increase by about four minutes to move a fully naïve subject to fully rational beliefs. Note that this is a very implausible magnitude: the average response time of approximately rational subjects is just 2.2 minutes, and increasing the naïve types' response times by four minutes corresponds to an increase of about 3.5 standard deviations in the sample.[15] While the within-treatment relationship between response times and correlation neglect is correlational in nature, columns (5) and (6) document that the exogenous threefold increase in financial stakes had no effect on correlation neglect. This is noteworthy since the tripling of the stake size *did* significantly affect cognitive effort as proxied by response times [compare columns (9)–(10)]. That is, tripling the stakes increases cognitive effort, but does not affect the presence of the bias.

15. In addition, the correlation between response times and naïveté might be purely mechanical: conditional on having developed different solution strategies, rational subjects ought to take longer to solve the problem than naïfs because computing rational beliefs requires additional mathematical steps.

TABLE 4
Overview of main treatments

Treatment	Description	Purpose
Baseline correlated	Four Computers, three Intermediaries	Establish correlation neglect
Baseline uncorrelated	Four Computers, three Intermediaries	Establish correlation neglect
High stakes corr. & uncorr.; Reading time	Four Computers, three Intermediaries	Role of cognitive effort
Reduced complexity correlated	Two Computers, one Intermediary	Importance of complexity
Reduced complexity uncorrelated	Two Computers, one Intermediary	Importance of complexity
Many Stimuli	Two Computers, three Intermediaries	Importance of size of information structure
Alternating	Four Computers, three Intermediaries	Importance of conceptual aspect

Finally, columns (7) and (8) show that the relationship between correlation neglect and subjects' reading times is also weak. Neither in unconditional nor in conditional regressions is reading time significantly associated with the bias.[16]

In sum, (i) neglect types do not take less time to read the instructions; (ii) they take only slightly less time to work on the specific tasks; and (iii) exogenous increases in effort do not translate into better beliefs. While mostly descriptive in nature, these various pieces of evidence are suggestive that correlation neglect is not driven by laziness.

3. MECHANISMS AND DEBIASING

We investigate the mechanisms behind correlation neglect through a series of treatment variations, as summarized in Table 4.[17] We will refer back to this table as we move along. Table 6 in the Online Appendix provides a complete list of all treatments that are part of this study, including robustness checks and extensions.

3.1. *The role of complexity*

A common theme in the literature is that the degree of complexity of a decision problem exerts a substantial effect on the existence and magnitude of cognitive biases (*e.g*. Charness and Levin, 2009). To examine the effects of complexity on correlation neglect, we implemented a set of low-complexity treatments that were identical to the baseline conditions, except that we reduced the number of computers and intermediaries. In these low-complexity conditions, only two computers (A and B) generated unbiased i.i.d signals about the state μ, and only one intermediary was present (see Figure 3). In both treatments, subjects were provided with information from computer A as well as from the intermediary. In the *Low Complexity Uncorrelated* treatment, the intermediary directly transmitted the signal of computer B. In the *Low Complexity 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 (see also Table 4). We implemented the same ten belief formation tasks as in the baseline treatments using the

16. The R^2 in most of these regressions is fairly small (similarly so in Table 5), which directly results from the large amount of heterogeneity in the data, both across subjects and within subjects across experimental tasks.

17. Online Appendix E.2 investigates whether correlation neglect 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.

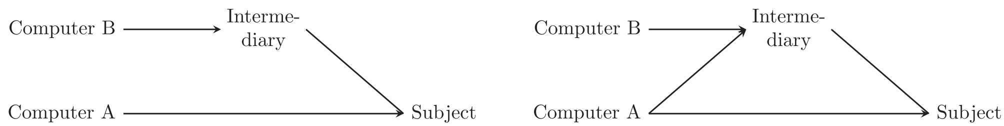

FIGURE 3

Low complexity uncorrelated (left panel) and correlated (right panel) information structure.

Notes: In the left panel, the intermediary directly transmit the signal from computer B. In the right panel, the intermediary takes the average of the signals of A and B, and transmits this average to the subjects.

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.

Result 2. *An extreme reduction in the environment's complexity eliminates the bias.*

Consistent with previous documentations of the role of complexity in different contexts, we find that correlation neglect disappears in our low-complexity treatments. Columns (1)–(2) of Table 5 present the results of OLS regressions of the naïveté implied in each belief of a given subject on a correlated treatment dummy.[18] While the point estimate is negative, it is very far from statistically significant. This finding is noteworthy because it shows that in (admittedly extremely) simple informational environments subjects do grasp the implications of correlated information structures; thus, correlation neglect depends not only on subject's updating type, but also on the properties of the informational environment.[19]

At the same time, in terms of pinning down the precise mechanisms that generate correlation neglect, these reduced complexity treatments do not provide a definitive answer because they manipulate a number of features of the experimental design at once, relative to the baseline conditions. To organize our discussion of the mechanisms underlying correlation neglect (and its dependence on complexity), we adopt a simple qualitative framework that clarifies the cognitive steps required to develop rational beliefs.

3.2. *Mechanisms: a framework*

Arguably, solving our experimental task requires three steps of reasoning:

(1) *Notice the problem:* subjects need to notice the presence of the correlation among signals and realize that averaging the correlated messages introduces a double counting problem.
(2) *Understand how to solve the problem:* subjects need to think through the problem and understand that it can be solved by backing out and averaging the underlying independent signals.
(3) *Solve the problem mathematically:* subjects need to be able and willing to solve the problem mathematically by setting up the corresponding equations and executing them.

18. The implied χ_i^j are computed using the same procedure across all low-complexity conditions:

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

19. Note, however, that our low-complexity environment 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.

TABLE 5
Mechanisms

	Dependent variable: *Naiveté* χ					
	Low complexity				*Corr. & Alternating*	
	(1)	(2)	(3)	(4)	(5)	(6)
0 if *Low complexity uncorr.*, 1 if *Low complexity corr.*	−0.14	−0.10				
	(0.10)	(0.08)				
0 if *Low complexity corr.*, 1 if *Many Stimuli*			0.34***	0.33***		
			(0.11)	(0.11)		
0 if *Correlated*, 1 if *Alternating*					−0.25***	−0.21***
					(0.08)	(0.08)
Constant	0.15***	0.18	0.0090	−0.55	0.54***	0.44**
	(0.04)	(0.31)	(0.09)	(0.54)	(0.05)	(0.19)
Additional controls	No	Yes	No	Yes	No	Yes
Observations	884	874	891	881	681	674
R^2	0.01	0.05	0.03	0.06	0.02	0.12

OLS estimates, robust standard errors (clustered at the subject level) in parentheses. The table analyses the mechanisms behind correlation neglect. The dependent variable is always subjects' naïveté as implied in a given belief. In columns (1)–(2), observations include all beliefs of subjects in the *Low complexity correlated* and *Low complexity uncorrelated* treatments. In columns (3)–(4), the sample includes all beliefs of subjects from the *Low complexity correlated* and *Many Stimuli* conditions. In columns (5)–(6), the sample includes subjects in the *Correlated* conditions (both baseline and high stakes) as well as in *Alternating*, where the set of beliefs is restricted to those tasks in which the *Alternating* treatment featured a correlated information structure. Additional controls include age, gender, cognitive skills, monthly income, marital status fixed effects, and task fixed effects. All regressions exclude extreme outliers with $|\chi_i^j| > 3$, but all results are robust to including these observations when employing median regressions. $^*p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

All of these steps are potentially affected by our complexity manipulation. For example, the number of mathematical steps that are required to solve the problem differs across the low-complexity and baseline conditions. Likewise, noticing and thinking through the double counting problem might be harder in our baseline treatments because the size of the information structure—captured by the number of signals and messages—increases.

In what follows, we narrow in on the mechanism. For this purpose, we follow a long line of work in cognitive psychology that divides mental operations into a *conceptual* and a *mathematical* or computational aspect.[20] In the conceptual part (steps 1–2 from above), subjects need to *develop* a solution strategy, while the mathematical part (step 3) requires them to *execute* that strategy. We focus our efforts on differentiating the steps of developing and properly executing a strategy.

For this purpose, we develop two treatment variations. The key idea behind both treatments is to hold the mathematical steps required to correctly solve the problem constant, yet manipulate the way people approach and think about the problem. For this purpose, we apply a basic idea reported in DellaVigna (2009) to our context, *i.e.* that the probability of attending to and understanding complex issues—here, the correlation—is a function of (at least) two aspects of the environment. First, people's propensity to notice and think through correlations might negatively depend on the size of the information structure, *i.e.* the total number of stimuli as proxy for the overall complexity of the problem. In particular, it is conceivable that a "bigger" problem either makes it harder for subjects to identify the double counting problem or makes it less obvious how to solve it. Secondly, attending to and understanding the double counting problem might be easier

20. For example, standard treatments of the computational theory of mind assume a distinction between representations and operations on those representations (*e.g.* Thagard, 1996; Horst, 2011).

if people are nudged to focus on and think about the correlation and the underlying independent signals.[21]

We test these two predictions by increasing the size of the information structure (Section 3.3.1) and nudging subjects to pay special attention to the double counting problem and the underlying independent signals (Section 3.3.2), respectively, while holding the computational steps constant.

3.3. *Conceptual versus mathematical problems*

3.3.1. Complexity and the Size of the Information Structure. Recall that the baseline and low-complexity environments differed not only in the size of the information structure *per se*, but also in the number of mathematical steps required to solve the problem. To isolate the pure effect of increasing the size (and hence the conceptual difficulty) of the problem, we designed a variation of the *Low Complexity Correlated* treatment. In treatment *Many Stimuli*, depicted in the right panel of Figure 4, again only two computers generated unbiased iid signals about the state of the world μ, but three intermediaries communicated with subjects. Intermediary 1 observed the signals of computers A and B and transmitted the average to subjects. Intermediary 2 observed the signal of computer A and reported it to subjects. Intermediary 3 observed the signals of A and B and transmitted $(3/4\times A+1/4\times B)$ to subjects. Taken together, subjects observed the signal of computer A twice, and also received two different linear combinations of the signals of A and B. Thus, this treatment manipulates the size of the information structure (via the number of messages), but requires exactly the same mathematical steps as those in *Low Complexity Correlated*: subjects could either simply copy the message of Intermediary 1, or invert the message of Intermediary 1 and then compute the average of A and B.

Forty-seven subjects participated in *Many Stimuli*, which lasted 80 minutes on average and yielded average earnings of €11.10.

Result 3. *Holding fixed the mathematical steps required to solve the updating problem, a larger number of messages induces more correlation neglect.*

Columns (3)–(4) of Table 5 report upon OLS regressions of the naïveté implied in subjects' ten beliefs on a treatment dummy which assumes a value of zero if subjects are in the *Low Complexity Correlated* condition and of one if participants are in *Many Stimuli*. The statistically significant point estimate suggests that the increase in the number of messages alone increases correlation neglect by about 0.34 units of naïveté (χ) on average, an effect size that is similar to our main treatment effect in the baseline conditions (0.38 units of χ, see column (1) of Table 3). Analogous regressions show that beliefs in *Many Stimuli* also significantly differ from those in *Low Complexity Uncorrelated* ($p<0.05$).

These results suggest that it is easier for subjects to notice and think through the correlated information structue when the size of the updating problem (as captured by the number of messages) is smaller. Arguably, this finding (i) provides our first piece of evidence that correlation neglect is to a large extent driven by conceptual as opposed to mathematical problems, but also (ii) suggests that people's propensity to notice and understand the correlation depends upon features of the environment, hence rationalizing the difference between the results in our baseline and low-complexity experiments.

21. Online Appendix E.1 presents a simple model in the spirit of DellaVigna (2009) that formalizes these ideas.

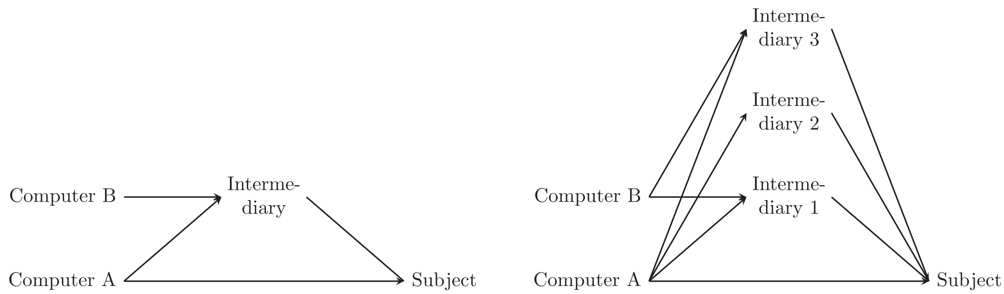

FIGURE 4

Information structure in *Low Complexity Correlated* (left panel) and *Many Stimuli* (right panel).

Notes: In the left panel, the intermediary directly transmit the signal from computer B. In the right panel, intermediary 1 observes the signals of computers A and B and transmits the average to subjects. Intermediary 2 observes the signal of computer A and reports it to subjects. Intermediary 3 observes the signals of A and B and transmits $(3/4\times A+1/4\times B)$.

3.3.2. Nudge evidence. We proceed by testing the idea that drawing subjects' attention to the double counting problem and the underlying independent signals may help them in noticing the correlation and understanding how to cope with it. To achieve this goal while holding the mathematical complexity of the problem constant, treatment *Alternating* (see Table 4) varies the nature of the information structure (correlated or uncorrelated) within subjects between tasks. The basic design of the *Alternating* treatment (*e.g.* the number of computers and intermediaries) is identical to the baseline conditions. The key difference is that 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. The instructions emphasized that subjects would have to pay special attention to the prevailing scenario. 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 workings of the intermediaries.

Arguably, alternating the correlated and uncorrelated information structure might manipulate both steps 1 and 2 of our framework above, *i.e.* (i) this treatment might make it easier for subjects to notice the double counting problem, but (ii) it may also help subjects understand that this problem can be solved by backing out the underlying independent signals because it makes the role of the underlying signals more salient. At the same time, the treatment does not provide any hints on how to *mathematically* compute the correct solution and hence leaves step 3 of the framework unaffected. Forty-seven subjects took part in the *Alternating* treatment and earned €13.10 on average.

Result 4. *Exogenously increasing subjects' focus on the correlation and the underlying independent signals reduces the bias.*

Columns (5)–(6) of Table 5 present the regression results. Here, we regress the naïveté implied in beliefs in *Correlated* and *Alternating* on a treatment dummy. The point estimate indicates that shifting subjects' focus reduces correlation neglect by 0.25 units of naïveté. To provide a complementary perspective, Online Appendix E.3 provides kernel density estimates of (median) beliefs in *Alternating*.

In sum, if subjects are nudged to focus on the correlation and the underlying independent signals, the bias is substantially reduced. Arguably, this treatment also provides evidence that people are in principle well capable of and sufficiently motivated to perform the calculations that are needed to develop rational beliefs—after all, treatment *Alternating* manipulates neither subjects' mathematical skills nor their incentives to solve the problem.[22]

3.4. *Discussion*

The analysis of mechanisms has revealed that people do not struggle so much with the mathematics involved in solving our experimental task, but more with the conceptual problem of noticing and thinking through the correlation. In addition, we have provided causal evidence that people are more likely to adequately process correlations if the information structure is "smaller", hence indicating the importance of complexity for noticing and understanding correlations. While the general idea behind these comparative statics effects may well extend to different environments, it should be acknowledged that the results are conditional on this particular type of (relatively simple) correlation.

The distinction between developing and executing a solution strategy appears potentially important for both policy and economic theory. For example, if people's shortcomings in processing correlations were closely linked to mathematical problems, then an obvious policy implication is to teach more basic math. On the other hand, if people's problems were predominantly conceptual in nature, meaning that they fail to notice or think through the correlations, then policy remedies may focus on making people aware of the correlation and nudging them towards the underlying independent signals. Crucially, in contrast to teaching math, such a policy will likely be *context-specific*: pointing people to the presence of correlations in one context would not necessarily help them in another context. In addition, from the perspective of theory, the distinction between developing and executing a strategy is potentially important as the former appears linked to having a wrong subjective model of the environment, while the latter could perhaps be formalized by modelling cognitive effort costs.

While the distinction between the conceptual and mathematical steps in our framework appears to be of direct economic relevance, this is arguably less the case for the distinction between steps 1 and 2. From the viewpoint of policy, treatment *Alternating* has shown that making the double counting problem and the underlying signals very salient largely eliminates the bias—whether this is the case because people do not even notice the double counting problem to begin with or do not understand how to solve it seems rather subtle and likely less relevant in practice. For example, in the context of the news media, a policy intervention that reminds people of the fact that many news articles rely on the same press agencies or journalists, is likely to manipulate both people's awareness of the correlation, and their propensity to develop the correct solution strategy. Likewise, it is difficult to foresee whether the rather subtle distinction between steps 1 and 2 could be productive for economic theory, especially given that both steps are related to the

22. That subjects appear to be sufficiently motivated confirms findings from Section 2.4.3 about the role of cognitive effort. Online Appendix E.4 presents the results from two further treatment variations that lend further credence to the insight that people struggle much more with noticing the double counting problem in the first place, rather than solving it mathematically. For example, when subjects receive the hint "…Think carefully about what the intermediaries do! What does that imply for the estimates of the intermediaries?", they notice and solve the double counting problem in a very similar fashion as in *Alternating*.

conceptual difficulty of processing correlations. For these reasons, we refrain from attempting to further disentangle steps 1 and 2 from our motivating framework.

4. EXTENSIONS AND CONCLUDING REMARKS

Using experiments with more than 1,000 subjects, this article provides clean evidence for people's tendency to neglect correlations in information sources when forming beliefs and the corresponding mechanisms. While we deliberately designed a tightly controlled and abstract information structure to obtain a clean view on the cognitive bias, it would be interesting to extend the analysis to more naturalistic information. In Online Appendix G, we explore one possible avenue by confronting subjects with real newspaper reports covering correlated information. 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 after they have read these correlated forecasts are indicative of correlation neglect.

Correlation neglect is likely to have implications in applied settings such as interactive social network or herding setups. In this respect, our findings contribute to an active empirical literature on naïve social learning, which has often identified updating patterns consistent with correlation neglect (Brandts *et al*., 2015; Chandrasekhar *et al*., 2016; Eyster *et al*., 2016; Grimm and Mengel, 2016).[23] Similarly, our paper also relates to work on financial decision-making in the presence of correlated asset returns (Kallir and Sonsino, 2010; Eyster and Weizsäcker, 2011). What sets us apart from all of these contributions is our focus on a simple and completely transparent information structure as well as a study of the underlying mechanisms. For example, our updating environment is stripped of the complexities that pervade experiments of social interactions in networks, such as lack of common knowledge of rationality, complicated updating processes after several rounds of communication, or the need to evaluate "indirect" information of uncertain origin (Golub and Jackson, 2010). Also, the correlation structure in our setup is very simple compared to the typically non-trivial correlations among financial assets. By documenting a "pure" form of correlation neglect, our findings arguably highlight the potential importance of this bias in a wide range of applied settings. In addition, our article provides evidence on the mechanisms underlying correlation neglect, and how the presence of the bias depends on features of the environment.

Markets are an additional obvious candidate to study correlation neglect. For instance, simple market interaction might already suffice to debias the neglect types: first, subjects could learn from feedback including profits and losses from trading; secondly, subjects might learn by observing others trade on evidently different beliefs. Alternatively or in addition, the correlation neglect types might be less certain about their beliefs than the rational types and hence engage in only moderate trading. If so, this would imply that despite the heterogeneity in updating types, the marginal (price-setting) trader would be rational. Online Appendix F reports upon experiments in which we embedded our correlated information structure into a standard continuous double auction in which subjects traded financial assets of *ex ante* unknown value. In these experiments, subjects received extensive feedback after each trading round, including the true state from the past period and the resulting profits and losses from trading. In addition, subjects could observe all bids from all market participants, which provided further opportunities to learn. The results show that experimental market interaction does not induce correlation neglect types to learn: in periods

23. Other studies, such as the ones by Corazzini *et al*. (2012), Möbius *et al*. (2013), and Weizsäcker (2010) find belief patterns that are less consistent with correlation neglect.

in which correlation neglect leads to overly optimistic beliefs, market prices in the correlated treatment are too high, and when neglecting correlations implies overpessimism, market prices are too low. In addition, within the correlated market treatment, subjects' propensity to ignore correlations predicts both individual trading behaviour and the degree of price distortions. In sum, correlation neglect also has predictable effects in simple experimental markets.

Acknowledgments. For valuable comments and discussions we are grateful to the editor, Botond Kőszegi, and three referees. We also thank Steffen Altmann, Sandro Ambuehl, Doug Bernheim, Stefano DellaVigna, Thomas Dohmen, Christine Exley, Erik Eyster, Armin Falk, Holger Herz, Alex Imas, Muriel Niederle, Matthew Rabin, Frederik Schwerter, Andrei Shleifer, Joel Sobel, Georg Weizsäcker, and Matthias Wibral. Helpful comments were also received from many seminar and conference audiences. Financial support from the Russell Sage Foundation, the International Foundation for Research in Experimental Economics, the Bonn Graduate School of Economics, and the Center for Economics and Neuroscience Bonn is gratefully acknowledged.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

AKERLOF, G. A. and SHILLER, R. J. (2009), *Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism* (Princeton, NJ: Princeton University Press).

BENJAMIN, D., MOORE, D. and RABIN, M. (2013), "Misconceptions of Chance: Evidence from an Integrated Experiment" (Mimeo).

BOHREN, J. A. (2016), "Informational Herding with Model Misspecification", *Journal of Economic Theory*, **163**, 222–247.

BORDALO, P., COFFMAN, K., GENNAIOLI, N. and SHLEIFER, A. (2016), "Stereotypes", *Quarterly Journal of Economics*, **131**, 1753–1794.

BRANDTS, J., GIRITLIGIL, A. E. and WEBER, R. A. (2015), "An Experimental Study of Persuasion Bias and Social Influence in Networks", *European Economic Review*, **80**, 214–229.

BROCAS, I., CARRILLO, J. D., WANG, S. W., *et al*. (2014), "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games", *Review of Economic Studies*, **81**, 944–970.

BUDESCU, D. V. and RANTILLA, A. K. (2000), "Confidence in Aggregation of Expert Opinion", *Acta Psychologica*, **104**, 371–398.

BUDESCU, D. V. and YU, H.-T. (2007), "Aggregation of Opinion Based on Correlated Cues and Advisors", *Journal of Behavioral Decision Making*, **20**, 153–177.

CAPLIN, A., DEAN, M. and MARTIN, D. (2011), "Search and Satisficing", *American Economic Review*, **101**, 2899–2922.

CHANDRASEKHAR, A. G., LARREGUY, H. and XANDRI, J. P. (2016), "Testing Models of Social Learning on Networks: Evidence from a Framed Field Experiment" (Mimeo).

CHARNESS, G. and LEVIN, D. (2009), "The Origin of the Winner's Curse: A Laboratory Study", *American Economic Journal: Microeconomics*, **1**, 207–236.

CHARNESS, G., KARNI, E. and LEVIN, D. (2010), "On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda", *Games and Economic Behavior*, **68**, 551–556.

CORAZZINI, L., PAVESI, F., PETROVICH, B., *et al*. (2012), "Influential Listeners: An Experiment on Persuasion Bias in Social Networks", *European Economic Review*, **56**, 1276–1288.

DANA, J., WEBER, R. A. and KUANG, J. X. (2007), "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness", *Economic Theory*, **33**, 67–80.

DELLAVIGNA, S. (2009), "Psychology and Economics: Evidence from the Field", *Journal of Economic Literature*, **47**, 315–372.

DEMARZO, P. M., VAYANOS, D. and ZWIEBEL, J. (2003), "Persuasion Bias, Social Influence, and Unidimensional Opinions", *Quarterly Journal of Economics*, **118**, 909–968.

ESPONDA, I. and VESPA, E. (2014), "Hypothetical Thinking and Information Extraction in the Laboratory", *American Economic Journal: Microeconomics*, **6**, 180–202.

EXLEY, C. L. (2015), "Excusing Selfishness in Charitable Giving: The Role of Risk", *Review of Economic Studies*, **83**, 587–628.

EYSTER, E., RABIN, M. and WEIZSÄCKER, G. (2016), "An Experiment on Social Mislearning". Available at SSRN: https://ssrn.com/abstract=2704746

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

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

EYSTER, E. and G. WEIZSÄCKER (2011), "Correlation Neglect in Financial Decision-Making" (DIW Discussion Papers, No. 1104).

FISCHBACHER, U. (2007), “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments”, *Experimental Economics*, **10**, 171–178.

GABAIX, X. (2014), “A Sparsity-Based Model of Bounded Rationality”, *Quarterly Journal of Economics*, **129**, 1661–1710.

GENNAIOLI, N. and SHLEIFER, A. (2010), “What Comes to Mind”, *Quarterly Journal of Economics*, **125**, 1399–1433.

GOLUB, B. and JACKSON, M. O. (2010), “Naïve Learning in Social Networks and the Wisdom of Crowds”, *American Economic Journal: Microeconomics*, **2**, 112–149.

GREINER, B. (2004), “The Online Recruitment System ORSEE 2.0 – A Guide for the Organization of Experiments in Economics”, *University of Cologne, Working Paper Series in Economics*, **10**, 63–104.

GRIMM, V. and MENGEL, F. (2016), “An Experiment on Belief Formation in Networks”. Available at SSRN: https://ssrn.com/abstract=2361007

HAISLEY, E. C. and WEBER, R. A. (2010), “Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior”, *Games and Economic Behavior*, **68**, 614–625.

HANNA, R., MULLAINATHAN, S. and SCHWARTZSTEIN, J. (2014), “Learning through Noticing: Theory and Experimental Evidence in Farming”, *Quarterly Journal of Economics*, **129**, 1311–1353.

HORST, S. (2011), “The Computational Theory of Mind”, *Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/entries/computational-mind/

JIN, G. Z., LUCA, M. and MARTIN, D. (2016), “Is No News (Perceived As) Bad News? An Experimental Investigation of Information Disclosure”.

KAHNEMAN, D. (2011), *Thinking, Fast and Slow* (Basingstoke, United Kingdom: Macmillan).

KAHNEMAN, D. and TVERSKY, A. (1973), “On the Psychology of Prediction”, *Psychological Review*, **80**, 237–251.

KALLIR, I. and SONSINO, D. (2010), “The Neglect of Correlation in Allocation Decisions”, *Southern Economic Journal*, **75**, 1045–1066.

LEVY, G. and RAZIN, R. (2015), “Correlation Neglect, Voting Behavior, and Information Aggregation”, *American Economic Review*, **105**, 1634–1645.

MAINES, L. (1990), “The Effect of Forecast Redundancy on Judgments of a Consensus Forecast’s Expected Accuracy”, *Journal of Accounting Research*, **28**, 29–47.

—— (1996), “An Experimental Examination of Subjective Forecast Combination”, *International Journal of Forecasting*, **12**, 223–233.

MÖBIUS, M. M., PHAN, T. and SZEIDL, A. (2013), “Treasure Hunt: Social Learning in the Field” (Working Paper).

NGANGOUE, K. and WEIZSÄCKER, G. (2015), “Learning from Unrealized Versus Realized Prices” (Working Paper).

ORTOLEVA, P. and SNOWBERG, E. (2015), “Overconfidence in Political Behavior”, *American Economic Review*, **105**, 504–535.

RUBINSTEIN, A. (2007), “Instinctive and Cognitive Reasoning: A Study of Response Times”, *Economic Journal*, **117**, 1243–1259.

—— (2016), “A Typology of Players: Between Instinctive and Contemplative”, *Quarterly Journal of Economics*, **131**, 859–890.

SCHWARTZSTEIN, J. (2014), “Selective Attention and Learning”, *Journal of the European Economic Association*, **12**, 1423–1452.

SELTEN, R. (1998), “Axiomatic Characterization of the Quadratic Scoring Rule”, *Experimental Economics*, **1**, 43–62.

SHILLER, R. J. (2000), *Irrational Exuberance* (Princeton, NJ: Princeton University Press).

SPIEGLER, R. (2016), “Bayesian Networks and Boundedly Rational Expectations”, *Quarterly Journal of Economics*, **131**, 1243–1290.

THAGARD, P. (1996), *Mind: Introduction to Cognitive Science*, Vol. 4 (Cambridge, MA: MIT Press).

WEIZSÄCKER, G. (2010), “Do We Follow Others When We Should? A Simple Test of Rational Expectations”, *American Economic Review*, **100**, 2340–2360.