Belief Formation Under Signal Correlation*

Tanjim Hossain
University of Toronto

Ryo Okui
Seoul National University

August 1, 2019

Abstract

Using a set of incentivized laboratory experiments, we characterize how people form beliefs about a random variable based on independent and correlated signals. First, we theoretically show that, while pure *correlation neglect* always leads to overvaluing of correlated signals, that may not happen if people also exhibit *overprecision*—perceiving signals to be more precise than they actually are. Our experimental results reveal that, while subjects do *overvalue* moderately or strongly correlated signals, they *undervalue* weakly correlated signals, suggesting concurrent presence of correlation neglect and overprecision. Estimated parameters of our model suggest that subjects show a nearly complete level of correlation neglect and also suffer from a high level of overprecision. Additionally, we find that subjects do not fully benefit from wisdom of the crowd—they undervalue aggregated information about others' actions in favor of their private information. This is consistent with models of overprecision where people do not properly incorporate the variance reducing power of averages.

Keywords: Correlated and independent signals, information processing, bounded rationality, correlation neglect, overprecision, belief elicitation, wisdom of the crowd.

JEL classification: C91, D81, D83, D84.

*We thank Andro Rilović, Clarice Zhao, Jiaying (Joey) Yang, and Yue Yu for excellent research assistance. Masaki Aoyagi, Guodong Chen, Zhenxing Huang, Gilat Levy, Takeshi Murooka, Imran Rasul, Eric Set, and Jie Zheng provided us with useful comments. We thank participants at a number of seminars and conference presentations for helpful comments and suggestions. A part of this research was conducted while Okui was at NYU Shanghai, Vrije Universiteit Amsterdam, and Kyoto University. We appreciate the Center for Research in Experimental Economics and Political Decision Making (CREED) at the University of Amsterdam for letting us to use their laboratory. All remaining errors are of course ours. We acknowledge financial support from the Japan Society for the Promotion of Science (JSPS) under KAKENHI Grant Nos. 25780151, 25285067, 15H03329 and 16K03598, a New Faculty Startup Grant from Seoul National University, the Housing and Commercial Bank Economic Research Fund from the Institute of Economic Research at Seoul National University, and SSHRC Grant No. 502502. Please direct all correspondence to Tanjim Hossain at tanjim.hossain@utoronto.ca.

1 Introduction

"In retrospect, a key mistake in the forecast updating that Kremp and I did, was that we ignored the correlation in the partial information from early-voting tallies," admitted statistician Andrew Gelman on his blog regarding his 2016 US presidential election forecasts.[1] When even an academic statistician like Gelman may not take correlation among information sources perfectly into account when aggregating them, it is unlikely that regular people would do so. Processing information or signals about the realization of a random variable from correlated and independent sources is a common, yet complex, problem. In their everyday lives, people frequently navigate through correlated information, which may arise in stock forecasts (Huang, Krishnan, Shon, & Zhou, 2017; Trueman, 1994), news reporting (Au & Kawai, 2012), social networks (Eyster & Rabin, 2014; Lee & Iyengar, 2016), and even in criminal trials (Harkins & Petty, 1987). There is growing evidence that people typically neglect correlation to some extent, which leads to overvaluing of correlated information (Enke & Zimmermann, 2019; Eyster & Weizsacker, 2016).

In this paper, we argue that investigating how people react to correlation is not a trivial task. The main challenge comes from the possibility that people incorrectly evaluate variance. By the nature of correlation, *correlation neglect* cannot be separately studied from *overprecision*, as described in Moore and Healy (2007) and Moore and Healy (2008), where one believes one's information to be more accurate than it actually is. Both correlation neglect and overprecision lead to a reduction in perceived variance of people's posterior belief and the interaction of these two behavioral biases create counterintuitive predictions. For example, correlation neglect may generate an apparent over-appreciation of correlation under some cases in the presence of overprecision, while this will not happen when only correlation neglect is present. Thus, identifying both correlation neglect and overprecision together requires a relatively rich experimental framework.

Using a comprehensive set of laboratory experiments, we investigate how people value signals regarding the realization of a random variable of varying levels of precision and (positive) correlation and process the information she receives to form her beliefs. Specifically, we analyze cases where subjects receive a combination of independent and correlated signals, where the level of correlation among correlated signals vary. We also analyze cases where signals are all independent or all correlated, with differing levels of precision and correlation. To give subjects a context for realization of random variables, we use realized returns from financial assets in our experiments, as described in Section 2. The underlying theoretical framework, nonetheless, is versatile enough to provide insights on information processing at a general level that can be useful for many different applications. Analyzing subjects' belief formation rules under a large set of information generation parameters allows us to identify the impact of correlation neglect and overprecision, something not possible when one only analyzes cases with very high level of correlation, as common in the

[1] November 8, 2016, http://andrewgelman.com/2016/11/08/election-forecasting-updating-error-ignored-correlations-data-thus-producing-illusory-precision-inferences/

existing literature. We find that the subjects *overvalue* moderately or strongly correlated signals, consistent with the findings from the received literature. Our new finding is that subjects *undervalue* signals that are weakly correlated. This finding can be rationalized by assuming that the level of people's correlation neglect depends on the level of correlation. However, using a theoretical model of behavioral bias that incorporates correlation neglect and overprecision, we show that such reversal in correlation neglect is not necessary to explain our findings. In fact, the same level of correlation neglect under all experimental settings can explain undervaluing of weakly correlated signals while overvaluing strongly correlated signals.

In our experiments, we provide subjects with ten signals about a random variable, where the signals are generated following the information structures in Dewan and Myatt (2008) and Myatt and Wallace (2012). We then elicit the mean of each subject's posterior belief using an incentive compatible scoring rule proposed in Hossain and Okui (2013). Each signal is unbiased and comes from one of two groups. Signals within a group are either positively correlated with each other or are independent of each other. Unconditional variance of any signal from a given group is the same, but they are different across groups. Signals from two different groups are independent of each other. We vary the unconditional variances and correlation between signals from the same group across the five treatments in our experiment. Under three treatments, signals from one group are independent while signals from the other are correlated. Depending on the correlation level of the correlated signals, we label these treatments *Strong*, *Moderate*, and *Weak*. Variance of signals from the independent group is the same across these treatments. Moreover, unconditional variance of signals from the correlated group is the same across treatments *Strong* and *Weak*, but the correlation levels are different. Under treatment *Zero*, both groups provide independent signals. The variance of signals from one group equals the variance of independent signals from the above-mentioned three treatments. The variance of signals from the other group equals the unconditional variance of the correlated signals in treatments *Strong* and *Weak*. Finally, in treatment *Both*, both groups of signals have the same unconditional variance and are correlated. The correlation level is strong for one group and weak for the other. Within a treatment, we elicit beliefs about independently drawn random variables across periods. Signal generation processes are unchanged across periods of a given treatment, but the combination of signals from the two groups vary across periods.

In each period, we ask subjects to predict the realized value of that period's random variable based on the private signals they receive. The subjects are fully informed of the signal generating process. We refer to the predicted realization as a subject's *initial prediction*. The optimal initial prediction based on the signals is a weighted average of the average of the signals from each group. The optimal weights depend on the numbers of the signals in each group and the variance and correlation structures of the two groups. Moreover, any weighted average (even with suboptimal weights) of the two averages is an unbiased prediction. From the initial predictions that the subjects report, we estimate the weights they put on the averages from two

groups in calculating their prediction.

While subjects in our experiments choose unbiased predictions, they typically choose weights sub-optimally. The direction of suboptimality, however, varies based on the level of correlation and the relative precision of the independent and correlated signals. Across the five treatments, we consistently find that subjects put sub-optimally *high* weight on strongly or moderately correlated signals, as found in the literature. However, a new finding is that they put sub-optimally *low* weight on weakly correlated signals. This may suggest that the presence of correlation neglect may depend on the correlation level. We also find that, when both groups provide independent signals, subjects put sub-optimally high weight on the less precise signals. This finding suggests that subjects may suffer from overprecision. Our reduced form estimates suggest that people are not "hyper-rational" and their "boundedly rational" behavior follows a specific pattern.

We propose a model of belief formation with both behavioral biases — correlation neglect and overprecision. In this model, subjects incorrectly compute the posterior variance of the average of signals from a given group. Based on this incorrectly calculated variance, they calculate the weights on the two group averages properly. To illustrate how overprecision and correlation neglect interact, we consider five different functional forms of the model. Under all of them, one treats the correlated part of variance for a group average as, to some extent, uncorrelated variance (correlation neglect) and also reduces the total variance of a group average in some form (overprecision). Under some models, overprecision also implies that they under-appreciate the variance reducing power of sample size. We numerically show that, even when people exhibit a high level of correlation neglect, in the presence of overprecision, they may undervalue weakly correlated signals and overvalue strongly correlated signals. Using our data, we estimate the overprecision and correlation neglect parameters for all five models. We find that a formation where overprecision also reduces the variance reducing power of sample averages fits our data the best. The parameters suggest that subjects almost completely neglect correlation and, roughly, treat standard deviation as variance in calculating the variance of their posterior belief.

We also investigate how subjects utilize aggregated information from other subjects using a second exercise. In each period, after they report their initial prediction based on their private signals, we report the average of initial predictions of all subjects in that session for that period's random variable. We refer to this average as the *session average* for that period. Note that, each subject receives a completely different set of private signals, generated using the same signal generation process, about the realization of the same random variable in a given period. Therefore, the session average is more informative than a subject's private signals. Subjects are then asked to submit a *revised prediction* based on their private signals and the publicly known session average. If a subject assumes that all other subjects' predictions are unbiased and have the same variance as her own prediction's variance, it is optimal to ignore her own initial prediction and report the session average (which includes her own initial prediction) as the revised prediction.

We find that subjects do take their initial prediction into account when calculating their revised prediction. Moreover, the weights they put on their own initial prediction are quite similar across treatments. Under all treatments, they put between 19%-26% weight on their initial prediction and around 74%-81% weight on the session average. This result can easily be explained by the behavioral model that fits our data on initial prediction the best. In that model, overprecision leads people to under-appreciate the variance reducing power of sample average by not incorporating the sample size properly in calculating variance of a sample average. Suppose subjects make the same mistake in calculating the variance of the session average. Using the parameter calculated with the initial prediction data, we can explain the weights used in revised prediction quite well.[2]

In the following subsection, we discuss how our work relates to the extant literature. Section 2 explains the setting of our experiments. Section 3 presents a theoretical framework based on our experimental design. Moreover, we discuss behavioral models of belief formation in this section. We examine the results of the experiments in Section 4. Section 5 concludes the paper. Proofs of the mathematical propositions are included in the Appendix.

Relation to the Literature: Correlation neglect has gained attention from researchers recently. Perhaps the most closely related paper is Enke and Zimmermann (2019), who demonstrate correlation neglect when correlation between signals arises due to repetition of some information. They suggest that correlation neglect may be remedied by making subjects notice the presence of correlation. We create a richer experimental setting that allows us to explore the extent of correlation neglect more comprehensively. We argue that we need to incorporate the effect of overprecision in order to study correlation neglect properly. While we find subjects overvalue strongly or moderately correlated signals, which is consistent with their findings, we also find that subjects undervalue weakly correlated signals. Considering overprecision in addition to correlation neglect is needed to explain this apparent reversal of how people treat correlation.

Our theoretical framework and experimental findings suggest that it is difficult to distinguish correlation neglect from overprecision under strong correlation, which is a common setting in the literature. Under overprecision, subjects under-appreciate the uncertainty attached to their beliefs. That is, they perceive the variance attached to their beliefs to be smaller than they actually are. Overprecision can lead to overconfidence regarding the precision level of one's beliefs. Moore and Schatz (2017) provide a comprehensive review of the psychology and management literature on this well-documented behavioral bias. More recently, overprecision has also been discussed in economics literature. For example, see Grubb (2015) and Daniel and Hirshleifer (2015).

Maines (1996) documents correlation neglect in an experimental setting where, similar to ours, subjects

[2]Our finding is consistent with that of Nöth and Weber (2003) and others who find that people underweight public information in favor of their private information and our behavioral model provides a mechanism behind their overconfidence interpretation.

evaluate forecasts by financial analysts. Luan, Sorkin, and Itzkowitz (2004) study how people process information using a small scale experiment and find that participants overvalue signals with high accuracy and high correlation. In psychology, Soll (1999) explains correlation neglect or even a preference for redundant information using different intuitive theories of information. In the context of financial markets, Eyster and Weizsacker (2016) provide subjects with multiple assets where the returns from some of them are independent and the rest are correlated and investigate their portfolio choice. They find evidence that subjects treat correlated variables as uncorrelated and give equal weights to all assets. We provide more generalizable insights regarding information processing by directly analyzing belief formation rules. We discuss the relation to that paper in more detail in Section 4.1.4. Kroll, Levy, and Rapoport (1988) and Kallir and Sansino (2009) also investigate the role of correlation in asset allocation.

There have been several attempts to characterize choices with misperception of correlation under a decision theoretic framework. Ellis and Piccione (2017) provide an axiomatic framework to represent asset choice under such settings. Levy and Razin (2018) consider situations in which correlation structure among signals is ambiguous. They show that correlation neglect is observed if each signal is precise and otherwise, people behave as if correlation were high. Their theory is not directly applicable to our experimental setting as the correlation structure is very clearly described in our experiments. Moreover, that paper's argument may be interpreted as that overprecision creates apparent correlation neglect. We argue that combining correlation neglect and overprecision, rather than considering that one implies the other, is necessary to explain our experimental findings.

There are papers which consider both correlation neglect and overprecision but they do not consider them as separate behavioral biases. Levy and Razin (2015) analyze the case when voters ignore correlation in signals about the state of the world. Ortoleva and Snowberg (2015) show that correlation neglect may give rise to overprecision, which can be connected to ideological extremeness in politics. These two papers use correlation neglect as a device to generate overprecision. We argue that we treat them as separate behavioral biases and it is important to analyze the effects of interactions between two biases.

2 Experimental Design

We ran a set of laboratory experiments to empirically investigate how people incorporate independent and correlated signals in forming their beliefs about a random variable. Subjects received information about the earning per share (EPS) from a different fictitious stock in each period of a session. We elicited their beliefs about the EPS for each stock using the binarized scoring rule (BSR) proposed in Hossain and Okui (2013), which is incentive compatible independently of a subject's risk preference. Elicited beliefs allow us to directly investigate how people process information that differ in precision and correlation structure. An experimental session consisted of 50 periods. The true EPS was independently drawn in each period from

a normal distribution with mean 500 and variance 25,000.[3] In each period, subjects received 10 forecasts about the EPS of a stock. For the rest of the paper, we will use forecasts and signals interchangeably. Below we discuss how each signal is generated.

Suppose a subject receives n_l signals from group $l \in \{A, B\}$, where the total number of signals is N, which equals 10 in our experiment. A signal equals the true EPS, T, plus two error terms. Specifically, signal $j \in \{1, 2, \ldots, n_{l-1}, n_l\}$ from group l is denoted by X_j^l where $X_j^l = T + \epsilon_{lIj} + \epsilon_{lC}$. Subjects observe only the signal X_j^l and not individual components (T, ϵ_{lIj}, and ϵ_{lC}) within the signal. Here, ϵ_{lIj} is an independently drawn error term that is different for each signal j from group l and ϵ_{lC} is an error term that is common to all the signals from group l. Note that errors ϵ_{lIj} and ϵ_{lC} are drawn from normal distributions with mean 0 and variances of σ_{lI}^2 and σ_{lC}^2, respectively. Thus, all signals are unbiased. If $\sigma_{lC}^2 > 0$, then all signals from group l, conditional on T, are correlated with each other. On the other hand, if $\sigma_{lC}^2 = 0$, then all signals from group l, conditional on T, are independent of each other. By varying the variances σ_{AI}^2, σ_{AC}^2, σ_{BI}^2, and σ_{BC}^2, we can vary the precision and correlation levels of the two groups of signals. We refer to these variances and the number of signals from the two groups as the *information generation parameter set*—$(\sigma_{AI}^2, \sigma_{AC}^2, \sigma_{BI}^2, \sigma_{BC}^2, n_A, n_B)$.

Subjects were completely aware of this signal generation process, including the precision level and correlation structure, used in their session. Given the parameters chosen, we had no case where the true EPS or any of the generated forecasts were negative. All subjects within a session received information about the same stock in a given period. However, for each subject, we generated a completely different set of forecasts based on the above process. We clearly informed them that information received in one period was not informative about the realized values of the random variables in the other periods. For each period, we showed subjects the 10 signals and the averages of group A, group B, and all signals.

A *treatment* of our experiment is represented by the set of variance parameters, $(\sigma_{AI}^2, \sigma_{AC}^2, \sigma_{BI}^2, \sigma_{BC}^2)$. Table 1 presents the five treatments that we used. We ran two sessions of each treatment. Our main goal is to analyze how subjects value correlated signals versus independent signals. In four of the treatments, signals from one group, say group A, conditional on the true EPS, are independently drawn with a variance of 500 ($\sigma_{AI}^2 = 500, \sigma_{AC}^2 = 0$). Signals from the other group, conditional on the true EPS, are potentially correlated with each other. We can divide the treatments into two sets of treatments. In the first set of three treatments, the total variance of a group B signals equal 265. That is, $\sigma_{BI}^2 + \sigma_{BC}^2 = 265 < 500 = \sigma_{AI}^2$. However, the level of correlation varies across the three treatments. In *Strong*, the correlated component of variance for group B signals was more than 94% of the total variance ($\sigma_{BI}^2 = 15, \sigma_{BC}^2 = 250$). Under *Weak*, σ_{BC}^2 equaled 15 and under *Zero,* σ_{BC}^2 equaled 0. That is, even group B signals are independent under the treatment *Zero.* Comparing the three treatments, we can investigate how subjects' belief formation rule changes as the level

[3]This implies that the true EPS if a stock will fall within $500 \pm 2.58 \times \sqrt{25000} \approx (92, 908)$ with likelihood greater than 99%.

Table 1: Set of Parameters under the Five Treatments

Treatment	σ^2_{AI}	σ^2_{AC}	σ^2_{BI}	σ^2_{BC}	Characterization
Strong	500	0	15	250	Group A independent and group B strongly correlated
Weak	500	0	250	15	Group A independent and group B weakly correlated
Zero	500	0	265	0	Both Groups A and B are independent
Moderate	500	0	250	250	Group A independent and group B moderately correlated
Both	250	15	15	250	Group A weakly correlated and group B strongly correlated

of correlation of group B signals changes. In the last two treatments, we choose the variances in a way that the total variance of a group A signal equaled the total variance of a group B signal, $\sigma^2_{AI}+\sigma^2_{AC} = \sigma^2_{BI}+\sigma^2_{BC}$, but the composition of the total variance was different across groups. In one of them, group A signals are independent and group B signals correlated with $\sigma^2_{AI} = 500, \sigma^2_{AC} = 0, \sigma^2_{BI} = 250$, and $\sigma^2_{BC} = 250$. As the correlation ratio of group B signals $(\sigma^2_{BC}/(\sigma^2_{BI} + \sigma^2_{BC}))$ is in between those of the treatments *Strong* and *Weak*, we label this treatment *Moderate*. Under the remaining treatment, treatment *Both*, both groups provide correlated signals, but the levels of correlation are different. Specifically, group A signals are weakly correlated $(\sigma^2_{AI} = 250, \sigma^2_{AC} = 15)$ and group B signals are strongly correlated $(\sigma^2_{BI} = 15, \sigma^2_{BC} = 250)$.[4] Table 1 summarizes the five treatments. Note that, under all treatments, the error terms of group A signals are independent of the error terms of group B signals.

In a given period, we asked a subject to report her belief about the realization of the EPS of the stock of that period. First, we asked her to predict the EPS based on the 10 forecasts she received using the following incentive scheme based on BSR: Let us denote this *initial prediction* by P. If the square of the deviation of this initial prediction from the true EPS, $(P - T)^2$, was below or equal to a random number K, generated from a uniform distribution on $[0, \bar{K}]$, she earned 100 points.[5] If $(P - T)^2$ was above K, she earned zero points. We informed the subjects that it is optimal for them to report the mean of their posterior belief about T based on the 10 forecasts. We did not tell them how to optimally construct their posterior belief, which is given in Proposition 1, as learning how subjects actually form their beliefs is the main objective of this experiment.

After entering her initial prediction P, each subject was informed of the *session average* for that period, i.e., the average of the predictions for that EPS by all subjects (including herself) in that session. Then we asked them to enter a *revised prediction*, denoted by P^r, for the same EPS. We again incentivized truthful revelation of belief using the BSR with a different random number K^r, generated from a uniform

[4] In four of our treatments, at least one group provided only independent signals. To maintain consistency across different treatments and to present signals from the two groups in a similar manner, we always presented the signals to subjects as the sum of the true EPS and two error terms. For correlated signals, we mentioned that one error term was common across all the signals from that group and the other one was independently drawn. For independent signals, we mentioned that both error terms for a signal were independently drawn (the sum of the variances of the two error terms equaled σ^2_{II}).

[5] We chose the values of $\bar{K}$ to equalize simulated earnings from hypothetical, yet plausible, suboptimal strategies across treatments. The values were 40, 30, 40, 40 and 50 for *Strong*, *Weak*, *Zero*, *Moderate*, and *Both* treatments, respectively.

distribution on $[0, \bar{K^r}]$.[6] The payoff from a revised prediction is 50 or zero points. After the subjects reported their initial and revised predictions, we reported the true EPS for that period, realizations of the relevant random numbers (K and K^r) and their income in points to the subjects. The random numbers K and K^r were redrawn in each period.

We divided a session into five 10-period long blocks. In the first session for each specification, the number of forecasts from group A analysts increased, with an increment of 2, from 1 to 9 in the five blocks of 10 periods each. In the second session, the number of forecasts from group A analysts decreased, with a deduction of 2, from 9 to 1 in these blocks. After the session was completed, we randomly chose one period from each of these five 10-period blocks, to determine subjects' incomes from the session.[7]

The sessions were run between December 2015 and September 2016 at CREED at the University of Amsterdam and the University of Toronto. The number of subjects in a session varied between 12 and 22. In total, we had 172 subjects.[8] The experiments were programmed and conducted with the software *z-Tree* developed by Fischbacher (2007). Instructions were provided in English.[9] Subjects were sent a pdf document describing statistical concepts such as mean, variance, and uniform and normal distributions, that are relevant for the experiment, after they signed up for a session. They were asked to ensure that they understood those concepts before coming to the session. They were given a copy of the document during the experiment. Moreover, we provided a brief statistical comprehension quiz and then provided the answers to the quiz to the subjects. Each subject had access to a calculator throughout her session. A sample experimental instruction and the document presenting definitions of statistical concepts are provided in Appendix A.4.

Our experiment is designed to provide insights on how people form their beliefs when they may simultaneously misperceive correlation and uncertainty associated with the information they receive. As a result, we need an informational setting that is rich enough to tease apart both biases. Specifically, our purpose requires a rich setting that would allow us to simultaneously vary signal mean and variance and correlation across signals. This may not be achieved with very simple setups. For example, when the random variable is a binomial event, the mean and variance cannot be separately manipulated. On the other hand, incorporating varying levels of correlation across signals while keeping variances fixed in a setting similar to that in Enke and Zimmermann (2019) would actually lead to a relatively complicated signal generating process. Normal distributions are typically familiar to university students, meets the above criteria, and are

[6] We chose the values of $\bar{K^r}$ to equalize simulated earnings across treatments. The values were 8, 6, 8, 8 and 10 for *Strong*, *Weak*, *Zero*, *Moderate*, and *Both* treatments, respectively.

[7] We also included 10 additional periods where subjects could choose the combination of the two groups of signals (i.e., they chose n_A) in every period. Given the focus of the current paper, we do not discuss subjects' choices of n_A in this paper.

[8] Four subjects left during the middle of the session for unexpected reasons. We do not include data from those subjects in our analysis.

[9] Most of the subjects at CREED are students of the University of Amsterdam. Both English and Dutch are used as media of instruction at the University of Amsterdam. In general, their English level is very high.

considered to be commonly observed in many real-life situations.[10]

We also note that the normal distribution has been used to model information circumstance people face in economics literature in both theoretical and empirical settings. The exact informational setup with correlated signals used in this paper has been successfully utilized in economics in Myatt and Wallace (2012) and subsequent papers. In empirical models of learning by individuals, signals being normally distributed around the true value is a common assumption starting from Erdem and Keane (1996). Since then, such models have been used in many papers including Ackerberg (2003), and Crawford and Shum (2005). Ching, Erdem, and Keane (2013) provide a nice summary of the literature. In structural estimations of common-value auction models, bidders are assumed to calculate their optimal bidding strategy where signals follow quite complicated distributions (e.g. Bajari and Hortaçsu (2003) use eBay auction data where auction participants are ordinary people). Thus, our experimental setup is no more mathematically complex than many empirical models.

3 Theoretical Framework

This section discusses the underlying theoretical framework for our experiment. First, we present the optimal posterior belief formation rule under full rationality for initial prediction. We then discuss boundedly rational or behavioral models of belief formation where people are Bayesian updaters, but calculate their posterior distribution incorrectly. We numerically illustrate how posterior beliefs under these models may differ from beliefs under rationality. Then, we discuss belief formation rules for revised predictions.

Based on the experimental design presented above, Setting 1 below summarizes the signal generation framework for our experiments.

Setting 1. *For a given information generation parameter set* $(\sigma^2_{AI}, \sigma^2_{AC}, \sigma^2_{BI}, \sigma^2_{BC}, n_A, n_B)$*, signals* X^l_j *for* $j = 1, \ldots, n_l$ *and* $l \in \{A, B\}$ *are generated by* $X^l_j = T + \epsilon_{lIj} + \epsilon_{lC}$*, where* $\epsilon_{lIj} \sim N(0, \sigma^2_{lI})$ *and* $\epsilon_{lC} \sim N(0, \sigma^2_{lC})$ *and* ϵ_{lIj} *and* ϵ_{lC} *are mutually independent.*

Based on receiving n_l signals from group $l \in \{A, B\}$ about the realization of T, X^l_j for $j \in \{1, \ldots, n_l\}$, the mean of group l signals, $\bar{X}_l = \sum_{j=1}^{n_l} X^l_j / n_l$, is distributed normally with mean T and variance $V_l = \sigma^2_{lI}/n_l + \sigma^2_{lC}$. Moreover, the most efficient estimator of T, which corresponds to the posterior mean because we let subjects minimize the quadratic loss, is a weighted average of $\bar{X}_A$ and $\bar{X}_B$.

Proposition 1. *Suppose that the prior of* T *is* $T \sim N(\mu_p, \sigma^2_p)$*. The mean of the limit of the posterior*

[10] Batanero, Tauber, and Sánchez (2004) found that students who were taught the properties of a normal distribution showed a good understanding of them.

distribution of T under $\sigma_p^2 \to \infty$ (i.e., when the prior tends to uninformative one) is

$$\frac{V_B}{V_A + V_B}\bar{X}_A + \frac{V_A}{V_A + V_B}\bar{X}_B.$$

The proof is in the Appendix, and it is a standard Bayesian posterior calculation. The optimal point estimate for the realized value of T is based on the averages of signals from each group. It is unbiased in the sense that the expected value given T is T because the weights sum to 1. All signals from a given group, including those whose realized values happen to be extreme, receive the same weight in the estimate. Moreover, for a given information generation parameter set, the optimal weights are independent of the values of the observed signals. Given our incentive compatible scoring rule, a subject should report this mean as her initial prediction.

We now propose behavioral models of belief formation where subjects incorrectly compute posterior variance V_l based on the signals they receive and then calculate the weights on $\bar{X}_l$ in a Bayesian manner based on the incorrect variances. First, we consider the possibility that subjects incorrectly incorporate the covariance between two signals from group l in calculating V_l. Specifically, similar to Ortoleva and Snowberg (2015), we assume that they incorrectly believe that the covariance between two group l signals is $d\sigma_{lC}^2$ for some $d \in [0, 1)$ while keeping the total variance unchanged. In other words, subjects wrongly allocate a certain portion of the common noise to independent noises. Then, subjects will exhibit *correlation neglect* with complete correlation neglect if $d = 0$. Moreover, the variance of $\bar{X}_l$ can be expressed by $(\sigma_{lI}^2 + (1-\gamma)\,\sigma_{lC}^2)/n_l + \gamma\sigma_{lC}^2$ for some $\gamma \in [0, 1)$. One can easily show that, if signals from one group are independent and signals from the other are correlated, a subject who neglects correlation (i.e., $\gamma < 1$) will *always put higher weight on the correlated signals* than what Proposition 1 suggests. However, she will calculate the variances of $\bar{X}_l$ and the weights perfectly if both groups provide independent signals.

Now we consider the possibility that subjects also suffer from misperception of precision; that is, they do not take variance into account correctly. To illustrate how precision misperception may affect belief formation, we will focus on the concept of *overprecision* where subjects perceive the variances to be smaller than they actually are. Formally, we will assume that subjects take various components of the variance of $\bar{X}_l$ to the power of $\rho < 1$. As there does not seem to be an established model of overprecision in the presence of multiple signals, we consider five different functional forms of potentially mistaken beliefs about V_l, which we denote by $\tilde{V}_l$, that combine correlation neglect and overprecision. We will refer to these formulations as Models 1 to 5 in the remainder of the paper.

$$\text{Model 1:} \quad \tilde{V}_l = \frac{\left(\sigma_{lI}^2\right)^\rho + (1-\gamma)\left(\sigma_{lC}^2\right)^\rho}{n_l} + \gamma\left(\sigma_{lC}^2\right)^\rho$$

$$\text{Model 2:} \quad \tilde{V}_l = \frac{(\sigma_{lI}^2 + (1-\gamma)\sigma_{lC}^2)^\rho}{n_l} + (\gamma\sigma_{lC}^2)^\rho$$

$$\text{Model 3:} \quad \tilde{V}_l = \left(\frac{\sigma_{lI}^2 + (1-\gamma)\sigma_{lC}^2}{n_l} + \gamma\sigma_{lC}^2\right)^\rho$$

$$\text{Model 4:} \quad \tilde{V}_l = \frac{\left(\sigma_{lI}^2 + (1-\gamma)\sigma_{lC}^2\right)^\rho}{(n_l)^\rho} + (\gamma\sigma_{lC}^2)^\rho$$

$$\text{Model 5:} \quad \tilde{V}_l = \frac{\left(\sigma_{lI}^2\right)^\rho + (1-\gamma)\left(\sigma_{lC}^2\right)^\rho}{(n_l)^\rho} + \gamma\left(\sigma_{lC}^2\right)^\rho$$

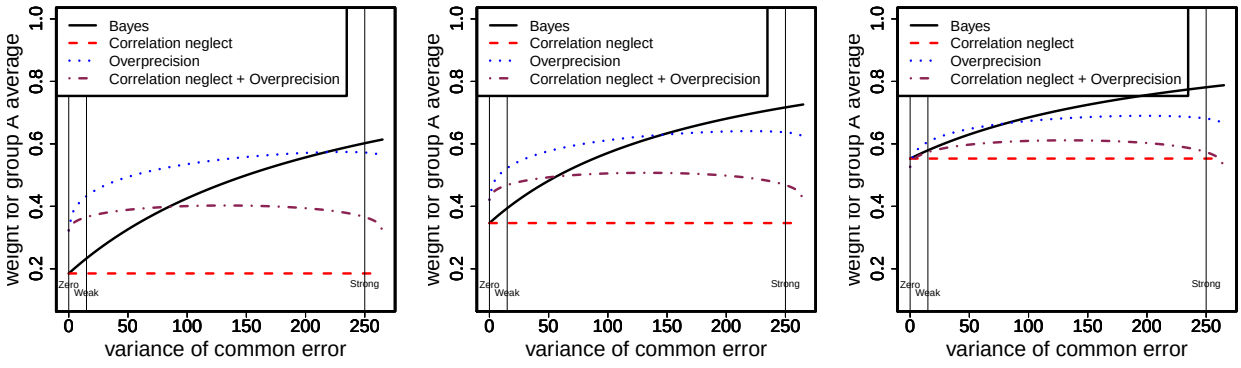
Note that in Models 1 and 2, parts of the (incorrectly calculated) variance for group l signal average is inversely proportional to the number of group l signals. In Models 3 to 5, on the other hand, we assume that this part of the variance is inversely proportional to $(n_l)^\rho$ instead. In this respect, Models 5 and 4 correspond to Models 1 and 2, respectively. If ρ is smaller than 1, the number of signals will have a smaller impact in reducing the variance under Models 3 to 5 than under Models 1 and 2.

Different values of parameters correspond to different behavioral types. For fully rational subjects, the values of the parameters are $\gamma = 1$ and $\rho = 1$. When a subject neglects correlation, she has $\gamma < 1$ and when she suffers from overprecision, she has $\rho < 1$. If a subject suffers from only correlation neglect (i.e., $\rho = 1$ but $\gamma < 1$), all five models are identical. If a subject suffers only from overprecision (i.e., $\gamma = 1$ but $\rho < 1$), the first two models are identical and the last two models are identical. We will use our experimental data to estimate the parameters under each model and find which one fits the data best. Of course, one can consider other variants of the model. Nonetheless, we will show that all our results can be well explained by some of the above models in both qualitative and quantitative senses.

Under Models 1 to 5, correlation neglect with correct perception of precision ($\rho = 1$ and $\gamma < 1$) would lead to overweighting of correlated signals in the treatments where only one group provides correlated signals. On the other hand, as the group A signals have higher overall variances and typically provide independent signals, overprecision will lead to underweighting of correlated signals. Thus, the overall impact of both correlation neglect and overprecision on weights put on the signals from the two groups will depend on the specific information generation parameter set. Note that, in treatment *Zero*, both groups provide independent signals. Then, correlation neglect has no impact on belief formation. Overprecision ($\rho < 1$) will lead to overweighting of group A signals, which have higher variance, while $\rho > 1$ will lead to underweighting of group A signals.

Figure 1 illustrates the impact of correlation neglect and overprecision on belief formation under varying degrees of correlation with examples based on Model 5. We set $\sigma_{AI}^2 = 500$, $\sigma_{AC}^2 = 0$, $\sigma_{BI}^2 + \sigma_{BC}^2 = 265$, and $N = 10$ as in treatments *Zero*, *Weak*, and *Strong*. The figure plots the relationship between σ_{BC}^2 and the weight on $\bar{X}_A$ under various behavioral assumptions. The solid black curves correspond to Bayesian beliefs ($\gamma = 1, \rho = 1$). Weight on $\bar{X}_A$ is increasing in σ_{BC}^2. The red dashed lines correspond to complete correlation

Figure 1: The weights for independent signals under various behavioral assumptions



(a) $n_A = 3$ **(b)** $n_A = 5$ **(c)** $n_A = 7$

neglect without any precision misperception ($\gamma = 0, \rho = 1$). It is flat because the degree of correlation does not affect one's perception about the usefulness of group B signals. The blue dotted curves correspond to overprecision and we use ($\gamma = 1, \rho = 0.5$). It is increasing in σ^2_{BC} as in the case of Bayesian. When the correlation is weak, it overweights independent signals with larger variance, but when the correlation is strong, it underweights them. Lastly, the violet dot-dashed curves correspond to a combination of correlation neglect and overprecision ($\gamma = 0, \rho = 0.5$). Correlated signals are overvalued when the correlation is weak, but are undervalued in the presence of strong correlation. Thus, even if subjects exhibited complete correlation neglects, they would behave as if they overvalue correlated signals if they also suffer from overprecision.

The figure also demonstrates the importance of examining settings with various degrees of correlation to identify the contribution of each of correlation neglect and overprecision. For example, when the correlation among group B signals is strong, both correlation neglect and overprecision predict a lower weight on $\bar{X}_A$ than that predicted under full rationality. We cannot separately identify whether a lower weight on independent signals comes from correlation neglect or overprecision. On the other hand, if we look only at cases with weak correlation, it does not allow us to separate a overprecision story and a story based on overreaction to correlation.[11] Combining results of experiments with varying levels of correlation makes it possible to separately identify the extent of correlation neglect and overprecision and the magnitudes of the effects of these two behavioral biases.

Finally, we analyze posterior beliefs regarding T after a subject learns the session average of initial predictions. Let $M \geq 2$ denote the number of subjects in a session. In a given period, each of these M subjects observe a completely different set of N signals about the same realized value, where the signals

[11] While we do not discuss overreaction to correlation or possibility of treating independent signals as correlated in detail in this paper, such a phenomenon is observed in the literature on gambler's fallacy and the law of small numbers. See, for example, Rabin (2002).

are generated using the same information generation parameter set. Optimal estimate of the realization of T with this public information depends on a subject's beliefs about the information content of the session average. Suppose a subject believes that the session average is unbiased and normally distributed with variance $\lambda\sigma^2$ where σ^2 is the variance of her own initial prediction. The following proposition presents a subject's belief formation rule for the revised prediction as a function of λ.

Proposition 2. *Suppose that a subject's initial prediction is unbiased, her beliefs follow a normal distribution and she believes that the session average is unbiased and follows a normal distribution with variance $\lambda\sigma^2$ where σ^2 is the variance of her own initial prediction. Then, the mean of her posterior belief equals $\frac{M(M-1)}{(M-1)^2+M^2\lambda-1}\bar{p} + \frac{M^2\lambda-M}{(M-1)^2+M^2\lambda-1}p^*$, where M is the total number of subjects, $\bar{p}$ is the session average and p^* is her own initial prediction.*

If a subject is Bayesian and believes that all subjects (including herself) provide unbiased and equally precise initial prediction, then λ would equal $1/M$. In that case, the subject's revised prediction should equal the session average. However, she will put a positive weight on her own initial prediction if she believes that $\lambda > 1/M$. Moreover, the weight on her own prediction would be increasing in λ. A large λ may come from overprecision. In particular, if people misunderstand the variance reducing power of taking an average as considered in Models 3-5, that will imply that λ will decrease slower than under Bayesian prediction as M increases. The parameter λ can be estimated using subjects' revised prediction and can be used to infer whether models designed to analyze initial predictions can have any extrapolation power.

In the following section, we will analyze our experimental data based on the theoretical framework provided in this section.

4 Results

The experimental sessions have provided us with a large data set of people's beliefs about the realized EPS under a comprehensive set of information generation parameters. Our main goal is to study how people incorporate independent and correlated signals in forming their beliefs, specifically the posterior mean based on the signals. Moreover, we also study how they incorporate additional information based on the actions of others.

The number of group A forecasts in the five 10-period blocks is increasing over time in half of the sessions and decreasing in the other half of the sessions. We find no effect of this difference in ordering.[12] Hence, we pool the results from the increasing and decreasing order sessions in our empirical analysis.

[12] Results are available from the authors upon request.

4.1 Initial Prediction with Only Private Information

We analyze data from initial predictions to examine how subjects respond to different signal structures in forming their beliefs. Specifically, we estimate the weights subjects put on the averages of the two groups of signals in their initial predictions using a linear regression model. Our main finding is that while initial predictions are unbiased, subject put sub-optimally high weights on strongly correlated forecasts and sub-optimally low weights on weakly correlated forecasts. This finding is consistent with subjects exhibiting substantial degrees of both correlation neglect and overprecision.

4.1.1 Econometric Specification and Some Hypotheses

As we have five different treatments with varying levels of correlation among signals, we analyze them separately when investigating how subjects process their private signals to form an initial prediction about the EPS. For each treatment, we elicit posterior means under a single set of variance parameters, but with five different combinations of the number of signals from each group. In a given period, the signal combination was the same across all subjects within the session. For each subject, we have 10 independent observations under the same information generation parameter set $(\sigma^2_{AI}, \sigma^2_{AC}, \sigma^2_{BI}, \sigma^2_{BC}, n_A, 10 - n_A)$.[13] An initial examination of the data indicates that initial predictions can be modeled to be linear in the means of group A and group B signals, $\bar{X}_A$ and $\bar{X}_B$, respectively. Specifically, neither nonlinear specifications, such as quadratic terms, nor extreme values, such as maximum or minimum forecasts, systematically determine initial predictions. We thus focus on specifications that are linear in $\bar{X}_A$ and $\bar{X}_B$.

For each information generation parameter set, we estimate the weights on $\bar{X}_A$ and $\bar{X}_B$ using the following equation:

$$(P_{it} - T_t) = \beta_0 + \beta_1(\bar{X}_{A,it} - T_t) + \beta_2(\bar{X}_{B,it} - T_t) + u_{it},$$

where, at period t, T_t is the true value of EPS, and for subject i, P_{it} is the initial prediction, $\bar{X}_{A,it}$ and $\bar{X}_{B,it}$ are the means of groups A and B signals, respectively, and u_{it} is the error term. We use deviations from the EPS to estimate the weights. In our experiment, the EPSs are generated from a very diverse distribution and most of the variations in predictions and forecasts come from variations in the EPS. Using values which are not centered around the EPS thus yields a very high coefficient of determination, which might make results difficult to interpret. Using deviations is a solution to this problem.[14] Note that centering the variables around the EPS would not change the results if predictions are unbiased.[15]

[13] We exclude observations where the initial prediction is above the maximum or below the minimum of the 10 forecasts a subject received as these observations are hardly justified by any economic theory and are likely to be typos or results of simple mistakes. These exclusions do not affect our results qualitatively.

[14] An alternative solution may be to include period fixed effects. Although not presented here, the model with uncentered variables but with period fixed effects generates similar results.

[15] Indeed, estimating the model with uncentered variables yields similar results except for the coefficients of determination

We now introduce a number of hypotheses regarding subjects' belief formation rules. For some null hypotheses, we present alternative hypotheses based on the five behavioral belief formation rules presented in Section 3, under reasonable parameter values. The null and alternative hypotheses relate to comparisons between the values of the coefficients, $(\beta_0, \beta_1, \beta_2)$ and optimal weights on groups A and B averages from Proposition 1, w_A^0 and w_B^0.

Hypothesis 1 (Unbiased prediction): An initial prediction being unbiased corresponds to $H_0 : \beta_0 = 0$ and $\beta_1 + \beta_2 = 1$. Unbiasedness is fundamental in investigating the validity of other hypotheses.

Hypothesis 2 (Bayesian prediction): Optimal initial predictions produced by Bayesian correspond to $H_0 : \beta_0 = 0$, $\beta_1 = w_A^0$ and $\beta_2 = w_B^0$.

There might be several ways in which Hypothesis 2 is violated and, assuming the unbiasedness, each violation has a specific interpretation. In particular, we consider the following alternatives.

H_1 **(Overprecision without correlation neglect):** In treatments *Strong*, *Moderate*, *Weak*, and *Zero*, $\beta_1 > w_A^0$ and $\beta_2 < w_B^0$. In treatment *Both*, $\beta_1 < w_A^0$ and $\beta_2 > w_B^0$. Note that the prediction for treatment *Zero* will hold even in the presence of correlation neglect.

While overprecision can be detected, irrespective of presence or absence of correlation neglect, by examining treatment *Zero*, how correlation neglect affects the results depends on the presence of overprecision.

H_1 **(Correlation neglect without overprecision):** In treatments *Strong*, *Moderate*, *Weak*, and *Both*, $\beta_1 < w_A^0$ and $\beta_2 > w_B^0$ and in treatment *Zero*, $\beta_1 = w_A^0$ and $\beta_2 = w_B^0$.

H_1 **(Correlation neglect with overprecision):** $\beta_1 < w_A^0$ and $\beta_2 > w_B^0$ in treatment *Strong*, but $\beta_1 > w_A^0$ and $\beta_2 < w_B^0$ in treatment *Weak*.

Lastly, we examine whether subjects' elicited posterior means are basically the simple average of the 10 signals they receive. This may result from completely ignoring variances. This hypothesis is related to the $1/N$ heuristic discussed in the context of portfolio choice by Benartzi and Thaler (2001) and Eyster and Weizsacker (2016).

Hypothesis 3 ($1/N$ heuristic): $w_l^0 = n_l/10$ for $l = A, B$.

Table 2: Initial Prediction, Treatment *Strong*

Dependent variable: *Initial Prediction*					
# of group A analysts	1	3	5	7	9
$\bar{X}_A$	0.284***	0.419***	0.547***	0.662***	0.786***
	(0.036)	(0.046)	(0.048)	(0.048)	(0.056)
$\bar{X}_B$	0.758***	0.502***	0.491***	0.386***	0.163***
	(0.031)	(0.047)	(0.030)	(0.046)	(0.037)
Constant	-0.744*	-0.281	0.562	0.212	0.945**
	(0.306)	(0.399)	(0.367)	(0.289)	(0.337)
Optimal weight on $\bar{X}_A$	0.355	0.602	0.717	0.781	0.827
Optimal weight on $\bar{X}_B$	0.665	0.398	0.283	0.219	0.173
F-stats:					
H_0: Unbiased	4.455	3.388	1.248	1.291	3.967
prediction	(0.019)	(0.046)	(0.300)	(0.289)	(0.029)
H_0: Bayesian	4.824	8.557	24.758	6.613	0.459
prediction	(0.015)	(0.001)	(0.000)	(0.004)	(0.636)
H_0: $1/N$	13.282	9.332	0.480	2.054	2.486
heuristic	(0.000)	(0.001)	(0.623)	(0.144)	(0.099)
$\bar{R}^2$	0.787	0.637	0.720	0.669	0.510
# of observations	309	339	332	336	333

Note: Estimated by OLS. Standard errors clustered at subject level are presented below the coefficients (within parentheses) and p-values are presented below the F statistics (within parentheses).

4.1.2 Detailed Discussion of Results for Treatments *Strong*, *Weak*, and *Zero*

Tables 2 to 4 present the estimates for treatments *Strong*, *Weak*, and *Zero*, respectively.[16] Signals across these three treatments have the same unconditional variances (500 for group A and 265 for group B). We also present a number of F-tests regarding the weights. Specifically, we test whether initial predictions can be considered unbiased (Hypothesis 1), whether the weights are optimal in the sense of Proposition 1 (Hypothesis 2), and whether the subjects follow the $1/N$ heuristic (Hypothesis 3).

The tables provide a comprehensive picture of how the subjects chose predictions using their private signals, which will be discussed below. First, we highlight one particular result: As seen in the literature, we find that subjects do not perfectly account for correlation. However, our result is more nuanced than the common finding that people overvalue correlated signals in their decision process. When signal correlation is moderate or strong, subjects put greater weight on correlated signals than optimal. On the other hand, they put sub-optimally *low* weights on the weakly correlated signals. In other words, subjects seem to over-compensate for correlation when correlation is weak. When both sets of signals are independent, subjects

and does not alter our conclusions qualitatively. These regression results are available in an online appendix at https://drive.google.com/open?id=1p3Y1nX95_JE_A35O24p2BxfiuyaHn-6p.

[16] Fixed effects estimation leads to the same results, qualitatively. They are available from the authors upon request. Note that a fixed effects regression model specifies that the intercept is individual specific. In our context, this means that the bias in the prediction is individual specific. It is another indication of unbiasedness of the predictions that inclusion of fixed effects does not alter the results qualitatively.

Table 3: Initial Prediction, Treatment *Weak*

Dependent variable: *Initial Prediction*					
# of group A analysts	1	3	5	7	9
$\bar{X}_A$	0.198***	0.339***	0.493***	0.647***	0.891***
	(0.030)	(0.027)	(0.025)	(0.028)	(0.037)
$\bar{X}_B$	0.884***	0.640***	0.476***	0.329***	0.104***
	(0.036)	(0.033)	(0.021)	(0.028)	(0.027)
Constant	-0.916*	-0.309	-0.247	-0.006	-0.525
	(0.351)	(0.273)	(0.300)	(0.234)	(0.364)
Optimal weight on $\bar{X}_A$	0.079	0.233	0.394	0.579	0.827
Optimal weight on $\bar{X}_B$	0.921	0.767	0.606	0.421	0.173
F-stats:					
H_0: Unbiased	6.428	0.998	1.121	0.261	1.095
prediction	(0.004)	(0.380)	(0.338)	(0.772)	(0.346)
H_0: Bayesian	11.193	10.007	33.405	6.778	4.281
prediction	(0.000)	(0.000)	(0.000)	(0.003)	(0.022)
H_0: $1/N$	8.922	1.811	0.623	1.944	0.033
heuristic	(0.001)	(0.179)	(0.543)	(0.159)	(0.968)
$\bar{R}^2$	0.634	0.709	0.741	0.699	0.618
# of observations	335	337	339	337	335

Note: Estimated by OLS. Standard errors clustered at subject level are presented below the coefficients (within parentheses) and p-values are presented below the F statistics (within parentheses).

put sub-optimally lower weights on more precise signals.

Now we discuss the main results regarding initial predictions in detail and also discuss and test a number of hypotheses. We cannot reject that the initial predictions are *unbiased* (Hypothesis 1) in most cases.[17] That is, initial prediction P can be expressed as $P = w\bar{X}_A + (1-w)\bar{X}_B$ for some $w \in [0,1]$. We have tested regression specifications where extreme values (maximum and/or minimum) of the signals (for both groups or separately) are added as explanatory variables. We rarely find any of those extreme values to be statistically significant. If we include them, initial predictions can still be characterized as a weighted average of the signals.[18] This may not seem surprising as we restrict attention to observations where the initial prediction is within the minimum and maximum of the signals a subject receives. It is, nonetheless, noteworthy because even simple but systematic mistakes such as over-weighting extreme values or habits such as always rounding up predictions may lead to biased predictions. Confirming that the initial predictions are unbiased is an important first step and the rest of the discussions rely on this finding.

While initial predictions are unbiased, they are typically suboptimal and Hypothesis 2 can frequently be rejected. Nonetheless, we notice a clear pattern in the departure from optimal prediction. For treatment *Strong*, subjects chose sub-optimally high weights for group B signals. For example, the estimated weight on

[17]The null hypothesis is rejected in two out of the 15 cases at the 5% level. However, because there are 15 cases, the testing procedure should be adjusted. A simple Bonferroni correction indicates that we should reject the null when we find a p-value less than $0.05/15 = 0.003$ but there is no such case.

[18]Results are available from the authors upon request.

Table 4: Initial Prediction, Treatment *Zero*

Dependent variable: *Initial Prediction*					
# of group A analysts	1	3	5	7	9
$\bar{X}_A$	0.141***	0.271***	0.435***	0.503***	0.726***
	(0.024)	(0.023)	(0.027)	(0.041)	(0.044)
$\bar{X}_B$	0.820***	0.692***	0.596***	0.491***	0.187***
	(0.062)	(0.050)	(0.040)	(0.042)	(0.037)
Constant	-0.184	0.242	-0.183	-0.423	-0.166
	(0.353)	(0.334)	(0.303)	(0.350)	(0.258)
Optimal weight on $\bar{X}_A$	0.056	0.185	0.346	0.553	0.827
Optimal weight on $\bar{X}_B$	0.944	0.815	0.654	0.447	0.173
F-stats:					
H_0: Unbiased	0.365	1.488	0.322	0.850	3.153
prediction	(0.697)	(0.242)	(0.727)	(0.437)	(0.057)
H_0: Bayesian	6.565	7.321	5.353	0.807	3.631
prediction	(0.004)	(0.002)	(0.010)	(0.456)	(0.038)
H_0: $1/N$	1.666	1.142	4.060	13.563	7.887
heuristic	(0.206)	(0.332)	(0.027)	(0.000)	(0.002)
$\bar{R}^2$	0.491	0.648	0.722	0.644	0.553
# of observations	316	318	317	316	317

Note: Estimated by OLS. Standard errors clustered at subject level are presented below the coefficients (within parentheses) and p-values are presented below the F statistics (within parentheses).

the mean of B signals is 0.502 under treatment *Strong* with three group A signals while the optimal weight is 0.398. On the other hand, for treatment *Weak* and *Zero*, they chose sub-optimally low weights for group B signals. For example, the estimated weight on the mean of B signals are 0.640 and 0.692 under treatments *Weak* and *Zero*, respectively, with three group A signals, while the optimal weights are 0.767 and 0.815, respectively.

These results are connected to the alternative hypotheses we have considered. First, results from treatment *Zero* indicate the presence of overprecision, as overweighting of group A signals is consistent with overprecision and inconsistent with correlation neglect without overprecision. On the other hand, because the weights are not proportional to the number of signals from a particular group, as observed in the F-tests ($1/N$ heuristic), subjects did not neglect variance completely. While treatment *Zero* shows the presence of overprecision, our finding that group A signals are underweighted in treatment *Strong* is inconsistent with overprecision without correlation neglect. This suggests the presence of both overprecision and correlation neglect. This is further supported by the finding that weakly correlated group B signals underweighted in treatment *Weak*. Thus, the three treatments together allow us to test for concurrent presence of overprecision and correlation neglect. To further support this supposition, we look at the weights on average of group A and group B signals in Tables 2 to 4. We find that the weights on the average of group B signals typically increase as we reduce the level of correlation among group B signals, as would be suggested by the behavioral models presented in Section 3. Table 5 presents pair-wise comparisons of the weights from the

Table 5: Test of Equivalence of Weights Used in Initial Predictions Across Treatments

# of A	Strong vs Weak	Strong vs Zero	Weak vs Zero
1	4.263	11.957	3.368
	(0.014)	(0.000)	(0.035)
3	4.427	8.385	2.480
	(0.012)	(0.000)	(0.084)
5	0.755	3.933	4.225
	(0.470)	(0.020)	(0.015)
7	1.223	5.508	11.141
	(0.295)	(0.004)	(0.000)
9	2.198	0.519	5.008
	(0.112)	(0.595)	(0.007)

Note: This table presents the results of F tests for the equivalence of coefficients on $\bar{X}_A$ and $\bar{X}_B$ between treatments in the regressions presented in Tables 2 to 4. In parentheses below F statistics are p-values.

three treatments considered here. The results demonstrate that subjects indeed changed weights on signals from the two groups based on the level of correlation. We thus find very strong evidence that subjects exhibit both overprecision and correlation neglect. Next, we examine the results from the remaining two treatments — *Moderate* and *Both* to gain some qualitative insights on which of the five behavioral models, that incorporate both correlation neglect and overprecision, match the data best. Finally, we will structurally estimate the overprecision and correlation neglect parameters based on those models.

4.1.3 Results from Treatments *Moderate* and *Both*

We now examine the results from treatments *Moderate* and *Both*. For both of these treatments, the marginal variance is identical for all signals. That is, $\sigma^2_{AI} + \sigma^2_{AC} = \sigma^2_{BI} + \sigma^2_{BC}$. Analyzing belief formation rules under these treatments provide insights not only about overprecision and correlation neglect, but also about the form of the interaction of these two behavioral biases.

Tables 6 and 7 present estimation results for treatments *Moderate* and *Both*, respectively. The results are in line with those from the other three treatments and the theoretical interpretation we have discussed above. Subjects put sub-optimally low weights on independent or weakly correlated signals and sub-optimally high weights on moderate to strongly correlated signals. These findings are consistent with concurrent presence of correlation neglect and overprecision.

Moreover, these treatments provide us with information about how subjects react to sample averages. The five behavioral models that we consider provide different predictions about how subjects react to sample averages of group A and group B signals. These five models can be categorized into two groups: In Models 1 and 2, subjects correctly understand that the variance of the sample average of independent signals is inversely proportional to the number of signals, while in Models 3 – 5, the variance reducing property of sample average is discounted because the ρ-th power of $1/n_l$ is used. To illustrate how these treatments are

Table 6: Initial Prediction, Treatment *Moderate*

Dependent variable: *Initial Prediction*					
# of group A analysts	1	3	5	7	9
$\bar{X}_A$	0.263***	0.405***	0.569***	0.699***	0.820***
	(0.027)	(0.033)	(0.041)	(0.035)	(0.047)
$\bar{X}_B$	0.752***	0.577***	0.447***	0.304***	0.141***
	(0.033)	(0.030)	(0.029)	(0.024)	(0.018)
Constant	0.148	-0.809*	0.282	-0.480	-0.416
	(0.437)	(0.301)	(0.239)	(0.291)	(0.411)
Optimal weight on $\bar{X}_A$	0.357	0.632	0.75	0.824	0.9
Optimal weight on $\bar{X}_B$	0.643	0.368	0.25	0.176	0.1
F-stats:					
H_0: Unbiased	0.247	3.819	0.770	1.362	1.155
prediction	(0.782)	(0.030)	(0.469)	(0.267)	(0.325)
H_0: Bayesian	7.353	27.735	23.495	16.801	3.027
prediction	(0.002)	(0.000)	(0.000)	(0.000)	(0.059)
H_0: $1/N$	19.244	8.419	1.908	0.012	3.027
heuristic	(0.000)	(0.001)	(0.161)	(0.988)	(0.059)
$\bar{R}^2$	0.757	0.745	0.747	0.688	0.516
# of observations	431	433	436	436	434

Note: Estimated by OLS. Standard errors clustered at subject level are presented below the coefficients (within parentheses) and p-values are presented below the F statistics (within parentheses).

useful in distinguishing these two categories of models in a simplified setting, we will assume that subjects exhibit complete correlation neglect in this subsection; that is, $\gamma = 0$. In treatment *Both*, where $\sigma^2_{AI} = \sigma^2_{BC}$ and $\sigma^2_{AC} = \sigma^2_{BI}$, even if subjects exhibit overprecision, they will put proportional (to the number of signals from the group) weights on $\bar{X}_A$ and $\bar{X}_B$ under Models 1 and 2. In other words, subjects' behavior would be equivalent to the $1/N$ heuristics. However, Table 7 shows that, for three out of the five combinations of signals, we can reject that the weights on $\bar{X}_l$ equals $n_l/10$, at the 5% level, while rejecting two at the 1% level. In particular, it indicates that Models 3 – 5 would better represent the actual behavior of subjects than Models 1 and 2. Treatment *Moderate* further strengthens the evidence against Model 2. Under that treatment, subjects' behavior would be equivalent to the $1/N$ heuristics under Model 2 because all signals have the same marginal variance. Table 6 shows that this hypothesis is rejected in two out of the five cases at the 5% level (three at the 10% level). Thus, both treatments *Moderate* and *Both* strongly suggest that subjects undervalue the variance reduction ability of sample averages.

Furthermore, comparing these two treatments allow us to further differentiate the behavioral models. Under complete correlation neglect, Models 3 and 4 are equivalent. In these two models, the marginal variance determines the weights but how each of the independent noise and correlated noise contributes to the overall variance does not matter. In contrast, Model 5 indicates that subjects treat the variance from independent noises differently from that from correlated noises even though they also mistakenly treat correlated noises as independent. Treatments *Moderate* and *Both* share a property that the marginal variances of all signals in

Table 7: Initial Prediction, Treatment *Both*

Dependent variable: *Initial Prediction*					
# of group A analysts	1	3	5	7	9
$\bar{X}_A$	0.250***	0.350***	0.530***	0.618***	0.643***
	(0.034)	(0.043)	(0.045)	(0.047)	(0.044)
$\bar{X}_B$	0.709***	0.573***	0.490***	0.408***	0.310***
	(0.044)	(0.063)	(0.021)	(0.037)	(0.041)
Constant	0.627	0.366	0.193	-0.012	-1.271**
	(0.425)	(0.470)	(0.342)	(0.317)	(0.419)
Optimal weight on $\bar{X}_A$	0.487	0.719	0.796	0.834	0.861
Optimal weight on $\bar{X}_B$	0.513	0.281	0.206	0.166	0.139
F-stats:					
H_0: Unbiased	1.661	1.808	0.219	0.154	4.861
prediction	(0.209)	(0.183)	(0.805)	(0.858)	(0.016)
H_0: Bayesian	26.132	37.870	95.085	24.201	15.616
prediction	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
H_0: $1/N$	13.686	2.096	0.268	4.527	22.382
heuristic	(0.000)	(0.142)	(0.767)	(0.020)	(0.000)
$\bar{R}^2$	0.688	0.639	0.780	0.790	0.677
# of observations	251	273	276	277	266

Note: Estimated by OLS. Standard errors clustered at subject level are presented below the coefficients (within parentheses) and p-values are presented below the F statistics (within parentheses).

a given treatment are the same. Hence, Models 3 and 4 predict that the belief formation rules should be the same for these two treatments. The F-test statistics for the equivalence of the weights between treatments *Moderate* and *Both* are 0.076 (0.927), 0.605 (0.546), 0.824 (0.439), 5.497 (0.004), and 14.561 (0.000), for $n_A = 1$, 3, 5, 7, and 9 respectively, where p-values are in parentheses. There is statistical evidence that the weights between the two treatments are different when the number of moderately or strongly correlated signals is small. This result favors Model 5 over Model 4.

4.1.4 Estimating the behavioral bias parameters

We now estimate the correlation neglect and overprecision parameters using the estimates of the weights. We use a minimum distance approach. We find that Model 5 yields the best fit. The estimated parameters indicate that subjects exhibit near-complete correlation neglect. Moreover, they incorrectly transform variances to about their square roots. That is, they almost treat standard deviations as variances. Below we describe how we estimate the model parameters.

Our five behavioral models are parameterized by γ (correlation neglect parameter) and ρ (overprecision parameter). Given the values of γ and ρ, each model predicts a unique combination of weights on $\bar{X}_A$ and $\bar{X}_B$ that sum up to 1 for each treatment and each signal combination. We thus estimate the values of γ and ρ by minimizing the distance between the estimated weight on $\bar{X}_A$ and those predicted by the model.

Table 8: Estimation results

	Model 1	Model 2	Model 3	Model 4	Model 5
fit	0.189	0.179	0.120	0.103	0.080
$\hat{\gamma}$	0.170	0.003	0.176	0.059	0.150
	(0.042)	(0.016)	(0.051)	(0.066)	(0.083)
$\hat{\rho}$	0.549	0.321	0.616	0.591	0.576
	(0.005)	(0.259)	(0.034)	(0.136)	(0.037)

Note: This table presents the results of the minimum distance estimation. For each model, the parameters are estimated by minimizing a distance between estimated weights on the average of group A signals and those predicted by the model. "fit" is the minimized value of the objective function. In parentheses are standard errors.

Let $\hat{w}_{A,n_A,D}$ be the estimated weight on $\bar{X}_A$ where subjects receive n_A group A signals under treatment $D \in \{Strong, Weak, Zero, Moderate, Both\}$. Let $w_{A,n_A,D}(\gamma,\rho)$ be the weight predicted by the model for a given (γ,ρ) combination. We then estimate (γ,ρ) by solving

$$(\hat{\gamma},\hat{\rho}) = \arg\min_{\gamma,\rho} \sum_{D\in\{Strong,\ Weak,\ Zero,\ Moderate,\ Both\}} \sum_{n_A\in\{1,3,5,7,9\}} (\hat{w}_{A,n_A,D} - w_{A,n_A,D}(\gamma,\rho))^2.$$

Table 8 presents the estimation results. The minimized value of the sum of distance squared for a given model determines the fit of the model. Thus, the smaller the value of the fit, the better the model explains our experimental data. The table demonstrates that Model 5 yields the best fit. The estimation results are incompatible with full rationality. For all models, both parameters are statistically different from 1 at the 5% level. The estimated correlation neglect parameters are quite small (for example, in Model 5, it is not statistically significantly different from zero at the 5% level), indicating the presence of nearly complete correlation neglect. Nonetheless, subjects undervaluing weakly correlated signals in favor of independent signals in treatment *Weak* shows that the presence of overprecision will not always lead to overvaluing of correlated signals even when people neglect correlation completely. The overprecision parameter is slightly above 0.5. A rough interpretation would be that subjects wrongly treat standard deviation as variance.

The fact that Model 5 best fits the data is consistent with the discussion in the previous subsection. This model exhibits the feature that the variance reduction power of sample averaging is downgraded and that the contribution of correlated noises to the overall variance is treated differently from that of independent noises. We quantitatively confirm that these features are important to have a good fit to the data.

The estimated values of the parameters are related to the results of Eyster and Weizsacker (2016). Their behavioral model has two parameters: one parameter represents the degree of correlation neglect and the other parameter refers to the degree of undervaluing precision. They also find a nearly complete level of correlation neglect and also that people's behavior would be consistent with the situations in which they treat standard deviations as variances. While their results are similar to ours, there are important differences. First, we examine belief formation directly while they examine portfolio choice. Second, their interpretation

of undervaluation of precision is that the degree of reaction to risk depends on the magnitude of risk. Indeed, it would be difficult to distinguish their interpretation from overprecision bias in their experiment. We avoid such a problem by directly eliciting beliefs using an incentive compatible scoring rule.

4.1.5 Robustness: Learning and Heterogeneity

Digging deeper into the data, we analyze how subject behavior changes over time under each treatment. Figures 2 to 6 in Appendix A.3 present time series plots of weights on the mean of group A signals over periods within a 10-period block for each analyst combination. We run cross-sectional regressions of prediction on the mean of group A signals and the mean of group B signals for each period (all variables are recentered around true EPS), and use the OLS estimate of the coefficient on the mean of group A signals as "weights on A." The five panels within a figure present the time-series diagram when the number of group A analysts is 1, 3, 5, 7, and 9, respectively. There is no clear pattern or systematic indication of learning within each block. There is also no indication that subjects' behavior converge to the theoretical optimal. Overall, we do not learn much more about subject behavior by separating behavior in each period over what is presented in Tables 2 to 4 and 6 to 7.

We also analyze how belief formation rules vary across subjects. Specifically, we ran time series regressions of prediction on the means of groups A and B signals for each 10-period block for each subject separately. Comparing the weights on group A signals, we find that the weights across subjects are, perhaps unsurprisingly, quite dispersed and there are some outliers. Nonetheless, we do not find any clear pattern in the dispersion of weights. These results are available from the authors upon request.

4.2 Revised Prediction with Additional Public Information

We now analyze the revised predictions submitted by subjects after they learn the session average for that period—the average of initial predictions in that period by all subjects in the session. First, we ensure that the session averages of initial predictions are indeed better than individual initial predictions. We find that the session averages yield about three times higher probabilities of receiving the reward than initial predictions do in all of the sessions.[19] While not surprising, this confirms that the average judgment of the crowd is better than the judgment of the individuals who make up the crowd. We then analyze whether people actually use this aggregated information appropriately. To that end, we use a linear regression model to estimate the weights subjects put on their own initial predictions and session averages. We find that they put non-zero weights on initial predictions, which we interpret as a result of incorrectly incorporating the variance reducing power of sample averages. Specifically, the results can be explained by subjects undervaluing the impact of the sample size in reducing the variance of an estimator. This also suggests that

[19] These results are available upon request.

even though there is wisdom in the judgment of the crowd overall, people do not fully utilize such wisdom. Nonetheless, we can extrapolate that, as the sample size becomes really large, the impact of such mistakes would be small.

4.2.1 Econometric Specification and Some Hypotheses

As each subject gets a completely different set of forecasts for a given stock, the session average indirectly provides information from many more signals than the subject received before submitting the initial prediction. It is optimal to report the session average as the revised prediction when a subject believes that all subjects' initial predictions are equally precise. However, if a subject under-appreciates the variance reducing power of average, as indicated by our findings from initial predictions in Section 4.1.4, she may rely on her private signals and her initial prediction in addition to the session average for her revised prediction.

To estimate the process subjects use to generate revised prediction, we relate a subject's revised prediction with her initial prediction and the session average. We report the results from linear regression models which are found to be reasonable after initial inspection of the data. Our estimation model is:

$$(RP_{it} - T_t) = \eta_0 + \eta_1(P_{it} - T_t) + \eta_2(SA_t - T_t) + u_{it},$$

where RP_{it} and P_{it} are the revised and initial predictions, respectively, by subject i at period t, T_t is the true EPS and SA_t is the session average at period t (note that they are common to all the subjects and there is no subject indicator), and u_{it} is the error term. The theoretical result in Proposition 2 does not depend on the information generation parameter set. Hence, we can pool all the periods within a specification in these regressions.[20]

The value of the coefficients, (η_0, η_1, η_2), relate to various hypotheses.

Hypothesis 4 (Unbiased prediction): A prediction is unbiased when $\eta_0 = 0$ and $\eta_1 + \eta_2 = 1$.

Hypothesis 5 (Optimal prediction): An unbiased prediction is optimal when $H_0 : \eta_1 = 0$ and $\eta_2 = 1$.

When subjects are Bayesian and believes that other subjects' predictions are as precise as hers, then λ equals $1/M$ in Proposition 2. In that case, her optimal revised prediction would be to report the session average as her revised prediction. However, if a subject believes that $\lambda > 1/M$ then she would put positive weight on her own initial prediction P_{it}. That leads to the following alternative,

H_1 **(Variance underweighting):** $\eta_1 > 0$ and $\eta_2 < 1$.

[20] We exclude observations where the revised prediction is above the maximum or below the minimum of the 10 forecasts she received and the session average. We also exclude periods in which at least one subject submitted an initial prediction that is above the maximum or below the minimum of the 10 forecasts she received because in such a period, the session average may not provide useful information. These exclusions do not affect our results significantly

Table 9: Estimation Results: Revised Prediction

Dependent variable: *Revised Prediction*						
	All	Strong	Weak	Zero	Moderate	Both
Initial Prediction	0.229***	0.187***	0.241***	0.264***	0.245***	0.211***
	(0.014)	(0.022)	(0.030)	(0.030)	(0.027)	(0.037)
Session Average	0.775***	0.794***	0.739***	0.823***	0.785***	0.763***
	(0.021)	(0.035)	(0.051)	(0.039)	(0.047)	(0.053)
Constant	0.013	-0.035	0.060	0.085	-0.121	0.133
	(0.046)	(0.092)	(0.068)	(0.105)	(0.090)	(0.168)
F-stats:						
H_0: Unbiased	0.064	0.433	0.650	2.809	1.872	0.460
prediction	(0.938)	(0.652)	(0.528)	(0.076)	(0.166)	(0.636)
H_0: Optimal	110.260	35.604	26.364	20.786	21.050	19.805
prediction	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.468	0.509	0.456	0.440	0.466	0.483
# of observations	7779	1351	1714	1626	1989	1099

Note: Estimated by OLS. Standard errors clustered at subject level are presented below the coefficients (within parentheses) and p-values are presented below the F statistics (within parentheses). Session Average is the average of the initial predictions made by all subjects.

4.2.2 Detailed Discussions of Results

Table 9 presents the estimation results for the five treatments, one in each column. Subjects typically choose a revised prediction that is in between their initial prediction and the session average, leading to an unbiased estimate. The weights on own initial prediction and the session average add up to one and we cannot reject Hypothesis 4 for any of the treatments at the 5% level. However, unlike what the theoretical optimal suggests, subjects put positive weight on their own initial prediction under all treatments. Moreover, the weights are relatively close to each other across treatments, ranging between 18.7% and 26.4%. Hence, we can reject Hypothesis 5 in support of subjects believing that $\lambda > 1/M$ (H_1 Variance underweighting) for all treatments.

Comparing these numbers across the five treatments, we find that the weights on own initial predictions are statistically significantly different at the 5% level only between treatments *Strong* and *Zero*, where the weights on initial prediction are the lowest and the highest, respectively. While subjects' initial prediction generation processes vary quite a bit across treatments, the revised prediction generation processes do not vary as much.

To explain the result that subjects typically put less than 80% weight on the session average to calculate their revised prediction, we go back to the behavioral models presented in Section 3. Specifically, Models 3 to 5, which fit the the initial prediction data better than the other model, suggest that subjects do not fully incorporate the sample size in calculating the variance of a sample average. If we assume that subjects behave as if $\lambda = 1/M^{\rho}$ to choose their revised prediction, then that would explain our findings well. Taking

this idea of incorrect incorporation of sample size seriously, we can even estimate alternative models that use n_l^ν in the denominators of Models 3 to 5 instead of n_l^ρ. The estimated parameters do not change dramatically in those cases and we find estimates of ν to be around 0.6. Moreover, the associated variation of Model 5 continues to fit the initial prediction data the best.[21]

Recall that we had 10 sessions in total for the five treatments. The number of subjects in these sessions varied between 13 and 22. However, the weights on the session average do not vary as wildly across treatments or sessions. This is also consistent with a model where subjects behave as if $\lambda = 1/M^\rho$ to choose their revised prediction. For example, when $\rho = 0.576$, weights on session average would vary between 86% and 89% between sessions with 13 and 22 subjects. Thus, weights on session averages not being very sensitive to the number of subjects in a session is consistent with our behavioral model. This empirical finding is not very consistent with models of overconfidence in one's ability in forming beliefs correctly (Eyster, Rabin, & Vayanos, 2019) or ambiguity about belief formation rules of other subjects (De Filippis, Guarino, Jehiel, & Kitagawa, 2017).

4.2.3 Robustness: Learning and Heterogeneity

To analyze how subjects' estimation process for revised prediction changes over time, Figure 7 in Appendix A.3 presents a time series plot of weights on session average. We run cross-section regressions of revised prediction on own initial prediction and session average for each period, and present the OLS estimate of the coefficient on the session average. Again, there is no clear pattern. Here we pool all the treatments together as the optimal weights are the same in all of the treatments. We find relatively low weights in periods 3 to 5 in this pooled picture and one might worry that this affects other results, or it indicates learning over time. However, looking at time series plots for each treatment separately, we did not find any clear pattern nor an indication of learning. Moreover, the main results regarding revised predictions do not change qualitatively if we drop the first 10 periods.[22]

We also analyze how heterogeneous the revised prediction generation processes is across subjects. Specifically, we ran time series regressions of revised prediction on initial prediction and session average for each subject. The weights across subjects are, as in the case of initial predictions, quite dispersed and there are some outliers. Nonetheless, we are unable to find any factor beyond gender and knowledge of statistics, that affects heterogeneity. These result are available from the authors upon request.

[21] Estimation results are available from the authors upon request.

[22] Separate time series plots for each treatment and regression results without the first 10-period data are available from the authors upon request.

5 Conclusion

We present results from a comprehensive set of experiments to ascertain how people treat correlated and independent information and information of different quality levels. We find that while people overvalue strongly correlated information, they undervalue weakly correlated information; such correlation over-adjustment is a new finding. This may suggest that whether people neglect correlation depends on the level of correlation. We theoretically show that the extent of correlation neglect being dependent on the correlation level is not necessary to explain our findings. Specifically, if people suffer only from correlation neglect, that would always lead to overvaluing of correlated signal. However, when people mis-perceive variance of their posterior belief to be lower than it actually is, that may lead to undervaluing of weakly correlated signals. All these results are consistent with our experimental findings. Thus, our experimental and theoretical results suggest that we need to take both correlation neglect and overprecision into account when predicting people's behavior under uncertainty.

We also find that people put sub-optimally high weight on information that they directly receive than those they indirectly receive through the actions of others even though such information aggregates a lot more signals. Our behavioral models suggest that people may not properly account for the variance-reducing power of sample averages. Taking this finding a step further, we show that incorrect incorporation of usefulness of aggregation of many signals explains this finding. This suggests that even when aggregated information exhibit wisdom of the crowd, people incorporate this wisdom at a lower rate than theoretically optimal.

Overall, our theoretical and experimental findings suggest that people show some sophistication in processing information. These results provide insights on the directions of bounded rationality in information processing. Signals in this experiment are generated through a complicated process, which may be considered a drawback. However, information in real life also follows complicated processes. It is important to figure out how people form beliefs in a complex environment; one they may not understand completely. Aided by our experimental results, the behavioral models presented in the paper provide insights for an "as-if" description of people's belief updating behavior.

Our results can be useful for many different applications. In particular, it is important to recognize that people perceive variance incorrectly in addition to ignoring correlation. As a result, people's responses to correlated information depend on the degree of correlation. As pointed out in the literature, people can be "fooled"by financial experts' suggestions when the experts have very similar view points by design (shared information source, common analytic tools, similar incentives) even when they are aware of the similarities. But this is not the whole story. They may not appropriately appreciate suggestions by financial experts who try to obtain independent sources of information but can get only weakly correlated sources. Combined with the findings from experiments on sequential actions by subjects designed to find herding behavior, our results on revised prediction may suggest that bank runs or excessive purchase of hot stocks (leading to

creation of a bubble) based on market activity may be tempered if enough people receive contrary private signals. Our results may also suggest that in e-commerce, one may overestimate the impact of click-through rates on eventual purchase when aggregating data from similar types of consumers and underestimate it when aggregating data from dissimilar types of consumers.

Our results also have implications on marketing and public opinion. In the literature on multiple source effect (Harkins & Petty, 1987), the same message from multiple media are considered more informative. While we obtain a similar implication under strong correlation and our argument provides a mechanism behind this effect, our results also imply that an opposite effect may appear under weak correlation: An effort to provide less correlated information may not be rewarded sufficiently. For example, political polls that are very weakly correlated may be "trusted"by voters much less than they are supposed to be, making them less useful to political operatives. They may instead prefer to report results from pollsters whose methodologies favor their candidate even if these polls are known to be more or less the same. Another example can be regarding public opinion about climate change. Climate scientists Anderson et al. (2013) provide an independent measure of global warming instead of using thermometer-based global surface temperature time series of 130 years which is widely used but provides correlated assessment. We would speculate that the information provided by these researches may not be appreciated by the public as much as one would expect because people may undervalue the effort to reduce correlation. More research needs to be conducted to examine how well such implications of our findings carry over to real applications.

This paper provides a framework for investigating belief formation under a large set of contexts and settings. One future direction would be to analyze how people form beliefs when information or signals arrive sequentially. One can also explore how people decide how much information to receive based on their beliefs. Another extension would be to allow for biased signals. This can be particularly relevant in the context of politics and media. There is some recent theoretical investigation on how correlation neglect may affect the way media or news outlets present information.[23] Thus, learning how people treat correlation among biased signals and form their beliefs using experiments will be useful. We can also vary the level of biases the consumers of news themselves have in their preferences and investigate how that interacts with potential misperception of correlation. In general, our experimental setup and results can open up a number of different research avenues.

References

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: An empirical examination. *International Economic Review*, *44*(3), 1007–1040.

[23]See, for example, Levy, de Barreda, and Razin (2017).

Anderson, D. M., Mauk, E. M., Wahl, E. R., Morrill, C., Wagner, A. J., Easterling, D., & Rutishauser, T. (2013). Global warming in an independent record of the past 130 years. *Geophysical Research Letters*, *40*(1), 189–193.

Au, P. H., & Kawai, K. (2012). Media capture and information monopolization in japan. *The Japanese Economic Review*, *63*(1), 131–147.

Bajari, P., & Hortaçsu, A. (2003). The winner's curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions. *The RAND Journal of Economics*, *34*(2), 329–355.

Batanero, C., Tauber, L. M., & Sánchez, V. (2004). Students' reasoning about the normal distribution. In *The challenge of developing statistical literacy, reasoning and thinking* (Chap. 11, pp. 257–276). Springer.

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American economic review*, *91*(1), 79–98.

Ching, A. T., Erdem, T., & Keane, M. P. (2013). Learning models: An assessment of progress, challenges, and new developments. *Marketing Science*, *32*(6), 913–938.

Crawford, G. S., & Shum, M. (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica*, *73*(4), 1137–1173.

Daniel, K., & Hirshleifer, D. (2015). Overconfident investors, predictable returns, and excessive trading. *Journal of Economic Perspective*, *29*(4), 61–88.

De Filippis, R., Guarino, A., Jehiel, P., & Kitagawa, T. (2017). *Updating ambiguous beliefs in a social learning experiment*. cemmap working paper.

Dewan, T., & Myatt, D. P. (2008). The qualities of leadership: Direction, communication, and obfuscation. *American Political Science Review*, *102*(3), 351–368.

Ellis, A., & Piccione, M. (2017). Correlation misperception in choice. *American Economic Review*, *107*(4), 1264–92.

Enke, B., & Zimmermann, F. (2019). Correlation neglect in belief formation. *The Review of Economic Studies*, *86*(1), 313–332.

Erdem, T., & Keane, M. P. (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science*, *15*(1), 1–20.

Eyster, E., & Rabin, M. (2014). Extensive imitation is irrational and harmful. *Quarterly Journal of Economics*, *129*(4), 1861–1898.

Eyster, E., Rabin, M., & Vayanos, D. (2019). Financial markets where traders neglect the informational content of prices. *Journal of Finance*, *74*(1), 371–399.

Eyster, E., & Weizsacker, G. (2016). *Correlation neglect in portfolio choice: Lab evidence*. mimeo.

Fischbacher, U. (2007). Z-tree - zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Grubb, M. D. (2015). Overconfident consumers in the marketplace. *Journal of Economic Perspective*, *105*(4), 9–36.

Harkins, S. G., & Petty, R. E. (1987). Information utility and the multiple source effect. *Journal of Personality and Social Psychology*, *52*(2), 260–268.

Hossain, T., & Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, *80*, 984–1001.

Huang, R., Krishnan, M., Shon, J., & Zhou, P. (2017). Who herds? who doesn't? estimates of analysts' herding propensity in forecasting earnings. *Contemporary Accounting Research*, *34*(1), 374–399.

Kallir, I., & Sansino, D. (2009). The neglect of correlation in allocation decisions. *Southern Economic Journal*, *75*(4), 1045–66.

Kroll, Y., Levy, H., & Rapoport, A. (1988). Experimental tests of the separation theorem and the capital asset pricing model. *The American Economic Review*, 500–519.

Lee, J. Y., & Iyengar, R. (2016). *A model of consumer search for experience goods*. mimeo.

Levy, G., de Barreda, I. M., & Razin, R. (2017). *Persuasion with correlation neglect: Media power via correlation of news content*. mimeo.

Levy, G., & Razin, R. (2015). Correlation neglect, voting behavior, and information aggregation. *American Economic Review*, *105*(4), 1634–45.

Levy, G., & Razin, R. (2018). *Combining forecasts in the presence of ambiguity over correlation structures*. mimeo.

Luan, S., Sorkin, R. D., & Itzkowitz, J. (2004). Weighting information from outside sources: A biased process. *Journal of Behavioral Decision Making*, *17*(2), 95–116.

Maines, L. A. (1996). An experimental examination of subjective forecast combination. *International Journal of Forecasting*, *12*(2), 223–233.

Moore, D. A., & Healy, P. J. (2007). Bayesian overconfidence. *Academy of Management Best Papers Proceedings*.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517.

Moore, D. A., & Schatz, D. (2017). The three faces of overconfidence. *Social and Personality Psychology Compass*, *11*(8), e12331.

Myatt, D. P., & Wallace, C. (2012). Endogenous information acquisition in coordination games. *Review of Economic Studies*, *79*, 340–374.

Nöth, M., & Weber, M. (2003). Information aggregation with random ordering: Cascades and overconfidence. *The Economic Journal*, *113*(484), 166–189.

Ortoleva, P., & Snowberg, E. (2015). Overconfidence in political behavior. *American Economic Review*, *105*(2), 504–535.

Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, *117*(3), 775–816.

Soll, J. B. (1999). Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psychology*, (38), 317–346.

Trueman, B. (1994). Analyst forecasts and herding behavior. *The Review of Financial Studies*, *7*(1), 97–124.

A Appendix

A.1 Proof of Proposition 1

Let $X = (X_1^A, \ldots, X_{n_A}^A, X_1^B, \ldots, X_{n_B}^B)^\top$. The log likelihood function of X is proportional to

$$(X - T\iota_N)^\top \Sigma^{-1} (X - T\iota_N),$$

where

$$\Sigma = \begin{pmatrix} \sigma_{AI}^2 I_{n_A} + \sigma_{AC}^2 \iota_{n_A} \iota_{n_A}^\top & 0_{n_A \times n_B} \\ 0_{n_B \times n_A} & \sigma_{BI}^2 I_{n_B} + \sigma_{BC}^2 \iota_{n_B} \iota_{n_B}^\top \end{pmatrix},$$

ι_a is the $a \times 1$ vector of ones, I_a is the $a \times a$ identity matrix and $0_{a \times b}$ is the $a \times b$ matrix of zeros.

Because the prior is $T \sim N(\mu_p, \sigma_p^2)$, the posterior is

$$T|X \sim N(\mu^*, (\sigma^*)^2),$$

where

$$\begin{aligned} \mu^* &= \left(\frac{1}{\sigma_p^2} + \iota_N^\top \Sigma^{-1} \iota_N \right)^{-1} \left(\iota_N^\top \Sigma^{-1} X + \frac{\mu_p}{\sigma_p^2} \right), \\ (\sigma^*)^2 &= \left(\frac{1}{\sigma_p^2} + \iota_N^\top \Sigma^{-1} \iota_N \right)^{-1}. \end{aligned}$$

By the rule $(I + AA^\top)^{-1} = I - A(I + A^\top A)^{-1} A^\top$ for an identity matrix I and a matrix A, we have

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_{AI}^2} I_{n_A} - \frac{\sigma_{AC}^2}{\sigma_{AI}^2} \frac{1}{\sigma_{AI}^2 + \sigma_{AC}^2 n_A} \iota_{n_A} \iota_{n_A}^\top & 0_{n_A \times n_B} \\ 0_{n_B \times n_A} & \frac{1}{\sigma_{BI}^2} I_{n_B} - \frac{\sigma_{BC}^2}{\sigma_{BI}^2} \frac{1}{\sigma_{BI}^2 + \sigma_{BC}^2 n_B} \iota_{n_B} \iota_{n_B}^\top \end{pmatrix}.$$

It therefore follows that

$$\begin{aligned} \iota_N^\top \Sigma^{-1} \iota_N =& \frac{n_A}{\sigma_{AI}^2} - \frac{\sigma_{AC}^2}{\sigma_{AI}^2} \frac{n_A^2}{\sigma_{AI}^2 + \sigma_{AC}^2 n_A} + \frac{n_B}{\sigma_{BI}^2} - \frac{\sigma_{BC}^2}{\sigma_{BI}^2} \frac{n_B^2}{\sigma_{BI}^2 + \sigma_{BC}^2 n_B} \\ =& \frac{n_A}{\sigma_{AI}^2 + \sigma_{AC}^2 n_A} + \frac{n_B}{\sigma_{BI}^2 + \sigma_{BC}^2 n_B}, \\ \iota_N^\top \Sigma^{-1} X =& \left(\frac{1}{\sigma_{AI}^2} - \frac{\sigma_{AC}^2}{\sigma_{AI}^2} \frac{n_A}{\sigma_{AI}^2 + \sigma_{AC}^2 n_A} \right) \sum_{j=1}^{n_A} X_j^A + \left(\frac{n_B}{\sigma_{BI}^2} - \frac{\sigma_{BC}^2}{\sigma_{BI}^2} \frac{n_B^2}{\sigma_{BI}^2 + \sigma_{BC}^2 n_B} \right) \sum_{j=1}^{n_B} X_j^B \\ =& \frac{n_A}{\sigma_{AI}^2 + \sigma_{AC}^2 n_A} \bar{X}_A + \frac{n_B}{\sigma_{BI}^2 + \sigma_{BC}^2 n_B} \bar{X}_B. \end{aligned}$$

Noting that $V_A = \sigma_{AI}^2/n_A + \sigma_{AC}^2$ and $V_B = \sigma_{BI}^2/n_A + \sigma_{BC}^2$, it holds that

$$\iota_N^\top \Sigma^{-1} \iota_N = 1/V_A + 1/V_B \quad \text{and} \quad \iota_N^\top \Sigma^{-1} X = \bar{X}_A/V_A + \bar{X}_B/V_B.$$

Therefore, when $\sigma_p \to \infty$ (uninformative prior), the posterior of T is a normal distribution with mean:

$$\left(\frac{1}{V_A} + \frac{1}{V_B}\right)^{-1} \left(\frac{\bar{X}_A}{V_A} + \frac{\bar{X}_B}{V_B}\right) = \frac{V_B}{V_A + V_B}\bar{X}_A + \frac{V_A}{V_A + V_B}\bar{X}_B$$

and variance:

$$\left(\frac{1}{V_A} + \frac{1}{V_B}\right)^{-1} = \frac{V_A V_B}{V_A + V_B}$$

A.2 Proof of Proposition 2

We note that this problem can be considered a Bayesian problem in which the prior is the subject's belief (which is the posterior obtained in the proof of Proposition 1 if the subject is truly Bayesian) and the observation is the average of predictions of all other subjects. Let $\bar{p}_- = (M\bar{p} - p^*)/(M-1)$ be the average of all other subjects' predictions.

The subject believes that the distribution of the session average is $N(T, \lambda\sigma^2)$. Therefore, $\bar{p}_- \sim N(T, (M^2\lambda - 1)\sigma^2/(M-1)^2)$. After observing the session average $\bar{p}$, the subject's posterior distribution can be described by

$$N\left(\frac{\sigma^2}{\sigma^2 + (M^2\lambda-1)\sigma^2/(M-1)^2}\bar{p}_- + \frac{(M^2\lambda-1)\sigma^2/(M-1)^2}{\sigma^2 + (M^2\lambda-1)\sigma^2/(M-1)^2}p^*, \frac{\sigma^2(M^2\lambda-1)\sigma^2/(M-1)^2}{\sigma^2 + (M^2\lambda-1)\sigma^2/(M-1)^2}\right).$$

The posterior mean is

$$\begin{aligned}
&\frac{\sigma^2}{\sigma^2 + (M^2\lambda-1)\sigma^2/(M-1)^2}\bar{p}_- + \frac{(M^2\lambda-1)\sigma^2/(M-1)^2}{\sigma^2 + (M^2\lambda-1)\sigma^2/(M-1)^2}p^* \\
=&\frac{M(M-1)}{(M-1)^2 + (M^2\lambda-1)}\bar{p} + \frac{(M^2\lambda - M)}{(M-1)^2 + (M^2\lambda-1)}p^*.
\end{aligned}$$

A.3 Additional Figures

Figure 2: Initial Prediction, Time Series of Weights on A, Treatment *Strong*

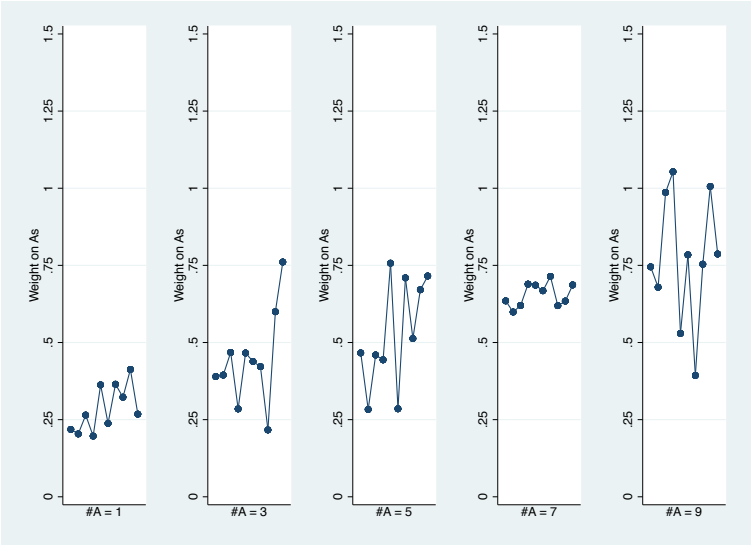


Figure 3: Initial Prediction, Time Series of Weights on A, Treatment *Weak*

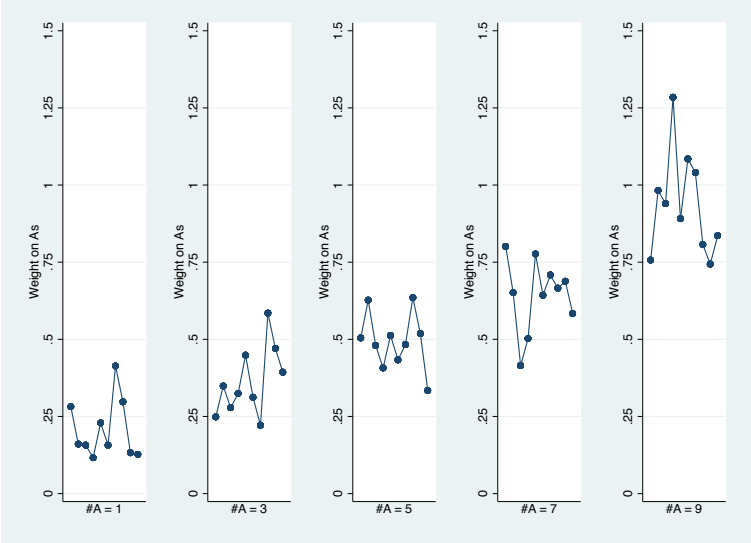


Figure 4: Initial Prediction, Time Series of Weights on A, Treatment *Zero*

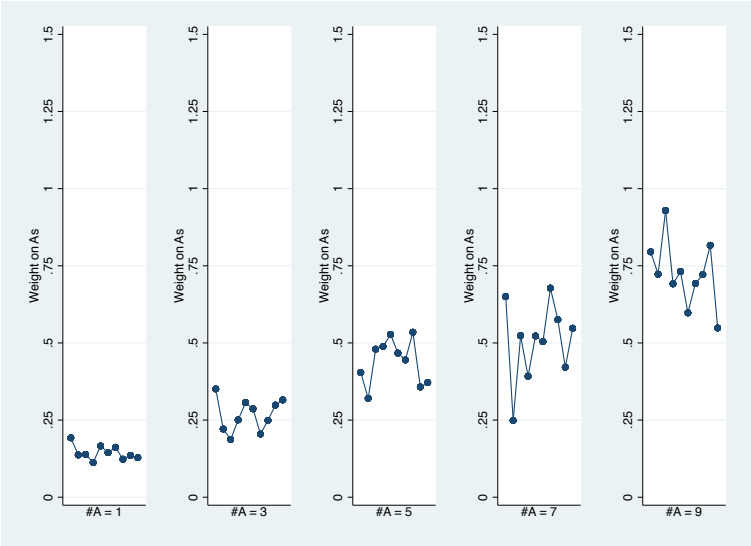


Figure 5: Initial Prediction, Time Series of Weights on A, Treatment *Moderate*

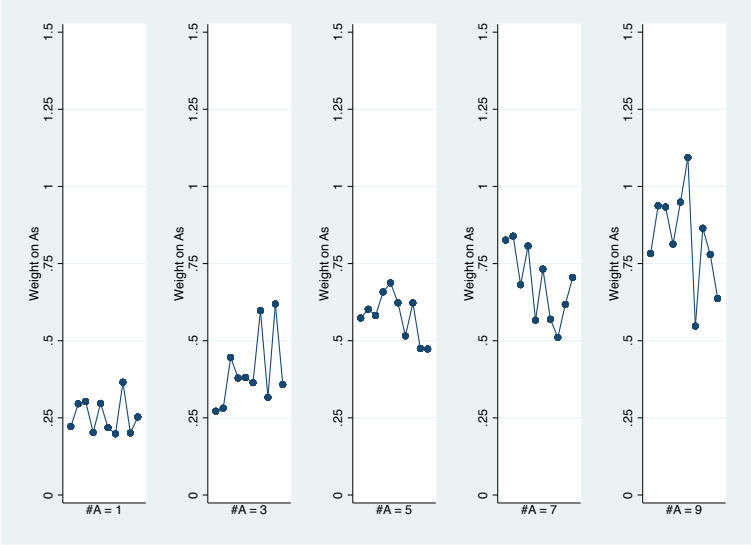


Figure 6: Initial Prediction, Time Series of Weights on A, Treatment *Both*

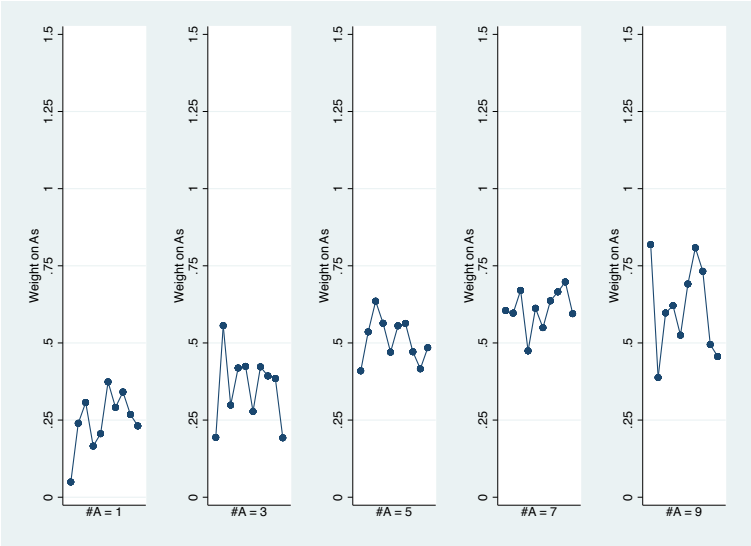
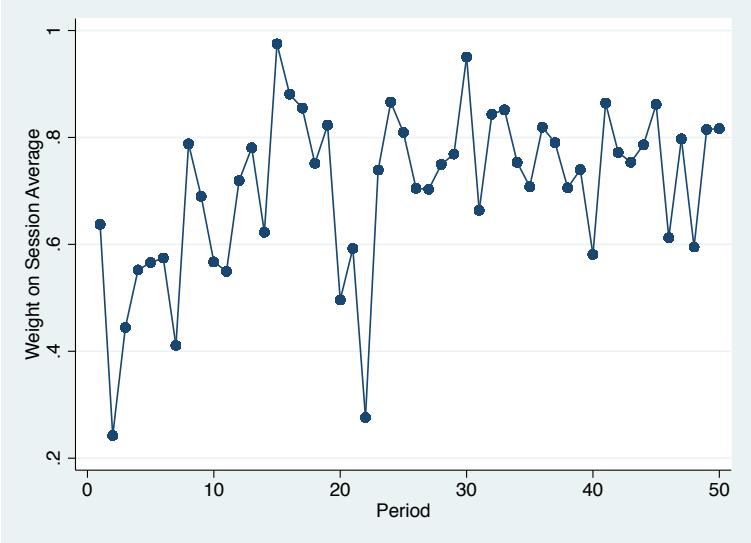


Figure 7: Revised Prediction, Time Series of Weights on Session Average, All Treatments



A.4 Sample Experimental Instructions and Statistical Definitions

Sample Experimental Instructions for the Treatment "Strong"

General Rules

This session is part of an experiment about how people aggregate multiple forecasts about returns from a financial asset to estimate the average return. If you follow the instructions carefully and make good decisions, you can earn a considerable amount of money. There are __ people (including you) in this room who are also participating as subjects in this session. They have all been recruited in the same way as you and are reading the same instructions as you are for the first time. It is important that you do not communicate with any other participant or discuss the details of the experiment with anyone during and after the session.

Setting

In each period of this session, you will receive forecasts from 10 stock analysts regarding the earning per share (EPS) for the stocks of a company. In every period, the information will be provided for a new company. Thus, the EPS in each period is independent of each other. In each period, your goal is to predict the true value of the EPS for that period's stock based on the information you receive.

For each stock, you will receive 10 forecasts about the EPS. An analyst cannot perfectly forecast the true EPS of a stock. Rather, she observes the EPS along with two error terms— Division 1 error and Division 2 error. Specifically, her observation equals the true EPS plus division 1 error and Division 2 error where any error term can be positive or negative. She reports this observation as her forecast.

In each period, some of the 10 forecasts will be from *Group A* analysts and the rest will be from *Group B* analysts. For all *Group A* analysts, the errors from both divisions are independent of each other. Suppose, in some period, the true EPS for that period's stock is *T*. If analyst *i* is a *Group A* analyst, then she observes forecast F_i where $F_i = T + \varepsilon_{1i} + \varepsilon_{2i}$. In this session, Division 1 errors for *Group A* analysts are drawn from a normal distribution with mean 0 and variance of 250 and Division 2 error for *Group A* analysts are drawn from a normal distribution with mean 0 and variance of 250. On the other hand, all *Group B* analysts have the same Division 1 error term but different Division 2 error terms. That is, if analyst *j* is a *Group B* analyst, then she observes forecast F_j where $F_j = T + \varepsilon_1 + \varepsilon_{2j}$. The error ε_1 is common for all *Group B* analysts. But, error ε_{2j} is different for each *Group B* analyst. Hence, forecast errors from *Group B* analysts are correlated but not the same. Note that, in this session, the Division 1 error for *Group B* analysts are drawn from a normal distribution with mean 0 and variance of 250 and Division 2 errors for *Group B* analysts are drawn from a normal distribution with mean 0 and variance of 15.

Description of a Period

You will receive 10 forecasts regarding the EPS for the stock. The forecasts will be generated according to the system described above. Observing the forecasts, you will be asked to enter your prediction for the true EPS (see Figure 1). Let us refer to your entered prediction as *P*. Once you have entered *P*, we will calculate the squared loss, which is defined as $(P - T)^2$ where *T* is the true EPS. We will also independently draw a number *K* randomly from a Uniform distribution on $[0,40]$. If *K* is larger than or equal to the squared loss, you will receive 100 points. If *K* is smaller than the squared loss, you will receive no point. Thus, it is optimal for you to report what you think the true EPS is, on average, based on the 10 forecasts you receive as your prediction.

In a given period, all other participants in the room also receive 10 forecasts for the same stock as you do. However, each of them observe a completely different set of forecasts. All the participants will submit a prediction for the true EPS based on the forecasts they receive as explained above. Once all of you have reported your predictions. We will report to everyone the average of predictions from all the participants in the room (including yours). Then we will ask you to submit a revised prediction (see Figure 2). We refer to this revised prediction as P^r. After you have entered P^r, we will calculate the squared loss for this entry $(P^r - T)^2$. Note that, P^r can be same as or different from your original prediction P. We will also independently draw a number K^r randomly from a Uniform distribution on $[0,8]$. If K^r is larger than or equal to $(P^r - T)^2$, you will receive 50 points. If K^r is smaller than $(P^r - T)^2$, you will receive no point. Thus, it is optimal for you to report what you think the true EPS is, on average, based on the 10 forecasts and the average of all participants' predictions as your revised prediction.

Differences between Periods

In each period of this session, we will consider a different company. The EPS for the stock of the company in each period will be independently drawn from a distribution with a very large variance. Thus, the true EPS across different periods are independent of each other and the EPS of the stock for one company does not provide information about the EPS of the stock for another company across different periods.

Recall that, in each period, some of the 10 forecasts will be from *Group A* analyst and the rest will be from *Group B* analysts. In periods 1 to 10, there will be 1 *Group A* and 9 *Group B* analysts. In periods 11 to 20, the number of *Group A* and *Group B* analysts will be 3 and 7, respectively. In periods 21 to 30, these numbers will be 5 and 5, respectively, and in periods 31 to 40, these numbers will be 7 and 3, respectively. In periods 41 to 50, the number of *Group A* and *Group B* analysts will be 9 and 1, respectively. That is, in the first 50 periods, all participants will have the same combination of *Group A* and *Group B* analysts, but the combination will change (for everyone) every 10 periods.

In all of the periods, first you will enter your prediction for the EPS of the stock in that period. Then you will receive the average of every participants' predictions and will be asked to enter a revised prediction for the EPS. After entering the prediction and then the revised prediction, you will be informed of the true EPS, the squared loss for both the prediction and revised prediction, the relevant random numbers (K and K^r) drawn from uniform distributions and your earnings (in points) from the prediction and the revised prediction.

Ending the Session

At the end of the session, you will see a screen displaying your point earnings from each period. In addition to the participating fee, you will earn an amount based on your point earnings from one randomly chosen period from each 10-period block. Points earned in these periods will be converted to money at the rate of __. You will be paid this amount in cash at the end of the session.

Some Statistical Concepts

Mean (Average) is calculated by summing the observed numerical values of a variable in a set of data and then dividing the total by the number of observations involved. If we have data set (or sample) with n data points and the data points are x_1, x_2,…, x_n then the sample average, $\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n}$.

Variance is the average of the squared differences between each of the observations in a set of data and the mean. Variance is used to indicate how possible values are spread around the mean. Then the sample variance, $s^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}$.

Example: Suppose our data set has 5 data points — $x_1 = 1$, $x_2 = 5$, $x_3 = 7$, $x_4 = 3$, $x_5 = 12$

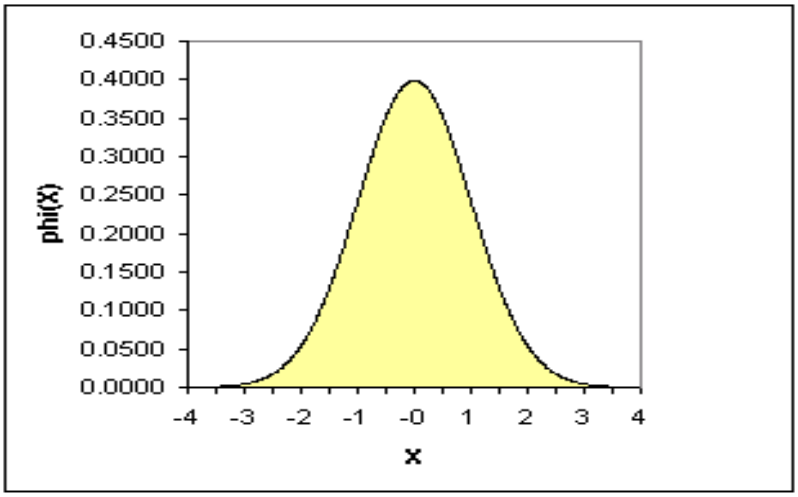
Mean: $\bar{x} = \frac{x_1+x_2+x_3+x_4+x_5}{5} = (1 + 5 + 7 + 3 + 12)/5 = 5.6$

Variance: $s^2 = [(1 – 5.6)^2 + (5 – 5.6)^2 + (7 – 5.6)^2 + (3 – 5.6)^2 + (12 – 5.6)^2]/5 = 14.24$

Distributions

If a variable is drawn from a **Uniform Distribution** on the interval [a, b] that means that all points on [a, b] are equally likely to be drawn.

Normal Distribution is pattern for the distribution of a set of data which follows a bell shaped curve. The following example is a normal distribution with mean of 0 and variance of 1. We refer to this distribution as the standard normal distribution. A picture appears below. The data points (x) can take values from negative to positive infinity and the probability density at point x is given by $\Phi(x)$.



The probability distribution of a normal distribution with mean M and variance S^2 is as follows. That is, if the random number x is drawn from such a distribution then:

Prob($x \le M - 1.28160\ S$) = 0.1	Prob($x \le M - 0.84162\ S$) = 0.2	Prob($x \le M - 0.52440\ S$) = 0.3
Prob($x \le M - 0.25335\ S$) = 0.4	Prob($x \le M$) = 0.5	Prob($x \le M + 0.25335\ S$) = 0.6
Prob($x \le M + 0.52440\ S$) = 0.7	Prob($x \le M + 0.84162\ S$) = 0.8	Prob($x \le M + 1.28160\ S$) = 0.9

Independent draws from the same distribution are those draws selected from the same distribution which have no effect on one another. That is, no correlation exists between the draws.