Overinference from Weak Signals and Underinference from Strong Signals

Ned Augenblick[*], Eben Lazarus[†], and Michael Thaler[‡]

First version: September 2021
This version: December 2022

Abstract

We use experimental and empirical data to show that whether people overreact or underreact to news depends on how informative the news is. Numerous experiments have studied how people respond to highly informative news; however, in many important settings, people typically receive only *weakly* informative news. In an experiment, we show that people systematically overinfer when signals are weak despite underinferring when signals are strong. Misperceptions are more severe for subjects who are less experienced and less cognitively sophisticated. These results align with a model of cognitive imprecision we adapt to study misperceptions about signal informativeness. We then use complementary empirical data from betting and financial markets to show how these patterns play out in high-stakes real-world settings. We develop an empirical method to test over- vs. underreaction in observed prices across differing signal-strength environments, using time to resolution as an observable correlate of signal strength. When signals are weak, prices exhibit excess movement, indicating overreaction; the pattern is reversed with stronger signals, in accord with our theory and experiment.

JEL classification: C91; D83; D91; G14; G41

[*]UC Berkeley Haas. Email: ned@haas.berkeley.edu.
[†]MIT Sloan. Email: elazarus@mit.edu.
[‡]University College London. Email: michael.thaler@ucl.ac.uk.
This paper was originally circulated as "Overinference from Weak Signals, Underinference from Strong Signals, and the Psychophysics of Interpreting Information" by Thaler in September 2021. We would like to especially thank Matthew Rabin for his advice throughout this project. We are also grateful to Roland Benabou, Ben Enke, Christine Exley, Alessandro Lizzeri, Terrance Odean, Pietro Ortoleva, Mike Woodford, Leeat Yariv, and seminar participants at Princeton for their helpful feedback. The preregistration plan is available at `https://aspredicted.org/ax4wg.pdf`. The experiment was approved by the IRB at Princeton University (13114-03).

1 Introduction

How do people react when they receive new information? Given the importance of this question, there is an extensive experimental, empirical, and theoretical literature documenting and explaining how people update their beliefs. The vast majority of the experimental literature focuses on people's reactions to signals that are relatively *strong*, with the common finding that people *underreact*. In the real world, meanwhile, people often receive a constant stream of new information that is only *weakly* informative about future outcomes. In this paper, we use experimental and empirical data to show that these patterns reverse when people are faced with weak signals, finding that individuals and markets both underreact to strong signals and overreact to weak signals.

To understand how both findings can arise, we present a simple model based on theories of cognitive imprecision that predicts that individuals will both underinfer to strong pieces of information and overinfer to weak pieces of information. To provide evidence for this simple theory, we first present evidence from a tightly controlled experiment in which we can experimentally vary signal strengths. We then construct a new empirical strategy to study updating in betting and financial markets (in which the signal strengths are unknown). In both cases, we provide new evidence that supports our hypothesis that people underinfer from strong signals but overinfer from weak signals. These findings help reconcile a set of seemingly contradictory findings between and within the experimental and empirical literatures.

We start with the observation that people often face many pieces of new information that are only weakly diagnostic about future events. For example, a person might receive information about a grade on a test. This is perhaps a strong signal of the grade a person will receive in the class, but it is a weak (though still meaningful) signal about the likelihood of graduating with a high GPA, the likelihood of obtaining a job, and so on. Similarly, day-to-day changes in current events or the stock market may have an impact on the likelihood of reelection for an incumbent candidate, but these effects are likely to be small far in advance of an election. We hypothesize that — contrary to the finding of underreaction in the experimental literature — people often overreact to these weak signals, causing their beliefs to bounce around excessively. These beliefs could lead people to choose career paths inefficiently, stereotype from weakly informative observable characteristics, and excessively buy and sell in markets. Likewise, they could lead to excessive demand for weakly informative tests or frequent news updates that report on individual anecdotes instead of aggregated data.

Such weak-signal settings are understudied in the experimental literature. The online appendix of Benjamin (2019), for example, lists over 500 experimental treatment blocks across 21 papers that study inference from binary signals about a binary state. In *none* of these papers do subjects see signals that have a diagnosticity p — the likelihood of seeing an "high" signal conditional on the "high" state — lower than 3/5. Belief updating in these settings often features underreaction to new information relative to the Bayesian benchmark, particularly when signals are very strong, and much of the past literature has settled on underinference as the dominant form of probabilistic reasoning.[1] Given the limited space of signals studied, however, the experimental literature has had little to say about the reaction to less-informative signals.

The conclusions are more ambiguous outside of experimental environments. One prominent strand of such work, which will be the focus of our empirical analysis, uses asset-price data to obtain observational evidence on belief formation.[2] Given that asset prices encode both beliefs and preferences, definitive evidence on the true source of observed price patterns (whether behavioral or rational) is difficult to come by. But even confining attention to the potential role of updating-related biases, the literature has advanced a range of different, often seemingly contradictory, explanations for various patterns observed across financial markets. Aggregate equity valuations often appear significantly more volatile than measures of fundamental value (Shiller 1981), which would seem to be consistent with overreaction to fundamental information.[3] For individual stocks, meanwhile, the robust empirical phenomenon of *post-earnings announcement drift* — in which a firm's stock price tends to continue drifting in the same direction as the initial change after an earnings announcement — might appear to be consistent instead with *under*reaction.[4] We seek to reconcile the apparently disparate findings across settings by focusing on the informativeness of the relevant signals. Market-level fundamental news is characterized by a constant stream of generally weak, uninformative signals; company earnings announcements, by comparison, are significantly more informative (Vuolteenaho 2002), similar to the signals often studied in experimental settings.[5]

Our core goal is to determine how reaction to information varies when signals are weak versus strong. This is relatively straightforward in an experimental setting, where we have full control of the data-generating process (DGP): the main treatment of our pre-registered experiment is designed to more comprehensively vary the strength of the signal in a simple environment. In particular, in the simplest treatment block in our experiment, we expose 500

[1]According to Benjamin's (2019) survey of the literature, for example, "Stylized Fact 1" is that "Underinference is by far the dominant direction of bias."

[2]Another line of work uses beliefs as elicited in surveys, often regarding macroeconomic outcomes; see Angeletos, Huo, and Sastry (2021) for a survey. Our focus for now is on asset-price data, as financial markets provide useful high-frequency proxies for beliefs over multiple horizons and signal-strength environments.

[3]See, for example, De Bondt and Thaler (1985) for a discussion of this possible connection; more recent work includes Giglio and Kelly (2018), Bordalo et al. (2022), and Augenblick and Lazarus (2022).

[4]See Bernard and J. K. Thomas (1989) for an early survey of the possible link between post-earnings drift and underreaction, and Kwon and Tang (2022) for a more recent discussion.

[5]Afrouzi et al. (2021) pursue a qualitatively similar explanation in an experimental forecasting setting.

people to 16 different signal strengths in a version of a commonly-used (e.g. Phillips and Edwards 1966) "bookbag-and-poker-chips" experimental design in which people receive one binary signal about a binary state.

The main result of our experiment is that people underinfer from strong signals but overinfer from weak signals. In the simple treatment block, subjects have a prior of $\pi_0 = 1/2$ on a binary state (G or V) and receive a signal s with diagnosticity $p > 1/2$ that would move a Bayesian to a posterior of $\pi_1 = p$. Consistent with nearly all existing evidence, for $p \geq 2/3$, we find robust evidence for underinference: subjects form posteriors that are significantly less than p. For $p \in [3/5, 2/3]$, subjects form posteriors of approximately p.[6] However, for $p \in (1/2, 3/5)$, we find clear and robust evidence for overinference: subjects form posteriors that are statistically significantly greater than p, and the effect is convex in p. For very weak signals, subjects act as if signals are twice as strong as they truly are.

To provide a theoretical framework to understand our results, we outline a theory using ideas from psychophysics (Fechner 1860).[7] Specifically, we use a meta-Bayesian model in which people are *cognitively imprecise* about the signal strength $\mathbb{S}$ (similar to models in Petzschner, Glasauer, and Stephan (2015) and Woodford (2020)). As a consequence, people's average perception of signal strength — $\hat{\mathbb{S}} = k \cdot \mathbb{S}^\beta$ for $\beta \in (0, 1)$ — is biased toward their prior.[8] For our data, we estimate that $\beta = 0.76$ and $k = 0.88$, and that there is a *switching point* of $p = 0.64$ at which people switch from overinference to underinference. We believe there are other plausible theories that predict a similar switching point: our main goal is not a tight test to differentiate between these theories, but rather to demonstrate this core effect. That said, we do find that the shape of the empirical updating response maps quite closely to the functional form from the theory.

We find further evidence that cognitive imprecision may explain these misperceptions by exploring heterogeneous treatment effects. Consistent with cognitive sophistication playing a role, subjects with higher scores on a cognitive reflection test (a modified version of the classic test in Frederick 2005) both infer less from weak signals and infer more from strong signals. Relatedly, subjects with more experience within the experiment misinfer less in both directions. Finally, consistent with imprecision affecting inference, we find that subjects

[6]As discussed in Benjamin (2019), evidence in this range is mixed, though often favoring underinference (e.g., Phillips and Edwards 1966 and Edwards 1968). However, there is evidence from recent work, such as Ambuehl and S. Li (2018) and others with asymmetric signals, that finds overinference for $p = 3/5$.

[7]The theory was originally used to explain stimuli such as the weight of an object and brightness of a light; it has been more recently used in economics to explain misperceptions of numbers, probabilities, and risk and ambiguity (Kaufman et al. 1949; Khaw, Z. Li, and Woodford 2021; Enke and Graeber 2021; Frydman and Jin 2020). We differ in using psychophysics to study informational signal strengths. In subsequent work, Ba, Bohren, and Imas (2022) use a similar model.

[8]Stevens (1946) first popularized the power function for misperceptions of physical stimuli, and Khaw, Z. Li, and Woodford (2021) use it for misperceptions of values and lotteries.

with higher variance in their perception of signals misinfer *more* in both directions. These heterogeneities provide further evidence that cognitive imprecision affects belief updating, expanding upon the work by Woodford (2020) and Enke and Graeber (2021) in different decision problems.

Next, we find that biases in demand for information tend to mimic biases in inference. Subjects are asked to purchase 0-3 signals of varying strengths. Compared to a payoff-maximizing benchmark, subjects purchase too many weak signals and too few strong signals, indicating that they are insufficiently sensitive to the quality of information.[9] We also find consistent evidence that subjects underreact to the *number* of signals they receive (as previously documented by Griffin and Tversky 1992; Benjamin, Rabin, and Raymond 2016; and Benjamin, Moore, and Rabin 2018), which may be another factor that leads people to deviate from payoff-maximizing information acquisition levels.

We use additional experimental treatments to explore mechanisms behind these findings. We show that our main results are driven by distorting signal processing, not by distorting probabilities (unlike, for instance, Kahneman and Tversky 1979; Tversky and Kahneman 1992; and Enke and Graeber 2021). We are also able to distinguish between signal processing distortions and probability distortions by considering how subjects infer from *asymmetric* signals. A signal that happens to have a likelihood of close to 0.5 does not systematically affect inference, suggesting that it is the signal that is misperceived as opposed to its associated probability.[10] Additionally, we show that, when faced with signals of an *uncertain* strength, results suggest that people act as if they misperceive *possible realizations* of the signal.

There are reasons to think that the behavior we observe in the experiment may not carry over to real-world situations. First, the benefit for getting the answer right in an experiment is relatively small, so one may think that people will correct their biases in higher-stakes environments. Second, the setting in which the experiment takes place is unfamiliar to subjects, so they may rely more on heuristics; in real-world settings with more sophisticated agents, perhaps familiarity and experience will drive beliefs towards true probabilities.

In light of these concerns, we next turn to evidence from more realistic high-stakes settings by studying the movement of market-implied probability distributions in betting and financial markets.[11] A key challenge with testing any theory of updating in real-world observational data is that we no longer have direct knowledge of the true DGP or signal strength. To

[9] These results are qualitatively consistent with those of Ambuehl and S. Li (2018), but with substantially more evidence for over-purchasing due to our emphasis on weak signals.

[10] We also consider other confounds that arise because of the particular design of the experiment (such as a general bias toward high or low numbers), but do not find evidence that these confounds drive results.

[11] One cost of this analysis is that we must use equilibrium prices (which reflect the beliefs and risk preferences of the marginal investor) rather than individual beliefs. We nonetheless find that market-implied beliefs align well with our theory, so these results serve as a useful external check on our experimental results.

overcome this challenge, we develop a new empirical method based on theoretical results from Augenblick and Rabin (2021) and Augenblick and Lazarus (2022). The core intuition of these papers is that, when a Bayesian is changing their beliefs over time about some event, they must be learning something and therefore (on average) must reduce their uncertainty correspondingly. This intuition can be formalized by defining *movement* as the sum of the squared deviations of changes in beliefs over time, and *uncertainty reduction* as the drop in perceived variance in the outcome.[12] While movement and uncertainty reduction may differ for a given signal realization, they must be equal in expectation across signal realizations, *regardless of the DGP*. This insight allows for a DGP-agnostic test of Bayesian updating in observational data. And crucially for our analysis, these statistics are intuitively and theoretically related to over- and underreaction: overreaction will lead to positive excess movement (movement greater than uncertainty reduction) on average, while underreaction will lead to too little movement relative to the reduction in uncertainty.

While this allows for an intuitive test of over- vs. underreaction with an unknown DGP, to test our theory, we also need to distinguish situations in which signals are weak versus strong. Given that the signal strength is also unobservable, we must instead turn to an ex ante known (and observable) separating variable that is related to signal strength. Our innovation here is to focus on *time to resolution* as our separating variable. Intuitively, when a person is predicting the value of the S&P 500 in three months, information today should generally not lead to much belief movement; meanwhile, market-related information today is highly informative for the value of the S&P tomorrow, and we should accordingly observe more movement of short-horizon beliefs in response to information.[13]

Our theory then suggests that there should be overreaction (and too much movement) at long forecast horizons, and underreaction (and too little movement) at short horizons. We first confirm, by simulating simple DGPs matching our empirical setting, that this pattern exists for an agent who misperceives according to our theory, and that this pattern differs from that under Bayesian updating or constant under- or overreaction. We then turn to the data. We first study sports-betting data, making use of over 5 million transactions from a large sports prediction market for five major sports (corresponding to more than 20,000 sporting events). We then study S&P 500 index option markets, using option-implied beliefs regarding the future value of the S&P from over 4 million daily option prices (corresponding

[12]For example, assume that a Bayesian's subjective prior that some outcome will take place is $\pi_0 = 0.25$. Then it must be the case that $\mathbb{E}[(\pi_1 - 0.25)^2]$ (expected movement, where the expectation is taken over possible date-1 signals) is equal to $0.25 \times 0.75 - \mathbb{E}[\pi_1(1 - \pi_1)]$ (uncertainty reduction).

[13]The relationship between the time horizon and Bayesian belief movement of course depends on the exact DGP. We show that the predicted positive relationship holds very strongly in simulations of game-like DGPs, and more importantly, it clearly holds in our empirical settings.

to 5,500 trading dates and 955 option expiration dates).[14]

We start by plotting the average movement and uncertainty reduction over time for the sports and financial data. Both uncertainty reduction and movement are generally increasing as the event continues and resolution approaches. However, movement is generally higher than uncertainty reduction at the beginning of the event and lower at the end of the event. For example, in the options data, there is very little daily uncertainty reduction until a few weeks before the contract expires, but beliefs consistently move back and forth, generating positive movement. In other words, news today appears to hold relatively little information about the value of the S&P in multiple months, but the market acts as if has more diagnosticity. However, within two weeks of a contract's resolution, the relationship reverses: movement is either less than or equal to uncertainty reduction. That is, as signals become stronger, the market begins to underreact. On net, total movement averaged over an entire option contract is too high, matching the finding of excess movement in Augenblick and Lazarus (2022). But this overall average masks meaningful heterogeneity as one varies the signal strength, with markets misperceiving in the manner predicted by our theory. The same broad pattern holds in the sports-betting data, for all sports (with the partial exception of hockey).

To statistically test this relationship, we regress average movement on average uncertainty reduction in a given time window (where time is measured as distance to resolution). While a Bayesian should have a constant of 0 and slope of 1 in such a regression, we show that the constant is significantly positive and the slope is significantly less than one (again with the exception of hockey).[15] The positive constant indicates that when signals are likely to be weak (uncertainty reduction is low), there is too much belief movement. The less than one-for-one slope indicates that as signals become stronger (uncertainty reduction increases), there is a point at which the relationship flips, and there is too little belief movement.

Given that we cannot observe true signal strength, we must rely on our indirect measure (time to resolution) to test the relationship between signal strength and over- vs. underreaction in real-world high-stakes settings. The strength of our experimental setting, meanwhile, is that these variables are fully observable, but the experiment is lower-stakes and less realistic. The two settings thus provide complementary evidence for our theory, whose predictions align well with both sets of data.

Our main results relate to many papers on the topic of belief updating, including several papers that present evidence for other forms of over- and underreaction to information.

[14] The option-implied probability distribution is also referred to as the *risk-neutral* distribution. It coincides with the marginal trader's subjective distribution only in the case that the trader is risk-neutral; otherwise it incorporates risk aversion. We discuss how we deal with this complication in detail in Section 5.

[15] We discuss a variety of econometric issues such as attenuation bias, and show that the richness of the data allows us to tightly bound the effect of such bias.

Classically, Griffin and Tversky (1992) show that subjects underreact to weight relative to strength, and Massey and Wu (2005) show that subjects overreact to a signal as compared to the environment that produces the signal. More recently, Gonçalves, Libgober, and Willis (2022) find underreaction to strong signals but even further underreaction to the retraction of those signals, while Bhuller and Sigstad (2021) find overreaction to the retraction of certain types of information that may be classified as weak. This paper also relates to the use of cognitive imprecision in inference problems, such as in Enke and Graeber (2021) and Toma and Bell (2022), that finds relationships between underreaction to strong pieces of evidence and cognitive imprecision. Since other forms of decision problems can lead to overreaction, as shown in papers like Fan, Liang, and Peng (2021) and Afrouzi et al. (2021), we keep the decision problem constant.

More recently, Ba, Bohren, and Imas (2022) run an experiment that confirms many of the patterns that we initially documented in our experiment: they also vary signal diagnosticity and find overreaction to weak signals and underreaction to strong signals and propose similar explanations. Using other treatments, they additionally find evidence that complexity plays an important role in increasing cognitive imprecision. By contrast, they do not analyze non-experimental data, whereas we provide additional empirical evidence from sports-betting and financial markets. Our results are thus further related to a large literature on asset prices and beliefs, as introduced above (on p. 2) and surveyed recently in Barberis (2018).

The paper proceeds as follows. Section 2 introduces the theory; Section 3 details the experimental design; Section 4 discusses the experimental results; Section 5 analyzes the market data; and Section 6 concludes. The Appendix contains additional results and empirical details; screenshots of the pages in the experiment are in the Online Appendix.

2 Theory

Setup

We consider individuals who may misinterpret information they receive. In the environment we study, people form a prior at time 0, then receive a signal s about a binary state $\theta \in \{0, 1\}$, and form a posterior at time 1. An individual has a prior $\pi_0 \equiv P(\theta = 1)$. The signal either sends a low message or a high message (s_L or s_H), with $P(\theta = 1|s_H) > P(\theta = 1|s_L)$. After seeing signal s, the individual forms a posterior belief $\pi_1(s) \equiv P(\theta = 1|s)$, which we typically shorten to π_1. Defining the logit function $\text{logit}(\pi) \equiv \log\left(\frac{\pi}{1-\pi}\right)$, a Bayesian would form a posterior such that $\text{logit}(\pi_1) = \text{logit}(\pi_0) + \log\left(\frac{P(s|\theta=1)}{P(s|\theta=0)}\right)$ (using the setup developed in Grether 1980). We consider a *misperceiving person* who may over- or underinfer from information and instead forms a posterior of $\text{logit}(\hat{\pi}_1) = \text{logit}(\pi_0) + w \log\left(\frac{P(s|\theta=1)}{P(s|\theta=0)}\right)$.

Misperception of Signal Strength

The strength of a signal is defined by its log odds ratio: $\mathbb{S} \equiv \left|\log\left(\frac{P(s|\theta=1)}{P(s|\theta=0)}\right)\right| \in [0, \infty)$. We study how w depends on s through $\mathbb{S}$. We posit that when people receive a signal of strength $\mathbb{S}$, they systematically distort their perception of its strength. Following Petzschner, Glasauer, and Stephan (2015), and Woodford (2020), individuals in our model misperceive $\mathbb{S}$ via a form of cognitive imprecision. First, as described by the classic Weber-Fechner law, they interpret changes in $\mathbb{S}$ in relative terms instead of in absolute terms (Fechner 1860). Second, they are uncertain as to what the true signal strength is.

Specifically, we follow Woodford (2020) in assuming that a misperceiving person represents signal strengths $\mathbb{S}$ with a value r, which is drawn from a normal distribution: $r \sim \mathcal{N}(\log \mathbb{S}, \eta^2)$.[16] They have a prior belief about how strong signals tend to be, and this is distributed log-normally: $\log \mathbb{S} \sim \mathcal{N}(\mu_0, \sigma^2)$ and update using Bayes' rule. When observing the signal, their perceived expected value of $\mathbb{S}$ will be equal to:

$$\hat{\mathbb{S}}(r) = \mathbb{E}[\mathbb{S}|r] = \exp\left[\left(\frac{\sigma^2}{\sigma^2+\eta^2}\right)\log\bar{\mathbb{S}} + \left(\frac{\eta^2}{\sigma^2+\eta^2}\right) r\right],$$

where $\bar{\mathbb{S}} \equiv \exp(\mu_0 + \sigma^2/2)$ is the prior mean over signal strengths.

Therefore, their perception $\hat{\mathbb{S}}$ of the true signal strength $\mathbb{S}$ is distributed log-normally with mean equal to

$$\mathbb{E}[\hat{\mathbb{S}}] = k \cdot \mathbb{S}^\beta, \tag{1}$$

where $k \equiv \exp(\beta^2\eta^2/2) \cdot \bar{\mathbb{S}}^{1-\beta}$ and $\beta \equiv \sigma^2/(\sigma^2+\eta^2) \in (0, 1)$.

Our main object of interest in this paper is the average level of over- and underinference; studying the rest of the distribution is beyond the scope of this paper. So, for simplicity, we will assume that people perceive $\mathbb{S}$ to be *exactly* the expectation given by Equation (1); that is, when they see a signal of strength $\mathbb{S}$, they always perceive it to be of strength $k \cdot \mathbb{S}^\beta$. This simplification makes the functional form equivalent to the classical one from Stevens (1946); it makes the analysis more straightforward, and results are not qualitatively affected by it.[17]

That is, on average, the misperceiving person interprets a signal of strength $\mathbb{S}$ as having strength $k \cdot \mathbb{S}^\beta$ for $\beta \in (0, 1)$. Another way to describe this misperception is to say that

[16] The Weber-Fechner law says that r has a mean of $\log \mathbb{S}$, and cognitive imprecision says that individuals face uncertainty about their perception of $\mathbb{S}$.

[17] But this assumption is not *quantitatively* innocuous, as we are taking nonlinear transformations of the function.

the misperceiving person is mis-weighting signals: by treating signal $\mathbb{S}$ as having strength $k \cdot \mathbb{S}^{\beta} = k \cdot \mathbb{S}^{-(1-\beta)} \cdot \mathbb{S}$, they are effectively weighting a signal with strength $\mathbb{S}$ by a factor of $k \cdot \mathbb{S}^{-(1-\beta)}$. There is a *switching point* $\mathbb{S}^* \equiv k^{\frac{1}{1-\beta}}$ for which this weight is below 1 (leading to overreaction) if $\mathbb{S} < \mathbb{S}^*$ and above 1 (leading to underreaction) if $\mathbb{S} > \mathbb{S}^*$.

Updating Given Symmetric Signals

We now focus on *symmetric signal* structures, which we use in the experiment, in which each signal has the same *precision* $p \equiv P(s_H|\theta = 1) = P(s_L|\theta = 0) > 1/2$, and therefore the strength of each signal is $\mathbb{S} = |\text{logit}(p)|$. When the prior is $\pi_0 = 1/2$ (as in our experiment), then the Bayesian posterior equals the signal precision $\pi_1 = p$. This leads to simplified formulas, which allows for a set of simple two-dimensional graphs that represent the core relationships in the model (we will reproduce these same graphs using experimental data in the next section).

While the Bayesian perceives signals as having the correct strength $\mathbb{S}$, the misperceiving person instead treats each signal as having a mean strength of

$$\hat{\mathbb{S}} = k \cdot (\text{logit(p)})^{-(1-\beta)}. \tag{2}$$

The top panel of Figure 1 plots the relationship between perceived and true strength (given the parameters k and β we estimate from the experiment). For comparison, we also plot the relationship for a Bayesian (in which they are equal) and a person who exhibits constant conservativism (for which perceived strength is always lower than reality). [18] The misperceiving person in our model overreacts to weak signals, but underreacts to small signals.

When a Bayesian observes a symmetric signal, they will update to either p or $1 - p$. Instead, the misperceiving person will update to either $\hat{p}$ or $1 - \hat{p}$ at the mean perceived signal strength, where:

$$\hat{p} = \frac{\exp(k \cdot (\text{logit(p)})^{\beta})}{1 + \exp(k \cdot (\text{logit(p)})^{\beta})}. \tag{3}$$

That is, $\hat{p}$ can be seen as the perceived precision of the signal. The middle panel of Figure 1 plots the mapping between the perceived and true precision, which is another way of capturing the core error. The misperceiving person believes that signals with little precision are more precise than reality, and signals with high precision are less precise. Note that, as signals become arbitrarily informative or uninformative, their weighting function becomes arbitrarily

[18] Note that constant conservatism, which corresponds to $\beta = 1$ (and $k < 1$), is not achievable with this functional form, since $\beta = 1$ implies $\eta = 0$, but $\eta = 0$ implies $k = 1$.

small or large, but their posteriors converge to the Bayesian's.

Following the discussion above, a final interpretation is that the individual is mis-weighting signals. Defining $\hat{w}(p)$ as the weighting function given precision p yields:

$$\hat{w}(p) \equiv k \cdot (\text{logit}(\text{p}))^{-(1-\beta)}, \tag{4}$$

The bottom panel of Figure 1 plots the weighting function relative to the precision p, again given our experimental parameters. The misperceiving person places too much weight on weak signals and too little weight on strong signals.

Model Predictions

Figure 1 shows some properties of the weighting function and in particular shows the agent switching from overinference to underinference as signal strength increases. Proposition 1 shows that this is a general property.

Proposition 1 (Misweighting information)
The weighting function $\hat{w}(p)$ from Equation (4) and the precision $\hat{p}$ from Equation (3) have the following properties:

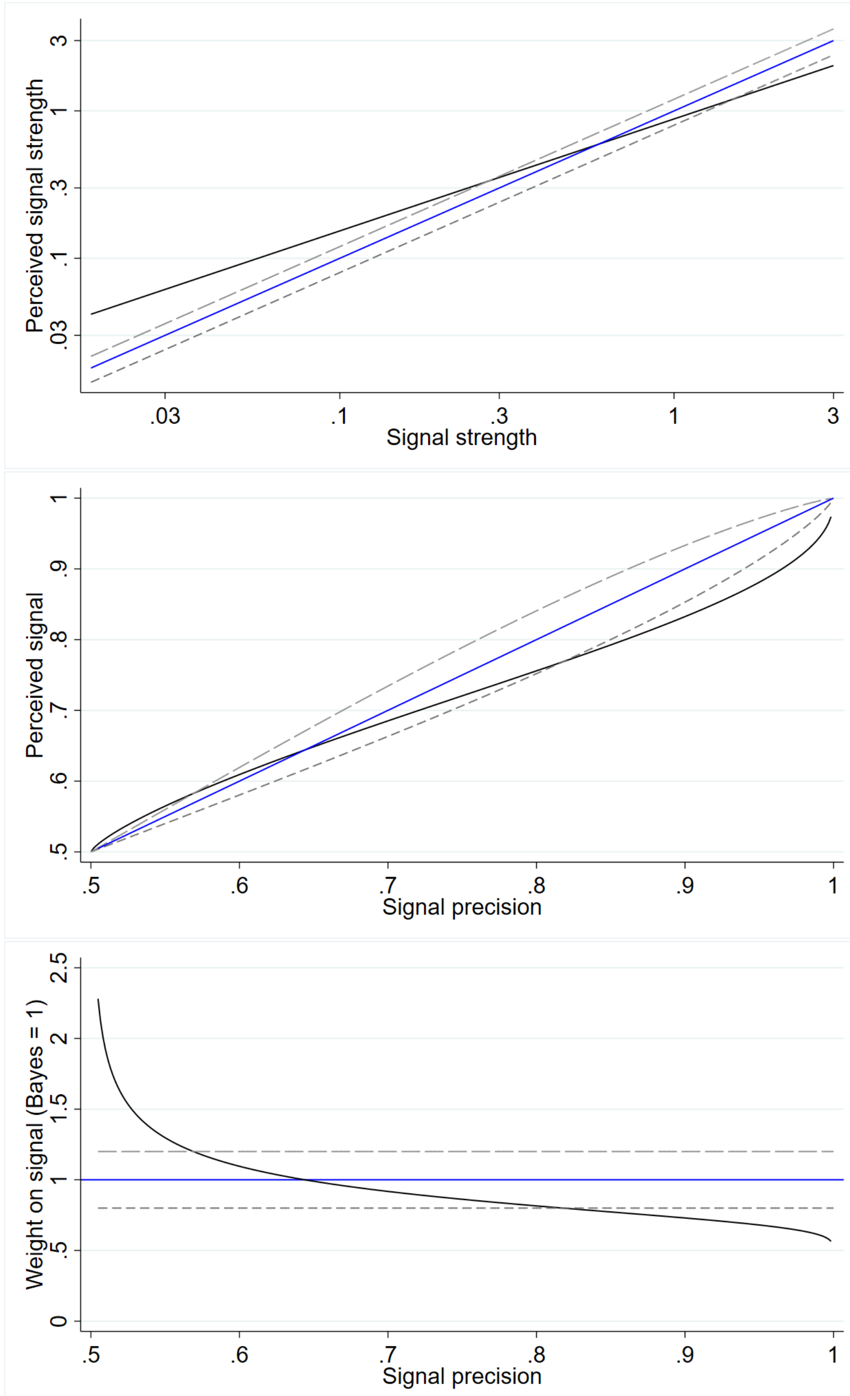
1. *$\hat{w}(p)$ is decreasing in p.*
2. *There exists a switching point $p^* \equiv \frac{k^{\frac{1}{1-\beta}}}{1+k^{\frac{1}{1-\beta}}}$ such that:*
   - *If the Bayesian perceives the precision to be $p < p^*$, $\hat{p} > p$ and $\hat{w}(p) > 1$.*
   - *If the Bayesian perceives the precision to be $p > p^*$, $\hat{p} < p$ and $\hat{w}(p) < 1$.*

Increasing cognitive imprecision, as measured by σ^2, moves people further from Bayesian updating for both strong and weak signals. For sufficiently large p, $\frac{\partial w(p)}{\partial \sigma} < 0$, and for sufficiently small p, $\frac{\partial w(p)}{\partial \sigma} > 0$.[19]

The misperceiving person's instrumental value for a signal tracks signal perceptions. Suppose that after receiving the signal, the person is asked to take an action a. They receive utility $1-(1-a)^2$ if $\theta = 1$ and $1-a^2$ if $\theta = 0$. In this setup, a person maximizes expected utility by taking the action corresponding to their belief. Given this choice, a person with belief $\hat{p}$ will expect to earn utility $\hat{p}(1-(1-\hat{p})^2)+(1-\hat{p})(1-\hat{p}^2) = 1-\hat{p}(1-\hat{p})$. An individual who is aware that they will perceive a signal using Equation (3) will therefore instrumentally value a signal that moves beliefs from 1/2 to $\hat{p}$ at $1/4 - \hat{p}(1-\hat{p})$. When individuals start with a prior of 1/2, this means that overinference will lead to excess demand for information and underinference will lead to too little demand, and there will be a switching point at p^*.

[19]Note that both k and β are functions of σ, so p^* will also change.

Figure 1: Theoretical Predictions of Over- and Underinference by Signal Strength



Notes: These figures provide three representations of the core deviation in our model. Blue lines correspond to Bayesian updating (correct perception of signal strength $\mathbb{S}$), short dashed lines to underinference (with perceived signal strength $0.8 \cdot \mathbb{S}$), long dashed lines to overinference (perceived signal strength $1.2 \cdot \mathbb{S}$) and black lines to the misperception in our model (perceived signal strength $k \cdot \mathbb{S}^{\beta}$ with $k = 0.882$ and $\beta = 0.760$). The top panel plots signal strength perception as a function of signal strength on a log-log scale. The middle panel plots the perceived precision as a function of the true precision. The bottom panel plots the weight put on signals as a function of the true precision. All figures show that misperceiving individuals overweight weak signals and underweight strong signals.

3 Experiment Design

This section overviews the design and the sample, then outlines the timing and treatment blocks, and lastly discusses further details of each treatment block. The Online Appendix contains screenshots of each type of page.

3.1 Overview and Data

Subjects were told that a computer has randomly selected one of two decks of cards — a Green deck or a Purple deck — and both decks are equally likely ex ante. Each deck has Diamond cards and Spade cards: the Green deck has D_1 Diamonds and $N - D_1$ Spades, and the Purple deck has D_2 Diamonds and $N - D_2$ Spades. While subjects see the numbers themselves, we will discuss results in terms of shares: Green has share p_1 Diamonds and $1 - p_1$ Spades, and Purple has share p_2 Diamonds and $1 - p_2$ Spades.

On most questions, subjects draw one (or multiple) cards from the selected deck, observe their suit, and are asked to predict the percent chance of the Green deck and Purple deck after observing the cards drawn. These probabilities are restricted to be between 0 and 100 percent and the sum of the percent chances is required to equal 100. One signal corresponds to a draw of one card. On other questions, they are asked to purchase cards prior to seeing their realization. This design is adapted from Green, Halbert, and Robinson (1965).

The experiment used monetary incentives to elicit subjects' true beliefs, as incentives have been shown to improve decision-making in these settings (e.g. Grether 1992). We implement a version of the binarized scoring rule (Hossain and Okui 2013) that is easier for subjects to comprehend: *paired-uniform scoring* (Vespa and Wilson 2017).[20] Subjects' answers determine the probability that they win a high bonus as opposed to a low bonus.[21]

The experiment was conducted in March 2021. Subjects were recruited on the online platform Prolific (prolific.co). Prolific was designed by social scientists in order to attain more representative samples online; it has been shown to perform well relative to other subject pools (Gupta, Rigotti, and Wilson 2021). 552 subjects completed the experiment. Of these, 500 subjects (91 percent) passed the attention check. As preregistered, analyses are restricted to these 500 subjects.

[20] In general, binarized scoring rules are better able to account for risk aversion and hedging (Azrieli, Chambers, and Healy 2018).

[21] Specifically, five subjects were randomly chosen to win bonuses. If they won the high bonus, they received \$100; if they won the low bonus, they received \$10. All subjects received a \$3 show-up fee and the average bonus earnings for the selected subjects was \$82.

3.2 Timing

Subjects saw the following five treatment blocks: (1) one symmetric signal, (2) one asymmetric signal, (3) three symmetric signals, (4) demand for information, (5) uncertain signals. Details of each are in the subsequent subsections. The ordering of when subjects saw each treatment block was as follows:

Rounds	**Treatment Block**	**Frequency**	**Observations**
1-12	One symmetric signal	67 percent	4,036
1-12	One asymmetric signal	33 percent	1,964
13	Attention check	100 percent	500
14-18	Three symmetric signals	100 percent	2,500
19-23	Demand for information	100 percent	2,500
24-25	One uncertain signal	100 percent	1,000

Excluding the attention check, there are 8,000 observations. The 4,036 symmetric signals include 3,964 informative and 72 completely uninformative signals. Except when noted, analyses are restricted to the 7,928 informative-signal observations.

Questions within each treatment block were randomized for each subject. The ordering of treatment blocks (besides "one symmetric" and "one asymmetric") were fixed for ease of subject comprehension.[22] The fixed order makes cross-treatment comparisons more difficult than within-treatment comparisons.

3.3 Treatment Blocks

One Symmetric Signal

This is the case where $p \equiv p_1 = 1 - p_2$. The Bayesian posterior is equal to p or $1 - p$.

There were 32 possible values of p within the range [0.047, 0.495] or [0.505, 0.953]. These values correspond to 16 possible signal strengths ($\mathbb{S} = |\text{logit } p|$) in the range $\mathbb{S} \in [0.02, 3.00]$.[23] On each question, we randomized whether the Green deck or Purple deck had more Diamonds or Spades, which suit was chosen, and whether the deck consisted of 1665 cards or 337 cards.

[22]For instance, subjects do not see the "demand for information" treatment until they have played rounds in which they inferred from one signal and from multiple signals. Uncertain signals come after demand because they do not reflect the signals that subjects would purchase.

[23]More specifically, we chose whole numbers of cards such that signal strengths would be closest to the following values: $\{0.02, 0.05, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50, 3.00\}$.

To further ensure that answers are sensible, some subjects also received an uninformative signal in which they draw a Diamond or Spade when p is exactly 1/2.[24]

One Asymmetric Signal

This is the case where $p_1 \neq 1 - p_2$. The Bayesian posterior is $\frac{p_1}{p_1+p_2}$ given a Diamond and $\frac{1-p_1}{(1-p_1)+(1-p_2)}$ given a Spade.

One of the probabilities was fixed at 0.505 or 0.495 and the other probability was randomly chosen from $\{0.18, 0.35, 0.65, 0.82\}$. We randomized which of p_1 or p_2 corresponds to the near-0.5 probability, which deck had more Diamonds, which suit was chosen, and whether the deck consisted of 1665 or 337 cards.

Three Symmetric Signals

In this treatment block, subjects saw three simultaneous symmetric signals, each of strength $\mathbb{S}$, but possibly in different directions. For a given p, subjects randomly saw (3 Diamonds, 0 Spades), (2 Diamonds, 1 Spade), (1 Diamond, 2 Spades), or (0 Diamonds, 3 Spades) with equal probability.

There were 14 possible values of $p \in [0.047, 0.495] \cup [0.505, 0.953]$, which corresponded to 14 possible signal strengths. These values were chosen so there would be overlap between the posteriors for five (3,0) signals and five (2,1) signals for a Bayesian, allowing for a sharper comparison. Suit distribution and the card drawn were randomly varied as above; the deck size was fixed for this treatment block at 1665.

Demand for Information

In this treatment block, subjects were given the distribution of cards, but not given a draw of a card. Signals were (known to be) symmetric, so the expected signal strength was p regardless of the signal itself. We fix $p < 0.75$. Unlike before, subjects were asked to choose how many cards they would like to purchase, with costs being a convex function of the number of cards. In particular, they were given the option to:

- Draw 0 cards for a cost of \$0;
- Draw 1 card for a cost of \$0.50;
- Draw 2 cards for a cost of \$1.50; or
- Draw 3 cards for a cost of \$3.

[24] Subjects indeed treat this signal as uninformative; 96 percent of subjects give a posterior of exactly 50 percent.

These numbers were chosen such that purchasing 0 cards is optimal for all $p \in [0.5, 0.57]$, purchasing 1 card is optimal for $p \in [0.57, 0.62]$, purchasing 2 cards is optimal for $p \in [0.62, 0.67]$, and purchasing 3 cards is optimal for $p \in [0.67, 0.75]$.

We set five possible values of $p \in \{0.512, 0.525, 0.550, 0.622, 0.731\}$ such that optimal level for purchasing ranges from 0 to 3. Suit distribution and the card drawn were randomly varied as above and the deck size was fixed at 1665.

Uncertain Signals

In this treatment block, subjects saw a signal with p that was either equal to p_L or p_H with equal probability. Subjects randomly saw one of the following sets of distributions: $\{(p_L = 0.495, p_H = 0.817), (p_L = 0.505, p_H = 0.807), (p_L = 0.495, p_H = 0.193), (p_L = 0.505, p_H = 0.183)\}$. Note that a Bayesian would have the same perceived signal strength in each of these cases.

Suit distribution and the card drawn were randomly varied as above and the deck size was fixed at 1665.

4 Experiment Results

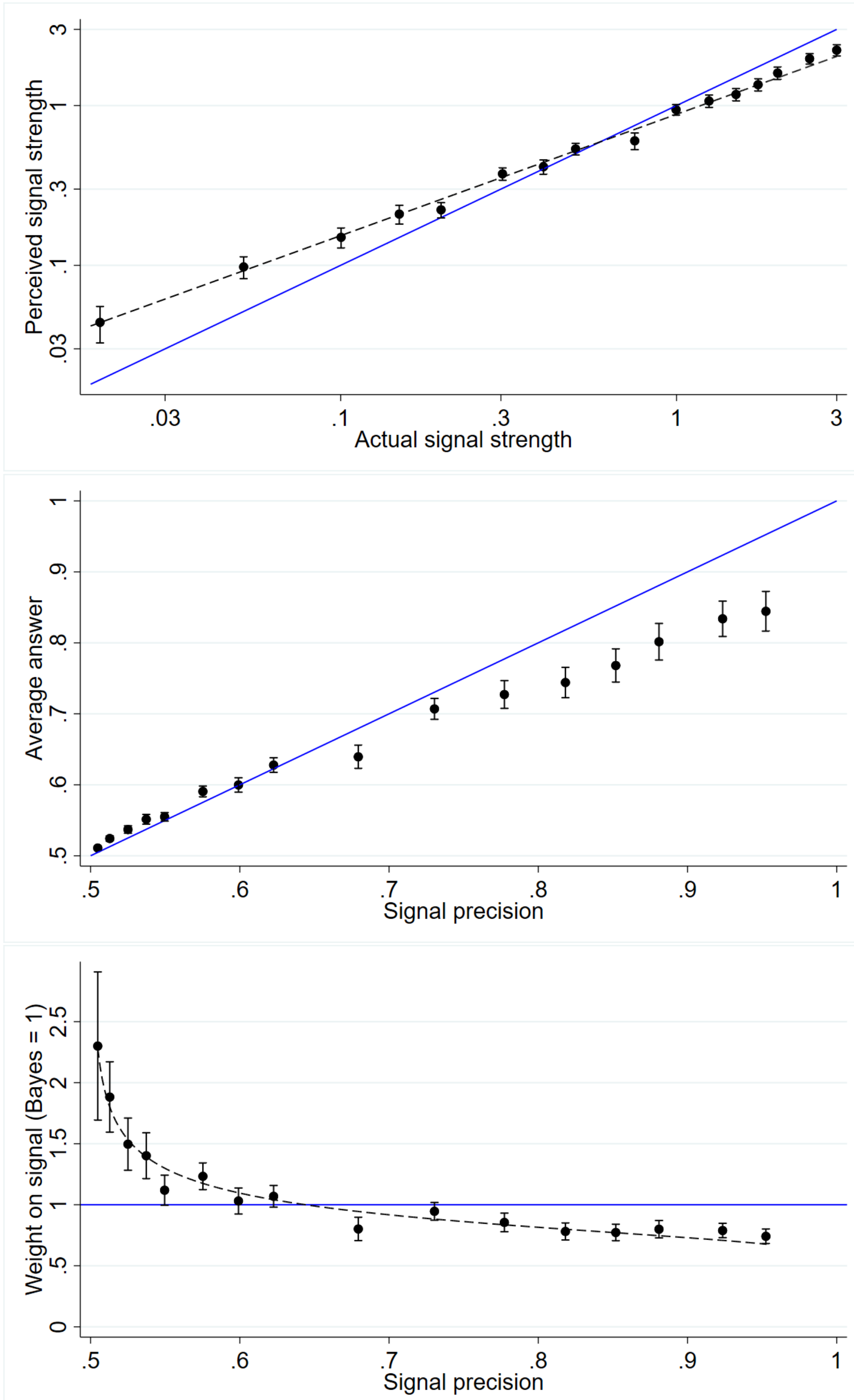
This section discusses the results for each treatment block. The first subsection presents the main results from inference from one symmetric signal and explores heterogeneity. Next, results about inference from multiple signals and demand for information are presented. After exploring potential mechanisms, and how the other treatment blocks help disentangle them, we discuss robustness exercises to probe the strength of the main results.

4.1 Main Results

First, we consider how subjects infer from one symmetric signal. As is consistent with the theory, subjects have a systematic pattern of overinference from weak signals, underinference from strong signals, and a switching point. The figure shows nonlinearity in weighting as a function of signal strength and that a power function fits the data well.

Figure 2 shows that subjects systematically overinfer from weak signals and underinfer from strong signals and shows that the data clearly reject models in which weight is not a function of the Bayesian posterior. The top panel shows that overinference is approximately linear in signal strength using a log-log plot, which suggests that the power law model is a good fit. The middle panel shows that beliefs are increasing but concave in signal precision. The bottom panel provides more evidence that the effect of signal precision on weight is

Figure 2: Over- and Underinference of Symmetric Signals by Signal Strength



Notes: The top panel plots signal strength perception as a function of true signal strength on a log-log scale. The middle panel plots the signal strength average answer subjects give as a function of the Bayesian posterior. The bottom panel plots the average weight subjects put on signals relative to a Bayesian. Blue lines indicate Bayesian behavior. The dashed line fits the data with a power weighting function using nonlinear least squares. All panels show that subjects overweight weak signals and underweight strong signals. Error bars indicate 95% confidence intervals, clustered at the subject level.

nonlinear. Besides the qualitative explanatory power, the model in Section 2 predicts the levels of overinference and underinference well.

p^* and β are estimated using nonlinear least squares for signal types s:

$$\text{Weight}_s = (\text{logit } p^*)^{-(1-\beta)} \cdot |\text{logit } p_s|^{-(1-\beta)} \quad (5)$$

The estimated value for p^* is 0.64 (s.e. 0.01) and for β is 0.76 (s.e. 0.03). The value of β is statistically significantly less than one (p-value < 0.001). These values correspond to an estimate of k of 0.88 (s.e. 0.02). All standard errors are clustered at the subject level.[25]

4.2 Heterogeneity

Evidence of heterogeneity in treatment effects suggests that cognitive sophistication and imprecision affect misperceptions. The three-item cognitive reflection test (CRT) from Frederick (2005) provides our barometer for cognitive sophistication.[26] The top panel of Figure 3 plots patterns of over- and underinference by CRT score and signal strength.

Subjects who score higher on the CRT infer less from weak signals ($p < 0.6$) and more from strong signals ($p > 0.7$). Moving from a CRT score of 0/3 to 3/3 is associated with a decrease of 0.52 weight on weak signals (s.e. 0.18, p-value = 0.004). Meanwhile, this change is associated with an increase of 0.15 weight on strong signals (s.e. 0.043, p-value < 0.001).[27]

Subjects may become more cognitively precise as they learn how to interpret more signals, so we next consider how subjects behave over the course of this part of the experiment. Results suggest that there is evidence for learning over the course of the experiment and the middle panel of Figure 3 plots over/underinference over the course of the twelve rounds. An increase in one round of the experiment is associated with a decrease of 0.074 weight on weak signals (s.e. 0.018, p-value < 0.001) and an increase of 0.008 weight on strong signals (s.e. 0.004, p-value = 0.039).[28] Appendix Figure A3 shows similar patterns when comparing subjects who have a higher self-reported level of news consumption, where the effects are driven by difference in overinference of weak signals.

Finally, one direct proxy for cognitive precision is the variance in subjects' answers; over the course of the experiment, subjects who have weights that have higher variance are likely

[25]The first column of Table A1 shows similar results using a linear model. There is a statistically significant negative relationship between signal strength and overinference, and the regression model predicts overinference when $\mathbb{S} < 0.308/(1.420 - 1) = 0.700$, which corresponds to a switching point at $p = 0.668$.

[26]We modify the text and answers for the items in the test in case subjects have previously seen the classic version of the CRT. See the Online Appendix for the exact questions.

[27]These regressions include controls for exogenous variation in treatments: round number, signal strength, and deck size.

[28]This regression includes controls for signal strength and deck size.

to be less cognitively precise, leading them to be less sensitive to the true strength of the signals. The bottom panel of Figure 3 ranks subjects by the variance in their weights in the experiment and plots over/underinference by this ranking. Moving from the 75th percentile in variance to the 25th percentile in variance is associated with a decrease of 0.93 weight on weak signals (s.e. 0.11, p-value < 0.001) and an increase of 0.17 weight on strong signals (s.e. 0.03, p-value < 0.001).[29] Note that variance is endogenous to levels, so some of this effect may be mechanistic. Both the CRT and news heterogeneities were preregistered, while the variance and round tests were conducted ex post.

In columns (2)-(5) of Table A1, we show that the heterogeneities are similar when we use a linear model. The interaction between signal strength and CRT score is positive, indicating that the higher the subject's CRT score, the more they infer as signal strength increases. Similarly, the interaction between signal strength and round number is positive, between signal strength and standard deviation of guesses is negative, and between signal strength and self-reported news consumption is positive. All effects are statistically significant at the 1-percent level.

4.3 Multiple Signals

When subjects receive multiple signals, they weight the signals less than if they receive one signal, as is consistent with past literature (as discussed in Benjamin 2019 and Benjamin, Rabin, and Raymond 2016). Appendix Figure A1 shows that subjects put less weight on multiple signals than on one signal, holding the Bayesian posterior fixed. This comparison holds both when signals are all in the same direction and when they are mixed.[30]

These results, as those in Griffin and Tversky (1992), are consistent with subjects misperceiving both the strength and the quantity of the signals. When misperceptions about quantity play a larger role than misperceptions about signal strength, there is a downward shift of the curves when people receive multiple signals.

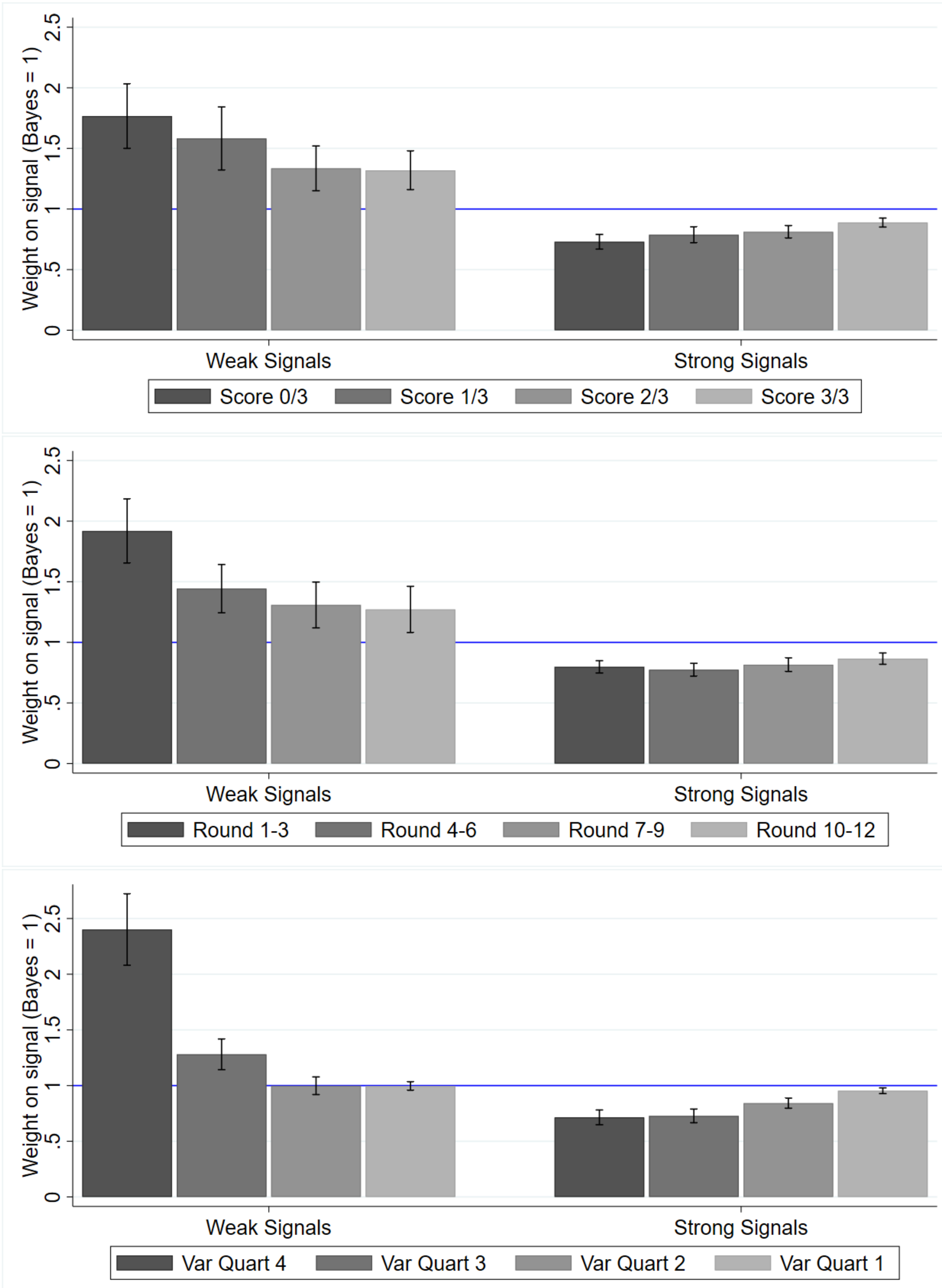
4.4 Demand for Information

Patterns of overinference and underinference also lead to demand that is too high or too low relative to the optimum. Figure 4 plots the average number of signals purchased as a

[29]These regressions include controls for exogenous variation in treatments: round number, signal strength, and deck size.

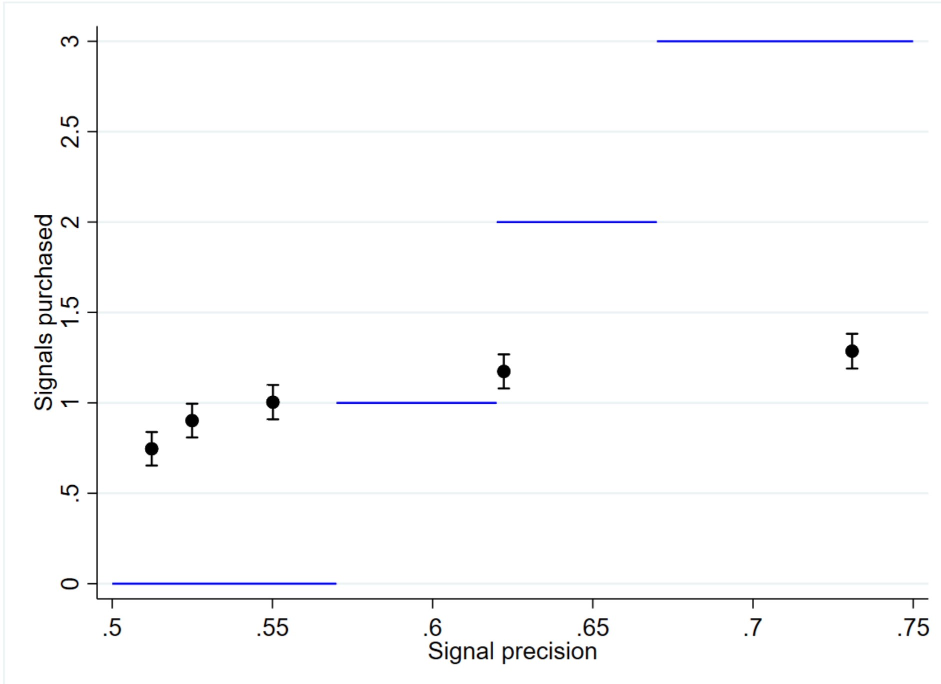
[30]It is worth mentioning that subjects always see the "one signal" treatment block before the "three signals" block. As such, an alternative explanation for these results is that subjects decrease inference over the course of the experiment.

Figure 3: Heterogeneity in Inference from Weak Signals and Strong Signals



Notes: In all three panels, the y-axis is weight subjects put on signals. Weak signals have precision between 0.5 and 0.6; strong signals have precision between 0.7 and 1. The top panel plots weight by CRT score. The middle panel plots weight by round in the experiment. The bottom panel plots weight by how much variance the subject has in their answers. Subjects are ranked by variance in weight and split into quartiles; Quart 1 has the least variance. Blue lines indicate Bayesian updating. This figure shows that higher CRT scores, more experience, and lower variance are associated with less weight on weak signals and more weight on strong signals. Error bars indicate 95% confidence intervals.

Figure 4: Number of Signals Purchased



Notes: This figure plots the number of signals purchased as a function of signal precision. The blue lines correspond to the payoff-maximizing number of signals purchased. This figure shows that subjects over-purchase weak signals and under-purchase strong signals. Error bars indicate 95% confidence intervals.

function of each signal strength, comparing subject behavior to the optimal choice if subjects were Bayesian and only valued signals for their instrumental value.

Figure 4 shows that subjects systematically over-purchase weak signals and under-purchase strong signals. The cost of a signal that leads a Bayesian to form a posterior of less than 0.57 outweighs its benefit; however, the majority of subjects purchase at least one signal when $p = 0.55$ and $p = 0.525$. Additionally, 81 percent of subjects purchase fewer than the optimal level of three signals when $p = 0.73$.

There are other reasons for purchasing signals that are orthogonal to inference. For instance, subjects may be curious to learn the card draw itself, leading them to demand information for its own sake (as in Golman et al. 2021), raising the curve. Subjects' answers are constrained to be between 0 and 3, so they may prefer to select interior answers, flattening the curve.[31]

These effects do not fully explain the results but may contribute to the extremeness. For instance, more subjects purchase zero or one signals when $p = 0.73$ than purchase two or three, which cannot be fully explained by curiosity or interior guesses. There is also suggestive

[31] A preference for stating interior answers would cause the average inference to move closer to 0.5 and in particular would not lead to overinference from weak signals.

evidence that differences in signal purchasing is partly explained by differences in inference at the individual level: subjects who purchase more signals on average have posteriors that move 0.49 pp further (s.e. 0.32 pp, $p = 0.124$) on all other questions.

4.5 Mechanisms

Overreaction and underreaction could be driven by misperceptions of *signal strength* or driven by misperceptions of *probabilities* that happen to be close to 1/2.[32] For instance, if $p_1 = 0.505$ and $p_2 = 0.495$, overinference may be due to misperceptions of weak signals or due to the fact that p_1 is only slightly greater than $1 - p_1$. To disentangle these hypotheses, we compare behavior when $p_1 = 0.505$ and $p_2 = p_H$ to behavior when $p_1 = 0.495$ and $p_2 = p_H$, where $p_H \gg 0.5$.

We find that misperceptions of signal strength explains these findings better. Subjects do not systematically infer differently when $p_1 = 0.505$ and $p_1 = 0.495$. When $p_1 = 0.505$, the average β is 0.54 (s.e. 0.05). When $p_1 = 0.495$, the average β is 0.60 (s.e. 0.05). The difference is not statistically significant (-0.06, s.e. 0.07, p-value = 0.404) and the point estimate is in the opposite direction of that of the alternative model.

Next, we consider how people respond to signals of uncertain strength. Misperceiving people may apply their weighting function and signal combination in one of two ways. They might either first combine the signals and then weight the *combined* signal, or they may first weight the signals *separately* and then combine the weighted components.

Consider a signal that either leads a Bayesian to a posterior of $p_L \gtrless 1/2$ or $p_H > \max\{p_L, 1 - p_L\} > 1/2$, each with probability 1/2. Denoting by $\hat{p}$ the agent's posterior, the "combined signals" model predicts $\text{logit}(\tilde{p}) = K \cdot (\text{logit}[(p_L + p_H)/2)]^{\beta}$. The "separate signals" model predicts $\text{logit}(\tilde{p}) = 1/2 \cdot \left[k * (\text{logit}(p_L/2))^{\beta} + K * (\text{logit}(p_H/2))^{\beta}\right]$ when $p_L > 1/2$ and $\text{logit}(\tilde{p}) = 1/2 \cdot \left[-k * (\text{logit}((1 - p_L)/2))^{\beta} + k * (\text{logit}(p_H/2))^{\beta}\right]$ when $p_L < 1/2$. These models make noticeably different predictions when p_L is close to 1/2. The combined signals model predicts smooth behavior and a higher $\tilde{p}$, while the separate signals model predicts a kink at $p_L = 1/2$ and a lower $\tilde{p}$.

In the experiment, subjects receive ($p_L = 0.505$, $p_H = 0.807$) or ($p_L = 0.495$, $p_H = 0.817$). The Bayesian model predicts posteriors of 0.656. Using the estimated parameters as before ($\beta = 0.760$ and $k = 0.882$), the combined signals model predicts near-Bayesian posteriors of 0.653 in each case. Meanwhile, the separate signals model predicts underinference and asymmetry, with a posterior of 0.636 when the signals are both larger than 0.5 and 0.628 when they are in opposite directions.

[32] Results may also be due to misperceiving the *difference* between priors and posteriors, which would produce similar results to signal misperceptions.

Empirically, the separate signals model fits the data best. The fit is better for both levels and differences. Subjects' posteriors are 0.620 (s.e. 0.057) when the signal directions are aligned and 0.609 (s.e. 0.054) when they are misaligned. In fact, subjects underinfer slightly *more* than the separate signals model predicts (see Liang 2021 for one possible explanation). Underinference is suggestively more severe when the directions are misaligned (p-value = 0.106) and the point estimate for this difference is similar to that of the theoretical prediction.

4.6 Robustness

This section considers three alternative hypotheses for these results that are unrelated to over- or underinference.

First, it is possible that subjects dislike stating "50 percent" even when signals are very uninformative. There is no systematic evidence for this. Among the 72 subjects who see a completely uninformative signal (where the decks each have 832 Diamonds and 832 Spades), 69 of them (96 percent) give an answer of exactly 50 percent.

Second, it is possible that subjects are systematically more inclined to prefer Green over Purple, or attend more to Diamonds over Spades, or vice versa. If this were the case, then subjects may overinfer in one direction and underinfer in the opposite direction by a different amount, leading to average misinference. However, there is no evidence of color asymmetry; subjects' average estimate of P(Green) is 0.503 (s.e. 0.002).[33] There is also no evidence for suit asymmetry; on average, subjects' signal weight is 1.156 (s.e. 0.031) when they see a Diamond and 1.134 (s.e. 0.035) when they see a Spade.

Third, it is possible that something about the particular number of cards in the deck leads subjects to misperceive signal strength. For instance, they may have a left-digit bias (M. Thomas and Morwitz 2005). Recall that for the one-signal questions, the deck size was randomly either 1,665 or 337. The deck size leads to a statistically-significant shift in the estimated switching point p^* but not in the sensitivity β: p^* for the larger deck is 0.68 (s.e. 0.02) and β is 0.75 (s.e. 0.03); p^* for the smaller deck is 0.61 (s.e. 0.01) and β is 0.77 (s.e. 0.04). With each deck size, β is statistically significantly less than 1 (p-values < 0.001).[34]

It is possible that these factors have modest quantitative effects on the main estimates but they cannot entirely explain the results.

[33]On the screen, subjects always are asked for P(Green) before P(Purple), so this also indicates little question-location asymmetry.

[34]One explanation for the shift in p^* is that subjects put some weight on the *differences* between numbers of cards, and don't only focus on the ratios. Such a story is possible but beyond the scope of this paper.

5 Evidence from Finance and Sports Betting

While the experimental setting is tightly controlled, its abstract features naturally raise questions of external validity. To test the implications of our theory in more realistic high-stakes settings, we therefore now consider evidence from a set of sports betting markets and financial markets. Departing from the experimental environment comes at a cost, as we no longer have direct knowledge of the true DGP or the informativeness of specific signals to which prices are responding. But we choose a set of markets to analyze so as to obtain close proxies for signal informativeness under minimal assumptions: we consider the prices of binary bets (with payouts of either \$0 or \$1) with known terminal horizon, for which we can construct proxies for signal informativeness under the null of Bayesian updating. By considering price movements across informativeness regimes, we test whether the patterns of overinference vs. underinference documented in the experiment apply in this real-world financial-market setting. We first describe our setting and estimation strategy in more detail, before turning to our empirical results.

5.1 Conceptual Setup

To set the stage for our empirical analysis, we first describe the conceptual framework that we then take to the data. This section builds closely on Augenblick and Rabin (2021) (AR 2021) and Augenblick and Lazarus (2022) (AL 2022) to generalize Section 2 to handle multiple periods with arbitrary signals. Time is indexed by $t = 0, 1, 2, \ldots, T$. As before, there is a binary state $\theta \in \{0, 1\}$. Each period, a person observes a signal s_t drawn from arbitrary distribution $DGP(s_t \,|\, \theta, H_{t-1})$, where $H_t \equiv \{s_1, s_2, ...\}$ is the history of signal realizations. The person's belief in state 1 at time t given the DGP and history H_t is denoted by $\pi_t(H_t)$ (or π_t for short). The *belief stream* $\boldsymbol{\pi}$ refers to the collection of the person's beliefs over time.

While we cannot directly test for overinference vs. underinference without knowledge of the DGP, keeping track of the following two objects will allow for well-motivated indirect tests. First, define the *movement* of a belief stream from period t_1 to $t_2 > t_1$ as the sum of squared changes of beliefs over these periods:

$$m_{t_1,t_2}(\boldsymbol{\pi}) \equiv \sum\nolimits_{\tau=t_1}^{t_2-1} (\pi_{\tau+1} - \pi_\tau)^2.$$

Then, defining the *uncertainty* of belief at period t as $u_t(\boldsymbol{\pi}) \equiv (1 - \pi_t)\pi_t$, we define *uncertainty reduction* from period t_1 to period $t_2 > t_1$ as:

$$r_{t_1,t_2}(\boldsymbol{\pi}) \equiv \sum\nolimits_{\tau=t_1}^{t_2-1} (u_\tau(\boldsymbol{\pi}) - u_{\tau+1}(\boldsymbol{\pi})) = u_{t_1}(\boldsymbol{\pi}) - u_{t_2}(\boldsymbol{\pi}).$$

For each variable, we define the concomitant random variable in capital letters (e.g., M_{t_1,t_2}).

Under the null model of Bayesian updating, beliefs satisfy $\pi_t(H_t) = \mathbb{E}[\theta \mid H_t]$ for all H_t, where $\mathbb{E}$ is the expectation under the true (*physical*) measure. As in Section 2, we denote the equivalent objects for a person under a non-Bayesian alternative with a hat (e.g., $\hat{\pi}_t$).

The Equality of Movement and Uncertainty

As in AR (2021), the martingale property of beliefs under the null implies that, *regardless of the DGP*, expected Bayesian belief movement from any period t_1 to period t_2 must equal expected uncertainty reduction:

Proposition 2 (Movement and Uncertainty Reduction)
Assume $\pi_t(H_t) = \mathbb{E}[\theta \mid H_t]$. For any DGP and for any periods t_1 and t_2, $\mathbb{E}[M_{t_1,t_2}] = \mathbb{E}[R_{t_1,t_2}]$.

This result formalizes the "correct" amount of belief volatility (or movement) under rationality, without the need to know the true unobservable DGP. One can then follow AR (2021) to use this result as the basis for a statistical test for Bayesian updating. The basic logic is to calculate the empirically-observable difference between movement and uncertainty reduction (which they call "excess movement") from a set of belief streams and then use a simple means test to see if the difference is statistically different from zero. If so, one can reject — with a certain confidence level — that the beliefs arose from Bayesian updating. We emphasize the crucial feature that this test is valid regardless of the DGP.

Underreaction and Overreaction

There is a natural connection between excess movement and overreaction: people who overreact are intuitively moving around "too much." AR (2021) formalize this connection. In particular, Proposition 6 of that paper demonstrates that in a two-period environment, a person with a correct prior who overinfers from signals will exhibit a positive excess movement statistic, while a person who underinfers will exhibit a negative statistic. Proposition 7 then demonstrates that the same relationship holds over many periods in a simple symmetric binary-signal environment (even though the person no longer necessarily has the correct beliefs when in later periods). However, with other DGPs it is challenging to provide general analytic results. Consequently, we instead use simulations to confirm the intuition that the same types of intuitive relationships hold given our updating model in environments that more closely map to our empirical setting.

5.2 Simulated Belief Streams

We use simulations to understand the basic patterns in movement and uncertainty reduction of a person who updates using the model in Section 2 when forming beliefs about an outcome of a sporting event or the future level of the stock market. In these types of events, similar information is received at each period and the aggregate of that information determines the final state. To transparently capture these situations, we consider a simple DGP in which there are two "teams" representing the two states, exactly one team scores in each of 27 periods, each team has equal probability of scoring in each period, and the final state is which team has the highest score after the final period. For example, if a team is up by one score with 2 periods left, they have a 75% chance of being the final winner because they win if they score in one of the future 2 periods.[35]

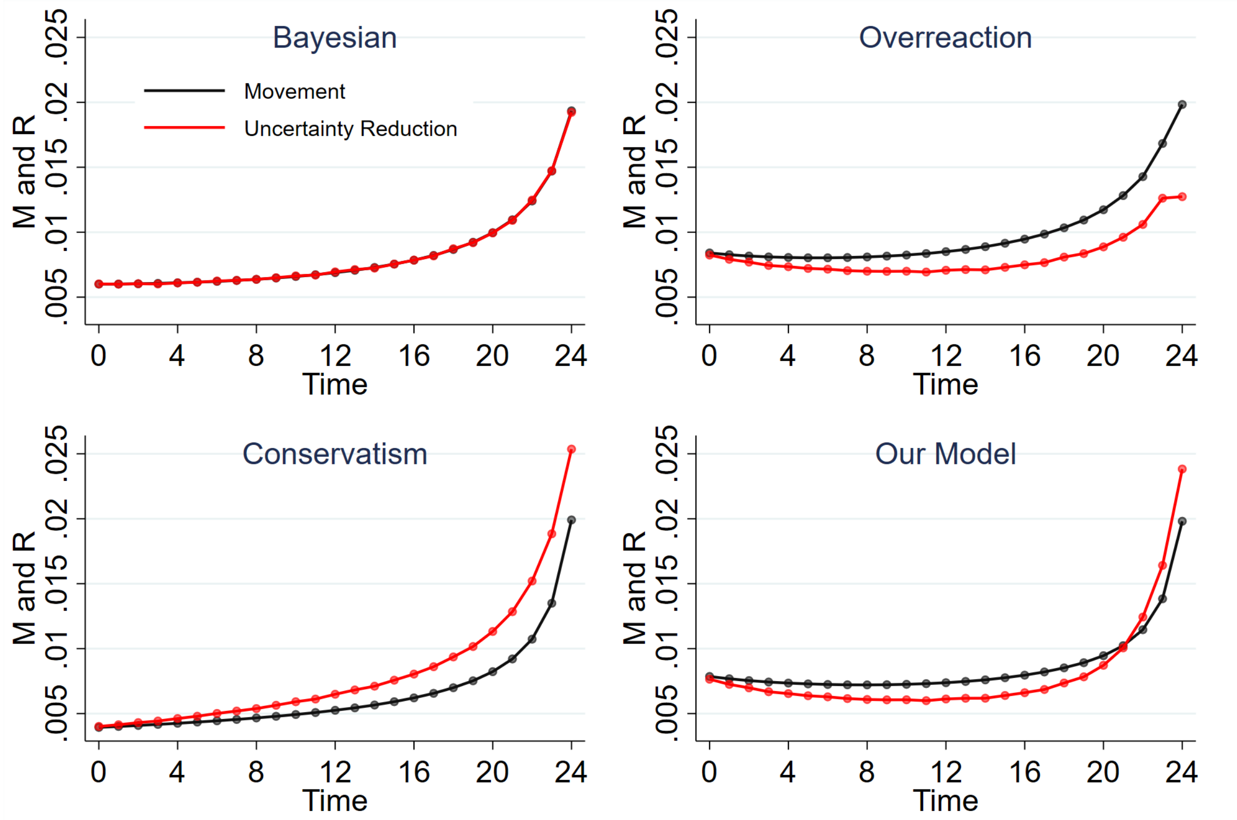
The top-left panel of Figure 5 shows the expected movement and uncertainty reduction statistics over time for a Bayesian (calculated from one million simulations of this DGP). First, note that, as predicted by Proposition 2, the statistics are equal at each period. Next, note that both statistics are rising as the resolution of the game approaches. The initial periods always hold very little information, while the later periods sometimes hold no information (because one team has an insurmountable lead) and sometimes hold a large amount of information (because the scores are very close). Overall though, the average movement, uncertainty reduction, signal strength, and log signal strength are rising over time. As we show shortly, this pattern will also hold in all of our empirical settings.

What do the statistics look like for misperceiving people? Following intuition and the theoretical results from simpler DGPs, overreaction (top-right panel) leads to positive excess movement in every period, while the opposite is true for underreaction (bottom-left panel). The bottom-right panel displays the results for our model with the parameters estimated in Section 4.1 ($\beta = 0.760, k = 0.882$). In the early periods, average signal strength is low, leading to overreaction, which in turn causes excess movement. However, in later periods, the amount of average revealed information is higher, leading to underreaction, which in turn causes too little movement. Consequently, at some midpoint, there is a crossing point at which average movement equals uncertainty reduction. This pattern is consistent with our model, but does not occur in Bayesian updating or when there is universal overreaction or underreaction.

There are multiple ways to determine if the pattern is statistically meaningful given a set

[35] There is one complication in this simple setup. In the first period, beliefs always start at 50%. However, *regardless of the updating rule*, movement and uncertainty reduction are always equal when beliefs start at 50%. Similarly, in the final period, the state has either been determined (in which case both movement and uncertainty reduction are equal) or beliefs are again 50%. Therefore, given that any updating rule will produce zero excess movement for these periods, we drop them.

Figure 5: Simulated Movement and Uncertainty Reduction Over Time For Different Models



Notes: This figure shows the average movement (black lines) and uncertainty reduction (red lines) statistics over time for four different models, averaged over 1 million simulations of the game-like DGP discussed in the text. The updating models are (1) Bayesian updating (the signal strength is correctly perceived as $\mathbb{S}$), (2) conservatism (the perceived signal strength is $0.8 \cdot \mathbb{S}$), (3) overreaction (the perceived signal strength is $1.2 \cdot \mathbb{S}$), and (4) our model (the perceived signal strength is $k \cdot \mathbb{S}^{\beta}$ with $k = 0.882$ and $\beta = 0.760$). For Bayesian updating, these statistics are always equal. For conservatism, movement is always less than uncertainty reduction, and the opposite is true for overreaction. For the misperceiving person in our model, movement is greater than uncertainty reduction in early time periods (where signals are generally weak) and lower in later time periods (where signals are generally strong).

of belief streams. We focus on a simple test: regressing average movement in each period on average uncertainty reduction in each period (or in a collection of respective periods). Given the null hypothesis of Bayesian updating, these statistics should always be equal, so that the constant will be equal to 0 and the slope coefficient equal to 1. However, for a misperceiving person who updates according to our model, movement will be higher than reduction when reduction is low, but lower than reduction when reduction is high, such that the constant will be positive and the slope coefficient will be less than one.

Note, however, that this discussion ignores a variety of empirical challenges which we discuss after introducing the data, including the appropriate definition of a period and the concern of noise leading to attenuation bias. We return to these issues after introducing the data; as will be seen later, the very large number of observations available in our setting effectively obviates these concerns.

5.3 Sports Betting Data

Data Description

We start with data on sports betting. Our data comes from Betfair, which operates a large prediction market in which individuals are matched on an exchange to make opposing financial bets about the outcome of a sporting event. As in a standard centralized financial exchange, one can observe time-stamped transaction prices — determined by supply and demand — for a contract in which one party pays another a set amount given a particular realized outcome of the game. These are the same data as used in AR (2021), and we use the same sample (2006–2014) and data filters (discussed further below) as in that paper.[36]

Given our focus in this section on equilibrium bet-price data, we follow the literature that interprets these prices as "market beliefs."[37] An excess movement test based on Proposition 2, therefore, can be viewed in this setting as a test of the joint null that market prices may be interpreted as beliefs and that these beliefs are Bayesian. But while this might affect the interpretation of full-sample excess movement tests, it poses less of a problem for our purposes. We are *fixing* the environment (i.e., the particular betting market in question) and comparing excess movement as one varies the signal strength (proxied by time to maturity) within this environment. If we assume that the mapping from individual to market beliefs does not change systematically within a stream as one moves closer to maturity, our findings are at minimum directionally informative about both individual and market-level reactions to information across signal-strength regimes.

We focus on markets for five major sports (soccer, American football, baseball, basketball, and hockey), and we restrict attention to contracts over the final winner of the game (and omit more exotic contracts, such as which team will be winning at the midpoint and the

[36]See AR (2021) or Brown and Yang (2019) for a more detailed introduction to the Betfair sports betting exchange data.

[37]The interpretation of betting-market prices as averages of heterogeneous individual beliefs has been studied in a range of work. Gjerstad (2005) and Wolfers and Zitzewitz (2006) show the validity of such an interpretation when traders have log utility and do not trade for speculative purposes (cf. Manski 2006). Ottaviani and Sørensen (2015) consider beliefs' and prices' reaction to information; they show that in such a setting, prices in fact generally *underreact* to information (in contrast to the bulk of our evidence) due to the logic of market clearing. (Martin and Papadimitriou 2022 show the obverse in a setting with speculative trading.) In standard Bayesian settings with complete markets, meanwhile, such an interpretation is straightforward (see AL 2022).

number of goals scored).[38] We use observations only when the game is being played. To remove extremely high-frequency noise, we follow AR (2021) and keep at most one observation per minute, and also drop observations with less than 1% of average volume. Finally, we attempt to have similar timing in events by dropping (less common) events in a category for which the timing of the game is different (such as WNBA games, in which the game time is shorter than the NBA).

Graphs of Movement and Uncertainty Reduction

Figure 6 shows average movement and uncertainty reduction (as well as confidence intervals) across time for each sport. Observations occur in continuous time and therefore must be aggregated in some way. Our data contain clock-time ("1:31pm") rather than game-time ("4:50 through the third quarter") observations; we therefore consider average movement and uncertainty reduction for observations within 24 time chunks, each of which corresponds to 1/24 the length of an average game.[39] As in the simulations, average movement and uncertainty reduction are generally increasing over time (with the exception of mid-period breaks). Early in games for each sport, movement is greater than uncertainty reduction, and for each sport there is a time at which movement drops below uncertainty reduction. For four of the five sports, movement then continues to be lower than uncertainty reduction after this time (for hockey, movement stays lower than uncertainty until the final period). The market accordingly appears to overreact to the less-informative signals at the beginning of a game, and underreact to the more-informative signals at the end of a game.

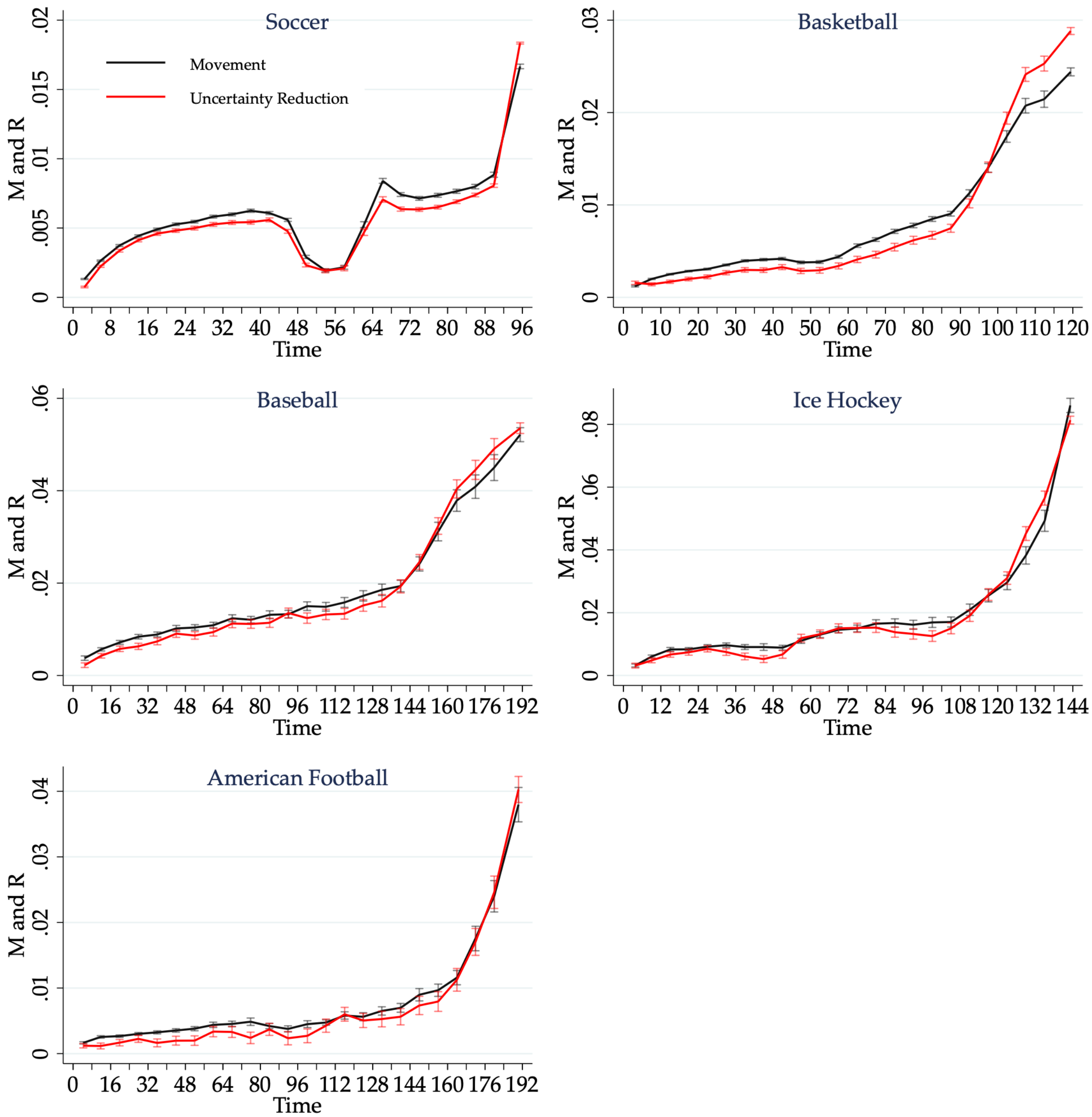
Statistical Tests

There are multiple ways to determine if the patterns above are statistically meaningful. The simplest is to test whether average movement is statistically higher than uncertainty reduction at the start of the game, and statistically lower at the end. The test at the start of the game (performed as a two-sided test) yields p-values below 0.001 except hockey at 0.003, which is perhaps not surprising given the difference and tight confidence intervals from the large amount of data. The test at the end of the game is similar but more heterogeneous, yielding p-values below 0.001 for soccer and basketball and 0.07 for football and baseball, but the difference in hockey has the opposite sign. While these tests provide some evidence

[38]If there are multiple contracts — e.g., one paying off if team A wins, another if team B wins — we use the contract for which the starting beliefs are closest to 0.5.

[39]For example, as the average basketball game lasts around 132 minutes, basketball games are broken into 24 chunks of 5.5 minutes. The final chunk then includes all observations that occur after 132 minutes. Results are similar if we use different numbers of chunks. Separately, in constructing confidence intervals, we assume observations are uncorrelated across contracts.

Figure 6: Movement and Uncertainty Reduction Over Time for Different Sports



Notes: This figure shows the average movement (black lines) and uncertainty reduction (red lines) statistics over time for four different sports (as well as 95% confidence intervals). In each case, movement is greater than uncertainty reduction in early time periods and then drops below in later time periods.

for our hypothesis, they only use a small percentage of the data. For example, although the difference for baseball is only moderately significant in the final period, the difference in the second-to-last period is very significant. Similarly, the second-to-last period for hockey is highly statistically significant in the predicted direction.

In order to use all of the data, we instead focus on another simple test introduced in the simulations above: regressing average movement in each period on average uncertainty reduction in each period. Given the null hypothesis of Bayesian updating, these statistics should always be equal, such that the constant will be equal to 0 and the slope coefficient equal to 1. However, for a misperceiving person who updates according to our model, average movement will be higher than average uncertainty reduction when reduction is low (and signals are weaker), but lower than uncertainty reduction when reduction is high (and signals are stronger), such that the constant will be positive and the slope coefficient will be less than one.

The results for these regressions are shown in the first five columns of Table 1. For four of the five sports, the slope and constant coefficients are highly statistically significantly different from the Bayesian benchmark in the direction predicted by the theory (and for hockey they are significant at the 10-percent level): we consistently observe evidence for overinference from weak signals (when average uncertainty reduction is low) and underinference from strong signals (when reduction is high).

To understand the magnitude of the constant estimates, note that a consistent movement of 3% up and then 3% down would produce movement of 0.0018 (close to the average constant coefficient) given no uncertainty reduction. Given this average constant, the average slope coefficient then implies that movement will cross uncertainty reduction when both are around 0.014, which corresponds to one belief movement of around 12%.

Note that we are regressing the *average* movement in a time chunk on the *average* uncertainty reduction. We do this because, for a Bayesian, the *expected* movement at any time $\mathbb{E}_t[M_{t,t+1}]$ must equal the *expected* uncertainty reduction $\mathbb{E}_t[R_{t,t+1}]$. However, any individual belief change does not yield movement and reduction equal to these expectations, but rather equal to these expectations plus a mean-zero error term. Therefore, if we simply regressed all the (unaveraged) movement statistics on the (unaveraged) uncertainty reduction statistics, there would be significant attenuation bias due to error in the independent variable. By taking the average over tens of thousands of belief changes at a given time, we are able to estimate the expectation at that time plus a tiny error term, thus effectively eliminating this bias. In our case, the estimated variance of the error term at each period is more than 100,000 times smaller than the estimated variance of the independent variable, such that the attenuation bias is negligible.[40] Note that the R^2 values in all cases are very close to 1:

[40]By averaging the movement and uncertainty reduction statistics over time chunks, we face the subjective question of how many chunks to use. The main results do not change significantly when using a different number of chunks. For example, the estimated slope coefficients for the five sports change to (0.839, 0.797, 0.903, 0.987, 0.912) when using 12 chunks (see Appendix Table A2) and to (0.847, 0.849, 0.883, 0.925, 0.920) when using 36 chunks (unreported). The p-values for all sports but hockey remain very highly significant.

Table 1: Regressions of Movement on Uncertainty Reduction

Dep Var:	Sports					Finance	
Movement	Soccer	Basketball	Baseball	Hockey	Football	Raw	Risk-Adj.
Uncert. Red.	0.918	0.806	0.889	0.945	0.912	0.897	0.921
	(0.030)	(0.021)	(0.017)	(0.032)	(0.014)	(0.019)	(0.018)
Constant	0.0009	0.0018	0.0026	0.0018	0.0015	0.0045	0.0048
	(0.0002)	(0.0002)	(0.0005)	(0.0009)	(0.0002)	(0.0009)	(0.0008)
R^2	0.977	0.985	0.995	0.995	0.995	0.990	0.992
Time Chunks	24	24	24	24	24	24	24
Events	6,584	5,176	3,927	4,123	1,390	955	955
Observations	4,589,289	867,567	166,346	109,751	86,193	58,864	58,864
p-val.: $\beta_1 = 1$	0.013	0.000	0.000	0.098	0.000	0.000	0.000
p-val.: $\beta_0 = 0$	0.000	0.000	0.000	0.042	0.000	0.000	0.000

Notes: This table presents the results from OLS regressions (with robust standard errors in parentheses) of average movement in a time period on average uncertainty reduction for four sports and the options data.

variation in movement can be explained almost entirely by variation in uncertainty reduction, but only when the coefficient is less than one.

5.4 Index Options Data

Data Description

The sports betting data provide a useful lab for studying beliefs in an incentivized setting, but they come naturally with some limitations: trading budgets are capped,[41] and payoffs are unrelated to macroeconomically relevant outcomes. So we now consider market-implied beliefs data from a large-scale financial market of first-order importance: options on the S&P 500 index, which are effectively bets on the value of the market index as of some fixed future expiration date.[42] As in AL (2022), we obtain data from OptionMetrics on S&P index option prices with all available strike prices and expiration dates traded on the Chicago Board Options Exchange (CBOE). The sample of trading dates is 1996–2018. We use daily (end-of-day) data, and we implement exactly the same data filters as described in AL (2022);

[41] As of late 2022, for example, Betfair imposes payout limits ranging from £50,000 to £250,000 for single-game bets on American sports.

[42] Denoting the market index price by V_t, a call option with strike price K and expiration date T has a date-T payoff of $\max(V_T - K, 0)$. (Conversely, a put option has payoff $\max(K - V_T, 0)$.)

see that paper for further details.

Converting Option Prices to Market-Implied Beliefs

On any given trading date t, there are prices for a wide range of S&P options with the same expiration date T. The options differ only in their strike prices K (for a call option, the minimum S&P index value at which the option will obtain a positive payoff at expiration). The set of option prices for a given (t, T) pair can thus be translated, using standard methods following Breeden and Litzenberger (1978), into a market-implied (or *risk-neutral*) probability distribution over the future S&P price on the option expiration date.[43]

Unlike in the case of sports betting data, index options have payoffs that are tied (by construction) to the value of aggregate wealth. Option prices therefore reflect risk aversion in addition to subjective probability assessments about the future index value. This is the main complication in using option-implied probability distributions: they do not, in general, correspond to any notion of aggregate subjective beliefs. (They are equivalent to subjective beliefs only in the case of risk neutrality over the index value, which motivates referring to them as risk-neutral beliefs.) For example, suppose that there are two possible date-T macroeconomic states that are perceived by investors as equally likely. If investors value a marginal dollar in the "bad" state (when the market is low) more than in the "good" state (when the market is high), they will be willing to pay more for the option that pays off in that state. If these risk preferences are not taken into account, one will (falsely) conclude that investors believe that the bad state is more likely.

Addressing this issue is the main theoretical task taken up in AL (2022). That paper shows that under certain assumptions, one can place a *bound* on excess movement in risk-neutral (RN) beliefs under the null that underlying subjective beliefs are rational. The bound is tight in the space of possible DGPs — that is, one can construct a DGP under which it holds exactly — but it is not necessarily tight under the true real-world DGP.[44] We therefore provide two sets of results in the current analysis, (1) using the raw RN beliefs, and alternatively (2) translating these beliefs to a set of physical (subjective) beliefs under an assumption on risk aversion. For (2), we consider multiple possible assumptions in translating from risk-neutral to physical beliefs, detailed in Appendix B. While the hundreds of possible assumptions and parameterizations matter in determining the physical belief estimated for a given risk-neutral belief, their effect on our belief-updating statistics (movement and uncertainty reduction) is

[43]This mapping requires only the minimal assumption that there are no arbitrage opportunities in option markets. Intuitively, as options allow for bets over the future index price, their prices allow us to back out a probability distribution over this future price.

[44]While the bound is sufficient for the full-stream tests considered in that paper, it might not be here: we wish to understand how "true" excess movement evolves with signal informativeness within a stream.

so small as to be nearly indetectable.[45] We therefore report estimation results here under our main translation, which assumes a representative investor with power utility over the terminal index value. For robustness, we estimate the statistics under a wide range of alternative parameterizations in the right panel of Appendix Figure A5, finding that these choices have little effect on our results.

To implement our measurement of RN beliefs, we follow AL (2022) and partition the state space of possible date-T index values into discrete ranges. To maintain the same set of states across expiration dates T, we set the states to correspond to ranges for the log return on the S&P 500 from the first observable option trading date to the expiration date. There are 9 potential terminal return states θ_j, with θ_1 realized if the log return in excess of the risk-free rate is less than -0.2 (roughly -20%); θ_2 if the log excess return is in the range $(-0.2, -0.15]$; θ_3 if $(-0.15, -0.1]$; and so on, so that each state (aside from the extreme states θ_1 and θ_9, which we discard in our analysis) corresponds to a five-percentage-point range for the S&P's excess return. The RN belief distribution is then the option-price-implied conditional (date-t) probability distribution over each of these 9 states. We measure this distribution for each trading date t until option expiration, at which point we assign probability 1 to the observed realized return state for the S&P 500. We again aggregate the data into 24 time chunks by trading days to expiration, and we average movement and uncertainty reduction statistics using beliefs over all states θ_j for $j = 2, \ldots, 8$.[46]

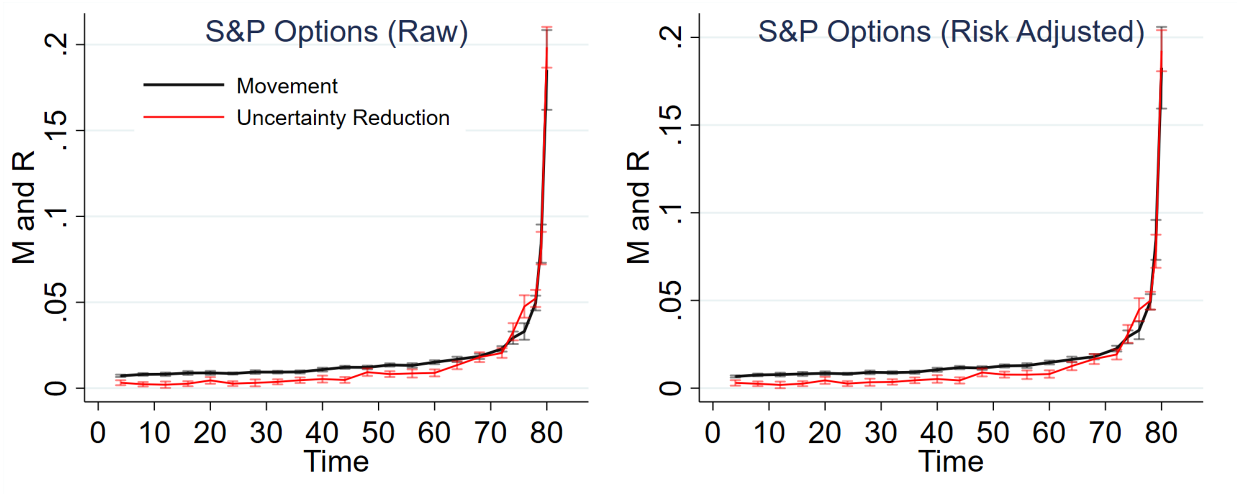
Graphs of Movement and Uncertainty Reduction

Figure 7 shows average movement and uncertainty reduction over time in the options data, analogous to Figure 6. The left panel shows the average movement and uncertainty reduction statistics for the raw risk-neutral beliefs, and the right panel shows the statistics for physical beliefs obtained under the main risk adjustment procedure. Time is in trading days, with date 0 normalized to roughly 4 months (85 trading days) from expiration. In both cases, movement is consistently above uncertainty reduction relatively far from expiration, when signals are only very weakly informative and uncertainty reduction is statistically indistinguishable from zero. Uncertainty reduction increases dramatically closer to expiration (when market movements are more informative regarding the true index value at the expiration date); while option-implied belief movement increases alongside uncertainty reduction, it appears to do

[45] This is intuitive, as risk aversion is unlikely to be changing meaningfully from day to day. In addition, the main point of interest for this analysis is that the basic patterns found in the experimental data and in the sports betting data are also observed in the finance data, regardless of the RN beliefs correction used.

[46] See AR (2021, Prop. 3) for formal justification of this many-state belief aggregation. We use only the interior states $j = 2, \ldots, 8$, as in AL (2022), because the mapping from risk-neutral to physical beliefs is likely to be unstable for the extreme states θ_1 and θ_9 given that they contain beliefs over a wide range of possible returns.

Figure 7: Movement and Uncertainty Reduction Over Time for Finance Data



Notes: This figure shows the average movement (dark solid line) and uncertainty reduction (lighter dotted line) statistics over time for the beliefs implied by option data (with 95% confidence intervals). The left-hand panel uses the implied unadjusted (risk-neutral) beliefs. The right-hand panel uses a risk adjustment described in the text. Movement is clearly greater than uncertainty at the start of the contract. At the later stages, movement is lower than uncertainty reduction.

so less than one for one, with uncertainty reduction crossing above movement roughly 10 trading days from expiration. The patterns observed in this high-stakes financial market are strikingly similar to those in the sports betting data for many sports plotted in Figure 6.

Statistical Tests

We conclude this analysis by conducting the same formal tests as in the previous case, regressing average movement on average uncertainty reduction in each time chunk. The results are shown in the final two columns of Table 1. For both the raw and risk-adjusted data, the estimated slope and constant are again highly statistically significantly different from the Bayesian benchmark in the direction predicted by our theory. The positive constant again indicates overreaction when signal informativeness (uncertainty reduction) is low, as movement is significantly positive in these cases; meanwhile, the slope being less than one (and numerically nearly identical to the estimated slope in the sports betting data) indicates underreaction for high enough levels of signal informativeness. The market therefore appears to misperceive in the way predicted for individuals modeled in Section 2. More broadly, the consistent results from the lab and from observational data indicate that there seems to be a unifying explanation for underreaction and overreaction that applies across settings.

6 Conclusion

The results in this paper have shown robust evidence for overinference from weak signals and underinference from strong signals. They have also shown why it is critical to study inference from weak signals in order to draw general patterns about how people process new information. Data and heterogeneities fit theories of cognitive imprecision that predict that people noisily misperceive weak signals as being too strong and strong signals as being too weak. These misperceptions can also help explain how people demand information, make inferences from multiple signals or from uncertain signals, and how prediction markets and financial markets exhibit different responses to information in different environments.

Further work can relate these patterns of misinference to more applied forms of news consumption and to better understand how to correct these biases. For instance, does overinference from weak signals predict greater susceptibility to the representativeness heuristic and stereotyping (e.g. Grether 1980; Bordalo et al. 2016) at an individual level? Are people with a high β better at distinguishing the quality of informed versus uninformed advisors?

Lastly, results also suggest that increasing cognitive precision (as in Enke and Graeber 2021) may both decrease perceptions of weak signals and increase perceptions of strong signals. As such, feedback interventions may be a promising way to move people's inference processes towards Bayes' rule, leading demand for information to be more sensitive to the quality of the information.

References

Afrouzi, Hassan, Spencer Kwon, Augustin Landier, Yueran Ma, and David Thesmar (2021). "Overreaction in Expectations: Evidence and Theory." In: *Working Paper*.

Ambuehl, Sandro and Shengwu Li (2018). "Belief updating and the demand for information." In: *Games and Economic Behavior*.

Angeletos, George Marios, Zhen Huo, and Karthik A. Sastry (2021). "Imperfect macroeconomic expectations: Evidence and theory." In: *NBER Macroeconomics Annual 2020*.

Augenblick, Ned and Eben Lazarus (2022). "A New Test of Excess Movement in Asset Prices." In: *Working Paper*.

Augenblick, Ned and Matthew Rabin (2021). "Belief Movement, Uncertainty Reduction, and Rational Updating." In: *The Quarterly Journal of Economics*.

Azrieli, Yaron, Christopher Chambers, and Paul Healy (2018). "Incentives in Experiments: A Theoretical Analysis." In: *Journal of Political Economy*.

Ba, Cuimin, J. Aislinn Bohren, and Alex Imas (2022). "Over- and Underreaction to Information." In: *Working Paper*.

Barberis, Nicholas (2013). "Thirty years of prospect theory in economics: A review and assessment." In: *Journal of Economic Perspectives*.

Barberis, Nicholas (2018). "Psychology-Based Models of Asset Prices and Trading Volume." In: *Chapter for the Handbook of Behavioral Economics*.

Beach, Lee Roy, James Wise, and Scott Barclay (1970). "Sample proportions and subjective probability revisions." In: *Organizational Behavior and Human Performance*.

Benjamin, Daniel (2019). "Errors in Probabilistic Reasoning and Judgment Biases." In: *Chapter for the Handbook of Behavioral Economics*.

Benjamin, Daniel, Don Moore, and Matthew Rabin (2018). "Biased Beliefs about Random Samples: Evidence from Two Integrated Experiments." In: *Working Paper*.

Benjamin, Daniel, Matthew Rabin, and Collin Raymond (2016). "A Model of Nonbelief in the Law of Large Numbers." In: *Journal of the European Economic Association*.

Bernard, Victor L. and Jacob K. Thomas (1989). "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" In: *Journal of Accounting Research*.

Bhuller, Manudeep and Henrik Sigstad (2021). "Do Judges Learn? The Causal Effects of Reversals on Judicial Decision-Making." In: *Working Paper*.

Bliss, Robert R. and Nikolaus Panigirtzoglou (2004). "Option-Implied Risk Aversion Estimates." In: *Journal of Finance*.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2016). "Stereotypes." In: *Quarterly Journal of Economics*.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (2022). "Belief Overreaction and Stock Market Puzzles." In: *Working Paper*.

Breeden, Douglas T. and Robert H. Litzenberger (1978). "Prices of State-Contingent Claims Implicit in Option Prices." In: *Journal of Business*.

Brown, Alasdair and Fuyu Yang (2019). "The wisdom of large and small crowds: Evidence from repeated natural experiments in sports betting." In: *International Journal of Forecasting*.

Dave, Chetan and Katherine Wolfe (2003). "On confirmation bias and deviations from Bayesian updating." In: *Working Paper*.

De Bondt, Werner F. M. and Richard Thaler (1985). “Does the Stock Market Overreact?” In: *Journal of Finance*.
Donnell, Michael and Wesley Du Charme (1975). “The effect of Bayesian feedback on learning in an odds estimation task.” In: *Organizational Behavior and Human Performance*.
Edwards, Ward (1968). “Conservatism in human information processing.” In: *Formal Representation of Human Judgment (Ed: Kleinmuntz, B.)*
Enke, Benjamin and Thomas Graeber (2021). “Cognitive Uncertainty.” In: *Working Paper*.
Fan, Tony, Yucheng Liang, and Cameron Peng (2021). “Belief Updating: Inference versus Forecast Revision.” In: *Working Paper*.
Fechner, Gustav (1860). *Elements of Psychophysics*. Breitkopf & Hartel (English translation: Holt, Rinehart, and Winston, 1966).
Frederick, Shane (2005). “Cognitive Reflection and Decision Making.” In: *Journal of Economic Perspectives*.
Frydman, Cary and Lawrence Jin (2020). “Efficient Coding and Risky Choice.” In: *Working Paper*.
Giglio, Stefano and Bryan Kelly (2018). “Excess Volatility: Beyond Discount Rates.” In: *Quarterly Journal of Economics*.
Gjerstad, Steven (2005). “Risk Aversion, Beliefs, and Prediction Market Equilibrium.” In: *Working Paper*.
Golman, Russell, George Loewenstein, Andras Molnar, and Silvia Saccardo (2021). “The Demand for, and Avoidance of, Information.” In: *Working Paper*.
Gonçalves, Duarte, Jonathan Libgober, and Jack Willis (2022). “Learning versus Unlearning: An Experiment on Retractions.” In: *Working Paper*.
Green, Paul, Michael Halbert, and Patrick Robinson (1965). “An Experiment in Probability Estimation.” In: *Journal of Marketing Research*.
Grether, David (1980). “Bayes Rule as a Descriptive Model: The Representativeness Heuristic.” In: *The Quarterly Journal of Economics*.
Grether, David (1992). “Testing Bayes rule and the representativeness heuristic: Some experimental evidence.” In: *Journal of Economic Behavior and Organization*.
Griffin, Dale and Amos Tversky (1992). “The Weighing of Evidence and the Determinants of Confidence.” In: *Cognitive Psychology*.
Gupta, Neeraja, Luca Rigotti, and Alistair Wilson (2021). “The Experimenters’ Dilemma: Inferential Preferences over Populations.” In: *Working Paper*.
Holt, Charles and Angela Smith (2009). “An update on Bayesian updating.” In: *Journal of Economic Behavior and Organization*.
Hossain, Tanjim and Ryo Okui (2013). “The binarized scoring rule.” In: *The Review of Economic Studies*.
Kahneman, Daniel and Amos Tversky (1979). “Prospect theory: An analysis of decision under risk.” In: *Econometrica*.
Kaufman, Edna, Miles Lord, Thomas Whelan Reese, and John Volkman (1949). “The Discrimination of Visual Number.” In: *The American Journal of Psychology*.
Khaw, Mel Win, Ziang Li, and Michael Woodford (2021). “Cognitive Imprecision and Small-Stakes Risk Aversion.” In: *The Review of Economic Studies*.

Kraemer, Carlo and Martin Weber (2004). "How Do People Take into Account Weight, Strength and Quality of Segregated vs. Aggregated Data? Experimental Evidence." In: *Journal of Risk and Uncertainty*.

Kwon, Spencer Yongwook and Johnny Tang (2022). "Extreme Events and Overreaction to News." In: *Working Paper*.

Lazarus, Eben (2022). "Horizon-Dependent Risk Pricing: Evidence from Short-Dated Options." In: *Working Paper*.

Liang, Yucheng (2021). "Learning from Unknown Information Sources." In: *Working Paper*.

Manski, Charles F. (2006). "Interpreting the predictions of prediction markets." In: *Economics Letters*.

Martin, Ian W. R. and Dimitris Papadimitriou (2022). "Sentiment and Speculation in a Market with Heterogeneous Beliefs." In: *American Economic Review*.

Massey, Cade and George Wu (2005). "Detecting regime shifts: the causes of under- and overreaction." In: *Management Science*.

Ottaviani, Marco and Peter Norman Sørensen (2015). "Price Reaction to Information with Heterogeneous Beliefs and Wealth Effects: Underreaction, Momentum, and Reversal." In: *American Economic Review*.

Peterson, Cameron and Alan Miller (1965). "Sensitivity of subjective probability revision." In: *Journal of Experimental Psychology*.

Petzschner, Frederike, Stefan Glasauer, and Klaas Stephan (2015). "A Bayesian perspective on magnitude estimation." In: *Trends in Cognitive Sciences*.

Phillips, Lawrence and Ward Edwards (1966). "Conservatism in a simple probability inference task." In: *Journal of Experimental Psychology*.

Sasaki, Shunichiro and Toshiji Kawagoe (2007). "Belief Updating in Individual and Social Learning: A Field Experiment on the Internet." In: *Working Paper*.

Shiller, Robert J. (1981). "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" In: *American Economic Review*.

Stevens, Stanley Smith (1946). "On the Theory of Scales of Measurement." In: *Science*.

Thomas, Manoj and Vicki Morwitz (2005). "Penny Wise and Pound Foolish: The Left-Digit Effect in Price Cognition." In: *Journal of Consumer Research*.

Toma, Mattie and Elizabeth Bell (2022). "Understanding and Improving Policymakers' Sensitivity to Impact." In: *Working Paper*.

Tversky, Amos and Daniel Kahneman (1992). "Advances in prospect theory: Cumulative representation of uncertainty." In: *Journal of Risk and Uncertainty*.

Vespa, Emanuel and Alistair Wilson (2017). "Paired-Uniform Scoring: Implementing a Binarized Scoring Rule with Non-Mathematical Language." In: *Working Paper*.

Vuolteenaho, Tuomo (2002). "What Drives Firm-Level Stock Returns?" In: *Journal of Finance*.

Wolfers, Justin and Eric Zitzewitz (2006). "Interpreting prediction market prices as probabilities." In: *Working Paper*.

Woodford, Michael (2020). "Modeling Imprecision in Perception, Valuation, and Choice." In: *Annual Review of Economics*.

A Appendix: Additional Results

Table A1: Linear Effect of Signal Strength on Over/Underinference

	(1) Main Effects	(2) By CRT Score	(3) By Round	(4) By SD	(5) By News
Signal Strength	-0.308	-0.481	-0.578	0.149	-0.519
	(0.031)	(0.064)	(0.072)	(0.031)	(0.093)
Strength x CRT Score		0.102			
		(0.028)			
Strength x Round Number			0.042		
			(0.009)		
Strength x SD of Guesses				-0.430	
				(0.038)	
Strength x News Cons					0.334
					(0.129)
Constant	1.420	1.421	1.416	1.398	1.420
	(0.030)	(0.030)	(0.030)	(0.021)	(0.030)
Subject FE	Yes	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes
Observations	3964	3964	3964	3964	3964
R^2	0.23	0.23	0.24	0.29	0.23

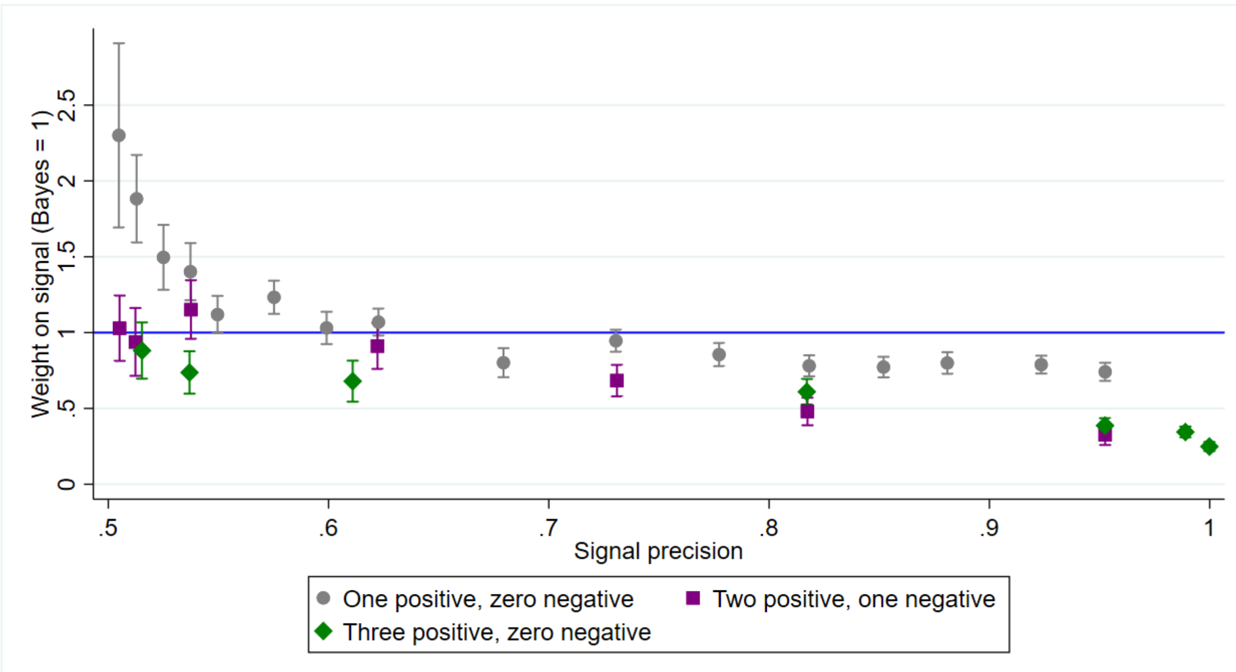
Notes: OLS, with standard errors (in parentheses) clustered at subject level. Dependent variable is the weight put on the signal compared to a Bayesian (as defined in Equation (2).) Weights greater than 1 correspond to overinference; weights less than 1 correspond to underinference. Column (1) shows that subjects overinfer from weak signals (since the constant is greater than 1) and that overinference declines for stronger signals. Columns (2)-(5) show heterogeneity across CRT score, round number, standard deviation of over/underweighting, and news consumption, respectively. CRT score ranges from 0 to 3. Round number ranges from 1 to 12. News consumption ranges from 0 to 1.

Table A2: Regressions of Movement on Uncertainty Reduction: Fewer Time Periods

Dep Var:	Sports					Finance	
Movement	Soccer	Basketball	Baseball	Hockey	Football	Raw	Risk-Adj.
Uncert. Red.	0.839	0.797	0.903	0.987	0.912	0.932	0.903
	(0.033)	(0.024)	(0.017)	(0.032)	(0.016)	(0.020)	(0.022)
Constant	0.0014	0.0018	0.0024	0.0013	0.0015	0.0045	0.0047
	(0.0003)	(0.0003)	(0.0005)	(0.0009)	(0.0002)	(0.0008)	(0.0009)
R^2	0.984	0.991	0.996	0.990	0.997	0.991	0.992
Time Chunks	12	12	12	12	12	16	16
Events	6,584	5,176	3,927	4,123	1,390	955	955
Observations	4,589,289	867,567	166,346	109,751	86,193	27,138	27,138
p-val.: $\beta_1 = 1$	0.000	0.000	0.000	0.685	0.000	0.046	0.000
p-val.: $\beta_0 = 0$	0.000	0.000	0.000	0.162	0.000	0.000	0.000

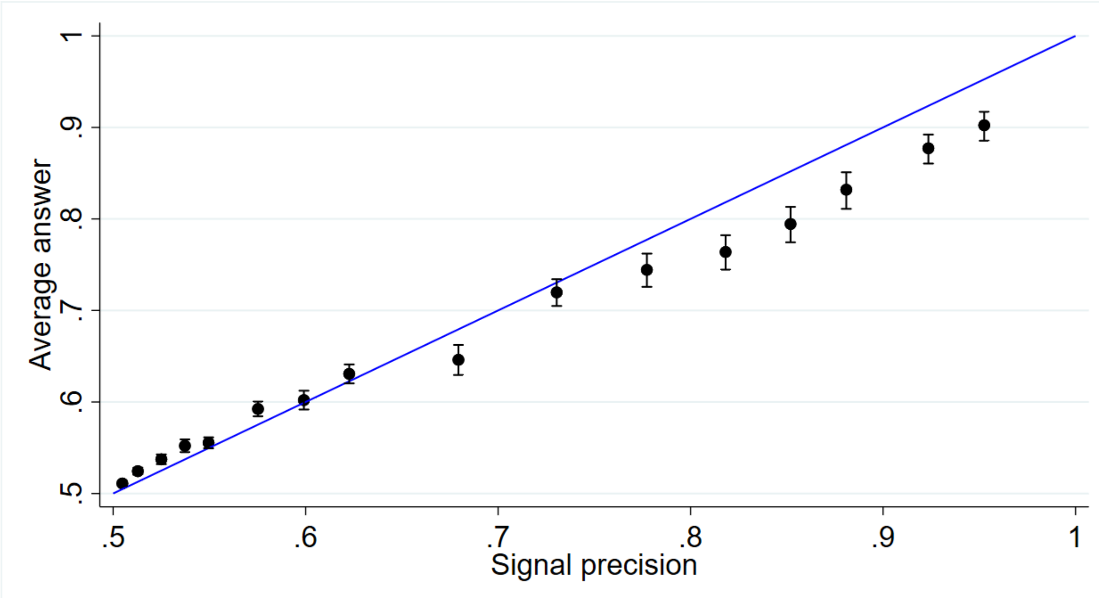
Notes: This table presents the results from OLS regressions of average movement in a time period on average uncertainty reduction for four sports and the financial data. In this table, each contract is split into fewer sub-periods (12 in the case of sports betting data, and 16 in the case of options data) than the 24 periods considered in the main text.

Figure A1: Over- and Underinference by Number and Type of Signals



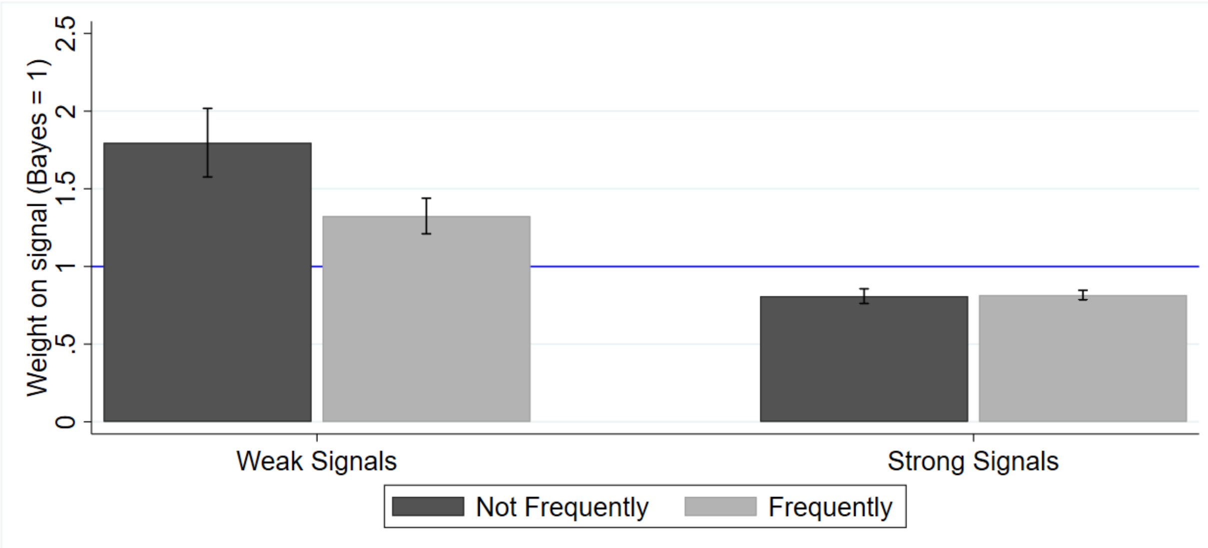
Notes: This figure plots the average weight subjects put on signals relative to a Bayesian (indicated by the blue line), split by signal distribution. Grey circles correspond to one signal of strength S (as in Figure 2); purple squares correspond to two signals of strength S in one direction and one signal of strength S in the opposing direction; green diamonds correspond to three signals of strength $S/3$ in the same the direction. This figure shows that subjects put less weight on three signals as compared to the weight they put on one signal. Error bars indicate 95% confidence intervals.

Figure A2: Over- and Underinference of Symmetric Signals by Signal Strength: Alternative Approach



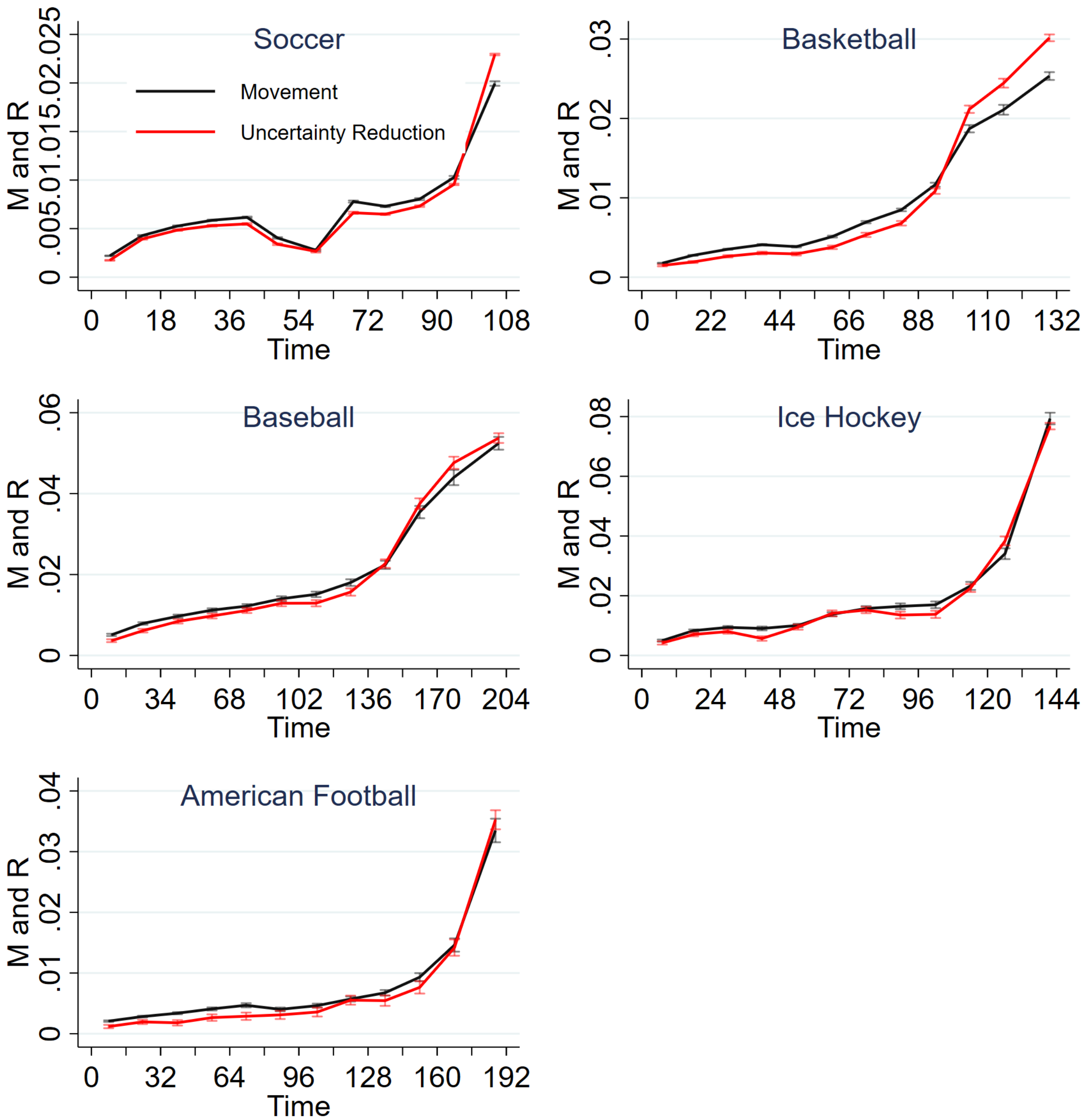
Notes: This figure plots the average answer subjects give as a function of the Bayesian posterior. It differs from Figure 2 by first transforming subjects' answers to be signal strengths, then averaging them, and then transforming back into precision estimates (as opposed to just averaging). The blue line indicates Bayesian behavior. The figure shows that subjects overweight weak signals and underweight strong signals. Error bars indicate 95% confidence intervals, clustered at the subject level.

Figure A3: Heterogeneity in Inference from by News Consumption



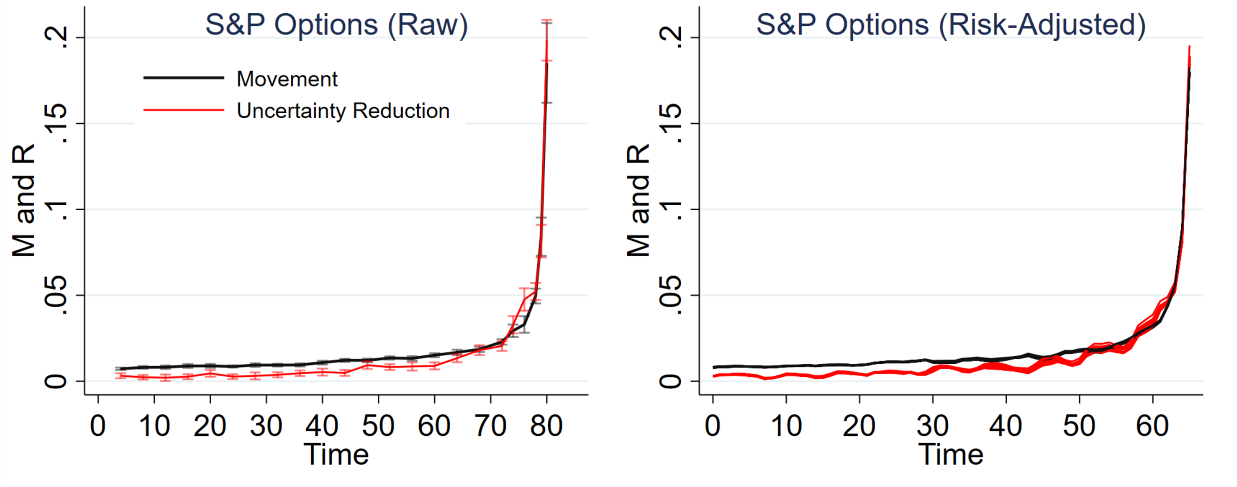
Notes: The y-axis is weight subjects put on signals. Weak signals have precision between 0.5 and 0.6; strong signals have precision between 0.7 and 1. This figure plots weight by self-reported news consumption. Individuals who consume news "Never," "Rarely," or "Occasionally" are categorized as "Not Frequently." Individuals who consume news "Somewhat Frequently," "Frequently," or "Very Frequently" are categorized as "Frequently." This figure shows that news consumption is associated with less weight on weak signals and similar weight on strong signals. Error bars indicate 95% confidence intervals.

Figure A4: Movement and Uncertainty Reduction in Betting Data with 12 Periods



Notes: This figure shows the average movement (black lines) and uncertainty reduction (red lines) statistics over time for four different sports (as well as 95% confidence intervals), with games split into 12 periods rather than 24. In each case, movement is greater than uncertainty reduction in early time periods and then drops below in later time periods.

Figure A5: Movement and Uncertainty Reduction for Options: Alternative Adjustments



Notes: This figure shows the average movement (black lines) and uncertainty reduction (red lines) statistics over time for the beliefs implied by option data (with 95% confidence intervals). The left-hand panel again uses the implied unadjusted (risk-neutral) beliefs. The right-hand panel uses the range of risk adjustments described in Appendix B. Movement is clearly greater than uncertainty at the start of the contract. At the later stages, movement is lower than uncertainty reduction.

B Appendix: Implementation Details for Finance Data

This Appendix describes the translation from risk-neutral to physical beliefs, as introduced for the finance data in Section 5.4, in greater detail. Assume there exists a representative investor ("the market") with time-separable utility over the market index value.[47] Assume that the state space (the set of possible terminal index values V_T) is discrete, with states indexed by j ($V_T = \theta_j$ for $j = 1, 2, \ldots, J$), and denote terminal utility by $U(V_T)$. The physical belief regarding the likelihood of state j is $\pi_{t,j}$, and the risk-neutral belief is $\pi^*_{t,j}$. The two are related as follows:[48]

$$\pi^*_{t,j} = \frac{U'(\theta_j)\pi_{t,j}}{\sum_k U'(\theta_k)\pi_{t,k}}. \tag{B1}$$

Our main translation assumes that $U'(V_T) = V_T^{-\gamma}$, corresponding to the assumption of power utility over the terminal index return, with constant relative risk aversion coefficient of γ. We then follow Bliss and Panigirtzoglou (2004) in estimating γ as the value under which the physical beliefs over the S&P 500 value at the one-month horizon are well calibrated (i.e., unbiased); see Bliss and Panigirtzoglou (2004) for details on the maximum likelihood estimation procedure.

We then consider hundreds of generalizations of this basic framework. First, we reparameterize (B1) in terms of the *ratio* of marginal utilities (or SDF realizations) across adjacent index states ϕ_j, by substituting $U'(\theta_j) = \phi_j U'(\theta_{j-1})$. We then make a range of assumptions on the function ϕ_j. We assume that ϕ_j varies by state j, either linearly or quadratically in V_T, and we estimate ϕ_j by maximum likelihood for each state; we assume that ϕ_j varies over time (either linearly or quadratically) or by horizon to expiration (as in Lazarus 2022); and then we consider interactions in which ϕ_j varies both by bin j and over time. In all cases (as can be seen in Figure A5, the right panel of which contains one line for each parameterization), the movement and uncertainty reduction statistics are close to unchanged. (This is in contrast to the *physical probabilities*, which do change depending on the parameterization; it is their evolution over time that is unchanged.)

[47] These illustrative assumptions aid in the interpretation of our risk-aversion assumptions, but they are stronger than needed in general; see AL (2022) for a discussion.

[48] This is a multi-state generalization of eq. (5) of AL (2022), or see eq. (7) of Bliss and Panigirtzoglou (2004).

C Online Appendix: Study Materials

Princeton Experimental Laboratory
for the Social Sciences

Overview and Bonus Payment

On the following pages, you will be asked to make a series of choices that can impact your bonus payment.

After all participants complete the study, 5 participants will be chosen at random to receive a bonus payment of up to $100 based on their choices. The high bonus is because it is important for us that you take this study seriously.

You will see between 25 and 35 question pages, most of which are similarly styled. These questions are divided into sections. There will be an "attention check" question in the study. The answer to this question will be obvious to anyone paying attention. Participants who do not answer the attention check question correctly will still receive their show-up payment, but **will not be eligible for a bonus payment**.

Instructions for the study are on the following page.

Princeton Experimental Laboratory for the Social Sciences

Instructions

On the following several pages, you will be asked to predict which jar a randomly-chosen ball comes from.

For instance, you will see questions like the following:

You draw a card from one of two modified decks of cards; a **Green** deck or a **Purple** deck.

The **Green** deck has **15 Diamonds** (♦) and **10 Spades** (♠).
The **Purple** deck has **10 Diamonds** (♦) and **15 Spades** (♠).

*The card you draw is a **Diamond** (♦).*

What do you think is the percent chance that your **Diamond** (♦) came from the **Green** deck vs. the **Purple** deck?

Please answer between 0 and 100 for each question and have your answers sum up to 100, where higher numbers mean you think that deck is more likely. Your answer may include decimals.

Percent chance that your **Diamond** (♦) came from the **Green** deck:

_______ percent

Percent chance that your **Diamond** (♦) came from the **Purple** deck:

_______ percent

We have carefully chosen the payment rule so that you will earn the most bonus payment on average if you give a guess that you think is the true likelihood. If you are interested, further details on the payment rule are below.

Payment for your prediction:

If you are selected to receive a bonus payment, to determine your payment the computer will randomly choose a question and then randomly draw two numbers. For each draw, all numbers between 0 and 100 (including decimal numbers) are equally likely to be selected. Draws are independent in the sense that the outcome of the first draw in no way affects the outcome of the second draw.

If the selected Jar is Jar 1 and the number you picked is larger than either of the two draws, you will get a bonus payment of $100.

If the selected Jar is Jar 2 and the number you picked is smaller than either of the two draws, you will get a bonus payment of $100.

Otherwise, you will get a bonus payment of $10.

Princeton Experimental Laboratory
for the Social Sciences

You are on Question 1 of 15 in this section.

You draw a card from one of two modified decks of cards; a **Green** deck or a **Purple** deck.

The **Green** deck has **853 Diamonds** (♦) and **812 Spades** (♠).
The **Purple** deck has **812 Diamonds** (♦) and **853 Spades** (♠).

*The card you draw is a **Spade** (♠).*

What do you think is the percent chance that your **Spade** (♠) came from the **Green** deck vs. the **Purple** deck?

Please answer between 0 and 100 for each question and have your answers sum up to 100, where higher numbers mean you think that deck is more likely. Your answer may include decimals.

Percent chance that your **Spade** (♠) came from the **Green** deck	0	percent
Percent chance that your **Spade** (♠) came from the **Purple** deck	0	percent
Total	0	percent

Princeton Experimental Laboratory
for the Social Sciences

You draw a card from one of two modified decks of cards; a **Green** deck or a **Purple** deck.

The **Green** deck has **867 Diamonds** (♦) and **5309 Spades** (♠).
The **Purple** deck has **5309 Diamonds** (♦) and **867 Spades** (♠).

*The card you draw is a **Spade** (♠).*

This question is not like the previous ones. It is a check to make sure you are paying attention. Instead of answering the question normally, please answer 3 for the first question and 14 for the second one. You will not be eligible for a bonus payment if you get this question incorrect.

Percent chance that your **Spade** (♠) came from the **Green** deck	0 percent
Percent chance that your **Spade** (♠) came from the **Purple** deck	0 percent
Total	0 percent

You draw a card from one of two modified decks of cards; a **Green** deck or a **Purple** deck.

This time, you do not know the composition of the two decks for sure.

The **Green** deck **either** has:

- **841 Diamonds** (♦) and **824 Spades** (♠)

  OR

- **448 Diamonds** (♦) and **1217 Spades** (♠).

The **Purple** deck **either** has:

- **824 Diamonds** (♦) and **841 Spades** (♠)

  OR

- **1217 Diamonds** (♦) and **448 Spades** (♠).

Both compositions are equally likely.

*The card you draw is a **Diamond** (♦).*

What do you think is the percent chance that your **Diamond** (♦) came from the **Green** deck vs. the **Purple** deck?

Please answer between 0 and 100 for each question and have your answers sum up to 100, where higher numbers mean you think that deck is more likely. Your answer may include decimals.

Percent chance that your **Diamond** (♦) came from the **Green** deck	0 percent
Percent chance that your **Diamond** (♦) came from the **Purple** deck	0 percent
Total	0 percent

Princeton Experimental Laboratory for the Social Sciences

You are on Question 1 of 5 in this section.

On the previous questions, you were given a card and asked to predict what deck it came from.

On this question, you will choose *how many* cards to draw in order to help you with your predictions. Every card is drawn with replacement.

You may draw up to 3 cards at a cost. If you are selected for the bonus and this round is chosen for payment, you can win up to $100. The first card costs $1, the second card costs an additional $2 ($3 total), and the third card costs an additional $3 ($6 total).

- $1 roughly corresponds to moving from being 50% sure that you know the deck to being 60% sure;
- $3 corresponds to moving from being 50% sure that you know the deck to being 67% sure;
- $6 corresponds to moving from being 50% sure that you know the deck to being 74% sure.

If you think the cards are helpful in distinguishing the **Green** and **Purple** decks, you should draw more cards. If you think the cards are unhelpful in distinguishing the **Green** and **Purple** decks, you should not draw cards.

The **Green** deck has **853 Diamonds** (♦) and **812 Spades** (♠).
The **Purple** deck has **812 Diamonds** (♦) and **853 Spades** (♠).

You can now choose how many cards to draw from the deck.

- ◯ Do not draw any cards
- ◯ Draw 1 card (Total cost: 1 point)
- ◯ Draw 2 cards (Total cost: 3 points)
- ◯ Draw 3 cards (Total cost: 6 points)

The **Green** deck has **853 Diamonds** (♦) and **812 Spades** (♠).
The **Purple** deck has **812 Diamonds** (♦) and **853 Spades** (♠).

You chose to draw 2 cards.

What do you think the percent chance (between 0 and 100) is that the card is from the **Green** deck if the cards drawn are:

0	2 Spades (♠)
0	2 Diamonds (♦)
0	1 Diamond (♦) and 1 Spade (♠)

→

Princeton Experimental Laboratory for the Social Sciences

What is your age?

What is your gender?

- Female
- Male
- Other / Prefer not to answer

In politics today, do you consider yourself a Republican, a Democrat, or an Independent?

- Democrat
- Republican
- Independent

What is your highest level of education?

- Did not graduate high school
- High school graduate, diploma, or equivalent (such as GED)
- Began college, no degree
- Associate's degree
- Bachelor's degree
- Postgraduate or professional degree

What is your race/ethnicity?

- ◯ American Indian
- ◯ White
- ◯ Asian
- ◯ Black or African American
- ◯ Hispanic or Latino
- ◯ Two or more of these
- ◯ Other / Prefer not to answer

PEXL
Princeton Experimental Laboratory
for the Social Sciences

Quiz 1

A bat and a ball cost $10.50 in total. The bat costs $10.00 more than the ball.

How much does the ball cost?

- ◯ $0.50
- ◯ $10.00
- ◯ $10.25
- ◯ $0.25

Quiz 2

If it takes 7 machines 7 minutes to make 7 widgets, how long would it take 70 machines to make 70 widgets?

- ◯ 0.1 minutes
- ◯ 490 minutes
- ◯ 70 minutes
- ◯ 7 minutes

Quiz 3

In a community, there is a rapidly-spreading virus. Every day, the virus infects twice as many people. If it takes 42 days for the virus to infect the entire community, how long would it take the virus to infect half the community?

- ◯ 21 days
- ◯ 42 days
- ◯ 2 days
- ◯ 41 days