The Value of and Demand for Diverse News Sources

by

Evan M. Calford

and

Anujit Charkraborty

ANU Working Papers in Economics and Econometrics
688

June 2022

JEL Codes: D85, D91, D81

ISBN: 0 86831 688 1

THE VALUE OF AND DEMAND FOR DIVERSE NEWS SOURCES

EVAN M. CALFORD* AND ANUJIT CHAKRABORTY**

Abstract. We study the value of and the demand for instrumentally-valuable information in a simple decision environment where signals are transparently biased. We observe remarkable sophistication in information aggregation and acquisition. A majority of our subjects (63%) made unbiased reports even when faced with biased signals and the few subjects who made biased reports were split between under- and over-correcting for the signal bias. When allowed to buy pairs of opposite or similarly biased information sources, subjects actively shopped for diverse information at personal costs, and their demand for diverse information reacted rationally to its value and cost. Subjects who were worse at aggregating information, were more likely to purchase diverse signals, perhaps in an attempt to make their inference problem easier. Our results advocate for greater transparency in media bias, so that individuals can choose the right portfolio of information to make better choices.

JEL codes: D85, D91, D81,

* Research School of Economics, Australian National University; evan.calford@anu.edu.au.
** Department of Economics, University of California, Davis; chakraborty@ucdavis.edu.
Funding for subject payments was provided by the Krannert School of Management, Purdue University.
The experiments were approved by the Purdue IRB, protocol number 1706019387. Funding was provided by the Krannert School of Management, Purdue University. We thank seminar audiences at UC Davis and Queensland University of Technology.

In the 2020 Gallup poll, 83 percent of Americans recognized "a great deal" or "a fair amount" of political bias in news coverage, a number that has risen steadily over the last 20 years.[1] Media sources with contradictory biases can create completely different impressions of what actually happened through the selective omission of details and the choice of words. As an example of media bias, Gentzkow and Shapiro [2006] discuss the media coverage of the firefight in the Iraqi city of Samarra on December 2, 2003. Fox News, a conservative US based news channel, began its story with the following paragraph:

> In one of the deadliest reported firefights in Iraq since the fall of Saddam Hussein's regime, US forces killed at least 54 Iraqis and captured eight others while fending simultaneous convoy ambushes Sunday in the northern city of Samarra.

And the English-language website of the Qatar-based Al Jazeera (AlJazeera.net) began its report on the same incident with:

> The US military has vowed to continue aggressive tactics after saying it killed 54 Iraqis following an ambush, but commanders admitted they had no proof to back up their claims. The only corpses at Samarra's hospital were those of civilians, including two elderly Iranian visitors and a child.

The last two US elections have further increased the focus on media bias in the coverage of political issues and candidates. Donald Trump, who served as the 45th president of the United States, routinely claimed that liberal media bias was undermining his political campaign.[2] The 44th president, Barrack Obama,

[1] See the report here.

[2] For example, Trump commented "If the disgusting and corrupt media covered me honestly and didn't put false meaning into the words I say, I would be beating Hillary by 20%,"

has expressed similar views about the bias of conservative news sources.[3] The media bias extends to the coverage of financial news: Niessner and So [2018] find that a financial news story is approximately 22 percent more likely to be covered if it is negative, creating an overall negative media bias in financial news coverage.

Readers who exclusively access media sources that share identical bias would only get an incomplete account of the whole reality. As creatures that often make quick judgments based on heuristic-driven biases[4], for example hot hand bias (marquis de Laplace [1840]), Gambler's Fallacy (Chen et al. [2016a]), law of small numbers (Rabin [2002]) and base rate neglect (Kahneman and Tversky [1973]) to name a few, *how do we acquire and aggregate instrumental information when information sources are transparently biased? Do biased sources lead to severe and systematic errors? Do we diversify information sources when we have access to a market for information?* This paper uses experimental methods to find an answer.

We immerse subjects in a controlled environment of selective information-omission to compare the accuracy of their opinions under diverse versus biased information sources. Every round, subjects are asked to guess an objective state, the average of seven i.i.d random draws from the simple uniform distribution {1,2,3,..100}. To incentivize subjects, more accurate guesses earned a higher expected payment. In our first two treatments, subjects were *randomly* shown only three of the seven random draws as *signals* beforehand. We call signals between 51-100 high and those between 1-50 low. The bias in observed signals is *extreme* when all three observed signals are high, thus creating a high-bias, or when all three observed signals are low, creating a low-bias. This is similar to exclusively

[3] Obama has commented "If I watch Fox News, I wouldn't vote for me."

[4] Psychology studies have long proposed a dual-process model for two modes of information-processing: a "fast, associative" one "based on low-effort heuristics", and a "slow, rule based" one that relies on "high-effort systematic reasoning" (Kunda [1990]).

accessing media sources that share the same reporting bias on an issue with bipolar political divide. Instead, if the observed signals are a mix of high and low, then we will describe the signals as *diverse*. Our data-generating process is simple and the bias of observed signals is *completely transparent* to subjects.

In the Baseline treatment (called **No Colors** treatment), the subjects do not have any information about the unobserved signals, other than knowing that they are equally likely to be any number between $\{1,2,3,..100\}$. Given the observed three signals were randomly chosen from the seven draws, subjects might assume that their observed sample is "representative" of the seven draws and over-infer from it. If subjects ignored the information from the unobserved signals and simply calculated the average of the three observed signals (sample-average), their reports would be biased towards the polarity of extreme signals. Instead if they fell for the Gambler's fallacy, inferring mistakenly that observing three low (or high) signals increases the likelihood of the unobserved signals being of the opposite polarity, their reports would be biased in the opposite direction of the extremity of their signals. Contrary to our prior expectations, we find that subjects are remarkably accurate in adjusting their opinions for extreme signal bias: more than 60% of our subjects are on average within 5 points of the optimal Bayesian report. Digging deeper, we find that only 16% of subjects exhibited a consistent bias towards the sample-average heuristic and only 20% exhibited a consistent bias towards the Gambler's fallacy. The majority (64%) subjects are able to counteract the effect of biased information when bias is transparent!

As a benchmark for an easier decision environment, we designed the **Colors** treatment, where the subjects are also informed how many of the four unobserved signals lie between 1-50 (called Blue signals) or 51-100 (called Red signals). At a minimum, this alerts the subjects that the observed signals might not be informative about the unobserved signals, and at a maximum, it helps them account for the unobserved signals. As expected, this additional information improves the accuracy of subjects' reports without introducing any bias. We also find that playing the simpler Colors treatment before the Baseline (No Colors) treatment does not improve the performance in the Baseline treatment.

In our next two treatments, **Active Choice** and **Average**, instead of being assigned to signals randomly, subjects "buy" the signals they observe. We place subjects in a *market-place for information* to study how subjects, facing information bias, opt into or opt out of particular informational environments. Information in this setting is instrumentally valuable as it helps subjects better guess the payoff relevant state of a world. We use a pricing mechanism to vary the cost of different signal combinations, usually making diverse signals costlier than extreme signals. In the Active Choice treatment, subjects are told the price of observing a signal of each color, and they *choose* how many of their observed three signals should be Red or Blue. They see the realization of those three signals and the colors of the other four signals, before guessing the average. Thus the treatment is similar to the Colors treatment, but subjects can actively purchase three signals of the same color or a mixture of signals of opposite bias. Each transaction reveals when subjects believed that the instrumental value of diverse signals was higher than its cost. When diverse and extreme signals are equally costly, we find that approximately 95% of subjects choose a diverse portfolio of signals. The demand for diverse signals persists as the cost of diverse information increases. As the price for diverse signals increases, subjects become less likely to buy them, but even at the most extreme price differentials 20% of subjects continue to purchase diverse signals. Notably, we also find evidence of metacognition in signal choice: subjects who were worse at aggregating information were also more likely to purchase diverse signals, perhaps in an attempt to make their inference problem easier.

In our **Average** treatment, we measure if subjects can identify when the value of diverse information objectively increases, and if they react to it in their marketplace transactions. To exogenously increase the value of diverse signals, we impose the rule that after subjects choose a signal portfolio, a sample-average heuristic would calculate the report on their behalf. Our intervention significantly increases the value of diverse information as the accuracy of the sample-average heuristic is severely hurt under extreme information. The subjects know that accurate reports pay more, and thus, are incentivized to select signals that preserve accuracy under the imposed sample-average rule. In our experiments, most of the subjects realize that the accuracy premium from diverse signals is even higher under the sample-average rule and it is reflected in their signal choices. The demand for diverse signals is even higher in the Average rounds, increasing by more than 50% at intermediate information costs.

Our results advocate for greater transparency in media bias, so that individuals can choose the right portfolio of information to make better choices. For example, websites like Allside.com or Adfontesmedia.com that explain and measure various dimensions of media bias can be extremely useful in the quest for transparency. In particular, All Sides is a news website that presents multiple sources side by side in order to provide the full scope of news reporting. It also provides a Bias Ratings page that allows a visitor to filter a list of news sources by the bias on the political spectrum (left, center, right).[5]

[5]See their media bias chart here

Our paper is related to information acquisition and a large literature, pioneered by Tversky and Kahneman [1971, 1974], that studies the prevalence of biases and heuristics in probabilistic decisions. For clean identification, our experimental environment excludes the scope for motivated reasoning (Kunda [1990]), the tendency of people to conform assessments of information to some goal or end extrinsic to accuracy. Similarly, it also excludes mistakes originating in the failure of hypothetical or contingent reasoning; any event that subjects should condition on is clearly and explicitly displayed. We discuss the related literature in detail in Section 1.

1. Related Literature

The Sample-average rule, that subjects might find naturally attractive in our decision environment, belongs to the class of simplistic heuristics that overweight one type of information over other available information.[6] For example, experimental subjects frequently discard or under-weight base-rate information because it is not relevant to judgements of representativeness (Kahneman and Tversky [1973]), or because the likelihood information is more "vivid, salient, and concrete" (Nisbett and Ross [1980]). Subjects in our treatments might similarly overweight the information from the three signals that they observe, thinking it is representative of the seven signals, or because the information it provides is more salient and concrete. A recent literature in cognitive psychology connects such behavior to people behaving like "naive intuitive statisticians" who despite being skilled in making judgements based on memory-stored frequencies, often naively assume that their information samples are representative and that sample properties can be directly used to estimate population analogs (Fiedler and Juslin [2006]).

[6] Benjamin [2019] provides a detailed literature review.

The Gambler's Fallacy (GF in short) is one of the oldest documented probabilistic biases. marquis de Laplace [1840] described people's belief that the fraction of boys and girls born each month must be roughly balanced, so that if more of one sex has been born, the other sex becomes more likely. Rabin [2002] and Oskarsson et al. [2009] review the extensive literature that documents the GF in surveys and experiments. Chen et al. [2016a] finds consistent evidence of negative autocorrelation in decision making that is unrelated to the merits of the cases considered in three separate high-stakes field settings: refugee asylum court decisions, loan application reviews, and Major League Baseball umpire pitch calls. They link it to people underestimating the likelihood of sequential streaks occurring by chance—leading to negatively autocorrelated decisions that result in errors. Similarly, experimental participants playing a game with a unique mixed-strategy Nash Equilibrium or tennis players making a serve switch their actions too often (Rapoport and Budescu [1997], Gauriot et al. [2016]), and this excessive switching could reflect the mistaken GF intuition for what random sequences look like.

Studies on information extraction are also closely related to the growing literature on contingent reasoning or hypothetical thinking. For example, to avoid the winner's curse, bidders in a common value auction should extract information from their private signal, while conditioning on the hypothetical event of winning the auction. Similarly, voters should extract information from their private signal, while conditioning on the hypothetical event of being pivotal to the outcome. Esponda and Vespa [2019], Araujo et al. [2021] find that experimental subjects routinely fail to perform such contingent reasoning while processing their private information. Enke [2020] finds that when subjects are exclusively shown information consistent with their initially reported prior, they often behave as if the sample selection does not even come to their mind. Enke [2020] provides further causal evidence that the frequency of such incorrect mental models is a function of the computational complexity of the decision problem. To disentangle the value of and demand for diverse information from these other behavioral forces, we offer subjects a simple decision problem where the difference between extreme and diverse signals is transparent, even to subjects who cannot reason through hypothetical events.

Our paper is also related to the literature on the value of instrumental information. Duffy et al. [2019, 2021] study how subjects choose between social and private information sources that vary in relative quality. Charness et al. [2021] and Montanari and Nunnari [2019] study how subjects update their beliefs about a payoff-relevant state of the world while choosing *exactly one of two information sources* (signals) which have mutually opposite biases, and thus are the closely related. In both studies, subjects had to guess the probability that a single ball drawn randomly from an urn would be of a particular color, and also guess the color of the ball. To inform their guesses, subjects first chose one of a pair of computerized advisors, from which to receive an informative signal about the ball drawn. Subjects were fully informed of the probabilities with which each advisor would provide each signal as a function of the true color of the ball drawn from the urn. Charness et al. [2021] find that the fraction who choose information optimally and the fraction who use a mistaken confirmation-seeking rule are roughly equal. In Montanari and Nunnari [2019], when the two information sources are equally reliable, subjects select information optimally. But, when the source less supportive of the prior belief is more informative, subjects display a dis-confirmatory pattern of information acquisition that is not always consistent with the theoretical predictions. Even in cases where information is not instrumentally valuable, subjects might have preferences over how information is disclosed (Zimmermann [2015], Ganguly and Tasoff [2017], Masatlioglu et al. [2017], Nielsen [2020]). An emerging literature on motivated reasoning addresses how we often seek particular information and stay willfully ignorant of other information, because we wish to arrive at our desired conclusion (Festinger [1962]). We deliberately frame our experimental tasks to remove the scope for motivated reasoning, as we want to to study to the demand for information with *purely* instrumental value.

Acquisition of instrumental information has also been studied in applied settings. Fuster et al. [Forthcoming] use an experimental survey instrument to determine which pieces of economic data subjects prefer to consult when predicting house price movements. Burke and Manz [2014] asked subjects to forecast inflation in a simulated laboratory economy, and provided subjects with a choice of viewing historical information on inflation, interest rates, unemployment, population growth, or price changes of specific commodities, before making their forecast. In both environments choices of more informative sources were correlated with measures of economic sophistication. Mikosch et al. [2021] study how information acquisition about the future development of the exchange rate is related to the exposure to and uncertainty about exchange rate risk, and the perceived information acquisition costs. Roth et al. [2022] find that a higher personal exposure to unemployment risk during recessions increase the demand for an expert forecast about the likelihood of a recession. Capozza et al. [2021] provide a detailed review the emerging literature on information acquisition in field settings.

Most of the lab-experimental literature on information-acquisition compares empirical information choice to a theoretical optimal calculated for the Bayesian subjects. This allows for sharp theoretical predictions, but the test for optimal signal choice becomes a joint test of optimal choice and subjects updating using Bayes Rule. The information-choice data can only reject optimal choice when information-updating behavior is consistent with Bayes Rule. There are, however, some exceptions in the literature. Charness et al. [2021] use a treatment with exogenously assigned signals to show that the Bayesian optimal information choice was also optimal for the behavioral non-Bayesian types in their subject pool (i.e. the optimality of information choices is not dependent on an assumption of Bayesian updating). Ambuehl and Li [2018] measure the value of information against both a Bayesian benchmark and an empirical updating benchmark and find that subject's value for information is higher, but still lower than optimal, when measured against the empirical benchmark. We instead test if subjects react optimally to an increase in the value of diverse signal choice and we do not require subjects to be Bayesian. All we require is that diversification is more valuable under the sample-average rule than whatever updating rule subjects actually use. As we show later in Figure 4.3, this is indeed true for all-but-one individual participant in our study. Further, in our Average treatment we exogenously impose a sub-optimal updating rule on subjects which allows us to measure, unconfounded by updating ability, subject propensity to select information that offsets updating biases.

2. Experimental Design

To study the acquisition and aggregation of extreme information, we conduct 4 treatments: No Colors (NC), Colors (C), Active Choice (AC), Average (Avg). The first two treatments randomly allocate signals (information) to study information aggregation in isolation. The last treatment fixes the information aggregation procedure exogenously to study information acquisition in isolation, and the AC treatment combines both information aggregation and information selection tasks.

2.1. **Baseline/ No Colors treatment (NC):.** The first treatment is No Colors (NC henceforth). Each round in the NC treatment has the following structure:

(1) **Signal realization stage:** A set of seven signals are independently drawn from the numbers 1 to 100. Each signal, s_i for $i \in \{1, \ldots, 7\}$ is assigned a color based on its realization: Blue if $s_i \leq 50$ and Red if $s_i \geq 51$.

(2) **Information stage:** In the experiment, we framed s_1 as the subject's own signal and each other signal $\{s_2, .., s_7\}$ as belonging to a passive computer "player". Subjects see the realization of s_1, and two other randomly chosen signals (say s_2 and s_3).[7] We use this framing to imitate information flow in the subject's social network. Subjects are not informed about the colors of the remaining 4 signals. For example, in a particular round, if the signal realization were $\{\underbrace{10}_{s_1}, \underbrace{20}_{s_2}, \underbrace{30}_{s_3}, 40, 50, 60, 70\}$, and s_2 and s_3 were shown to the subject, then the subject would observe $\{10, 20, 30\}$ and but they would *not know* the color composition of the unobserved 4 signals. Thus the name No Colors treatment.

(3) **Updating/ aggregation stage:** Subjects state their best estimate of the average value $\overline{s} = \frac{\sum_{i=1}^{7} s_i}{7}$ given the information above.

One crucial feature of our design is that we did not provide any feedback to the subjects in between the rounds, in any of our treatments. Subjects were provided with a hand-held calculator, although few subjects elected to use the calculators; we did not want to test the ability of subjects to do basic arithmetic.

[7]Given the signals are ordered randomly, this is equivalent to showing subjects any three signals randomly.

At the end of the experiment, one of the rounds was randomly chosen as the round for which subjects were paid. For the chosen round, the guessing error was calculated as the absolute difference between the reported guess and the realized $\bar{s} = \frac{\sum_{i=1}^{7} s_i}{7}$. Each subject won a large prize worth 200 points with $\max(100 - 6 \times error, 0)\%$ chance, and a small prize worth 50 points with the complementary probability. This payment function ensures that reports are unaffected by the curvature of the subject's utility function over money. The linear loss function was chosen as it is simple and it incentives reporting the expected value of $\bar{s}$ truthfully under an assumption that the belief about $\bar{s}$ is symmetric and single peaked.[8]

2.2. **Colors treatment (C):.** The Colors treatment is identical to the No Colors treatment, except, in the **Information stage**, they are also shown the colors of all the remaining 4 remaining signals. For example, in a particular round, suppose the signal realizations were $\{\underbrace{10}_{s_1}, \underbrace{20}_{s_2}, \underbrace{30}_{s_3}, \underbrace{40}_{\leq 50}, \underbrace{50}_{\leq 50}, \underbrace{60}_{>50}, \underbrace{70}_{>50}\}$. If the realizations s_2 and s_3 were shown to the subject, then the subject observed $\{10, 20, 30\}$ and additionally they would know that among the unobserved 4, there are 2 Blue ($<= 50$) and 2 Red (> 50) signals. The Colors treatment makes the importance of unobserved signals salient.

2.3. **Active choice treatment (AC):.** The AC treatment is identical to the Colors treatment, except, in the **Information stage**, subjects make an active choice about the signals they observe. They observe s_1 by default. They can choose to observe the realizations of two Blue signals or two Red signals or one signal of each color. For example, in a particular round, if the signal realization were $\{\underbrace{10}_{s_1}, \underbrace{20, 30, 40, 50}_{Blue}, \underbrace{60, 70}_{Red}\}$, then the subject only sees $\{10\}$ by default. If she

[8]This assumption obviously holds for Bayesian subjects, but can hold more generally.

chooses to see 2 Blue signals, then she is randomly shown exactly two signals from $\{\underbrace{20, 30, 40, 50}_{Blue}\}$ and she is told that among the unobserved 4, there are 2 Blue and 2 Red signals. If she instead chooses to see 2 Red signals, then she is shown both $\{60, 70\}$ and she is told that all the unobserved 4 signals are Blue. If she instead chooses to see 1 signal of each color, then she is shown one of $\{20, 30, 40, 50\}$, one of $\{60, 70\}$and she is told that among the unobserved 4, there are 3 Blue and 1 Red signals.

The subjects made their signal selection without knowing if their choice were available. In the improbable event that the selected signals are not available (for example, all but one of $\{s_2, s_3, .., s_7\}$ are Blue but the subject requests two Red signals), then the subject is shown a combination of signals that is as close as possible to matching the subjects requested combination of colors (in this example, the subject would be shown one each of Red and Blue signals). The subject is always informed of the colors of any unobserved signals (in this example, there are 4 unobserved Blue signals).

Further, each round, subjects are paid p_1 for each signal they observe that is the same color (Blue/ Red) as s_1 and p_2 for each observed signal that is of the opposite color from s_1. We vary p_1 and p_2 to generate the price-differences

$$\Delta p = (p_1 - p_2) \in \{-6, -2, 0, 2, 4, 6, 8, 10, 14, 20\}$$

When $\Delta p > 0$, subjects have an explicit monetary incentive for choosing more signals that are of the same color as s_1. When $\Delta p < 0$, they have the opposite incentive, and when $\Delta p = 0$, there is no explicit incentive and subjects should choose the set of signals that they subjectively view as the most informative.

2.4. **Average treatment (Avg):** The Average treatment is identical to the Active Choice treatment in the Information stage: subjects choose the colors of

Features	Treatments			
	Color (C)	No Color (NC)	Active Choice (AC)	Average (Avg)
Signal acquisition	Random	Random	Active choice	Active choice
Information Aggregation	Active choice	Active choice	Active choice	Sample-Avg
Colors of unobserved balls	Available	Unavailable	Available	Not Applicable

TABLE 1. Features of different treatments.

the signals, subject to the available price-difference Δp. But, subjects know that their report in the Updating/ aggregation stage is *exogenously constrained to be the sample average* of their three observed signals. Thus, if a subject observes $\{10, 20, 30\}$, then the software would report 20 on their behalf. Similarly, if they observe $\{10, 20, 60\}$, then the software would report 30 on their behalf. Thus the name Average treatment. This treatment manipulation increases the value of diversified signals for all subjects, given that the sample-average updating rule suffers a huge bias when all observed signals are extreme. As before, we vary p_1 and p_2 to generate the price-differences

$$\Delta p = (p_1 - p_2) \in \{-6, -2, 0, 2, 4, 6, 8, 10, 14, 20\}$$

In the Appendix, we report on a fifth treatment. We do not include this treatment in the main text because, as discussed in the Appendix, there is evidence of substantial subject confusion in the novel part of the extra treatment. The sessions that included the fifth treatment also included some rounds of the Colors and Avg treatments as well. We exclude all sessions that included the fifth treatment from the main text, but include the data from those sessions in the Appendix. All our results are robust to including the fifth treatment sessions.

Session name	Treatment(s) run		#Sessions
	Rounds 1-20	Rounds 21-40	
C1+NC2	Colors	No Colors	2
NC1+C2	No Colors	Colors	2
AC+Avg	Active Choice	Average	2

Table 2. Number of sessions conducted with each pairwise combination of treatments. C1 and C2 mean Colors treatment were run in the first and second half of the session respectively. Similarly, NC1 and NC2.

2.5. **Sessions.** Each session was run with around 15 subjects. Sessions lasted 40 rounds, grouped into 2 treatment-blocks of 20 rounds each. In the table below we summarize the treatment composition of the sessions

Subjects were paid for the sum of points earned during one randomly selected round. At the end of the experiment, points were converted to US Dollars at an exchange rate of $0.07 per point. Thus, the larger prize (200 points) was worth $14 and the smaller prize (50 points) was worth $3.50. They were also paid a $5.00 show up fee in addition to any money they earned during the experiment. We conducted 6 sessions in total, with a total of 89 subjects. Our experiments were conducted in-person at the Purdue University, using student subjects drawn from Purdue's implementation of the ORSEE subject database [Greiner, 2015], during 2018. The experiments were programmed using oTree [Chen et al., 2016b].

3. Hypotheses

3.1. Information aggregation.

Given the information subjects have while making a guess, the Bayesian report for both Color and No Color rounds can be calculated as follows. For the Color treatment the Bayesian report is given by

$$R_C^* = \frac{s_1 + s_2 + s_3 + 25.5 \times (\#Blue) + 75.5 \times (\#Red)}{7}$$

where $\#Blue$ and $\#Red$ are the number of unobserved Blue and Red signals, respectively, the subscript C denotes the Colors treatment, and the superscript $*$ denotes the optimal Bayesian report. In the NoColor treatment the Bayesian report is

$$R_{NC}^* = \frac{s_1 + s_2 + s_3 + 50.5 \times 4}{7}$$

where the subscript NC denotes the NoColor treatment. The Color treatment provides subjects with more information about the unobserved signals and will, on average, lead to better performing reports in the Color treatment than the NoColor treatment. Our first set of hypotheses concern the effects of the distribution of the three observed signals on the quality of reports. We call a set of three signals *extreme* if they are all Red (51 or above) or all Blue (50 or below). If a set of signals is not *extreme* then it is *diversified*. For example, $\{10, 20, 30\}$ would be a set of extreme signals, and so would $\{95, 60, 70\}$. We initially focus on two behavioral biases that are ex-ante plausible in the NoColor rounds.

1) **Sample average**: An intuitive but incorrect decision rule would be to completely disregard any information present in the unobserved signals and simply report the sample average of the three observable signals. That is, a naive, sample-average report would be to report

$$\underbrace{R_{NC}^{SA} = \frac{s_1 + s_2 + s_3}{3}}_{\text{sample average}}$$

where the subscript SA stands for sample-average. Given the observed three signals were randomly chosen from the seven draws in the NC treatment, subjects might assume that their observed sample is "representative" of the seven draws and be attracted to such a heuristic. In this case, when the signals are extreme, the guess would be more inaccurate on average and would be *biased towards the signal-extremity.* For example, when all the three observed signals are Blue, subjects would, on average, under-report by a large margin

$$\mathbb{E}[R^{SA}_{NC}|(0R,3B)] = 25.5 < \mathbb{E}[R^{*}_{NC}|(0R,3B)] = \frac{25.5 \times 3 + 50.5 \times 4}{7} = 39.79$$

Similarly, when all the signals are Red, subjects would over-report by a large margin. Instead when the observed signals are mixed, for example two Blue and one Red, the margin of error is much smaller:

$$\mathbb{E}[R^{SA}_{NC}|(1R,2B)] = 42.2 < \mathbb{E}[R^{*}_{NC}|(1R,2B)] = \frac{51 + 75.5 + 50.5 \times 4}{7} = 46.92$$

2) **Gambler's Fallacy**: The gambler's fallacy (GF) is the common, but mistaken, belief that that i.i.d. random variables are "self-correcting towards the mean" and hence exhibit negative serial correlation.[9] In our setting, the gambler's fallacy implies that subjects mistakenly believe that each observed Red (or Blue) signal reduces the likelihood of the unobserved signals being that color. For simplicity, one can think of this as the rule

$$R^{GF}_{NC} = \frac{s_1 + s_2 + s_3 + 4 \times (75.5p + 25.5(1-p))}{7}$$

9

The fallacy earns its name from the story of a gambler who, after observing a run of black numbers at a roulette table, exclaims "We are due for a red number next!"

where p depends on the colors of the signals s_1, s_2, s_3. Thus, if n is the number of Red balls among s_1, s_2, s_3, then, according to the GF,

$$p(n = 0) > p(n = 1) > \underbrace{.5}_{\text{under Bayes}} > p(n = 2) > p(n = 3)$$

Thus, when the signals are extreme (either $n = 0$ or $n = 3$), the guess would be further from the Bayesian estimate on average and would be *biased in the opposite direction of the signal-extremity.* One could construct hybrid rules by taking the following convex combinations of the non-Bayesian heuristics with the Bayesian rule, that is, $\alpha R^*_{NC} + (1-\alpha)R^{SA}_{NC}$ and $\alpha R^*_{NC} + (1-\alpha)R^{GF}_{NC}$. For any value of $\alpha \in [0, 1)$ these hybrid rules would inherit the directional bias of their parent non-Bayesian rule. We construct our null hypothesis under the generalized sample-average rule $\alpha R^*_{NC} + (1-\alpha)R^{SA}_{NC}$ for $\alpha \in [0, 1)$.

Hypothesis 1 (Error and Bias, Sample Average)**.** *In the No Colors and Colors treatments, extreme signals create reports that are further away from the Bayesian report and biased towards the corresponding extremity.*

An alternative hypothesis, using the Gambler's Fallacy, would suggest a bias towards the opposite direction under extreme signals.

The Gambler's Fallacy, as described above, is ruled out in the Colors treatment. A heuristic like sample-average is also unlikely when subjects are explicitly informed about the colors of the unobserved realizations. Thus, subjects are more likely to make mistakes in the No Colors rounds, which is our next hypothesis.

Hypothesis 2 (Color vs No Color)**.** *Bias is higher when the color information of unobserved balls is unavailable (No Colors treatment).*

3.2. **Learning:** Subjects get no feedback between rounds. But, consider a sub-population of subjects who use the Sample average rule R_{NC}^{SA} in the No Color treatment. They completely disregard any information that might be present in the unobserved signals. Their play might be influenced by the order in which they play both the Color and No Color treatments. In particular, any prior experience in the Color treatment might make it salient for them that the unobserved signals play a significant role in determining the target $\overline{s} = \frac{\sum_{i=1}^{7} s_i}{7}$.[10] This experience might move them away from a naive sample-average rule to some rule that accounts for the unobserved signals (for e.g, $\alpha R_{NC}^{*} + (1-\alpha) R_{NC}^{SA}$) and thus improve their quality of information aggregation when they get to the No Color treatment. To test this, we could compare the data from subjects who experienced Colors before the NoColors treatment (i.e. using the data from the C1+NC2 sessions) to those who did not (from the NC1+C2 sessions).

Hypothesis 3 (Learning)**.** *Compared to the NC1 condition, subjects in the NC2 condition aggregate information more accurately.*

3.3. **Information choice.** For any subject i who uses an aggregation strategy $\hat{R}_C^i$, the optimal choice of signal depends jointly on the signal-prices and i's beliefs about how $\hat{R}_C^i$ interacts with signal choice. The variation in signal prices Δp helps us measure the demand for signals without assuming any structure on $\hat{R}_C^i$. When $\Delta p = 0$, all signal compositions are equally expensive. Thus, any subject i from the Active Choice treatment should choose a signal combination that she believes would deliver the most accurate report, given $\hat{R}_C^i$. If they believe that extreme signals might result in more a more inaccurate report, then they would prefer

[10] Recall that subjects were never provided feedback, so this learning can only occur if subjects realize the connection between colors and the sample average introspectively.

diverse signals. It is important to note that, i's choices are guided by her *beliefs* about which signal choices lead to more accuracy, i.e, the perceived value of information, which might not be equal to the actual value of information. In the range $\Delta p > 0$, as Δp increases, diversification gets increasingly expensive, while its perceived value stays the same. Thus, they should choose diverse signals less often.

Hypothesis 4 (Demand for diversity at zero cost)**.** *In the AC treatment, subjects prefer diverse signals over extreme signals for* $\Delta p = 0$*.*

The sample-average rule imposed in the Average treatment, disregards all information about the unobserved signals. Thus, unless the subjects themselves are using the sample-average rule in the Colors treatment, which is unlikely, the actual value of diversification increases significantly when the sample average rule is imposed.[11] Does the perceived value of diversification react to this change? Our next hypothesis is about how the perceived value of diversification, as measured through the demand for diverse signals, changes when the sample-average rule is imposed:

Hypothesis 5 (Higher demand at higher value)**.** *Compared to the AC treatment, subjects in the Avg treatment are more likely to choose diverse signals over extreme signals for all* $\Delta p > 0$*.*

To summarize, hypotheses 1, 2, and 3 presume that subjects use a naive aggregation rule (for e.g, sample-average or Gambler's fallacy) to predict comparative statics over signal diversity, experience, or informativeness of treatments (C versus NC treatments). Sophisticated aggregation behavior

[11]Recall that we expect the sample-average rule to be more prevalent in the NoColors treatment, rather than the Colors treatment.

would result in their rejection. Our last two hypotheses (4 and 5) posit sophisticated signal choice favoring diverse signals, and would be rejected if subjects are naive in their signal choice.

4. Results

We address Hypotheses 1, 2 and 3 using the data from all rounds of C1+NC2 and NC1+C2 sessions. For ease of exposition, we define a subject to be inexperienced during rounds 1-20, and experienced when they are in rounds 21-40, having played a different treatment previously in rounds 1-20. In Table 3, we isolate the effects of diverse signals, the observation of colors, and experience. We regress the absolute deviation from Bayesian reports on indicator variables for extreme signals, the C treatment, and Experience, plus all interaction terms, clustering standard errors at the subject level. The baseline observations are the NC1 rounds where inexperienced subjects observed diverse signals. From the top left panel, we see that observing extreme signals increases the average absolute reporting error by 2.08 units ($p < 0.001$) for inexperienced NC subjects. For experienced subjects who play the NC treatment in the second 20 rounds the relative effect of extreme signals is slightly smaller, increasing average error by 2.08-.37=1.71 units ($p < 0.01$) relative to diverse signals. To place this effect size in context, given the incentive structure of the experiment, observing diverse signals in the NC treatment increases expected earnings by approximately \$1.26 or \$1.07 for experienced or inexperienced subjects, respectively. Further, the standard errors indicate that the effects of extreme signals are estimated with reasonable precision (see also the top left panel in Table 4). The standard errors, of around half a unit on the 100 point scale used in the experiment, are approximately 28 times smaller than the expected absolute difference between the Bayesian estimate and the Sample Average heuristic when extreme signals are observed.

	Absolute deviation from Bayesian report
$\mathbb{1}_{Extreme}$	2.08***
	(0.51)
$\mathbb{1}_{Color}$	-1.06
	(0.74)
$\mathbb{1}_{Extreme} \times \mathbb{1}_{Color}$	-1.33
	(0.78)
$\mathbb{1}_{Experienced}$	-1.48
	(0.77)
$\mathbb{1}_{Extreme} \times \mathbb{1}_{Experienced}$	-0.37
	(0.71)
$\mathbb{1}_{Color} \times \mathbb{1}_{Experienced}$	2.10
	(1.47)
$\mathbb{1}_{Extreme} \times \mathbb{1}_{Color} \times \mathbb{1}_{Experienced}$	0.16
	(1.19)
Constant	4.87***
	(0.50)
N	2360

TABLE 3. The effects of Extreme signals, the observability of colors, and subject experience, on the absolute deviation of subject reports from the Bayesian report. The omitted category is observations from the No Colors rounds 1-20 where subjects observed diverse signals. Standard errors clustered at subject level are reported in parentheses (59 clusters). $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$. Data includes all rounds of C1+NC2 and NC1+C2 sessions.

To aid the interpretation of these results we also report, in Table 4, the interaction effects from this regression. The bottom left panel of 4 shows the effects of observing Colors on the average absolute deviation from Bayesian reports. A statistically significant effect is only found for inexperienced subjects who observe extreme signals: the improvement in reports, when observing extreme signals, for subjects who are participating in the C treatment (relative to those participating in the NC treatment) and have not yet experienced the other treatment is 2.40

units ($p < 0.05$).[12] The effects of Experience are shown in the top right panel. There is an improvement in reports in the NC treatment for subjects who have already had experience in the C treatment (relative to those who play the NC treatment first), but the effects are not significant at the 5% level.

Result 1. (a): Diverse signals improve reports, relative to the Bayesian benchmark, only when color information is not available. (Qualified support for Hypothesis 1.)

(b): Observing colors improves reports, relative to the Bayesian benchmark, only when signals are extreme and subjects are inexperienced. (Qualified support for Hypothesis 2.)

(c): Prior experience with the Colors treatment does not cause a statistically significant improvement in reports in the NoColors treatment. (Fails to support Hypothesis 3.)

Result 1 documented the effects of extreme signals and observing colors on the absolute error of subject reports, but is silent on whether errors are generated by biased reports or are generated by unbiased variance in reports.

We define *bias towards the extremity* as instances where the report was lower than (Bayesian estimate $-$ 1) when all signals were low, or the report was higher than (Bayesian estimate + 1) when all signals were high. Conversely, we define *bias against the extremity* as in the report was higher than (Bayesian estimate + 1) when all signals were low, and lower than (Bayesian estimate $-$ 1) when all signals were high. We allow the ± 1 tolerance band around the Bayesian estimate to allow for inconsistencies between how the computer and subjects rounded fractions, and our results are robust to

[12] That is, this statistic is a between subject measure of the effect of observing the colors of unobserved signals among inexperienced subjects.

Extreme signals	Rounds 1-20	Rounds 21-40
Colors	0.75	0.54
	(0.59)	(0.56)
NoColors	2.08***	1.71**
	(0.51)	(0.49)

Experience	Diverse signals	Extreme signals
Colors	0.61	0.41
	(0.82)	(1.20)
NoColors	-1.48	-1.85
	(0.77)	(1.06)

Colors	Rounds 1-20	Rounds 21-40
Diverse signals	-1.06	1.04
	(0.74)	(0.84)
Extreme signals	-2.40*	-0.14
	(1.11)	(1.14)

TABLE 4. The top left table measures the effect of moving from diverse signals to extreme signals, at each level of Colors and Experience, on the absolute deviation from the Bayesian report. The bottom left table measures the effect of moving from the NC treatment to the C treatment at each level of Extreme and Experience, on the absolute deviation from the Bayesian report. The top right table measures the effect of experience at each level of Colors and Extreme, on the absolute deviation from the Bayesian report. All values are calculated from a regression of the absolute deviation from the Bayesian report on indicators for Extreme signals, Colors, and Experience, plus all interaction terms, with standard errors clustered at the subject level. Standard errors are in parenthesis. $^{*}p < 0.05,^{**} p < 0.01,^{***} p < 0.001$

alternative tolerance bands, for example, ± 1.5 or ± 2. Misreporting towards the extremity is consistent with subjects employing the generalized sample-average ($\alpha R^*_{NC} + (1-\alpha) R^{SA}_{NC}$), and misreporting against the extremity is consistent with the generalized Gambler's fallacy ($\alpha R^*_{NC} + (1-\alpha) R^{GF}_{NC}$).

Under systematic misreporting towards the extremity, subjects would be more likely to over-report with respect to the Bayesian paradigm under all-high signals. Similarly, subjects would be more likely to under-report under all-low signals. In columns [1] and [2] of Table 5, we report a multinomial probit regression of whether the subjects under or over-reported, on whether the signals were all-high or all-low. We find that extreme signals increase the probability of both under and over reporting, implying noisier reports rather

	$\mathbb{1}_{Underreport}$	$\mathbb{1}_{Overreport}$
	[1]	[2]
	NC1,NC2	NC1,NC2
$\mathbb{1}_{Red}$	0.43*	0.36
	(.21)	(.20)
$\mathbb{1}_{Blue}$	0.25	0.52*
	(.20)	(.22)
Constant	-0.06	-0.07
	(0.16)	(0.16)
N	1180	1180

TABLE 5. Multinomial probit regression of the probability of underreporting (column [1]), correctly reporting (base group, not shown), and overreporting (column [2]) on indicator variables for observing all Red signals or all Blue signals (with diverse signals as the base group). Standard errors clustered at subject level are reported in parentheses. There were 59 clusters. $^{*}p < 0.05,^{**} p < 0.01,^{***} p < 0.001$.

than systematic under/ over reporting. although the effect is stronger and significant for high signals causing under reporting (and vice versa). Importantly, there is no significant difference in the proportion of under (or over) reporting when signals are all-high as compared to all-low. If extreme reports caused biased reports, we would expect the rate of under reporting to be substantially different, and also to differ in sign, when facing all-high as compared to all-low signals.

In the Appendix we include an alternative analysis that focuses on the magnitude, rather than the probability, of misreporting. The conclusions are the same: there is no evidence of bias in our sample.

Result 2: In the baseline (non-color rounds), aggregate behavior is inconsistent with systematic misreporting biased towards the extremity. (Fails to support Hypothesis 1.)

To dig deeper, we conduct a subject-level analysis. For each subject who faced extreme signals in the No Color rounds 3 or more times, we calculate how frequently they misreported towards and against the extremity of their observed signals. In Figure 4.1, we bubble-plot these fractions of misreports that were towards or against the extremity, for these same subjects. There are 9 (out of 54) subjects who misreport in the direction of the sample-average heuristic but never misreport in the direction of the Gambler's fallacy, and 11 subjects who do the opposite. Thus, 9 and 11 subjects are, respectively, consistent with sample-average and Gambler's fallacy, and the remaining 34 subjects do not show a systematic bias under extreme signals.[13] The data overall is slightly biased towards the lower right of the figure, suggesting that mistakes a la Gambler's fallacy were marginally more prevalent than Sample average rule.

We also calculate, at the subject level, the average absolute deviation from the Bayesian report and the naive sample-average report across each of the Color and No Color rounds, and then plot this data in Figure 4.2a. We plot the average absolute distance from the Bayesian report on the x-axis, and the distance from the sample-average report on the y-axis. Subjects who are positioned above the 45-degree line are closer to the Bayesian average, and subjects below the 45-degree are closer to the sample-average report, with distance from the 45-degree line giving an indication of the size of the advantage of one rule over the other. It is immediately visually apparent that (i) most subjects are above the 45-degree line and, therefore, on average, closer to the Bayesian report than the sample-average

[13]Alternative classification procedures lead to similar conclusions. For example, we could classify a subject as exhibiting the Gambler's fallacy if 80% of their choices are biased in the direction of the fallacy (and similarly for the sample-average heuristic). In this case, we would classify 8 subjects as exhibiting the Gambler's fallacy, 6 subjects as exhibiting the sample-average heuristic, and 40 subjects as being unbiased.

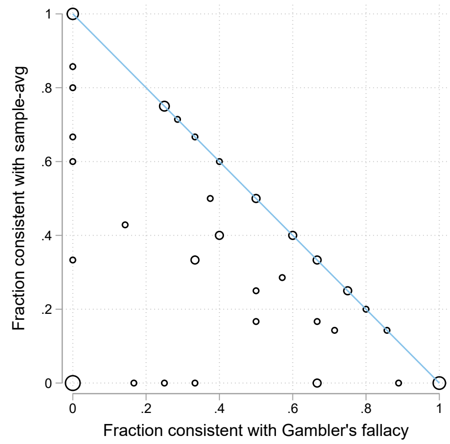
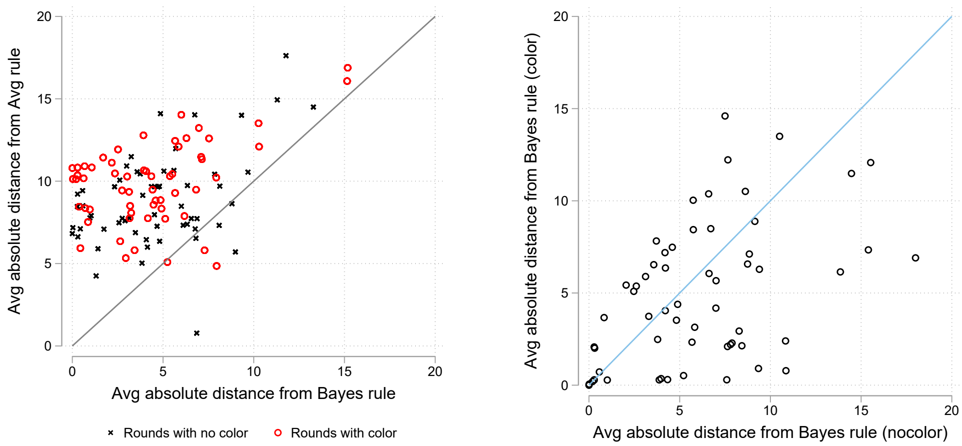


FIGURE 4.1. Proportion of choices, by subject, that are consistent with either the gambler's fallacy or the sample average bias for extreme signals in the NC treatment. Restricted to subjects who observed extreme signals at least three times. The size of each bubble represents the number of subjects at each point.

report (ii) there is no strong relationship between the observability of colors and average deviations from either rule. In fact, more than 60% of subjects were, on average, within 5 points of the Bayesian update for both treatments.

In figure 4.2b we drill further down into the distinction between the C and NC treatments. In this panel, we plot the subject-level average absolute deviation from Bayes rule in the NoColor and Color rounds on the x and y axis, respectively. Here, subjects above the 45-degree line provide better reports in the NC treatment, and subjects below the 45-degree line provide better reports in the C treatment, relative to the Bayesian optimal report.

Result 3: Reports are substantially closer to the Bayesian paradigm than to the naive sample-average paradigm, for both the C and NC treatments. (Fails to support Hypothesis 2.)



(A) Average absolute deviation from the Sample Average rule plotted against the average absolute deviation from the Bayesian optimal report, at the subject level, for the C treatment (red circles) and the NC treatment (black crosses).

(B) Average absolute deviation from the Bayesian optimal report in the C treatment plotted against the average absolute deviation from the Bayesian optimal report in the NC treatment, at the subject level, when signals were extreme.

FIGURE 4.2. Heterogeneity

Hypothesis 5 predicts that subjects will have a stronger preference for diverse signals in the Average treatment than the in the Active Choice treatment. This hypothesis relies on the implicit assumption that diverse signals are actually more valuable in the Avg treatment than in the AC treatment: an assumption that is testable in our data. Not only is the assumption supported on average, across the subject population, we find that that it holds individually for all but one subject. For each subject we calculate the average absolute difference of their guesses from the true $\bar{s}$ for all Active choice rounds, separately for diverse and extreme signals, and interpret their difference as the loss from choosing extreme signals. In Figure 4.3 we plot the CDF of the subject specific losses from extreme signals. It is clear from the figure that, while the aggregation rule used by the median subject experiences essentially no gain from diversity, there is substantial

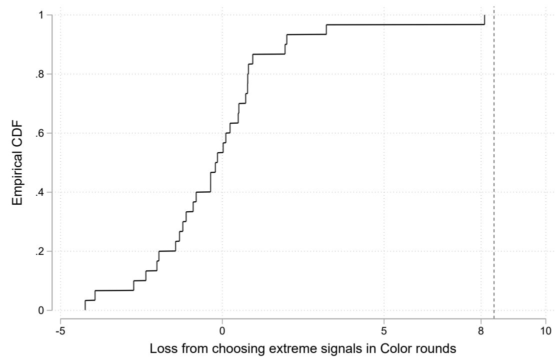


FIGURE 4.3. CDF of the average gain (at the subject level) from diversification over all Color rounds. The vertical dashed line denotes the average gain from diversification observed in the Average rounds.

heterogeneity across subjects. We also calculate and plot the average loss from extreme signals in the Average rounds (under the sample-average aggregation) as the vertical line at, approximately, 8.4. Despite the heterogeneity, the gain from diverse signals in the AC treatment is less than 8.4 units for all subjects.

Result 4: Diverse signals improve reports (relative to $\bar{s}$) in the Average rounds more than they do in the Active Choice rounds.

An unfortunate programming constraint, which was not noticed until after the experiments were run, allows us to only observe the signals finally received by the subjects and not the chosen or requested signals. The received signals can differ from the requested signals when, for example, the subject requests two Blue signals but only one Blue signal is available in $\{s_2, s_3, .., s_7\}$. In the rounds where $\{s_2, s_3, .., s_7\}$ contained at least two signals of either color, the signals requested and received are guaranteed to be identical. We use only data from these rounds for our following results on signal choice. Given subjects made their signal choice without knowing the composition of $\{s_2, s_3, ..s_7\}$, conditioning on the composition does not bias the analysis of the signal choice in any conceivable way.[14] For comparison, we repeat the analysis with the full data set, containing all rounds, in Appendix A. The results are similar.

In Figure 4.4, we plot the proportion of extreme signal choices against Δp, the difference between the price of own signal and the price of the other signal. We do this separately for when subjects aggregate on their own (Active Choice rounds) versus when they aggregate under sample-average rule (Average rounds), pooling similar price differences to simplify the figure.

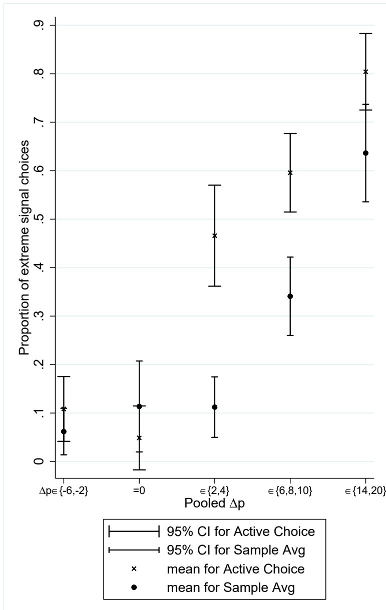
When $\Delta p = 0$, approximately 10% of signal choices are extreme as most subjects prefer a diverse portfolio of signals. As the price-difference increases, subjects become more likely to choose extreme signals. In the Active Choice treatment, even at the highest price-difference group ($\Delta p \in \{14, 20\}$), approximately 20% of decisions are in favor of diverse signals and this figure is higher still in the Average treatment (approximately 35%). In the Average treatment, where there is a clear objective benchmark for optimal behavior, it is optimal to choose diverse signals at all values of Δp, which indicates that some subjects are over-reacting to the price of signals.[15]

Result 5: Subjects rarely choose extreme signals when $\Delta p = 0$. (Support for Hypothesis 4.)

As seen in Figure 4.4, at every $\Delta p > 0$, subjects are less likely to choose extreme signals under Average than under Active Choice. Table 4.4 presents

[14] $\{s_2, s_3, ..s_7\}$ are drawn independently of s_1, and hence s_1 is uninformative about the other 6 signals.

[15] In expectation, a one-unit improvement in guess accuracy is worth $(200 - 50) \times 0.06 = 9$ points, which implies that the value of diverse signals is substantially larger than the largest value of $\Delta p = 20$.



Probit (Active choice + Avg)		
	Dependent variable	
	1 if signal choice was Extreme, 0 otherwise	
	[1]	[2]
Δp	0.10***	0.10***
	(0.02)	(0.02)
$\Delta p \times \mathbb{1}_{SmplAvg}$	-0.00	-0.01
	(0.01)	(0.02)
$\mathbb{1}_{SmplAvg}$	-0.61**	-0.92**
	(0.20)	(0.28)
Dev		-0.11**
		(0.03)
Dev$\times \mathbb{1}_{SmplAvg}$		0.08*
		(0.04)
Constant	-0.65***	-0.19
	(0.14)	(0.19)
N	909	909

FIGURE 4.4. The left hand panel plots the proportion of extreme signal choices against Δp, the difference between the price of an own colored signal and an other colored signal, separately for when subjects update on their own (Active Choice rounds) versus when they update under sample-average rule (Average rounds) with 95% confidence intervals. The right hand panel presents a probit regression of extreme choice on Δp, with a dummy variable for the Average rounds with standard errors clustered at the subject level (30 subjects). In regression [2], we additionally control for "Dev", which is calculated at the subject level as the average deviation from the Bayesian update across the first 20 Active Choice rounds and restricted to rounds with diverse signals. Both panels restrict the data to only include rounds where there were at least two red signals and at least two blue signals in $\{s_2, s_3, s_4, s_5, s_6, s_7\}$

two regressions designed to study the determinants of extreme signal choices. In column [1], which controls for Δp and the treatment (Average or Active Choice) we observe a negative coefficient on the dummy for Average. In column [2] we add a control, Dev, which captures subject-level guess accuracy in the Active Choice rounds when the subject observed diverse signals. Thus, the higher the value of Dev, the worse was the quality of information aggregation by the subject.[16]

From column [2] of Table 4.4 we conclude, given the negative coefficient on Dev, that subjects who are worse at aggregating information are more likely to select diverse signals in the AC treatment. To provide some context for the estimated value of -0.11, in the Active Choice treatment, at the sample average $\Delta(p)$ and average value of Dev, a one unit improvement in signal aggregation ability leads to a 4 percentage point decrease in the likelihood of choosing extreme signals. That is, subjects who are worse at aggregating exhibit some self-awareness of this and respond by giving themselves an easier updating problem. For the Avg treatment, the effect of Dev is $-0.11 + 0.08 = 0.03$ and statistically insignificant, suggesting that the choice of signals is independent of aggregating ability in the Avg treatment. This forms our final result, and suggests that subjects are able to separate information acquisition from information processing. The estimates of $\Delta(p)$ and Dev in Table 4.4 are also rather precise, with standard errors of only 0.02 and 0.03, respectively.

Result 6: Subjects choose extreme signals less frequently (i) in the Average treatment and (ii) when they are poor at aggregating information in the Active Choice treatment. (Support for Hypothesis 5.)

Result 7: Signal choices in the Average treatment are independent of guess accuracy in the Active Choice treatment.

[16] If the Dev variable was calculated using rounds with both diverse and extreme signals, then there would be a potential endogeneity problem: subjects who choose diverse signals more often might have systematically different deviations from the Bayesian update. We checked the robustness of column [2] by recalculating Dev using either all rounds with diverse signals or all rounds with extreme signals, finding that the results are qualitatively unchanged.

5. Conclusion

In this paper we study the value of and the demand for diverse information sources in a simple decision environment where information-processing does not require contingent reasoning. We find that subjects are remarkably resistant to making mistakes when receiving news with transparent bias. Subject reports are unbiased even when signal bias is extreme. We find little evidence for subjects following a naive sample-average rule or committing the Gambler's paradox. Most importantly, subjects are willing to pay non-negligible amounts to observe diversified, rather than biased, signals and subject demand for diverse information reacts rationally to the value and cost of diverse information. Remarkably, subjects who perform poorly when aggregating information appear to be cognizant of their limitations and exhibit a stronger demand for diversified information. Finally, when we exogenously impose a naive sample-average aggregation rule the subject level demand for diversified information sources is, rationally, not dependent on subject level aggregation ability.

Previous research (e.g. Enke [2020]) has identified conditions under which information bias can lead to ex-post polarization. Our results, conversely, demonstrate that when information bias is transparent, and motivated reasoning is not present, that subjects are surprisingly good at constructing a balanced portfolio of signals and then constructing unbiased estimates of the true state of the world. Our results advocate for greater transparency in media bias, so that individuals can choose the right portfolio of information to make better choices.

We believe that there are three additional factors that future work could integrate into our design. First, in our experiment, demand for diversity was measured when information was objectively biased. It would be interesting to extend this measurement to the cases where the bias is ambiguous or subjective. For example, some may believe that the Huffington Post is fairly balanced (AllSides rates Huffington Post as far Left) while others might see Fox as fair and balanced (the AllSides media bias rating for Fox is Lean Right). Second, the bias from the two different colors were symmetrically opposite, mutually exclusive and complementary. Subjects might not subjectively believe in the existence of symmetrically opposite biased sources at all. In the 2019 Gallup study, around 72% of Democrats only 31% of Republicans agreed that there were enough sources to be able to sort out the facts. Information choice might be affected by such beliefs. Third, it would be interesting to see if information fatigue induces reversion to the choice of extreme signals. In response to feeling overwhelmed by the abundance of news sources in the current media environment, a plurality of Americans (39%) reported they only pay attention to one or two trusted sources, while 30% try to consult a variety of sources to see where they agree.

Appendix A. Supplementary Tables

A.1. **Magnitude of bias.** In the main text, we evaluate bias in reports by evaluating the probability that a subject over or under reports as a function of observing all high or all low signals. Here, we provide a robustness check by examining the magnitude of bias as a function of observing all high or all low signal in Table A.1. The first column of Table A.1 presents a restricted version of the regression contained in Table 3, while the second and third tables estimate the bias of reports. The second column uses only rounds where the sample average heuristic lies above the Bayesian estimate, and the third column uses rounds where the sample average heuristic lies below the Bayesian estimate. If reports are biased, for either extreme or diverse signal observations, then either the constant or the coefficient on Extreme must be different from zero. As the Table shows, all coefficients in both regressions are close to zero and not statistically significant, indicating that there is no evidence of bias in our sample. Note that the second and third column use only data from the NC treatment, given that the sample average heuristic is unnatural in the C treatment.

A.2. **Robustness of the information selection results.** As described in the main text, our data only allows us to observe the signals received by the subjects (and not the signals requested by the subjects). In the main text we restrict attention to a subset of rounds for which we know that requested and received signals must be the same. Here, we provide a robustness test by including data from all rounds. Figure A.1 is a robustness check on Figure 4.4.

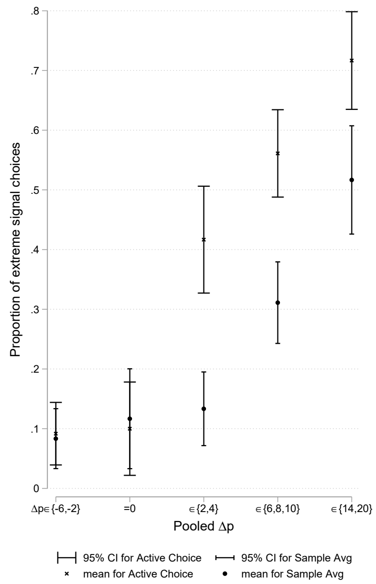
A.3. **The fifth treatment.** In a fifth treatment, we gave subjects the opportunity to construct an algorithm that would calculate the subject's report of $\bar{s}$ automatically given the signals and colors that the subject observed. Despite our

Sample	$\lvert deviation \rvert$ Full C+NC	$deviation$ avg>Bayes NC only	$deviation$ avg<Bayes NC only
$\mathbb{1}_{Extreme}$	1.83***	0.53	-0.78
	(0.37)	(1.00)	(0.86)
$\mathbb{1}_{Color}$	-0.07		
	(0.30)		
$\mathbb{1}_{Extreme} \times \mathbb{1}_{Color}$	-1.19*		
	(0.48)		
Constant	4.21***	-0.24	0.06
	(0.39)	(0.46)	(0.37)
N	2360	615	565

TABLE A.1. The first column regresses the absolute deviation of subject reports from the Bayesian benchmark on an indicator for Extreme signals and a Color treatment indicator, using the full sample (all rounds of C1+NC2 and NC1+C2 sessions). The second and third columns regress the deviation of subject reports from the Bayesian benchmark on an indicator for Extreme signals, using only data from the NC treatment, using samples restricted to upwards and downwards biased signals. Standard errors clustered at subject level are reported in parentheses (59 clusters). ${}^{*}p < 0.05, {}^{**}p < 0.01, {}^{***}p < 0.001$.

attempts to design an interface that would be intuitive and easy for subjects to understand, the algorithms that subjects constructed demonstrated that subjects did not understand the algorithm construction process sufficiently. Sessions that contained the algorithm treatment consisted of 20 rounds of the Colors treatment, followed by 10 rounds of the algorithm treatment, followed by 10 rounds of the Sample Average treatment. In the remainder of this subsection, we repeat some of the analysis from the main text with the inclusion of data from Colors and Sample Average treatments in these sessions. The results are similar and more precise as the standard errors shrink further.

Table A.2 is a robustness check on Table 3, and Table A.3 is a robustness check on the right hand panel of Figure 4.4.



Probit (Active choice + Avg)		
	Dependent variable	
	1 if signal choice was Extreme, 0 otherwise	
	[1]	[2]
Δp	0.09***	0.09***
	(0.01)	(0.01)
$\Delta p \times \mathbb{1}_{SmplAvg}$	-0.01	-0.01
	(0.01)	(0.01)
$\mathbb{1}_{SmplAvg}$	-0.45**	-0.79**
	(0.16)	(0.24)
Dev		-0.08**
		(.03)
Dev$\times \mathbb{1}_{SmplAvg}$		0.08*
		(0.04)
Constant	-0.71***	-0.32
	(0.12)	(0.17)
N	1200	1200

FIGURE A.1. Robustness check for the information selection results from Figure 4.4, including data from all rounds.

	Absolute deviation from Bayesian report
$\mathbb{1}_{Extreme}$	2.08***
	(0.51)
$\mathbb{1}_{Color}$	-0.53
	(0.79)
$\mathbb{1}_{Extreme} \times \mathbb{1}_{Color}$	-1.49*
	(0.70)
$\mathbb{1}_{Experienced}$	-1.48
	(0.77)
$\mathbb{1}_{Extreme} \times \mathbb{1}_{Experienced}$	-0.37
	(0.71)
$\mathbb{1}_{Color} \times \mathbb{1}_{Experienced}$	1.57
	(1.43)
$\mathbb{1}_{Extreme} \times \mathbb{1}_{Color} \times \mathbb{1}_{Experienced}$	0.32
	(1.12)
Constant	4.87***
	(0.50)
N	2900

Table A.2. (Robustness check on Table 3 when we add the Color rounds from the fifth treatment.) The effects of Extreme signals, the observability of colors, and subject experience, on the absolute deviation of subject reports from the Bayesian report. The omitted category is observations from the No Colors rounds 1-20 where subjects observed diverse signals. Standard errors clustered at subject level, including subjects from the Algorithm sessions, are reported in parentheses (86 clusters). $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$.

Probit (Active choice + Avg)		
	Dependent variable	
	1 if signal choice was Extreme, 0 otherwise	
	[1]	[2]
Δp	0.10^{***}	0.10^{***}
	(0.02)	(0.02)
$\Delta p \times \mathbb{1}_{SmplAvg}$	-0.00	-0.01
	(0.01)	(0.02)
$\mathbb{1}_{SmplAvg}$	-0.55^{**}	-1.03^{***}
	(0.17)	(0.23)
Dev		-0.12^{**}
		(.03)
Dev$\times \mathbb{1}_{SmplAvg}$		0.13^{**}
		(0.04)
Constant	-0.65^{***}	-0.13
	(0.14)	(0.19)
N	1116	1116

TABLE A.3. (Robustness check on the right hand panel of Figure 4.4 when we add the AC rounds from the fifth treatment.) Probit regression of extreme choice on Δp, with a dummy variable for the Average rounds with standard errors clustered at the subject level, including subjects from the Algorithm sessions (57 subjects). In regression [2], we additionally control for "Dev", which is calculated at the subject level as the average deviation from the Bayes rule across the first 20 Active Choice rounds.

References

Sandro Ambuehl and Shengwu Li. Belief updating and the demand for information. *Games and Economic Behavior*, 109:21–39, 2018.

Felipe A Araujo, Stephanie W Wang, and Alistair J Wilson. The times they are a-changing: experimenting with dynamic adverse selection. *American Economic Journal: Microeconomics*, 13(4):1–22, 2021.

Daniel J Benjamin. Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186, 2019.

Mary A. Burke and Michael Manz. Economic literacy and inflation expectations: Evidence from a laboratory experiment. *Journal of Money, Credit and Banking*, 46(7):1421–1456, 2014.

Francesco Capozza, Ingar Haaland, Christopher Roth, and Johannes Wohlfart. Studying information acquisition in the field: A practical guide and review. *Available at SSRN 3952561*, 2021.

Gary Charness, Ryan Oprea, and Sevgi Yuksel. How do people choose between biased information sources? evidence from a laboratory experiment. *Journal of the European Economic Association*, 19(3):1656–1691, 2021.

Daniel L Chen, Tobias J Moskowitz, and Kelly Shue. Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *The Quarterly Journal of Economics*, 131(3):1181–1242, 2016a.

Daniel L. Chen, Martin Schonger, and Chris Wickens. oTree – an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016b.

John Duffy, Ed Hopkins, Tatiana Kornienko, and Mingye Ma. Information choice in a social learning experiment. *Games and Economic Behavior*, 118:295–315, 2019.

John Duffy, Ed Hopkins, and Tatiana Kornienko. Lone wolf or herd animal? information choice and learning from others. *European Economic Review*, 134: 103690, 2021.

Benjamin Enke. What you see is all there is. *The Quarterly Journal of Economics*, 135(3):1363–1398, 2020.

Ignacio Esponda and Emanuel Vespa. Contingent thinking and the sure-thing principle: Revisiting classic anomalies in the laboratory. Technical report, Working paper, 2019.

Leon Festinger. *A theory of cognitive dissonance*, volume 2. Stanford university press, 1962.

Klaus Fiedler and Peter Juslin. Taking the interface between mind and environment seriously. *Information sampling and adaptive cognition*, pages 3–29, 2006.

Andreas Fuster, Ricardo Perez-Truglia, Mirko Wiederholt, and Basit Zafar. Expectations with endogenous information acqusiition: An experimental investigation. *Review of Eocnomics and Statistics*, Forthcoming.

Ananda Ganguly and Joshua Tasoff. Fantasy and dread: The demand for information and the consumption utility of the future. *Management Science*, 63 (12):4037–4060, 2017.

Romain Gauriot, Lionel Page, and John Wooders. Nash at wimbledon: evidence from half a million serves. *Available at SSRN 2850919*, 2016.

Matthew Gentzkow and Jesse M Shapiro. Media bias and reputation. *Journal of political Economy*, 114(2):280–316, 2006.

Ben Greiner. Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125, July 2015.

Daniel Kahneman and Amos Tversky. On the psychology of prediction. *Psychological review*, 80(4):237, 1973.

Ziva Kunda. The case for motivated reasoning. *Psychological bulletin*, 108(3): 480, 1990.

Pierre Simon marquis de Laplace. *Essai philosophique sur les probabilités*. Bachelier, 1840.

Yusufcan Masatlioglu, A Yesim Orhun, and Collin Raymond. Intrinsic information preferences and skewness. *Ross School of Business Paper*, 2017.

Heiner Mikosch, Christopher Roth, Samad Sarferaz, and Johannes Wohlfart. Uncertainty and information acquisition: Evidence from firms and households. *Available at SSRN*, 2021.

Giovanni Montanari and Salvatore Nunnari. Audi alteram partem: An experiment on selective exposure to information. *Unpublished manuscript*, 2019.

Kirby Nielsen. Preferences for the resolution of uncertainty and the timing of information. *Journal of Economic Theory*, 189:105090, 2020.

Marina Niessner and Eric C So. Bad news bearers: The negative tilt of the financial press. *Available at SSRN 3219831*, 2018.

Richard E Nisbett and Lee Ross. Human inference: Strategies and shortcomings of social judgment. 1980.

An T Oskarsson, Leaf Van Boven, Gary H McClelland, and Reid Hastie. What's next? judging sequences of binary events. *Psychological bulletin*, 135(2):262, 2009.

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

Amnon Rapoport and David V Budescu. Randomization in individual choice behavior. *Psychological Review*, 104(3):603, 1997.

Christopher Roth, Sonja Settele, and Johannes Wohlfart. Risk exposure and acquisition of macroeconomic information. *American Economic Review: Insights*, 4(1):34–53, 2022.

Amos Tversky and Daniel Kahneman. Belief in the law of small numbers. *Psychological bulletin*, 76(2):105, 1971.

Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131, 1974.

Florian Zimmermann. Clumped or piecewise? evidence on preferences for information. *Management Science*, 61(4):740–753, 2015.