How Do People Choose Between Biased Information Sources? Evidence from a Laboratory Experiment

Gary Charness Ryan Oprea Sevgi Yuksel

March, 2020

Abstract

People in our experiment choose between two information sources with opposing biases in order to inform their guesses about a binary state. By varying the nature of the bias, we vary whether it is optimal to consult information sources biased towards or against prior beliefs. Even in our deliberately-abstract setting, there is strong evidence of confirmation-seeking and to a lesser extent contradiction-seeking heuristics leading people to choose information sources biased towards or against their priors. Analysis of post-experiment survey questions suggests that subjects follow these rules due to fundamental errors in reasoning about the relative informativeness of biased information sources.

We would like to thank Larbi Alaoui, Giuseppe Attanasi, Pierpaolo Battagali, Roland Bénabou, Doug Bernheim, Juan Carrillo, Martin Dufwenberg, Benjamin Enke, Drew Fudenberg, Uri Gneezy, PJ Healy, Frank Heinemann, Steffen Huck, Muriel Niederle, Salvatore Nunnnari, Theo Offerman, Antonio Penta, Jacopo Perego, Collin Raymond, Julian Romero, Aldo Rustichini, Charles Sprenger, Séverine Toussaert, Emanuel Vespa, Georg Weizsäcker, Leeat Yariv, Florian Zimmermann and seminar participants at Bocconi, TU Berlin, University of Arizona, Chapman, Ohio State, Stanford, UCSD Rady, Princeton, University of Maryland, Lyon, Pompeu Fabra, UPenn, Columbia, USF, USC, PSE, NYU, Caltech, Oxford, ESA conferences in Berlin and Richmond, SITE, Predictive Game Theory conference at Northwestern, Munich Belief Workshop, CoCoLab Workshop in Nice, Workshop on Communication in Amsterdam and the Max Planck Laboratory Inaugural Conference at Bonn for helpful comments.
Charness: Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 93106, charness@econ.ucsb.edu; Oprea: Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 93106, roprea@gmail.com; Yuksel: Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 93106, sevgi.yuksel@ucsb.edu.

1 Introduction

Modern decision-makers frequently must choose between competing sources of information (e.g. news sources, policy analysts, medical or financial advisors, scientific papers, product reviews) – a task that is complicated by the fact that in many (perhaps most) contexts, available information sources are *biased* in some way (i.e. in favor of some ideology, product, theory etc.). The question of *how* people choose between information sources in the face of such bias has become an important topic of discussion in recent years and a popular answer is that people tend to consult sources that are biased in support of their own prior beliefs. This type of "confirmation-seeking" behavior can lead to distorted posterior beliefs and is often blamed for contemporary problems like information-bubbles, echo-chambers, conspiracy theories and product lock-in.[1]

In this paper we report an experiment designed to gather some direct evidence on how (and how well) subjects choose between biased sources of instrumentally- valuable information and to understand why they make mistakes. Errors like confirmation-seeking in the field are often attributed to motivated reasoning (i.e. people get utility from receiving information that confirms their prior) or reputational concerns (i.e. people trust information sources aligned with their prior) in the economics literature. We do not doubt that motivated reasoning and reputations are important, but our aim is to study whether such errors are also driven by more fundamental mistakes in how people *reason* about what makes information valuable.[2]

For this reason, our experiment is simple and deliberately abstract, removing the scope for motivated reasoning and reputational concerns in the design and thereby providing a lower bound for the prevalence of confirmation-seeking behavior. We provide subjects with a prior over a stochastic state of the world ("green" or "orange") and pay them for correctly guessing the state. Before guessing, subjects first choose one of two computerized information structures from which to receive a signal about the state.

Identifying basic errors in reasoning about information is important because such errors have the

[1] Prior (2007), Pariser (2011) and Sunstein (2018) provide a discussion of the literature on the topic in the context of political information. Gentzkow & Shapiro (2011) and Iyengar & Hahn (2009) and recently Jo (2017) show evidence of selective exposure. Empirical literature studying the determinants of media bias find bias to be mostly demand driven (Gentzkow & Shapiro (2010)); and there is evidence to suggest that biased news sources can have an impact on voting behavior. (See DellaVigna & Kaplan (2007), Martin & Yurukoglu (2017), Adena et al. (2015) and Durante et al. (2017) .)

[2] Both motivated and unmotivated forms of confirmation bias have been discussed and documented in the psychology literature (see Nickerson (1998) for a review). We discuss our contribution in relation to this literature at the end of this section.

potential to produce behaviors like confirmation-seeking even in environments in which agents have little motivated attachment to their prior beliefs or concerns about information quality. They also have the potential to *intensify* the effects of motivated beliefs in settings where agents *are* attached to their priors. Finally, to the degree that such basic reasoning mistakes exist, policy responses beyond those designed to combat motivated reasoning and reputational bias will be required to remove behaviors like confirmation seeking and their consequences.

A key feature of our design is that subjects can see that each information structure provided to them is biased towards one of the two states, and we vary the nature of this bias across problems: in one set of problems, we induce bias by *commission* (generated by the possibility that the signal is misleading), while in others, we induce bias by *omission* (generated by the possibility that a signal revealing the state is not produced).[3] By varying the nature of the bias, we are able to identify patterns in how decision makers seek out information. As we show in Section 3, an optimizing decision maker chooses the information structure biased towards her prior when the bias is by commission, but does the exact opposite when the bias is by omission. In contrast, we identify a confirmation-seeking decision maker as one who consistently chooses the information structure biased towards her prior. Other salient decision rules such as "contradiction-seeking" (always choose the information structure biased against your prior) and "certainty-seeking" (choose the information structure most likely to give unambiguous signals) are likewise naturally identifiable under our design.

We find that subjects frequently make costly mistakes in choosing between biased information structures, but that these mistakes are not random. Indeed, clustering analysis, finite-mixture models and simple classification exercises suggest that most mistakes are made by subjects consistently employing confirmation-seeking, contradiction-seeking or certainty-seeking rules. The most common of these is confirmation-seeking which is employed about as frequently as the optimal rule in the subject population even in our neutral and abstract design. Contradiction-seeking and certainty-seeking rules are employed substantially (and significantly) less often.

We included a number of diagnostic features in the design that allow us to examine the mechanism behind this pattern of results.

First, as mentioned above, our experiment was designed to rule out, *ex ante*, the most common explanations offered for confirmation-seeking behavior such as *motivated beliefs* (people have a

[3]See Gentzkow et al. (2015) for further discussion of categorization of bias. Bias by commission is related to cheap-talk games as in Crawford & Sobel (1982) and persuasion games as in Kamenica & Gentzkow (2011); bias by omission is related to disclosure games as in Milgrom & Roberts (1986)).

preference for reinforcing closely held beliefs) or *reputational concerns* (people trust information sources that conform with their prior beliefs). By studying an environment in which (i) prior beliefs are over abstract states (the color of a ball drawn from an urn) and (ii) change radically from decision to decision over the course of the experiment (so that subjects are unlikely to be attached to beliefs about any one state over the course of the experiment) we effectively eliminate the scope for motivated beliefs to generate confirmation-seeking behavior. Likewise, the design removes reputational concerns by providing subjects with exact signal distributions for each information structure they might choose.

The data strongly suggest these design choices were successful in removing both motivated reasoning and reputational concerns: subjects show no aversion to making guesses or forming beliefs that contradict their priors (even subjects who exhibit strong confirmation-seeking behavior in selecting information structures).

Second, our results indicate that the use of sub-optimal decision rules is not due to subjects being confused about the instructions or underwhelmed by the incentives. When we conduct control tasks in which incentives and framing are identical to those in the main treatments but information structures are both biased in the *same direction* (hence Blackwell ranked), subjects have little difficulty making optimal choices.

Third, in addition to measuring choices over information structures and guessing behavior about the state, we elicit subjects' beliefs about the state as a function of the signals. We also have sessions where we include an additional sequence of decision problems in which subjects are *exogenously* assigned each of the possible information structures and asked to report guesses and beliefs for every possible signal (we will call these exogenous assignments "EX" decisions). These additional problems allow us to characterize how subjects value each of the information structures presented to them in the earlier *endogenous* ("END") decision problems (in which subjects chose between information structures).

With these data, we are able to assess and ultimately rule out errors in responding to signals as the primary driver of mistakes in choices over information structures. In particular, we find that all types make more accurate guesses when they are assigned the optimal information structure and are not much more likely to choose optimal information structures when their beliefs are accurate enough to correctly rank information structures than when they are not. The explanatory power of guesses and beliefs over patterns of choices is dwarfed by the explanatory power of the direction of the bias (towards or against the prior) in information structures. That is, when subjects make mistakes, they seem to be responding to the biases themselves (relative to priors) rather than to guesses and beliefs induced by information structures.

Fourth, a set of incentivized cognitive questions, conducted post-experiment, suggest that use of non-optimal decision rules is linked to reasoning ability. Subjects who perform well on cognitive tests that measure logical reasoning (Wason selection tasks and Belief Bias tasks) are significantly more likely to use optimal decision rules.

Finally, at the end of the experiment we elicit incentivized advice from subjects in order to understand how and why they made their decisions. Subjects are often able to accurately describe the rules they employed in the experiment and often provide highly-sophisticated (but typically mistaken) justifications for their use of these rules. These descriptions and justifications indicate that subjects find decision rules like confirmation-seeking normatively appealing and that the resulting patterns of mistakes in choosing between information sources are founded in simple errors in reasoning.

Taken together, our results are quite clear on the sources of mistakes in our data. Sub-optimal decision rules like confirmation-seeking emerge here because *it is difficult to correctly reason through information valuation problems*, even in our deliberately simple setting. Subjects who make systematic mistakes typically describe their own decision process in ways that are insightful but incomplete. For instance confirmation-seeking types reason correctly about the importance of accuracy in the more likely state, but confuse the direction of bias for accuracy. Contradiction-seeking types correctly identify the importance of instrumentally-valuable information but underweight accuracy. These mistakes are systematic, leading subjects to employ rules that match (or anti-match) biases in information to prior beliefs and arise because subjects mistakenly reason that these rules will result in good outcomes.

This finding suggests that errors like confirmation-seeking may be more fundamental and therefore more widespread than is commonly supposed.[4] For instance, while frequently-invoked sources of confirmation-seeking behavior such as motivated reasoning are likely to be important in environments in which people are highly attached to their prior beliefs, our experiment shows that serious mistakes are prevalent even in environments in which they have no attachment to their priors at all. This suggests that patterns like confirmation seeking may arise in contexts in which we would not expect motivated reasoning to be a serious concern. For instance consumers often consult review sources like Amazon, Yelp, Metacritic, healthgrades.com to guide their purchases

[4] Confirmation bias is historically discussed mostly in relation to established beliefs and emotionally significant issues. See Rabin & Schrag (1999) for a discussion of this literature. Charness & Dave (2017) also provides a recent overview.

(e.g. for consumer goods, restaurants, movies or doctors) and have a choice of whether to focus on positive or negative reviews. While, clearly, there are cases in which consumers may have strong affective attachments to their priors about the best purchasing decisions, it is easy to imagine cases in which they likely have little motivation other than to make the best choice. Likewise, doctors who have patients at risk of some disease (given age, family history etc.) often must choose among medical tests (with different type 1 and 2 error rates) to determine a diagnosis or path of treatment. Although doctors may sometimes have strong attachment to their priors over the diseases they are diagnosing, it seems probable that doctors often care little about the diagnosis per se but are motivated only to diagnose correctly. However, even in this latter case, our results suggest that reasoning errors can lead to significant confirmation-seeking. Very similar examples naturally arise for investors, managers and policy makers.[5] Our results suggest that decision makers in such settings may be quite vulnerable to serious errors like confirmation-seeking simply because of the way they reason about the value of information.

The remainder of the paper is organized as follows. In Section 2 we review prior literature and in Section 3 we describe the theoretical setting and generate a set of predictions. In Section 4 we discuss our experimental design and in Section 5 we describe how this design allows us to identify distinct decision rules in the data at the individual level. In Section 6 we present the main results of the experiment, reporting the optimality of choices over information structures and evidence for the prevalence of different decision rules. In Section 7 we use diagnostic features of our design to evaluate explanations for the use of sub-optimal decision rules. We close the paper in Section 8 with a Conclusion.

2 Prior Literature

Our work contributes to several literatures. Several related papers study choices over sources of information with instrumental value, but tackle fundamentally different questions. Most recently, Montanari & Nunnari (2019) follow a similar experimental design to ours (by studying choices over biased information structures), but their focus is on understanding how the reliability of

[5]Nonetheless, we emphasize that our findings hardly rule out factors like motivated reasoning as important drivers of confirmation-seeking behavior in the field. Indeed, it seems likely that *both* mechanisms operate simultaneously in the field in some contexts, particularly in highly-affectively-loaded settings such as politics, ideology and religion where motivated reasoning is most plausibly at work. Similarly, it is not difficult to imagine difficulties in reasoning interacting with reputational problems to induce more widespread incidence of confirmation-seeking behavior than either mechanism operating alone.

signals interacts with bias to influence choices. Ambuehl (2017) studies how incentives affect choices between information sources and focuses on a very different choice problem (the decision of whether or not to eat an insect). Likewise Duffy et al. (2017, 2018) study how subjects choose between social and private information sources that vary in relative quality. Also related is a recent literature on the choice between information structures in settings in which, in contrast to our setting, information has no instrumental value. Nielsen (2018), Zimmermann (2014) and Falk & Zimmermann (2017) study preferences over the timing and concentration of information and how that can change with one's prior. Masatlioglu et al. (2017) find evidence of strong preference for *positive skewness* in a setting with non-instrumental information: subjects prefer information structures which rule out more uncertainty about the desired outcome (while tolerating uncertainty about the undesired outcome) compared to those that rule out more uncertainty about the undesired outcome (while tolerating uncertainty about the desired outcome). By contrast, in our setting, choices over information structures have clear payoff consequences, and subjects have no *ex-ante* reason to differentiate between the states.

This paper is part of a larger literature in economics studying biases in learning and demand for information, mostly focusing on deviations from the Bayesian paradigm in belief updating.[6] Eil & Rao (2011), Burks et al. (2013) and Mobius et al. (2011) find that subjects asymmetrically update beliefs in response to objective information about themselves, over-weighting positive feedback relative to negative.[7,8,9] Ambuehl & Li (2018) study demand for information and find that individuals differ consistently in their responsiveness to information.[10] In field settings involving medical and financial decisions, Oster et al. (2013) and Sicherman et al. (2015) find evidence for information avoidance, where agents trade-off instrumental information with the desire to hold on to optimistic

[6] See Camerer (1998) and Benjamin et al. (2016) for more comprehensive literature reviews. Bénabou & Tirole (2016) and Bénabou (2015) provide reviews on motivated beliefs.

[7] Eliaz & Schotter (2010) identify a *confidence effect*: the desire to increase one's posterior belief by ruling out "bad news" even when information has no instrumental value. In another setting in which information has no instrumental value, Loewenstein et al. (2014) study diverse motives driving the preference to obtain or avoid information. Burks et al. (2013) find that it is people who think that they did well who want to purchase information about how they did, while people who didn't think they did well don't want to know. Zimmermann (2018) studies motivated beliefs in the presence of feedback.

[8] Charness & Levin (2005) study how people make choices in environments where Bayesian updating and reinforcement learning push behavior in opposite directions.

[9] In social settings, Weizsäcker (2010), Andreoni & Mylovanov (2012) and recently Eyster et al. (2018) study failures in learning and persistence of disagreement.

[10] Ambuehl & Li (2018) find undervaluation of high-quality information, and a disproportionate preference for information that may yield certainty. Although they make up a small share of our data, we also find some such certainty-seeking behavior in our environment.

beliefs. Focusing on the endogenous design of information structures, Fréchette et al. (2018) study the role of commitment. Also relevant is a recent literature highlighting subjects' difficulties updating beliefs in the face of selection issues in signal distributions which have a relationship to some of our findings. This literature finds that many subjects neglect missing information (Enke (2017)), suffer from *correlation neglect* (e.g. Enke & Zimmermann (2017)) and show insufficient skepticism about failures by other subjects to disclose information (Jin et al. (2015)).[11]

Finally, our paper also relates to a literature in psychology on confirmation bias (see Nickerson (1998) for an overview), particularly a literature that studies biases in how people seek out evidence for hypothesis testing. For example, there is a large literature that studies variations on the Wason (1960) selection task in which subjects must decide what pieces of information will effectively test logical hypotheses of the "if p, then q" variety (see Klayman (1995) and Baron (2000) for reviews and Jones & Sugden (2001) for a contribution in the economics literature). Studies in this literature tend to find that people are drawn to uninformative evidence, and that this may cause them to neglect informative evidence that might usefully falsify the hypothesis. These experiments are usually conducted in settings that are quite different from ours (and from the information-theoretic setting of economic models of information acquisition).[12] Closer to our work are several experiments in this literature that document information acquisition patterns that reveal a tendency by subjects to search for evidence that is likely to be reinforcing of their initial hypothesis (Klayman & Ha (1987) Skov & Sherman (1986), Slowiaczek et al. (1992), Baron et al. (1988)).[13]

Our paper makes two main contributions relative to this literature. First, we show that, in the standard information-theoretic setting economists use to model information acquisition, confirmation-seeking behavior (and related mistakes) occur at a high rate due purely to failures to reason about the value of information structures. Specifically, our paper is the first to carefully evaluate the causal relationship between subjects' ability to (i) choose optimal sources of information and (ii) make effective use of information they receive in such settings. We do this by having subjects (i) choose information sources and (ii) respond to exogenously assigned information separately, and comparing behavior in each case within-subject. Second, our paper is the first to embed

[11]For further literature on this topic, we refer the reader to Eyster & Rabin (2005), Gabaix & Laibson (2006), Mullainathan et al. (2008), Heidhues et al. (2016), and Ngangoue & Weizsacker (2018).

[12]For example, in the standard formulation of the Wason selection task, subjects are not provided with a prior on the likelihood that the hypothesis tested is true, or information about how the evidence to be evaluated were generated.

[13]These studies also differ from our work in that the initial hypothesis subjects are asked to evaluate is not necessarily associated with the ex-ante most likely state, but instead is often induced purely by the framing of the question.

confirmation-seeking behavior in a taxonomy of related patterns of behavior and to conduct an experiment specifically designed to separate between these different types of behavior. We do this by varying the nature of the bias in the information sources subjects choose between, allowing us to identify systematic mistakes that would be difficult to observe in previous designs. Doing this lets us to not only identify confirmation-seeking subjects but also optimal-, contradiction- and certainty seeking subjects.

3 Theoretical framework

Suppose there is an unobserved state of the world $\theta \in \Theta := \{L, R\}$ (called the left and right state) and an agent must submit a guess $a \in \Theta$ concerning the state. That is, the preferences of the agent conditional on the state θ and action $a \in \Theta$ can be represented by the following utility function:

$$u(a \mid \theta) = \begin{cases} 1, & \text{if } a = \theta \\ 0, & \text{if } a \neq \theta \end{cases}$$

The agent has an *ex-ante* prior belief p_0 over the probability that $\theta = R$ (referred to as right-leaning if $p_0 > 0.5$) and receives a signal s from an information structure to inform her guess. An information structure σ is a stochastic mapping from the state space to a set of signals $S := \{l, n, r\}$.

In this section we analyze how an agent should assess the relative *value* (discussed in Section 3.2) of two information structures that differ in their relative *biases* (as defined in Section 3.1) in providing signals of the state.

3.1 Bias

We first operationalize the notion of bias using a partial order introduced by Gentzkow et al. (2015). Let $p(s \,|\, \sigma, p_o)$ denote the Bayesian posterior belief that $\theta = R$ conditional on receiving signal s from information structure σ for an agent with prior p_o. Two information structures, σ and σ', are said to be *consistent* if they have the same support, i.e. produce the same type of signals, and the signals are ordered in the same way in terms of the posteriors they generate. Formally, for any two signals s and s', $p(s \,|\, \sigma, p_o) > p(s' \,|\, \sigma, p_o)$ if and only if $p(s \,|\, \sigma', p_o) > p(s' \,|\, \sigma', p_o)$. Note that, conditional on the prior, any information structure can be associated with the distribution of posteriors it generates. Let $\mu(\sigma|\sigma')$ denote the distribution of posteriors when an agent believes signals come from σ when they are actually generated by σ'.[14]

[14]Note that Definition 1 could be extended to allow for continuous state and signal spaces.

	l	r
$\theta = L$	1	0
$\theta = R$	$1-\lambda$	λ

	l	r
$\theta = L$	λ	$1-\lambda$
$\theta = R$	0	1

Table 1: Two symmetrically-biased information structures with bias by commission *Notes: Each cell represents the probability of a signal being generated conditional on θ, $0 < \lambda < 1$. (For example, the information structure presented on the right-hand side produces signal r with probability 1 when the state is R, and with probability $1-\lambda$ when the state is L.*

Definition 1. *(Gentzkow et al. (2015))*
σ' is biased to the right (that is towards R) of σ if

(i) σ and σ' are consistent, and

(ii) $\mu(\sigma|\sigma')$ first-order stochastically dominates $\mu(\sigma|\sigma)$.

This definition provides only a partial order on information structures but, by focusing on the distribution of posteriors, it allows us to consider (and compare) different types of bias.

The literature has emphasized two main forms of bias.[15] First, information structures can be biased through the possibility of misleading signals. This is bias by *commission*, which is often observed in environments where information is non-verifiable (as in cheap-talk models of Crawford & Sobel (1982)). Table 1 provides an example of two information structures that can provide signals l or r that noisily indicate states L and R respectively. Bias arises in this case through the possibility that the information structure sends a "misleading" signal conditional on the state of the world (i.e. sending signal l when the state is R or signal r when the state is L). Notice that, following Definition 1, the information structure shown on the right-hand side in Table 1 is *biased to the right* of the information structure that is depicted on the left-hand side.[16]

Second, information can be biased through the possibility that the state will not be *revealed.* This is bias by *omission*, which is often observed in environments where information is verifiable

[15]See DellaVigna & Hermle (2017) for empirical analysis of bias by omission vs. commission in movie reviews. Gentzkow et al. (2015) provides an overview of the literature and includes discussion of a third type of bias, bias by filtering, which captures bias introduced by selection when information sources are constrained by the dimensionality of signal space.

[16]By construction, the information structure on the right-hand side is more likely to produce r signals (and hence less likely to produce l signals). With both information structures, the r signal produces a higher posterior than the l signal. Hence if an agent believed that signals were coming from the information structure on the left-hand side, switching to the right-hand side would lead to a first order stochastic shift in the distribution of posteriors.

	l	n	r
$\theta = L$	λ_h	$1 - \lambda_h$	0
$\theta = R$	0	$1 - \lambda_l$	λ_l

	l	n	r
$\theta = L$	λ_l	$1 - \lambda_l$	0
$\theta = R$	0	$1 - \lambda_h$	λ_h

Table 2: Two symmetrically-biased information structures with bias by omission *Notes: Each cell represents the probability of a signal being generated conditional on θ, $0 < \lambda_l < \lambda_h < 1$. (For example, the information structure presented on the right-hand side produces signal r with probability λ_h and n with probability $1 - \lambda_h$ when the state is R.)*

(as modeled in disclosure games (e.g. Milgrom & Roberts (1986)). The information structures depicted in Table 2 provide an example. Note that in both information structures, the signals r and l are fully revealing of states R and L respectively, while the n signal can be thought of as a failure to produce a signal. Differences in bias in this case arise from differences in how often the information structure reveals the state conditional on the state of the world. As in the *commission* case analyzed above, following Definition 1, the information structure that is depicted on the right-hand side in Table 2 is biased to the right of the information structure that is depicted on the left-hand side.[17,18]

3.2 Value of an Information Structure

How do the pairs of information structures depicted in Tables 1 and 2 differ in terms of *value*? That is, in each case, which information structure would an agent optimally choose if she could receive only one signal from one information structure? An agent's objective in consulting an information structure is to improve her *guessing accuracy* which we can write as $p_0 a_R + (1 - p_0) a_L$, where a_i is the probability of guessing i when the state is i. An information structure creates value for the agent if it allows her to condition her guesses on signals in a way that changes a_R and a_L (relative to the no-information benchmark) so that $p_0 a_R + (1 - p_0) a_L$ increases.

Our main interest is in analyzing how the optimal structure choice is related to the biases of the available information structures and the prior of the agent. Our experimental design builds on

[17] With both information structures, the r, n and l signals are ranked in the natural way in terms of the posteriors they generate. And, by construction, the information structure on the right-hand side, relative to the one on the left-hand side, shifts distribution of signals from l to n and from n to r.

[18] Note that for any λ and p_0, the two information structures in Table 1 are also ranked in terms of the Monotone likelihood ratio property. This is not true for every λ_h, λ_l, for the information structures presented in Table 2, but the parameters we choose will guarantee this ordering.

the insight that the answer critically depends on the nature of the bias.

Remark 1. *If two information structures are symmetrically-biased by commission, it is optimal to receive a signal from the structure biased in the same direction as one's prior.*

Importantly the reasoning behind Remark 1 does not require Bayesian inference or even probability calculations. In this simple setting an agent could increase her guessing accuracy only by following the "recommendation" of the information structure (i.e. guessing R in response to signal r and L in response to signal l). If the agent follows signals coming from the information structure on the left-hand side of Table 1, guessing accuracy conditional on state R is $a_R = \lambda$, and guessing accuracy conditional on state L is $a_L = 1$, while these would be reversed if she chose the information structure on the right.[19] There is thus a direct trade-off in terms of guessing accuracy in the two states: (a_R, a_L) is either $(\lambda, 1)$ or $(1, \lambda)$. Remark 1 follows simply from noticing that maximizing $p_0 a_R + (1 - p_0)a_L$ naturally involves resolving this trade-off by choosing the information structure that achieves higher guessing accuracy in the ex ante more-likely state. It is, for example, optimal for an agent with right-leaning prior to choose the information structure biased towards R (the one on the right-hand side in Table 1).[20]

Remark 2. *If two information structures are symmetrically-biased by omission, it is optimal to receive a signal from the structure biased in the opposite direction as one's prior.*

Again, the reasoning behind Remark 2 requires no Bayesian inference. Suppose that the agent has a right-learning prior, $p_0 > 0.5$. Since l and r signals are fully revealing, an optimizing agent will guess L (R) after receiving l (r), regardless of which information structure is selected from Table 2. To complete her guessing strategy, the agent determines which direction to guess conditional on the n signal. If she guesses in the direction of her prior (in this case R) (an optimal response to the n signal),[21] guessing accuracy conditional on state R is $a_R = 1$ regardless of which information structure from Table 2 is selected. But, guessing accuracy conditional on state L is $a_L = \lambda_h$ under

[19] We simply read the probability of receiving the r signal conditional on state R and the probability of receiving the l signal conditional on state L for this information structure from Table 1.

[20] Notice also that in the no-information benchmark such an agent would always guess R, implying $(a_R, a_L) = (1, 0)$. Thus, following signals coming from the information structure on the right-hand side of Table 1 increases a_L from 0 to λ without decreasing a_R.

[21] Notice that a *naive* agent who does not consider the n signal to carry any information would also automatically guess in accordance with her prior conditional on this signal, and hence correctly identify the optimal information structure as outlined above. Also, in our experimental design, p_0, λ_h and λ_l are chosen such that it is indeed optimal to always guess in accordance with the prior conditional on this signal regardless of which information structure was chosen.

the left-hand structure and $a_L = \lambda_l$ under the right-hand structure. Expected guessing accuracy $p_0 a_R + (1 - p_0) a_L$ is clearly maximized by choosing the left hand structure – the structure biased against the prior – because $\lambda_h > \lambda_l$. The only other guessing strategy that the agent could possibly consider is to guess against her prior (in this case L) upon receiving the n signal from the right-hand structure.[22] This produces guessing accuracies $(a_R, a_L) = (\lambda_h, 1)$ from the right-hand structure[23] compared to $(a_R, a_L) = (1, \lambda_h)$ from the left-hand structure. Again, the agent then maximizes $p_0 a_R + (1 - p_0) a_L$ by choosing higher guessing accuracy in the ex-ante more likely state, leading her to select the structure biased against her prior (the left-hand structure).

In summary, an agent choosing optimally between the information structures presented above will choose information sources biased in the same direction as her prior when bias is by commission, but will choose to do the opposite (i.e. choose sources biased in the opposite direction as her prior) when bias is by omission. This observation lies at the heart of our experimental design.

4 Design

The goal of our experiment is to (i) *measure* subjects' ability to differentiate between more and less informative information structures and (ii) to identify the types of decision rules individual subjects use in this type of choice. Our experimental design directly mirrors the decision setting described in the previous section. In each decision problem, subjects were shown an urn on their computer screen consisting of 20 balls colored orange or green in varying proportions and were asked to guess the color of a single ball drawn randomly from the urn. To inform their guesses, subjects first chose one of a pair of computerized advisors, from which to receive a signal ("orange", "green" or "null") 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.

The experiment consisted of two *decision blocks*. In each of the 14 rounds of the main "Endogenous Advisors" (or "END") block, we presented subjects with a pair of advisors that differed in their signal structures and asked the subject to choose an advisor from which to receive a signal. After choosing an advisor, subjects were then asked to guess the color of the ball and to submit

[22] Guessing L conditional on the n signal can never be optimal with the left-hand structure (for an agent with a right-leaning prior) because this signal is more likely to be generated conditional on the state being R, rather than L, which pushes the agent's posterior further to the right.

[23] The agent is guaranteed to guess correctly in state L: this is because only the n and l signals are generated in this case and the agent is assumed to guess L conditional on both. In state R, the agent guesses correctly only when the r signal is generated (which happens with probability λ_h.

a likelihood that the ball was green vs. orange as a function of *each possible signal* their chosen advisor might send (that is, we used a version of the strategy method, e.g. Brandts & Charness (2011)).[24] Over the course of the 14 rounds we varied the menu of advisors and the composition of the urn (and therefore the subject's prior beliefs). Subjects received no feedback from decision to decision and states and signals and payments were determined only at the end of the experiment. These design choices limit the degree to which intrinsic preferences over information driven by psychological motives can affect subject's choices over information structures.[25]

In six of the decision rounds of the END block subjects faced advisors with bias by commission, choosing between the advisors shown in Table 1, with $\lambda = 0.7$. In another six decision rounds, subjects faced advisors with bias by omission, choosing between the advisors shown in Table 2, with $\lambda_h = 0.7$ and $\lambda_l = 0.3$. In each case we varied the priors (likelihood of a green ball being drawn) between $p_0 \in \{\frac{4}{20}, \frac{5}{20}, \frac{6}{20}, \frac{14}{20}, \frac{15}{20}, \frac{16}{20}\}$.[26,27] We randomized the order of bias-type and, within type, the order of the prior thoroughly across sessions and subjects.[28] Finally, in the last two decision rounds of the END block, subjects faced what we refer to as the "Blackwell problems," designed to assess subjects' comprehension of the decision environment. In these problems, the two advisors presented to the subjects were biased in the same direction, and could be easily ranked via Blackwell ordering.[29]

[24] Belief elicitation has been combined with the strategy method in a number of prior information-response experiments, e.g. Cipriani & Guarino (2010), Agranov et al. (2018), Toussaert (2017).

[25] A growing theoretical literature including Kreps & Porteus (1978), Grant et al. (1998), Caplin & Leahy (2001), Brunnermeier & Parker (2005), Kőszegi & Rabin (2009), Dillenberger & Segal (2017) studies intrinsic preferences for the timing, concentration and skewness of information.

[26] We chose these values for the prior for the following reason. We wanted to maximize the difference between the informativeness of the two information structures biased in the opposite direction in the problems with bias by commission and omission questions. Clearly, this difference disappears as the prior converges to 0.5 since the agent has no reason to favor one information structure over the other. Similarly, as the prior converges to 1, the agent is able to guess almost perfectly even without any further information, so value of *any* information structure vanishes. We choose priors in this range to balance these counteracting forces.

[27] Alternatively (and equivalently), we might have instead varied values for $\lambda, \lambda_h, \lambda_l$ keeping p_0 constant. We chose to vary p_0 because (i) we hypothesized that it was easier for subjects to internalize information about the prior and (ii) in order to prevent subjects from becoming attached to their prior belief, potentially introducing scope for motivated reasoning.

[28] Specifically we randomized at the subject level whether the first set of 6 problems corresponded to the problems with bias by commission or the problems with bias by omission. Then, within question-type, we randomized (again at the subject level) the order of the prior the subject faced.

[29] In these questions, the advisors were both biased towards the orange color. Specifically, the probability of receiving the orange signal conditional on the color of the ball being orange was 1 for both advisors and the probability of receiving the orange (green) signal conditional on the color of the ball being green was 0.7 (0.3) for one advisor and

In order to measure how subjects' guesses and beliefs were shaped by their advisors, we ran half of our subjects through an additional diagnostic decision block called "Exogenous Advisors" (or "EX") following the END block. Instead of asking subjects to choose between advisors as in END, in each of the 12 rounds of EX we *assigned* subjects one of the four advisors from Table 1 and Table 2 (again with $\lambda = 0.7$, $\lambda_h = 0.7$ and $\lambda_l = 0.3$) and asked them to guess the color of the ball and submit likelihoods of green vs. orange for each possible signal the advisor might send. For each of the four advisors we varied the prior p_0 between $\{\frac{4}{20}, \frac{5}{20}, \frac{6}{20}\}$.[30] The order of the resulting 12 choices were again randomly sequenced for each session and subject.

At the end of the experiment, subjects were asked to complete a survey that included questions on demographic information, cognitive ability and over-confidence, political attitudes and media habits, as well as questions on the subjects' learning strategy in the experiment. We also asked subjects to write down (free-form text) advice to another subject who would participate in this experiment on how to choose between the different information structures.

Responses to the decision blocks and the survey were incentivized in the following way. The most significant portion of subjects' payoff was linked to their guesses about the state. One of the decision rounds from the END block was chosen randomly, and conditional on the state and the advisor chosen by subject, a random signal was generated. The subject earned \$10 if her guess for this signal matched the state (\$0 otherwise). Another information-choice problem was chosen randomly to determine payment for beliefs. The subject's belief response to this question (conditional on the state, the advisor chosen, and the realized state) determined the likelihood (according to the Binarized Scoring Rule) of winning \$1.[31] In the sessions that included the EX block, subjects faced additional incentives as in the END block, where they had a chance to earn

0.3 (0.7) for the other. These advisors could be Blackwell ranked in the sense that one could be written as a garbling of the other one. The prior on the color of the ball being green was either 0.3 or 0.7. The order of these questions was also randomized at the subject level.

[30] Priors in the END treatment are symmetrically arranged around 0.5 and we therefore only need half of these priors to assess subjects' ability to interpret information in the EX treatment. See Footnote 39 for more on this symmetry. Doing this has the advantage of allowing us to avoid fatiguing subjects with too many questions.

[31] The advantage of this over other traditional mechanisms is that it does not rely on risk neutrality to be incentive-compatible. We use the implementation outlined in Wilson & Vespa (2018); the method was first developed in Hossain & Okui (2013). We removed hedging motives between the belief responses and guesses by randomly determining the state independently for each case. We calibrated incentives for belief responses to be an order of magnitude smaller than incentives associated with guesses in order to prevent these elicitations from influencing subjects' reasoning about the optimality of information structure choices. Nonetheless, in every decision problem the information structure that maximized expected earnings from belief elicitation coincided with the optimal information structure which maximized expected guessing accuracy.

$10 based on their guess in a randomly-selected round, and $1 depending on their answers to the belief question in another randomly-selected round. In addition, subjects were paid $5 for show up, $2 for filling out the survey and could earn up to $2.5 from the cognitive ability questions in the survey. We also incentivized the advice question in the survey. Subjects were told that advice written down by three randomly-selected subjects would be shown to another subject in each session and the subject who wrote down the advice chosen to be "most useful" would receive an additional $1.[32]

We ran the experiment using 344 subjects over 18 sessions at the EBEL laboratory at UC Santa Barbara between January and June 2018. All of the sessions included the END decision block; 9 sessions (158 subjects) also included the EX block.[33] Detailed instructions with examples for different types of information structures were read out loud.[34,35] As a result of the payoff structure implemented, subjects earnings varied from $7.50 to $31.75 (average $21.5) for a 80-100 minute session.

5 Identifying Decision Rules Using the Design

The experiment outlined in Section 4 was designed not only to assess how well subjects choose between information structures, but also to facilitate identification of alternative sub-optimal behavioral rules like confirmation-seeking that are often raised in popular discussion.

As we outline in Section 3.2, a subject using an optimal[36] decision rule (an "Optimal" type subject) will choose the information structure biased in the same direction as her prior when the bias is by commission and will choose the structure biased in the opposite direction of her prior when it is by omission. Panel (a) of Figure 1 illustrates a subject who uses a pure optimal rule by plotting an example from our dataset. From left to right we plot the sequence of decisions the subject faced in the END block of the experiment (recall that this ordering is randomized across subjects) and below the x-axis we list the subject's prior for that decision. The y-axis represents the information structures biased towards the orange vs. green state. Hollow dots show the optimal

[32]Martínez-Marquina et al. (2018) uses a similar incentive structure for advice data.

[33]Sessions were computerized using *Qualtrics*. Sessions with only the END block lasted 80min, while those with the EX block went for 110min.

[34]See Online Appendix G for screenshots from the experiment and a copy of the instructions and the survey.

[35]We also highlighted to the subjects every time the set of possible advisors changed. Subjects were allowed to spend as much time as needed on each decision.

[36]Throughout the paper we define an information structure as "optimal" if it generates higher value (enables an agent to maximize guessing accuracy), as described in Section 3.2.

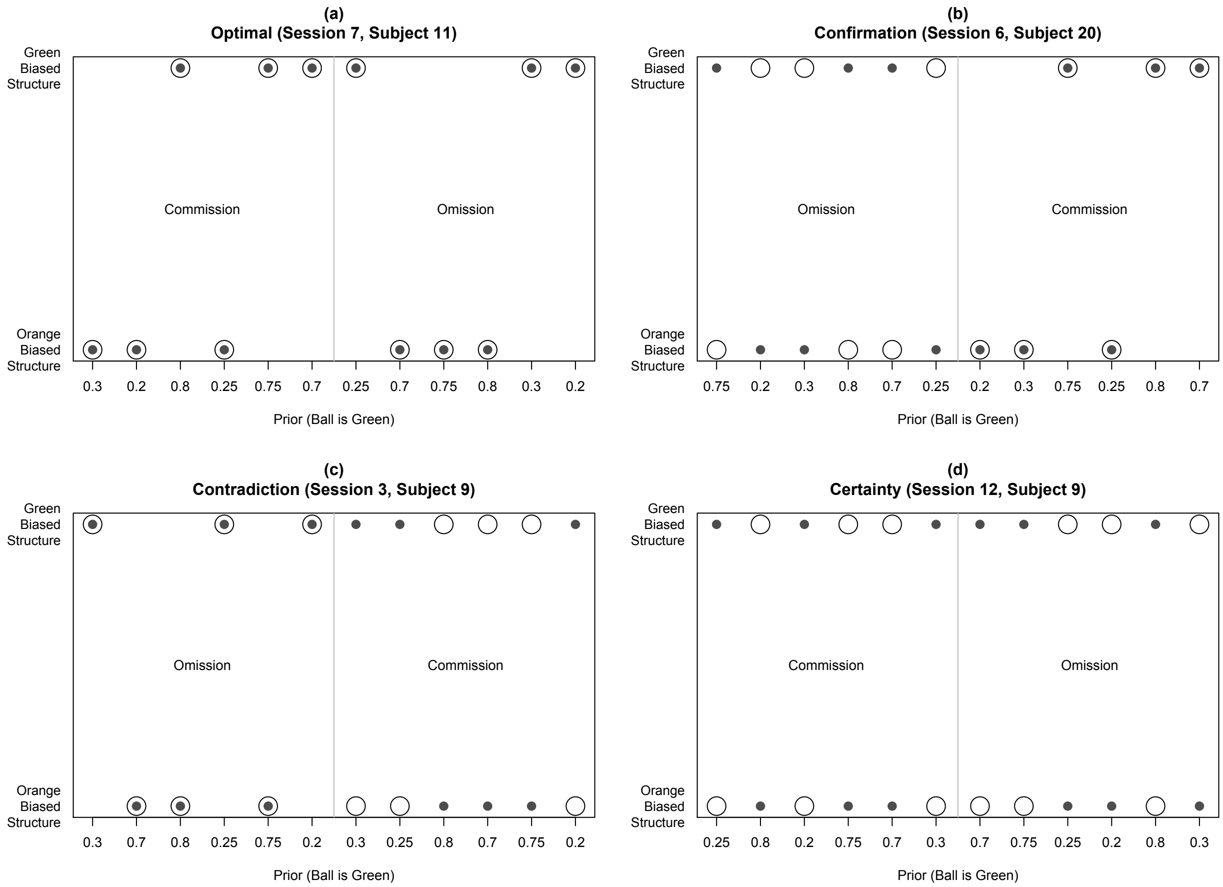


Figure 1: Identifying Decision Rules *Notes: Graphs compare actual choices over information structures to optimal behavior. The sequence of problems encountered by each subject is denoted on the x-axis (numbers indicate prior). Hollow dots show the optimal choice in each case and solid dots show the subject's actual choices.*

choice in each case and solid dots show the subject's actual choices. These two dots always overlap for an Optimal type.

The design also allows us to crisply identify several salient decision rules that depart from optimality. A subject using a consistent "confirmation-seeking" rule (a "Confirmation" type) will always – in problems both with bias by commission and omission – choose an information structure that is biased *in the same direction as her prior*. A subject using the opposite "contradiction-seeking" rule (a "Contradiction" type) will instead always choose an information structure biased *in the opposite direction* of her prior. Panels (b) and (c) show examples of subjects from our dataset employing each of these rules; note that each type makes perfectly optimal decisions for one bias-type (by commission or omission) and perfectly sub-optimal decisions for the other. It is important to emphasize that each of these rules leave no less distinctive a fingerprint than an optimal rule.

Finally, after designing the experiment, we discovered that a fourth decision rule was readily identified using our design. A subject seeking to maximize her chances of receiving a signal that identifies the states with certainty (a "Certainty" type) will do exactly the opposite of an Optimal decision maker, choosing an advisor that contradicts her prior in the problems with bias by commission and an advisor that confirms her prior in the problems with bias by omission.[37] Panel (d) shows an example subject from our dataset.

To summarize, we can identify types of subjects employing different decision rules in the experiment by examining how the information structure-bias favored by a subject (towards or against her prior) changes with the bias-type (commission or omission). We can identify four decision rules: Optimal, Confirmation-seeking, Contradiction-seeking and Certainty-seeking. These patterns are distinctive and they are extremely unlikely to arise by chance. Thus, the experiment is designed to allow us to distinguish the use of these decision rules from one another and from other behaviors such as random decision-making.

6 Main Results

In Section 6.1 we report aggregate results and provide evidence showing that (i) subjects frequently choose sub-optimal information structures, (ii) that these mistakes tend towards confirmation-seeking information structures and are highly non-random and (iii) that these mistakes come at a high cost to guessing accuracy. In Section 6.2 we type subjects according to the rules discussed in Section 5 and report that Confirmation types are as common as Optimal types.[38]

6.1 Optimality of choices over information structures

Figure 2 plots the aggregate rate at which subjects chose the optimal information structure (in the END decision block) for each condition (Bias by commission, Bias by omission and Blackwell) and

[37] Note that such a subject would choose the information structure that is most likely to generate a fully revealing signal in the ex-ante most likely state (which corresponds to the suboptimal choice in both the commission and omission problems). This type of behavior is consistent with some recent findings in the literature. Ambuehl & Li (2018) provides evidence that due to non-Bayesian updating patterns, subjects undervalue high-quality information that does not fully resolve uncertainty, while disproportionately preferring information that may yield certainty.

[38] All tests reported in the text are based on probit (for binary variables) or linear (for continuous variables) regressions clustered at the subject level.

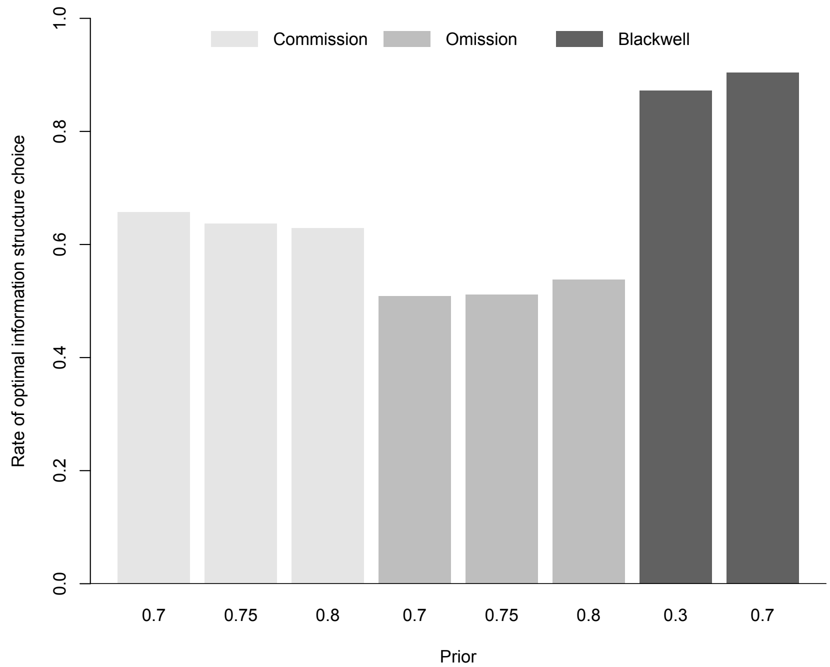


Figure 2: Frequency of choosing the optimal information structure by prior and question type

prior-strength (0.7,0.75 and 0.8).[39] In our control Blackwell problems, subjects choose the optimal information structure 89% of the time, significantly more often than under non-Blackwell problems ($p < 0.001$).

This high rate of optimal information-structure choice collapses when the two information structures cannot be Blackwell ranked. With bias by commission, subjects choose the optimal information structure only 64% of the time and under bias by omission, this rate drops significantly ($p < 0.001$) to 52%, which is not statistically different from 50%. These mistake rates do not vary much by the strength of the prior.[40]

Importantly, these high mistake rates come at a great cost in terms of guessing accuracy. To show this, Figure 3 plots average improvement in guessing accuracy relative to the no-information

[39] Priors are symmetrically arrayed around 0.5 in the design: $p_0 \in \{\frac{4}{20}, \frac{5}{20}, \frac{6}{20}, \frac{14}{20}, \frac{15}{20}, \frac{16}{20}\}$. The "prior-strength" normalizes the direction for commission and omission problems: $\max\{p_0, 1-p_0\}$. Regression results confirm that subject behavior is not different across state colors. Such a normalization is not possible in the Blackwell problems where both information structures were biased in the same direction and the prior was varied to be in the same or opposite direction of this bias (0.3 or 0.7).

[40] Furthermore, within omission and commission problems, the rate of optimal information structure choice does *not* increase with the associated relative gains (which depends on the prior).

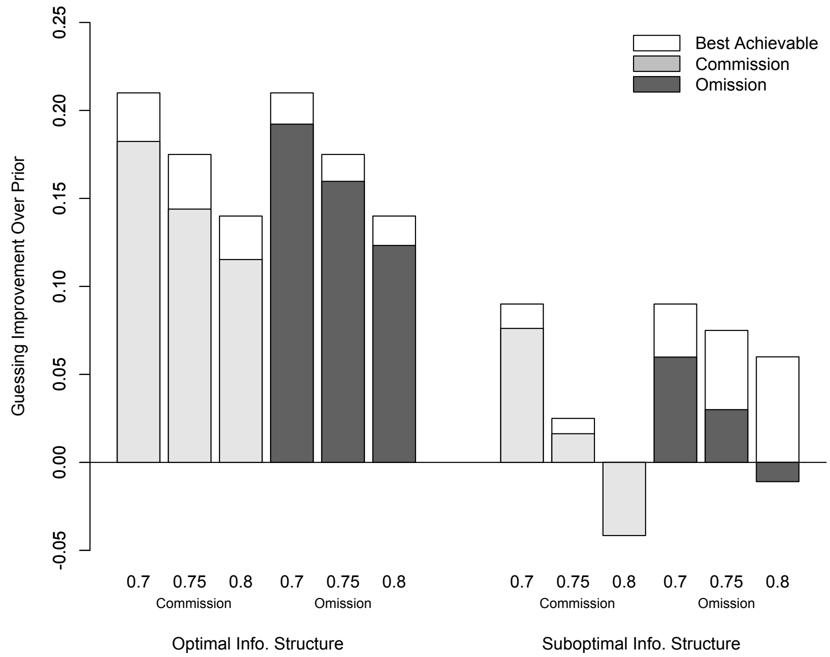


Figure 3: Guessing accuracy relative to prior by information-structure choice, prior and bias-type. *Notes: The Best Achievable benchmark was equal to 0 with the suboptimal information structure when the prior strength was 0.8 and the bias-type was commission: it was optimal to guess in accordance with one's prior regardless of the signal realizations.*

benchmark (guessing in accordance with the prior).[41] On the left side of the Figure we plot (broken down by bias-type and prior-strength) this measure for subjects who chose the optimal information structure and, on the right, for those who chose the sub-optimal one. In white we plot what this measure would be for a Bayesian subject – an upper bound for improvement in guessing accuracy conditional on information structure choice – while the shaded bars plot the actual improvement in guessing accuracy for subjects in the experiment.

The results show that subjects learn dramatically less (improve upon their priors to a much smaller degree) in the aggregate after choosing sub-optimal than after choosing optimal information structures. Part of this is a result of subjects making much less use of information provided by sub-optimal rather than optimal structures (the shaded bars are smaller relative to the white bars with sub-optimal than with optimal structures).[42] Indeed, in some decision rounds where the prior

[41]For this, we first calculate each subject's *expected* guessing accuracy in each problem given her guesses conditional on each signal (which we observe as a result of the strategy method). This accounts for the probability with which each signal would be generated conditional on each state and whether or not the subject's guess conditional on this signal is correct.

[42]The overall difference in difference – guessing accuracy relative to best achievable conditional on information structure choice – is highly significant ($p < 0.001$).

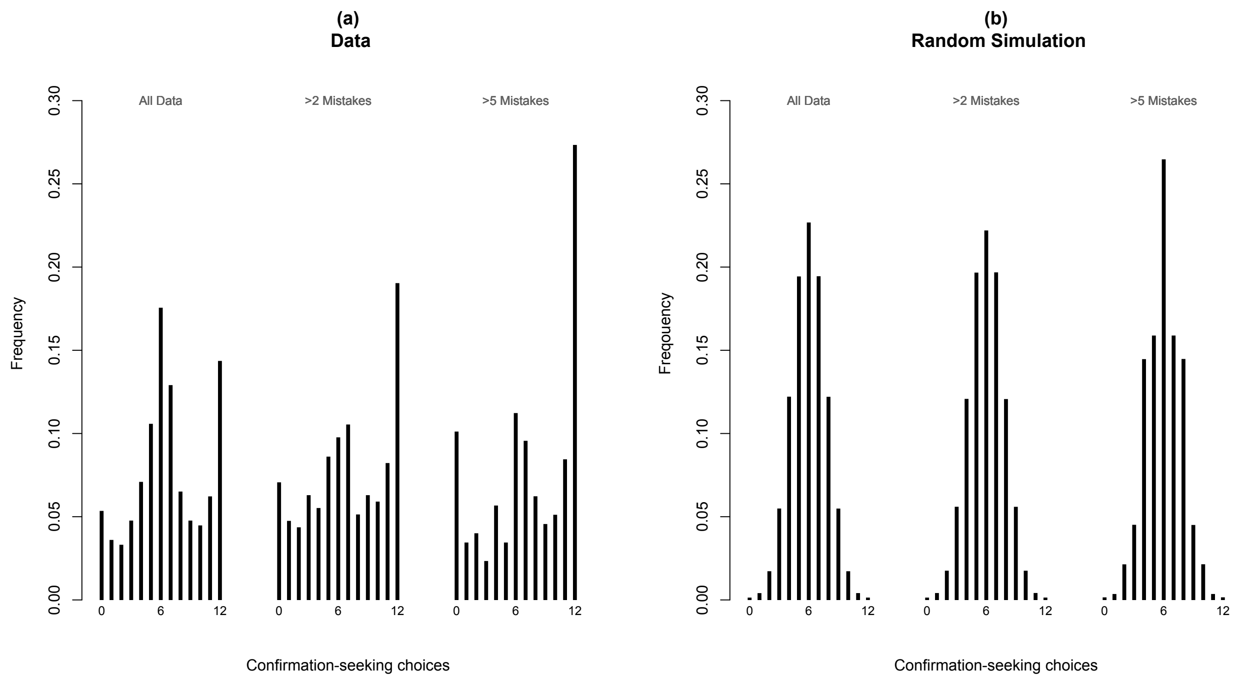


Figure 4: Frequency of information-structure choices that coincide with confirmation-seeking *Notes: "All Data" refers to the full data set. "> 2 Mistakes" ("> 5 Mistakes") is the subset of subjects that make more than 2 (5) mistakes relative to the optimum in choosing an information structure. Random Simulation is for* 10^7 *subjects who are assumed to choose randomly between information structures.*

strength was 0.8, subjects who choose the sub-optimal structures actually make worse guesses than they would by simply following their priors.[43,44]

We summarize the findings so far in a first Result:

Result 1. *Subjects frequently choose sub-optimal information structures, leading to severe failures in learning.*

Are these high mistake rates in choices over information structures driven by consistent decision rules (such as those outlined in Section 5), or are they driven by random errors in choice? To answer this question, we count the number of times each subject chooses information structures that are

[43] As the Best Achievable benchmark indicates, when the prior strength was 0.8 and the bias-type was by commission, the optimal guess was always equal to the ex-ante most likely state regardless of the signal from the sub-optimal information structure, implying that guessing accuracy couldn't improved with this information structure. In contrast, if the bias-type was omission, due to the fully revealing nature of the r and l signals, there was always opportunity for learning.

[44] Interpreting these results is complicated in part by self selection: subjects choose their information structures, which might lead to biased estimates of the causal impact of structures on learning. This is one of the motivations for the EX treatment, which removes this potential source of bias. See Section 7.4 below.

biased towards their priors. Panel (a) of Figure 4 plots histograms of this measure across subjects for the full dataset ("All Data") and for the subset of subjects that make more than 2 ("> 2 Mistakes") and more than 5 mistakes ("> 5 Mistakes") relative to the optimum, out of the 12 problems they encounter, in choosing an information structure.

Focusing first on "All Data," there are concentrations of subjects at 12 and 0, corresponding to subjects that consistently make confirmation-seeking and contradiction-seeking choices, respectively. There is also a concentration of subjects at 6, which could be driven by optimal choice (recall that optimal behavior requires confirmation-seeking choices in only the six problems of bias by commission) but could also be a result of random decision-making (subjects whose six confirmation-seeking choices are not concentrated in problems of bias by commission as they would be for an optimal decision-maker). In order to focus on the nature of *mistakes* we filter out near-optimal subjects by examining the subset of subjects that make at least 3 and at least 6 mistakes relative to the optimum. When we consider only subjects that make more than two mistakes in their choices over information structures, the mode at 6 shrinks and confirmation-seeking choice becomes the salient mode. When we consider subjects that make more than a handful of mistakes (> 5 mistakes) pure confirmation-seeking and contradiction-seeking behavior become dominant. This exercise suggests that the mode at 6 in the full dataset includes a number of near-optimizing subjects and that confirmation-seeking behavior is a particularly strongly-represented decision rule.

In order to better interpret these results, panel (b) of Figure 4 conducts the same exercise for thousands of simulated subjects, programmed to make iid random choices. The results here are strikingly different: pure confirmation-seeking and contradiction-seeking subjects are completely absent, with confirmation-seeking choices concentrated around 6 and the distribution changing little as we focus on subsets making mistakes.[45]

These aggregate results suggest that choices (and mistakes) in our experiment are far from random but are instead driven by heterogenous subjects using confirmation-seeking decision rules, optimal decision rules and, to a lesser extent, contradiction-seeking choice rules. We report this as a second result:

Result 2. *Mistakes in choices over information structures tend to be skewed towards pure confirmation-seeking or (to a much lesser extent) pure contradiction-seeking behavior.*

[45]Similar patterns emerge in simulations of (i) noisy optimal decision makers that implement the Optimal rule but with random mistakes and (ii) random utility maximizers that are stochastically less likely to make more costly mistakes. In both cases, pure confirmation-seeking and contradiction-seeking choices almost never occur for reasonable parameterizations that match aggregate optimal choice rates. Footnote 47 provides further details.

Type share among classified subjects (%)	Classification method			
	Perfect	$\leq$ 1 error	$\leq$ 2 error	Mixture model
Optimal	28 (*25*)	33 (*25*)	35 (*25*)	37
Confirmation	47 (*25*)	39 (*25*)	35 (*25*)	34
Contradiction	17 (*25*)	17 (*25*)	17 (*25*)	17
Certainty	9 (*25*)	11 (*25*)	13 (*25*)	12
Share classified in data	31 (*0.1*)	52 (*1.2*)	70 (*7.7*)	81

Table 3: Type shares *Notes: Each column presents results from one classification exercise. The values in parentheses show values estimated on a random sample. For example, focusing on the second column, 52% of the subjects choose between information structures in a way that differs from the signature pattern associated with one of the four decision rules described in at most one problem. Among these subjects, 33% are typed as Optimal. The corresponding values on a random sample are 1.2% (share classified) and 25% (typed Optimal among those classified).*

6.2 Types and Heterogeneity

Figure 4 suggests that subjects use well-defined decision rules but that these rules are quite heterogeneous across subjects. While some subjects seem to use near-optimal rules, some others systematically choose advisors biased towards their priors. In order to better understand the internal consistency of subjects' decision rules and study the prevalence of various rules in the subject population, we conduct an exercise to sort individuals into types. In particular we examine the degree to which subjects employ (perhaps with a small number of errors) rules from the taxonomy described in Section 5.

In Table 3, we classify subjects according to the types described in Section 5 based on their choices over information structures.[46]

In the first column, we look at the share of subjects whose choices are perfectly consistent with the four types of behavior we discussed in Section 5. We observe that behavior for 31% of the population fits perfectly one of these categories. As a benchmark, in the last row, we present what

[46]When we conduct a clustering analysis that endogenously determines groups of subjects with similar strategies (identifies exemplars among subjects and forms clusters of behavior around these exemplars), we find that the analysis picks as exemplars precisely the four types described in Section 5 (Optimal, Confirmation, Contradiction and Certainty seeking). For the clustering analysis, we implement affinity propagation in R using the APCluster package (Bodenhofer et al. (2011)). The algorithm takes two arguments: input data and a similarity function. As input data, for each subject, we feed in a vector $v_i \in \{0,1\}^{12}$ which specifies for each type of question (numbered from 1 to 12 to differentiate bias-type, direction of prior, and prior-strength) whether or not the subject chose the optimal information structure. For the similarity function we use the standard squared distance.

these shares would be on a random sample - a large set of simulated subjects who randomly pick between the advisors. In this case, in contrast, the categories considered capture less than 0.1% of the population.[47]

We replicate this analysis allowing for choices to differ from the signature behavior associated with the categories in one choice - second column - or two choices - third column. The share of subjects who are classified goes up to 52% with one error and 71% with two errors, although the associated values on a random sample remain very low.[48,49]

In order to validate this typing we estimate a finite-mixture model whose results are reported in the last column of Table 3. We parameterize the mixture model in the following way. We denote the population share of the different types by $\omega_{op}, \omega_{cf}, \omega_{ct}, \omega_{ce}$, and we allow the total share of these types to be weakly less than 1. The remaining share of the population is assumed to be randomly choosing between the information structures. We also allow for implementation noise denoted by $\kappa \in [0, 0.5)$. That is, each type, in each problem, chooses the information structure associated with his type's signature decision rule with $1 - \kappa$ probability. The advantage of this approach is that we do not need to take an ex-ante position on how flexible we should be with the classification method. This falls out of the estimation as an output - that is, we search for the implementation noise measure that best explains the data. In summary, we estimate $\omega_{op}, \omega_{cf}, \omega_{ct}, \omega_{ce}$ and κ on 344×12 decisions. The last column in Table 3 reports precisely the estimated values for the type shares. The estimated κ is 9.7%, broadly consistent with the 11.2% frequency for sub-optimal choice

[47] We consider two alternative benchmarks: The first is a Noisy Optimal agent – an Optimal type who makes an error in each choice with some probability p ($p = 0.5$ is just the iid random decision maker shown in Table 3). We pick p to match the aggregate optimal information structure choice rate. The second benchmark is a random utility maximizer whose choices are described by the logistic distribution. We pick the noise parameter associated with this distribution to match the aggregate optimal information structure choice rate. Both of these benchmarks generate type distributions that are fundamentally different from those observed in the data. For instance, with the two error classification, only 10% of the data would be classified under the Noisy Optimal benchmark. And among those who are classified, 62% would be classified as Optimal, and 17% as Confirmation and 17% as Contradiction. Similarly, only 11% of the data would be classified with the random utility benchmark; among those classified, 61% would be classified as Optimal, and 21% as Confirmation and 14% as Contradiction.

[48] Test of proportions show that the percentage of people classified as confirmation-seeking is significantly higher ($p < 0.001$) than the percentage classified as contradiction or certainty-seeking, using any of the error classifications.

[49] Although subjects received no feedback between rounds, there is evidence in the data of a small amount of learning from introspection. The high fraction of confirmation-seeking types reported in the data does not vary much with the order in which subjects were exposed to structures biased by commission vs. omission (38% and 32% respectively). However subjects are about 15 percentage points more likely to be classified as optimal types and 12 percentage points less likely to be classified as contradiction-seeking types if they experienced structures biased by omission before those biased by commission. We provide detailed analysis in Online Appendix F.

in the Blackwell questions.[50] Moreover, the estimated κ value (probability of deviating from the prescribed path) is slightly higher than the 6.9% we observe among those subjects who are classified (with ≤ 2 errors), which is consistent with a higher share of the population being classified with the mixture model.[51]

The type-classifying results reported in Table 3 reveal that, however we look at the data, the type distribution within classified subjects tells a clear story. There are as many subjects whose behavior is best explained as confirmation seeking as there are subjects displaying optimal behavior. Looking over the results from the different classification models represented by the different columns in this table, we see significant shifts in the share of the subject population that can be classified. But, among those who are classified, the share corresponding to either optimal or confirmation-seeking behavior is always high, making up 70-75% of the population. There is also some evidence for contradiction- and certainty-seeking behavior, although the fraction of subjects classified in these categories is consistently smaller relative to optimal and confirmation-seeking behavior.

Result 3. *Subjects are as likely to exhibit confirmation-seeking behavior as they are to exhibit optimal behavior, and these two decision rules jointly describe the majority of our subjects. Subjects are half as likely to exhibit contradiction-seeking and even more rarely certainty-seeking behavior.*

7 Mechanism

In this section we consider a number of possible explanations for our results and use features of our design and data to evaluate them, one-by-one.

7.1 Motivated Reasoning and Reputation

Behavior like confirmation-seeking is often attributed to *motivated reasoning*, but we designed the experimental task to remove this as a mechanism in order to explore other possible causes of such behavior. Specifically, the experimental design is deliberately abstract: states are colors of a ball drawn from an urn to remove any reasons for subjects to form an attachment to their prior, and

[50] If we look at this rate – frequency of choosing the sub-optimal advisor chosen in the Blackwell questions – among those subjects who are classified (with ≤ 2 errors), it goes down to 9.1%.

[51] For robustness, we tried several other specifications of this model, including versions with fewer types and a version in which the error rate κ is allowed to vary across types (this has no qualitative effect on our results). The version reported in the paper is the one selected as most preferred by the Bayesian Information Criterion (BIC).

the direction of the prior changed over the course of the experiment (between 0.2 to 0.8) in order to prevent subjects from anchoring on one state or the other.

The experimental results strongly suggest that we successfully eliminated attachments to priors in the design: even subjects who we typed as Confirmation-seeking usually guess against their priors after receiving prior-contradicting signals. Likewise, the design removes reputational concerns by providing subjects with exact signal distributions for each information structure they might choose. There is no indication that subjects doubted the veracity of these probabilities (see Section 7.2 below).

7.2 Confusion

We included a pair of information problems where the information structures could be Blackwell ranked. The high rate of optimal choice in this task (about 90%) assures us that subjects understand how to interpret the experimental interface and instructions, and are sufficiently motivated by the incentives to make considered decisions in the experiment.

7.3 Risk Preferences

The experiment is designed so that failure to choose information structures that maximize guessing accuracy cannot be driven by risk preferences. Given two binary lotteries that produce the same prizes, any subject who prefers the lottery that is more likely to produce the higher payoff prize will prefer the information structures that we call "optimal."

7.4 Errors in Interpreting Signals

A natural explanation for mistakes in choosing between information structures is that subjects make systematic mistakes in interpreting signals and that this has the upstream effect of causing subjects to under-value optimal sources of information. We designed the experiment to evaluate this explanation in several ways. Since we use the strategy method, we know how each subject in each decision round would guess conditional on each of the signals she could receive. This allows us to calculate expected guessing accuracy, α, for each subject for each round. We focus on subjects' *learning* – the improvement in guessing accuracy subjects achieve relative to simply guessing based on their prior beliefs. Figure 5 normalizes across bias-types and priors by plotting the average value

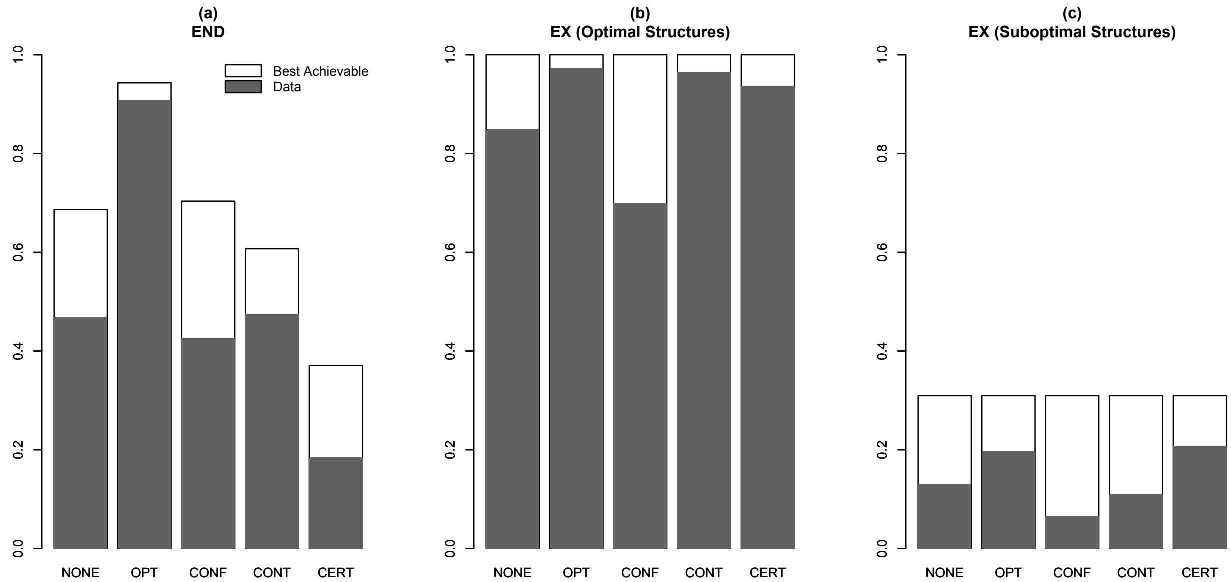


Figure 5: Learning by type

of

$$\frac{\alpha - p_0}{\alpha^{Opt} - p_0}$$

(where α^{Opt} is the expected guessing accuracy conditional on optimality of the information structure and guesses) in gray. This learning measure is maximized at 1 for subjects who both (i) receive signals from optimal information structures (defined on the pair of information structures presented in each round of the END block) and (ii) make optimal guesses for each signal they could receive from this information structure. In white we plot, for reference, the maximum value of this statistic achievable conditional on the information structure from which the subject received signals (averaged across subjects).[52]

Panel (a) of Figure 5 plots data from the END block (the same data plotted in Figure 3), broken down by type.[53] Focusing on the gray bars (actual, observed learning based on guessing behavior), it is clear that Optimal types learn considerably more than subjects using sub-optimal decision rules: Optimal subjects achieve 91% of possible improvement in guessing accuracy (relative to the prior) compared to the 43%-48% for Confirmation/Contradiction types and 19% for Certainty types. Examining the white bars, we find that much (though not all) of this difference is driven by the fact that Optimal types are learning from more informative structures – the white bar is much

[52]For the remainder of the paper, when comparing different types, we will use the ≤ 2 error classification and focus on the problems where the information structures cannot be Blackwell ranked.

[53]See Online Appendix B for regression analysis relating to statistical claims in this section.

higher for Optimal types than for other types. Indeed, the gray bar for Optimal types is greater than the white bars for all other types, meaning that even if other types were perfectly Bayesian they would learn less than Optimal types do! Importantly, however, the difference between the height of gray and the height of white bars in Figure 5 is much smaller for Optimal types than it is for other types (particularly Confirmation and Certainty types), suggesting that Optimal types also do a better job of interpreting and using their information. Despite these differences, values for learning are significantly different from zero in all cases implying that on average subjects are able to make use of signals to improve their guessing accuracy regardless of how they choose between information structures.[54]

Interpreting learning data from the END treatment is complicated by the fact that subjects self-select into the different information structures. In particular, the exercise tells us nothing about what subjects would have done with signals from the information structures they rejected. Panels (b) and (c) show guessing data from the EX block that was designed to overcome exactly this type of concern, by asking every subject to submit guesses for every one of the information structures available in the END block.[55] In each of these EX panels in Figure 5 we classify subjects according to their choices over information structures in the END block, and then examine their guessing behavior in the EX block when they are assigned optimal (panel b) vs. sub-optimal (panel c) information structures from reach round of the END block.[56,57]

The results show a striking pattern. First, *all types* learn substantially more when they are assigned the optimal information structure relative to the case when they are assigned the sub-optimal one (gray bars in panel b are much higher than gray bars in panel c). The Bayesian benchmark in white highlights that most of the decline in learning is due to the change in the information structure, not changes in guessing behavior conditional on the information structure. When assigned an optimal information structure, Contradiction and Certainty types make almost as much use of information as Optimal types but Confirmation types make considerably worse use of information.[58]

[54] The only exception is that confirmation-seeking types do not learn anything significant from suboptimal advisors (though they do learn significantly from optimal advisors).

[55] Recall that the EX task was assigned in only half of our sessions.

[56] Note that white bars mechanically extend to 1 in the former case and much lower in the latter.

[57] Conditioning on the information structure and controlling for the prior, there is generally no statistical difference between how different types behave (in terms of guesses and stated beliefs) between the decision rounds in the END and EX blocks. The few exceptions are: Certainty types learn more in the EX block from suboptimal information structures in the bias-by-omission problems; Contradiction types state less accurate beliefs in the EX block when assigned the optimal information structure in the bias-by-commission problems.

[58] Subjects of all types seem to have much more difficulty making use of information from sub-optimal than optimal

We can use data from the EX block to estimate the "true cost" of not choosing the optimal information structure in the END block, since it allows us to form a counterfactual measure of what learning would be like if subjects *had* selected the optimal information structures. In order to conduct this exercise, we replace each subject's guessing behavior, whenever they chose the sub-optimal information structure, with their guessing behavior for the optimal information structure (taken from their choices in the EX block). In this counterfactual, learning substantially improves for all types, with learning measures rising from 93% to 98% for Optimal, from 41% to 77% for Confirmation types, from 41% to 88% for Contradiction types and from 21% to 94% for Certainty types. This suggests that subjects are not avoiding optimal information structures because they correctly foresee personal difficulties in interpreting optimal structures – subjects of all types would have been substantially better off by choosing optimal information structures and they sacrifice significant guessing accuracy by failing to do so.

Overall, Figure 5 suggests that (i) most of the variation in learning across subjects is driven by variation in information structure (gray bars in panel c are much smaller than those in panel b), but (ii) Confirmation types make less use of information than other types of subjects (gray bars are lower for Confirmation types in panels b and c)[59]. In Online Appendix B, we use the EX block data to show that these patterns also hold separately for bias-by-commission and bias-by-omission problems (that is, the results are universal across problems).

We collect these observations as a further result:

Result 4. *All types learn significantly better from optimal than sub-optimal information structures. Confirmation-seeking types learn less from both optimal and sub-optimal information structures than other types of subjects.*

In addition to collecting subjects' guesses, we also elicited beliefs for each signal from each information structure. Beliefs and guesses were mutually consistent 96% of the time and provide additional information on how subjects interpret signals. In Online Appendix C we conduct a detailed analysis and show that while Optimal types on average form more accurate posteriors than other types (and Confirmation types deviate most strongly from the Bayesian benchmark), variation in these beliefs does very little to explain subjects' information structure choices. Specifically, we calculate the subjective expected value of each information structure for each subject using the

structures (gray bars are much more similar to white bars for optimal structures than for sub-optimal structures). Here, too, Confirmation types are an outlier, making worse use of information from sub-optimal structures than do subjects of other types.

[59] Confirmation types are not significantly different from Contradiction types in the END block and EX block when the suboptimal information structure is assigned.

beliefs they submit in the EX block and show that mistakes in choosing between information structures are not predicted by mistakes in these expected values.[60] The results thus reinforce our conclusions from the guessing data above that information structure choice is not driven by difficulty in interpreting signals.

7.5 Errors in Reasoning

Errors in selecting information structures are not driven in our experiment by confusion about the task, motivated reasoning, reputations, risk aversion or difficulty in interpreting signals. By elimination, we conclude that subjects make mistakes because they have difficulty reasoning about the relative value of information structures. In this section we use other features of our design to provide additional support for this conclusion and to characterize the nature of these reasoning errors.

7.5.1 Relationship with Other Reasoning Tasks

If difficulty in evaluating information structures is driven by difficulty with reasoning through the problem, we might expect suboptimal choices to be correlated with performance difficulties in other reasoning tasks. We administered a multi-part, incentivized cognitive test including the Wason selection task (which tests deductive reasoning and is often associated with confirmation bias), three Raven matrix questions (of varying difficulty, measuring abstract reasoning), three non-standard variations on cognitive reflection test questions (which measures tendencies to override initial responses to questions) and three Belief Bias questions (measuring ability to evaluate logical arguments).[61] We use probit regressions (detailed in Online Appendix D) to estimate how predictive these measures are of the likelihood that subjects employ each of the four decision rules (Optimal, Confirmation etc.). Most importantly, our results suggest high Belief Bias and Wason scores to

[60] That is, subjects are not more likely to choose the optimal information structure when their beliefs suggest they put higher value on these information structures.

[61] Thomson & Oppenheimer (2016) provides detailed discussion of these cognitive ability measures; Cosmides (1989) studies the Wason selection task.

be predictive of being typed as Optimal.[62,63,64] This suggests that subjects are less likely to use sub-optimal decision rules the stronger their cognitive abilities are.[65]

Result 5. *Subjects with high measured cognitive abilities are less likely to employ sub-optimal decision rules.*

7.5.2 Awareness and Intentionality in Advice Data

Subjects do not simply make mistakes in selecting information structures – they implement highly-consistent heuristic rules. How conscious are subjects of these rules and do they implement them because they believe they are payoff-enhancing choices? In order to get some insight on these questions, we included an advice task at the end of the experiment, asking subjects to provide free-form advice (including explanation/justification) to a prospective participant in the experiment.

Subjects provided strikingly detailed (and often sophisticated) descriptions of how they made their choices (about 70% of subjects gave us clear, operationalizable descriptions of how they made their choices). These descriptions reveal that most subjects choose information structures by (i) explicitly identifying the direction of bias of each information structure and (ii) choosing the information structure whose bias either matches (confirmation-seeking) or anti-matches (contradiction-seeking) the *ex-ante* more likely state (i.e. the prior provided for the task). Careful reading of the free-form advice text also suggests that subjects find heuristics based on "bias matching" to

[62] The statistical analysis presented in Online Appendix D adjusts for multiple hypothesis testing using the conservative Bonferroni correction. There is also some indication that having an analytical major of study is predictive of being a Contradiction-seeking type and against being a Confirmation type.

[63] We included two measures of self-perception. The first looks at self-confidence which combines answers to two questions on how well subjects think they have done in the cognitive ability questions: one is a simple guess on the number of questions they answered correctly, and the second is an assessment of how well they have done relative to others. We also have a second measure on how closely the subject identifies with the following statement: "It is very important to me to hold strong opinions." These measures are not predictive of a subjects' type.

[64] Ranking subjects in terms of their overall scores for the cognitive ability questions, we observe the share of Optimal and Confirmation types in the top (bottom) quartile to be 52% (13%) and 14% (30%) respectively. Furthermore, among those subjects who solved the Wason task correctly 51% were Optimal and 12% were Confirmation. These shares change to 21% and 27% for the remainder of the data.

[65] Our survey also included several questions about political attitudes and media habits. However, we had too little variation in these measures relative to our sample size to conduct credible hypothesis tests with this data. Combining all individual characteristics measured with the post-experimental survey, we also applied LASSO to select a subset of regressors and re-estimated our Probit regressions (again correcting for multiple hypothesis testing). Details are provided in Online Appendix D.

be normatively appealing and rational, with many subjects making attractive arguments for the probabilistic sophistication of their approaches (in a few cases subjects even explicitly claim that confirmation-seeking is Bayesian!).

Importantly, subjects' advice strongly matches their behavior in the experiment. In Online Appendix E, we report results from blind coding[66] of advice as CONF-consistent, CONT-consistent, OPT-consistent, CERT-consistent or NONE and show that, for most types, (i) this type-consistency strongly predicts the typing based on behavior reported in Section 6.2 (e.g. CONF-consistency strongly increases the odds of being a Confirmation type) but does not predict other types (e.g. CONF-consistency in fact decreases the odds of being an Optimal or Contradiction type).[67] As a robustness check, after the advice question we asked subjects to classify their information structure decisions by choosing one of the following from a multiple choice list: (1) "I mostly considered whether there were more orange or green balls in the basket and chose the advisor that gave this color advice most often."; (2) "I mostly considered whether there were more orange or green balls in the basket and chose the advisor that gave the opposite color advice most often."; (3) "Neither is a good description of how I chose an advisor."[68] These choices, too, are highly predictive of subject types.

In summary, these results confirm that subjects are aware of the decision rules they are using and, moreover, are using them because they believe they are payoff enhancing. We report this as a further result.

Result 6. *A substantial fraction of subjects intentionally use sub-optimal decision rules like confirmation-seeking and find these mistaken rules to be normatively appealing.*

7.5.3 Descriptions of Reasoning Process in Advice Data

What reasoning errors make heuristics like confirmation- and contradiction-seeking appealing to people? We get some important clues from the justifications provided in the advice data. In

[66]Two authors independently coded the advice data (blind to any further information on subject behavior) and this was later aggregated.

[67]Similarly, OPT-consistency increases the odds of being an Optimal type while decreasing the odds of being a Confirmation type.

[68]This robustness check is useful because the advice task features a "beauty contest" structure that may conceivably push subjects to craft advice that they believe will be popular with other subjects. The multiple choice list cannot suffer from this problem. The similarity of the results across the two types of strategy elicitation suggests that subjects are likely honestly reporting strategies in both cases.

particular, this data suggests that both confirmation and contradiction-seeking types incompletely reason through two important insights needed to identify optimal information sources.

Confirmation types incompletely reason about the importance of selecting information structures that are more accurate in the *ex ante* most likely state.[69] Echoing intuition provided in Section 3.2, Confirmation types argue that it is more important to achieve higher guessing accuracy in the *ex ante* most likely state relative to guessing accuracy in the *ex ante* less likely state. But, the advice data suggests that they make a crucial mistake by confusing the direction of this trade-off with the direction of the bias of the information sources. For instance a typical argument made (subject 10 in session 1) is: "When choosing advisors, select the one that tells you more about [the more-likely state],.. you want to go with the one which has higher accuracy for [this state], since the increased accuracy multiplies with the increased possibility of drawing [this state]."[70] Thus, Confirmation types tend to reason that they should focus on the guessing accuracy in the most-likely state, but do not seem to grasp how this depends on guesses conditional on *all* signals. Their advice conflates *guessing accuracy* with *signal accuracy* (the likelihood of receiving signal r conditional on state R, or l conditional on L). This naturally leads to mistakes in the omission problems where the n signal breaks the equivalence between the two.[71]

Contradiction types incompletely reason about the importance of seeking signals that are instrumentally useful. They correctly reason that signals only improve guessing accuracy (relative to receiving no information) if they cause them to make guesses that go against their prior. But (in stark contrast with Confirmation types) do not grasp the importance of seeking high guessing accuracy in the more-likely state. This advice (from subject 15 in session 4) is typical: "Choose the advisor who will MOST LIKELY (high percentage) give you the right answer for [the less-likely color] BECAUSE [this] is unlikely to be the chosen color so you want the advisor to tell if the color is ever chosen (in other words, you want to create a situation where if the unlikely color is

[69]This type of reasoning suggests a connection with other types of non-Bayesian behavior. For example, people who suffer from base rate neglect in updating their beliefs often conflate the accuracy of the signal (conditional on the state) with the posterior (about the state) conditional on the signal, thus ignoring the prior.

[70]Here are two more examples demonstrating the same argument: (1) (subject 20 in session 11) "The advice I have is to select that advisor that has the highest accuracy for [the more-likely state]. My reasoning is that there is a higher chance of answering correctly if the advisor is most accurate in [this state]"; (2) (subject 8 in session 1) "Choose an advisor based on the one that can offer you the most certain odds of being correct. If one advisor is always or nearly always correct about one color, especially if that color is most frequent, then that advisor is the more reliable choice. Since in this experiment the two advisors were always reflections of one another, it is best to choose the one more likely to be right about the color in the majority."

[71]Although we should note that when confronted with the n signal, guessing patterns do not suggest Confirmation types ignore the informational content of this signal. (See Online Appendix B for evidence.)

chosen, the advisor will tell you so / / - for the color with the most balls, it already has a high chance of being chosen so luck is on your side with that color."[72] Thus, Contradiction types tend to emphasize that signals generate value by inducing correct guesses on the *ex ante* less-likely state, but incorrectly infer from this that it is valuable to maximize guessing accuracy in the less-likely state.

In each of these cases, subjects' advice suggests that they reason only part of the way through the valuation problem and thereby systematically miss important insights, leading to internally consistent patterns of mistakes.[73] In contrast to these types, Optimal types often explicitly (and correctly) discuss the trade-off between guessing accuracy in the *ex ante* more and less likely states. A typical example is (subject 21 in session 1): "If there are fewer green balls than orange balls, choose the adviser that can give you a response so that you know for certain the ball is green... The fewer of one color there is, the less ambiguous you want the advice for that color to be. The more of a color there is, the more ambiguous it should be, as there is a greater chance of it being that color ball."[74] (We give more examples of advice given by different types in Online Appendix E).

In summary, although subjects often display statistical sophistication in describing their reasoning process, this reasoning is often incomplete. It is this incomplete reasoning that seems to lie behind the highly internally consistent mistake-making the predominates in our data. Although our task is perhaps the simplest possible information selection setting (e.g. binary states, binary

[72]Here are two more examples demonstrating the same argument: (1) (subject 18 in session 3) "If there are a significant amount more green balls to orange balls in the basket, the participant should choose the advisor that tells the truth more times when the ball is orange. Because there is a greater chance of the ball selected being green because there are more of them, I believe that it is more helpful to know the true color of the ball that is less likely to be selected, in this case the orange ball."; (2) (subject 6 in session 6) "If there are a lot of green, choose the advisor who is more likely to say orange when the ball is orange. This is because the ball is more likely to be the majority color and you would rather the advisor be more correct when the lesser probability color is chosen."

[73]We have more limited data on Certainty types, but some of these subjects do mention signals that reveal the state perfectly to be particularly useful. For instance, subject 20 in session 1 writes " Decide for each adviser if there is information that you can know with 100% certainty"; subject 2 in session 3 writes "You want to select the advisor who is going to provide you with a CERTAIN answer most often"; and subject 18 in session 2 writes "Always look for the guarantees."

[74]Two more examples from Optimal types: (1) (subject 1 in session 9) "The most helpful advice is the one that tells will tell you the color of the ball that has the less likely chance of getting picked... to be clear, this is not the advisor that will say orange [the minority color] most of the time. this is whichever advisor will only say orange when the ball is actually orange"; (2) (subject 2 in session 7) "1. Look at the advisors / 2. Pick the one that is gonna say the ball with the lower probability of being picked most accurately. / / If there are a lot of orange balls and not very many green pick the advisor you can trust most if the green gets picked because if he gives an answer you are unsure of then it is probably orange (or the one with the largest amount of balls in the question.)"

structures), properly valuing information nonetheless requires a sophisticated type of reasoning that does not come naturally to most people.

8 Conclusion

Our experiment provides experimental evidence on how people choose between biased information sources and the motivations behind those choices. Subjects make frequent (and costly) mistakes in choosing between biased information structures, but these mistakes are not random. Indeed, patterns in choices over information structures are distinctive enough that we can categorize most subjects as implementing one of a handful of decision rules. The most common of these is a confirmation-seeking rule that arises in the subject population about as frequently as the optimal rule. Contradiction-seeking and certainty-seeking rules are less common..

Diagnostic treatments and other features of the design allow us to rule out motivated reasoning, reputational concerns, confusion about the task, risk aversion and difficulty in interpreting signals as explanatory factors for the observed patterns of behavior. Cognitive tasks and elicited advice suggest instead that these errors arise from difficulties in reasoning correctly about the relative valuation of information structures. Subjects incompletely reason their way to appealing but mistaken conclusions about how to judge the relative usefulness of information structures leading them to persistently select information structures biased towards or against their priors.

Our results point to an additional cause of failures to select optimal sources of information, joining previously identified causes like motivated reasoning and reputational concerns. A particularly important implication of this new mechanism is that it suggests that patterns of behavior like confirmation-seeking may be much more widespread than motivated reasoning and reputational accounts, alone, would suggest. To the degree that these patterns arise from errors in reasoning we might expect them also to arise in less affectively-loaded domains such as choices between competing product reviews, financial advisors, academic disciplines and sources of medical advice.

Our experiment was partly motivated by popular concern that biases in selecting sources of political information (particularly confirmation-seeking biases) are responsible for the proliferation of echo chambers, information bubbles and ultimately increased political polarization in recent years. Our finding that confirmation-seeking behavior is, in part, rooted in a simple error in *reasoning*, may have implications for policy responses to these type of phenomena. On the negative side, to the degree that behaviors like confirmation-seeking are driven by reasoning errors, policies aimed at improving trust in news sources (for instance calls for the development of fact-checking websites) or at reversing motivated-reasoning (for instance by reframing policies in conciliatory ways) might not completely remove the negative social effects of these errors. On the positive side, to the degree that cognitive mistakes drive these behaviors, there may be scope for interventions that improve people's learning and reasoning habits – policies that would be ineffective if confirmation-seeking were driven purely by reputational and motivated-reasoning mechanisms.

Moreover, our results show a surprising de-linkage between people's ability to select optimal information sources and their ability to make use of optimal information sources: our subjects are capable of making effective use of optimal information sources even when they are unable to select optimal information sources in the first place. This may suggest that policies designed to expose people to information that they would not voluntarily seek out themselves might be particularly effective in encouraging the formation of accurate beliefs.[75]

There are several promising directions for future research that builds on our design in order to advance this research agenda. First, it is an important next step to study whether (and under what circumstances) cognitive training, provision of information, or learning through natural feedback can eliminate the types of errors identified in this paper.[76] Such studies could also provide insights on how the failure to identify optimal information structures is connected to other forms of non-Bayesian behavior. Second, it would be valuable to build on this design to study the degree to which the cognitive channel our results points towards plays a role in settings where beliefs can also be motivated or where there are reputational concerns. While the relative importance of these different forces should be expected to be context dependent, such findings would be important in designing policy responses. Third, as a counterpart to our abstract design, it would be valuable to test whether the use of behavioral decision rules (such as confirmation and contradiction seeking) are more or less prevalent in natural framings of the problem.[77] While our experiment demonstrates the appeal of these heuristics to subjects even in our deliberately abstract design, it is quite possible that underlying elements of the problem that do not influence value (such as framing and context)

[75] In a similar spirit, Sunstein has argued that social media platforms should adopt an "architecture of serendipity", creating chance encounters with people and ideas with which we might not choose to engage.

[76] We note that learning about optimality of information structures from natural feedback can be particularly difficult as the feedback agents receive is endogenous to their choices.

[77] Information structures are often much more complex in natural contexts making it arguably harder to make optimal choices. In particular, bias of an information structure can be easier to identify, potentially making it easier for agents to follow bias-matching heuristics. On the other hand, work in psychology (especially in relation to the Wason task) suggest that humans can reason through hypothesis testing better in natural contexts. See Cosmides (1989).

trigger different modes of reasoning influencing the use of such heuristics. For this reason, research examining the robustness and stability of these results across different environments is an important direction for future work.

References

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V. & Zhuravskaya, E. (2015), 'Radio and the rise of the nazis in prewar germany', *The Quarterly Journal of Economics* **130**(4), 1885–1939.

Agranov, M., Dasgupta, U. & Schotter, A. (2018), 'Preferences for the resolution of uncertainty and the timing of information'.

Ambuehl, S. (2017), 'An offer you can't refuse: Incentives change how we inform ourselves and what we believe', *Working Paper* .

Ambuehl, S. & Li, S. (2018), 'Belief updating and the demand for information', *Games and Economic Behavior* **109**, 21–39.

Andreoni, J. & Mylovanov, T. (2012), 'Diverging opinions', *American Economic Journal: Microeconomics* **4**(1), 209–32.

Baron, J. (2000), *Thinking and deciding*, Cambridge University Press.

Baron, J., Beattie, J. & Hershey, J. C. (1988), 'Heuristics and biases in diagnostic reasoning: Ii. congruence, information, and certainty', *Organizational Behavior and Human Decision Processes* **42**(1), 88–110.

Bénabou, R. (2015), 'The economics of motivated beliefs', *Revue d'économie politique* **125**(5), 665–685.

Bénabou, R. & Tirole, J. (2016), 'Mindful economics: The production, consumption, and value of beliefs', *Journal of Economic Perspectives* **30**(3), 141–64.

Benjamin, D. J., Rabin, M. & Raymond, C. (2016), 'A model of nonbelief in the law of large numbers', *Journal of the European Economic Association* **14**(2), 515–544.

Bodenhofer, U., Kothmeier, A. & Hochreiter, S. (2011), 'Apcluster: an r package for affinity propagation clustering', *Bioinformatics* **27**(17), 2463–2464.

Brandts, J. & Charness, G. (2011), 'The strategy versus the direct-response method: a first survey of experimental comparisons', *Experimental Economics* **14**(3), 375–398.

Brunnermeier, M. K. & Parker, J. A. (2005), 'Optimal expectations', *American Economic Review* **95**(4), 1092–1118.

Burks, S. V., Carpenter, J. P., Goette, L. & Rustichini, A. (2013), 'Overconfidence and social signalling', *Review of Economic Studies* **80**(3), 949–983.

Camerer, C. (1998), 'Bounded rationality in individual decision making', *Experimental economics* **1**(2), 163–183.

Caplin, A. & Leahy, J. (2001), 'Psychological expected utility theory and anticipatory feelings', *The Quarterly Journal of Economics* **116**(1), 55–79.

Charness, G. & Dave, C. (2017), 'Confirmation bias with motivated beliefs', *Games and Economic Behavior* **104**, 1–23.

Charness, G. & Levin, D. (2005), 'When optimal choices feel wrong: A laboratory study of bayesian updating, complexity, and affect', *American Economic Review* **95**(4), 1300–1309.

Cipriani, M. & Guarino, A. (2010), 'Herd behavior in financial markets: An experiment with financial market professionals', *Journal of the European Economic Association* **7**(1), 206–233.

Cosmides, L. (1989), 'The logic of social exchange: Has natural selection shaped how humans reason? studies with the wason selection task', *Cognition* **31**(3), 187–276.

Crawford, V. P. & Sobel, J. (1982), 'Strategic information transmission', *Econometrica: Journal of the Econometric Society* pp. 1431–1451.

DellaVigna, S. & Hermle, J. (2017), 'Does conflict of interest lead to biased coverage? evidence from movie reviews', *Review of Economic Studies* **84**(4), 1510–1550.

DellaVigna, S. & Kaplan, E. (2007), 'The fox news effect: Media bias and voting', *The Quarterly Journal of Economics* **122**(3), 1187–1234.

Dillenberger, D. & Segal, U. (2017), 'Skewed noise', *Journal of Economic Theory* **169**, 344–364.

Duffy, J., Hopkins, E. & Kornienko, T. (2017), 'Lone wolf or herd animal? information choice and social learning'.

Duffy, J., Hopkins, E., Kornienko, T. & Ma, M. (2018), 'Information choice in a social learning experiment'.

Durante, R., Pinotti, P. & Tesei, A. (2017), 'The political legacy of entertainment tv'.

Eil, D. & Rao, J. M. (2011), 'The good news-bad news effect: asymmetric processing of objective information about yourself', *American Economic Journal: Microeconomics* **3**(2), 114–38.

Eliaz, K. & Schotter, A. (2010), 'Paying for confidence: An experimental study of the demand for non-instrumental information', *Games and Economic Behavior* **70**(2), 304–324.

Enke, B. (2017), 'What you see is all there is', *Working paper* .

Enke, B. & Zimmermann, F. (2017), 'Correlation neglect in belief formation', *Review of Economic Studies* .

Eyster, E. & Rabin, M. (2005), 'Cursed equilibrium', *Econometrica* **73**(5), 1623–1672.

Eyster, E., Rabin, M. & Weizsäcker, G. (2018), 'An experiment on social mislearning', *Working paper* .

Falk, A. & Zimmermann, F. (2017), 'Beliefs and utility: Experimental evidence on preferences for information'.

Fréchette, G. R., Lizzeri, A. & Perego, J. (2018), 'Rules and commitment in communication', *Working Paper* .

Gabaix, X. & Laibson, D. (2006), 'Shrouded attributes, consumer myopia, and information suppression in competitive markets', *The Quarterly Journal of Economics* **121**(2), 505–540.

Gentzkow, M. & Shapiro, J. M. (2010), 'What drives media slant? evidence from us daily newspapers', *Econometrica* **78**(1), 35–71.

Gentzkow, M. & Shapiro, J. M. (2011), 'Ideological segregation online and offline', *The Quarterly Journal of Economics* **126**(4), 1799–1839.

Gentzkow, M., Shapiro, J. M. & Stone, D. F. (2015), Media bias in the marketplace: Theory, *in* 'Handbook of media economics', Vol. 1, Elsevier, pp. 623–645.

Grant, S., Kajii, A. & Polak, B. (1998), 'Intrinsic preference for information', *Journal of Economic Theory* **83**(2), 233–259.

Heidhues, P., Kőszegi, B. & Murooka, T. (2016), 'Inferior products and profitable deception', *The Review of Economic Studies* **84**(1), 323–356.

Hossain, T. & Okui, R. (2013), 'The binarized scoring rule', *Review of Economic Studies* **80**(3), 984–1001.

Iyengar, S. & Hahn, K. S. (2009), 'Red media, blue media: Evidence of ideological selectivity in media use', *Journal of Communication* **59**(1), 19–39.

Jin, G. Z., Luca, M. & Martin, D. (2015), Is no news (perceived as) bad news? an experimental investigation of information disclosure, Technical report, National Bureau of Economic Research.

Jo, D. (2017), 'Better the devil you know: An online field experiment on news consumption'.

Jones, M. & Sugden, R. (2001), 'Positive confirmation bias in the acquisition of information', *Theory and Decision* **50**(1), 59–99.

Kamenica, E. & Gentzkow, M. (2011), 'Bayesian persuasion', *American Economic Review* **101**(6), 2590–2615.

Klayman, J. (1995), 'Varieties of confirmation bias', *Psychology of learning and motivation* **32**, 385–418.

Klayman, J. & Ha, Y.-W. (1987), 'Confirmation, disconfirmation, and information in hypothesis testing.', *Psychological review* **94**(2), 211.

Kőszegi, B. & Rabin, M. (2009), 'Reference-dependent consumption plans', *American Economic Review* **99**(3), 909–36.

Kreps, D. M. & Porteus, E. L. (1978), 'Temporal resolution of uncertainty and dynamic choice theory', *Econometrica: journal of the Econometric Society* pp. 185–200.

Loewenstein, G., Sunstein, C. R. & Golman, R. (2014), 'Disclosure: Psychology changes everything'.

Martin, G. J. & Yurukoglu, A. (2017), 'Bias in cable news: Persuasion and polarization', *American Economic Review* **107**(9), 2565–99.

Martínez-Marquina, A., Niederle, M. & Vespa, E. (2018), 'Probabilistic states versus multiple certainties: The obstacle of uncertainty in contingent reasoning', *Working paper* .

Masatlioglu, Y., Orhun, A. Y. & Raymond, C. (2017), 'Intrinsic information preferences and skewness', *Working paper* .

Milgrom, P. & Roberts, J. (1986), 'Price and advertising signals of product quality', *Journal of Political Economy* **94**(4), 796–821.

Mobius, M. M., Niederle, M., Niehaus, P. & Rosenblat, T. S. (2011), Managing self-confidence: Theory and experimental evidence, Technical report, National Bureau of Economic Research.

Montanari, G. & Nunnari, S. (2019), 'Audi alteram partem: An experiment on selective exposure to information', *Working Paper* .

Mullainathan, S., Schwartzstein, J. & Shleifer, A. (2008), 'Coarse thinking and persuasion', *The Quarterly Journal of Economics* **123**(2), 577–619.

Ngangoue, K. & Weizsacker, G. (2018), 'Learning from unrealized versus realized prices'.

Nickerson, R. S. (1998), 'Confirmation bias: A ubiquitous phenomenon in many guises', *Review of general psychology* **2**(2), 175–220.

Nielsen, K. (2018), 'Preferences for the resolution of uncertainty and the timing of information'.

Oster, E., Shoulson, I. & Dorsey, E. (2013), 'Optimal expectations and limited medical testing: evidence from huntington disease', *American Economic Review* **103**(2), 804–30.

Pariser, E. (2011), *The filter bubble: How the new personalized web is changing what we read and how we think*, Penguin.

Prior, M. (2007), *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*, Cambridge University Press.

Rabin, M. & Schrag, J. L. (1999), 'First impressions matter: A model of confirmatory bias', *The Quarterly Journal of Economics* **114**(1), 37–82.

Sicherman, N., Loewenstein, G., Seppi, D. J. & Utkus, S. P. (2015), 'Financial attention', *The Review of Financial Studies* **29**(4), 863–897.

Skov, R. B. & Sherman, S. J. (1986), 'Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation', *Journal of Experimental Social Psychology* **22**(2), 93–121.

Slowiaczek, L. M., Klayman, J., Sherman, S. J. & Skov, R. B. (1992), 'Information selection and use in hypothesis testing: What is a good question, and what is a good answer?', *Memory & Cognition* **20**(4), 392–405.

Sunstein, C. R. (2018), *# Republic: Divided democracy in the age of social media*, Princeton University Press.

Thomson, K. S. & Oppenheimer, D. M. (2016), 'Investigating an alternate form of the cognitive reflection test', *Judgment and Decision Making* **11**(1), 99.

Toussaert, S. (2017), 'Intention-based reciprocity and signaling of intentions', *Journal of Economic Behavior and Organization* **137**, 132–144.

Wason, P. C. (1960), 'On the failure to eliminate hypotheses in a conceptual task', *Quarterly journal of experimental psychology* **12**(3), 129–140.

Weizsäcker, G. (2010), 'Do we follow others when we should? a simple test of rational expectations', *American Economic Review* **100**(5), 2340–60.

Wilson, A. & Vespa, E. (2018), 'Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language', *Working paper* .

Zimmermann, F. (2014), 'Clumped or piecewise? evidence on preferences for information', *Management Science* **61**(4), 740–753.

Zimmermann, F. (2018), The dynamics of motivated beliefs, Technical report, Working Paper.

Online Appendix for

How Do People Choose Between Biased Information Sources? Evidence from a Laboratory Experiment

Gary Charness   Ryan Oprea   Sevgi Yuksel

CONTENTS:

A Further analysis on design

Remark 3. *The two information structures in Table 1 are ranked in terms of the Monotone likelihood ratio property.*

Proof. Let π_s^1 (π_s^2) denote the probability of observing signal s from the information structure on the left (right) hand side of Table 1. Clearly, as cross-multiplication shows:

$$\frac{\pi_r^2}{\pi_r^1} = \frac{(1-p_0)(1-\lambda)+p_0}{p_0(1-\lambda)} > \frac{(1-p_0)\lambda}{(1-p_0)+p_0(1-\lambda)} = \frac{\pi_l^2}{\pi_l^1}$$

□

Remark 4. *The two information structures in Table 2 are ranked in terms of the Monotone likelihood ratio property for $(\lambda_h, \lambda_l) = (0.7, 0.3)$.*

Proof. Let π_s^1 (π_s^2) denote the probability of observing signal s from the information structure on the left (right) hand side of Table 2. We would like to show:

$$\frac{\pi_r^2}{\pi_r^1} = \frac{\lambda_h}{\lambda_l} > \frac{\pi_n^2}{\pi_n^1} > \frac{\lambda_l}{\lambda_h} = \frac{\pi_l^2}{\pi_l^1}$$

$$\frac{\pi_n^2}{\pi_n^1} = \frac{(1-p_0)(1-\lambda_l)+p_0(1-\lambda_h)}{(1-p_0)(1-\lambda_h)+p_0(1-\lambda_l)} = \frac{1-\lambda_l+(\lambda_l-\lambda_h)p_0}{1-\lambda_h+(\lambda_h-\lambda_l)p_0}$$

The first inequality can be written as $\lambda_h - \lambda_h^2 + (\lambda_h^2 - \lambda_h\lambda_l)p_0 > \lambda_l - \lambda_l^2 + (\lambda_l^2 - \lambda_h\lambda_l)p_0$ which is equivalent to $\lambda_h - \lambda_l > (\lambda_l^2 - \lambda_h^2)(1-p_0)$. Similarly, the second inequality can be written as $\lambda_h - \lambda_l > (\lambda_l^2 - \lambda_h^2)p_0$. Note $\lambda_h - \lambda_l > (\lambda_l^2 - \lambda_h^2)$ when $\lambda_h, \lambda_l = (0.7, 0.3)$ which is sufficient for the result. □

B Further analysis on guesses

Data from the EX block also allows us to see whether use of suboptimal decision rules is associated with problems in making use of specific kind of signals. Specifically, we can see whether Confirmation types who consistently choose suboptimal information structures only in the bias-by-omission problems have difficulty making guesses conditional on the n signal. We do not find such a link: guesses are optimal 82% of the time conditional on this signal for Confirmation types, which is not statistically different from the 87% for Optimal types or the 76% for Contradiction types. Moreover, we see that Confirmation types do not shy away from guessing against their prior in response to this signal: this rate is 55% for Confirmation types, 44% for Optimal types, and 73% for Contradiction types.[78] In conclusion, behavior in the EX block does not suggest use of the Confirmation-seeking decision rule to be driven by problems with internalizing the informational content of failing to receive signals.

	END	EX-OPT	EX-SUBOPT
Confirmation	0.242***	0.275***	0.132**
	(0.0335)	(0.0918)	(0.0628)
Contradiction	0.0971***	0.00777	0.0870***
	(0.0255)	(0.0380)	(0.0298)
Certainty	0.151***	0.0366	-0.0109
	(0.0249)	(0.0564)	(0.0224)
None	0.183***	0.124*	0.0660**
	(0.0383)	(0.0627)	(0.0301)
Constant	0.0345***	0.0265	0.112***
	(0.00892)	(0.0183)	(0.0131)
Observations	4128	948	948

Standard errors (clustered at the subject level) in parentheses.

Statistical significance with respect to omitted category (Optimal).

***1%, **5%, *10% significance.

END: Endogenous Block;

EX-OPT: Exogenous Block with Optimal Info. Structure;

EX-SUBOPT: Exogenous Block with Subptimal Info. Structure.

Table 4: OLS Regression on Difference between Best Achievable and Actual Learning

[78] Confirmation types are not statistically different in this respect from others, but Contradiction types are different from Optimal types ($p < 0.05$).

	END	EX-OPT	EX-SUBOPT
Optimal	0.481***	0.275***	0.132**
	(0.0346)	(0.0918)	(0.0628)
Contradiction	0.0484	0.267***	0.0450
	(0.0423)	(0.0959)	(0.0669)
Certainty	-0.242***	0.238**	0.143**
	(0.0411)	(0.105)	(0.0640)
None	0.0417	0.151	0.0660
	(0.0513)	(0.108)	(0.0671)
Constant	0.427***	0.699***	0.0651
	(0.0326)	(0.0899)	(0.0614)
Observations	4128	948	948

Standard errors (clustered at the subject level) in parentheses.

Statistical significance with respect to omitted category (Confirmation).

***1%, **5%, *10% significance.

END: Endogenous Block;

EX-OPT: Exogenous Block with Optimal Info. Structure;

EX-SUBOPT: Exogenous Block with Subptimal Info. Structure.

Table 5: OLS Regression on Learning

	END	EX-OPT	EX-SUBOPT
Optimal	0.909***	0.974***	0.197***
	(0.0116)	(0.0183)	(0.0131)
Confirmation	0.427***	0.699***	0.0651
	(0.0326)	(0.0899)	(0.0614)
Contradiction	0.476***	0.966***	0.110***
	(0.0269)	(0.0333)	(0.0268)
Certainty	0.185***	0.937***	0.208***
	(0.0250)	(0.0534)	(0.0182)
None	0.469***	0.850***	0.131***
	(0.0395)	(0.0600)	(0.0272)
Observations	4128	948	948

Standard errors (clustered at the subject level) in parentheses.

Statistical significance indicated learning is different from zero.

***1%, **5%, *10% significance.

END: Endogenous Block;

EX-OPT: Exogenous Block with Optimal Info. Structure;

EX-SUBOPT: Exogenous Block with Subptimal Info. Structure.

Table 6: OLS Regression on Learning

	OPT	CONF	CONT	CERT	None
Optimal Information Structure	0.776***	0.634***	0.856***	0.729***	0.719***
	(0.0216)	(0.0542)	(0.0298)	(0.0482)	(0.0500)
Constant	0.197***	0.0651	0.110***	0.208***	0.131***
	(0.0132)	(0.0619)	(0.0276)	(0.0187)	(0.0273)
Observations	432	468	192	228	576

Standard errors (clustered at the subject level) in parentheses.

***1%, **5%, *10% significance.

OPT: Optimal; CONF: Confirmation; CONT: Contradiction; CERT: Certainty

Table 7: OLS Regression on Learning (Exogenous Block)

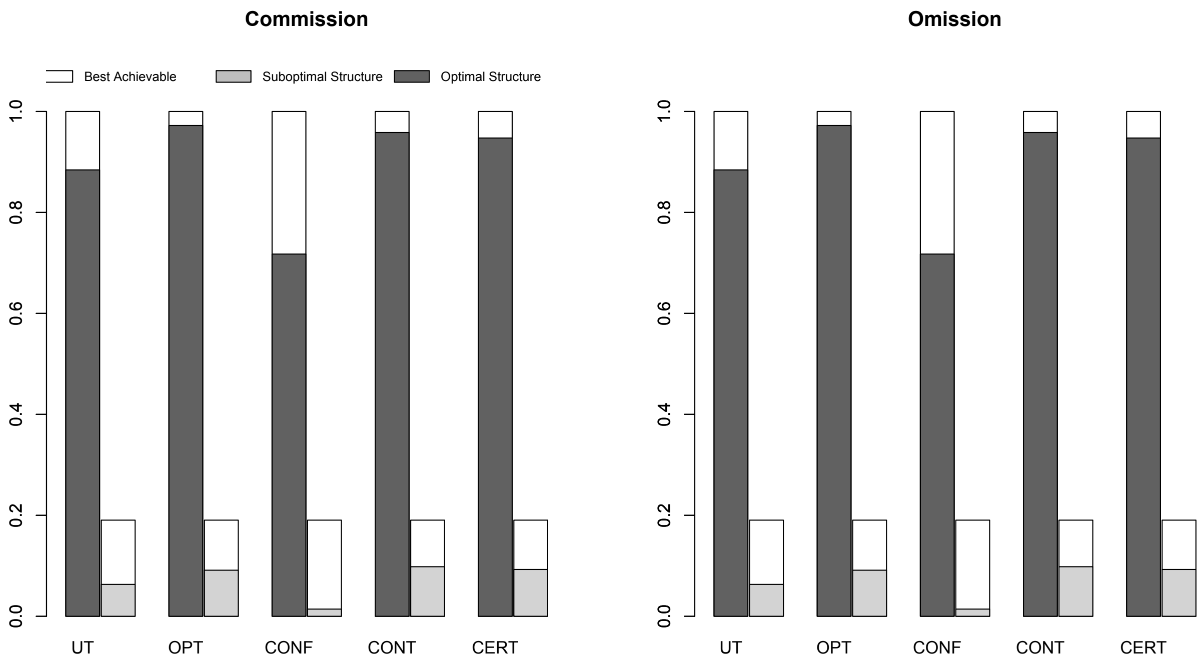


Figure 6: Learning by type separated by bias-type

C Further analysis on beliefs

In addition to guessing the state, subjects were incentivized to submit beliefs about the likelihoods of the state (in both END and EX blocks).[79] Panel (a) of Figure 7 shows the average (absolute) difference between these submitted beliefs and the beliefs a Bayesian would form upon receiving signals from the same information structure (for this we use data from the EX task where we have, for each subject, elicitation for every information structure).[80,81] As with the learning/guessing accuracy results from the previous subsection, Optimal types make better use of information than the other types and confirmation-seeking types stand out as forming the worst beliefs (those with the greatest deviation from Bayesian benchmarks).[82]

There is a relationship between mistakes in choices over information structures and mistakes in beliefs, but do the latter *cause* the former? Do subjects choose sub-optimal information structures *because* they mistakenly believe these structures will provide more useful signals? If so, we would expect variation in the accuracy of beliefs across subjects and decisions to predict when subjects make mistakes in their choices over information structures. To examine this, we can calculate the *expected value* of each information structure σ implied by the beliefs subject i submits for this structure in the EX block. Formally,

$$V_i(\sigma) = \sum_s \pi_s \max\{p_s, 1 - p_s\}$$

where π_s is the probability of receiving signal s and p_s is the stated posterior of the subject conditional on that signal.[83,84] Suppose among two information structures $\overline{\sigma}$ and $\underline{\sigma}$ the former is

[79] A useful "sanity check" on the elicited belief data is to examine whether these data are consistent with subjects' guessing data (that is, whether subjects tend to guess the state their elicited beliefs indicate to be more likely). In fact, subjects' guessing and belief data is consistent 96% of the time and this consistency hardly varies across types.

[80] Formally, the Figure plots $\sum_s \pi_s |p_s - p_s^{Bay}|$ where p_s^{Bay} is the Bayesian posterior and p_s is the stated posterior of the subject conditional on signal s and π_s is the probability of receiving signal s.

[81] Results are broadly similar for the END treatment but as we discuss above, self-selection into information structure makes these beliefs more difficult to interpret.

[82] Both Optimal and Confirmation types are statistically different from others ($p < 0.05$), and the difference between these two types is highly significant ($p < 0.001$). Nonetheless, there is substantial variation among all types. For example, focusing on the top quartile of the data (in terms of accuracy of beliefs), we see that among those only 34% of subjects are Optimal types. (The ratio goes up to 50% among classified types).

[83] The value is equivalent to the subject's expected guessing accuracy in that problem when receiving signals from this information structure. When subjects' stated beliefs coincide with the Bayesian posteriors, this value is equivalent to the Bayesian value.

[84] Recall that subjects directly observe the prior and the signal distribution conditional on each state when making their choices.

optimal and the latter sub-optimal. That is, an agent with Bayesian beliefs would consider $\overline{\sigma}$ to be of higher value than $\underline{\sigma}$. Under the hypothesis that mistakes in choices over information structures are driven by mistaken beliefs (deviations from Bayesian updating), we would expect $\underline{\sigma}$ to be chosen over $\overline{\sigma}$ much more frequently when subject i states beliefs such that $V(\underline{\sigma}) > V(\overline{\sigma})$ and less frequently when $V(\underline{\sigma}) < V(\overline{\sigma})$. Panel (b) of Figure 7 shows, for each subject type (Optimal, Confirmation etc.) and bias-type (commission or omission) the proportion of optimal choices (selections of $\overline{\sigma}$) in cases in which subjects' beliefs imply (i) a correct value ordering ($V(\underline{\sigma}) < V(\overline{\sigma})$, in black) versus (ii) an incorrect value ordering ($V(\underline{\sigma}) \geq V(\overline{\sigma})$, in gray).

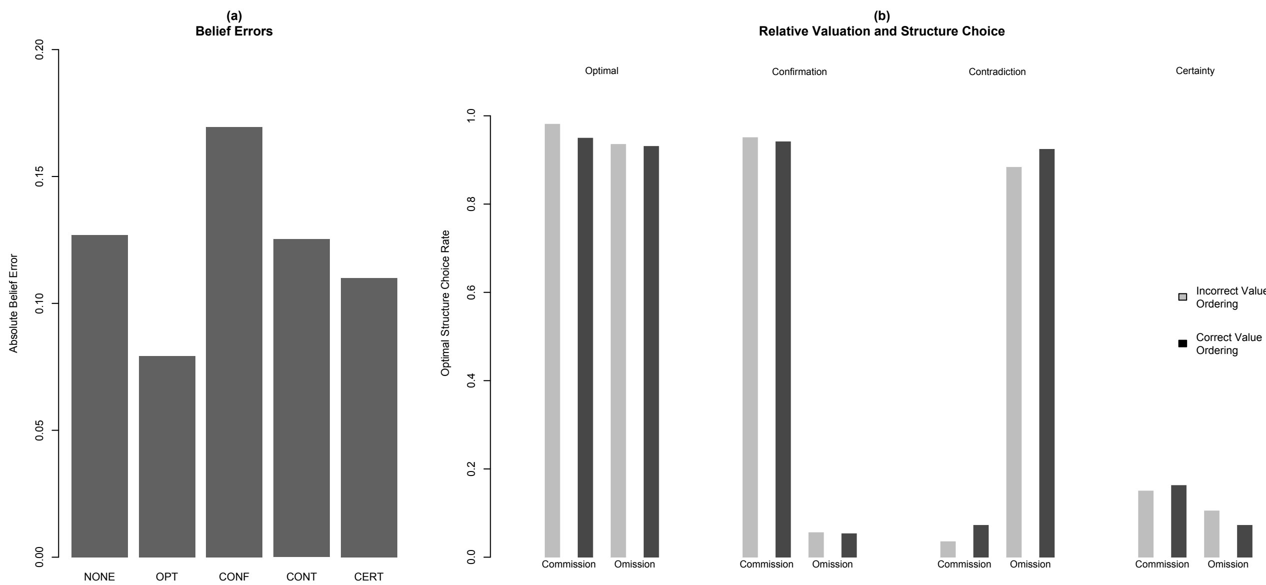


Figure 7: Belief Errors and Information Structure Choice by Type *Notes: Incorrect (Correct) Value Ordering refers to problems in which a subject's expected value for the optimal information structure (implied by the beliefs the subject submits for this structure in the EX block) is lower than the expected value for the sub-optimal information structure.*

The results cast serious doubt on the hypothesis that mistaken beliefs drive usage of sub-optimal decision rules (confirmation, contradiction and certainty seeking behavior). For both bias-type problems, the rates of optimal structure choice are *no higher* when beliefs generate a correct value ordering than when they generate an incorrect value ordering in the two most common types (Optimal and Confirmation).[85] This is true even though there is substantial variation in value orderings in each of these subject-type/bias-type combinations (a full 35% of Confirmation types have the correct value ordering even for bias-by-omission problems in which subjects almost never choose the correct structure). Contradiction types are slightly more likely to choose the correct

[85]This suggests that even Optimal types, to a large extent, don't rely on conditional beliefs to identify the optimal information structure. This is consistent with how subjects describe their decision rules in the Advice task.

advisor when value orderings are correct. However, even here these likelihood differences do almost nothing to explain the large drops in rates of optimal structure choice in the bias-by-omission case relative to the bias-by-commission case (or the reverse for Confirmation types).

We conclude that although types differ to some degree in the accuracy of their beliefs, these differences do little to explain observed patterns in subjects' mistakes in choosing between information structures.

Now we look at how biased stated beliefs are relative to the prior. A necessary condition for Bayesian updating is that the expected posterior should be equal to the prior. We calculate $\sum_s \pi_s p_s - p_0$ for each subject and problem to see how much it deviates from zero.[86] As noted before, in analyzing the data, we have relabeled the states in each of the questions such that subjects can be considered to start with a prior $p_0 > 0.5$ in all questions.[87] We call a subject to have *negative bias* in a question if $\sum_s \pi_s p_s < p_0$. A negative bias implies that in updating beliefs, relative to the Bayesian benchmark, a subject is overweighing signals that are contradictory to their prior relative to those that are reinforcing of their prior. We see overwhelming evidence for negtive bias in updating. Focusing on stated beliefs in the EX block, optimal types show negative bias in 74% of the questions. The corresponding values for confirmation and contradiction seeking types is 78% and 89%.[88] This finding indicates that, at least on how they state their beliefs, subjects are not attached to their prior and are willing to state opinions to the contrary. Furthermore, we observe the relative ranking of these biases to be linked nicely to the classification categories. Subjects displaying confirmation seeking behavior in their information-structure choice also form beliefs that are most likely to lean in the direction of their prior, and those displaying contradiction seeking behavior form beliefs that are most leaning in the opposite direction.

To understand why beliefs are generally biased and significantly different for all types from the Bayesian benchmark, we generate a measure of *responsiveness to information* for different types of

[86] We are again making use of the strategy method here which enables us to have data on posteriors conditional on each state.

[87] Without this relabeling, this analysis would be testing if subjects form beliefs that are systematically biased towards the *orange* or *green* state as presented in the design (we vary which state is more likely according to the prior.) We don't find any bias in this case which confirms that subjects did not treat these states differently, but made decisions based only on which state was more likely.

[88] For certainty seeking types, the value is 75%. For untyped subjects, the value is 79%.

Type	α_r	α_l	α_n
Optimal	0.81	1.07	0.91
Confirmation	0.30	0.92	0.55
Contradiction	0.50	1.04	1.08
Certainty	0.68	1.11	0.86

Table 8: Responsiveness to information

signals by calculating for each subject and problem the following variable:[89]

$$p_s = p_0 + \alpha_s(p_s^{Bay} - p_0)$$

Note that $\alpha_s = 1$ corresponds to perfect Bayesian updating, $\alpha_s < 1$ suggests under-responsiveness to information, and $\alpha_s > 1$ suggests over-responsiveness to information. Once again, we normalize the questions to always set $p_0 > 0.5$, so that we can interpret r to be the prior reinforcing signal ($p_r^{Bay} > p_0$), and l to be the prior opposing signal ($p_l^{Bay} < p_0$). We focus on data from the EX block, where subjects were assigned information structures to facilitate the comparison between the different types.

A few observations stand out in Table 8. First, consistent with our analysis on *negative bias* above, in relative terms, subjects are more responsive to prior opposing signals, that is $\alpha_l > \alpha_r$ for all types ($p < 0.01$). Second, Optimal types are most responsive to signals that are reinforcing of one's prior, and surprisingly Confirmation types are the least responsive to these signals.[90] Third, Confirmation types, at least in how they state their beliefs, display difficulty in internalizing the informational content of failing to receive signals.[91] While the value for α_n is significantly different from 0 ($p < 0.01$), suggesting that, in the aggregate, there is some learning as a consequence of

[89]The literature usually focuses on logistic representation of Bayes' rule to construct a measure of responsiveness. (See Mobius et al. (2011) and Ambuehl & Li (2018) and references cited there for an overview of this). However, the type of information structures we include in our experiment, where there are fully revealing signals which give log likelihood ratios of zero or infinity, are not conducive to this type of analysis. For these reasons, to illustrate the main features of the data, we use a very simple measure.

[90]For all types α_r is significantly different than 1 ($p < 0.01$). Optimal and Certainty types are over-responding to contradictory signals with α_l significantly different than 1 ($p < 0.05$ for optimal , and $p < 0.01$ for Certainty types). Confirmation types are marginally under-responding to these signals ($p < 0.10$).

[91]We should note that analysis looking at guessing patterns reported in Section 7.4 does not show this problem to be reflected in how these types make guesses. Guesses are optimal 82% of the time conditional on this signal.

receiving this signal, the value is also significantly different from 1 ($p < 0.01$).[92,93]

[92] For Optimal types, it is also marginally different from 1 ($p < 0.10$).

[93] Recent literature has documented similar problems with interpreting failure to receive signals in different settings. See Jin et al. (2015) and Enke (2017) for a discussion on this.

D Further analysis on survey

	OPT	CONF	CONT	CERT
Wason Task	0.663***	-0.538	-0.407	-0.0643
	(0.219)	(0.280)	(0.330)	(0.308)
Belief Bias Score	0.219**	-0.0605	0.0723	-0.0363
	(0.0826)	(0.0729)	(0.0889)	(0.0947)
Raven Score	0.146	-0.0881	0.0472	0.207
	(0.118)	(0.110)	(0.133)	(0.146)
CRT Score	0.0813	-0.135	-0.0190	0.0831
	(0.121)	(0.112)	(0.134)	(0.155)
Values Strong Opinions	0.00783	-0.0175	0.123	-0.183
	(0.0668)	(0.0662)	(0.0817)	(0.0837)
Male	0.0621	0.260	-0.311	0.0658
	(0.160)	(0.159)	(0.195)	(0.203)
Analytical Major	0.175	-0.383*	0.434*	0.0659
	(0.158)	(0.155)	(0.188)	(0.198)
Confidence	1.042	-0.434	-0.405	0.0537
	(0.479)	(0.459)	(0.540)	(0.596)
Constant	-2.594***	0.416	-1.633**	-1.460*
	(0.526)	(0.462)	(0.581)	(0.619)
Observations	344	344	344	344

Standard errors in parentheses. ***1%, **5%, *10% significance.

Statistical significance accounts for multiple hypothesis testing (adjusted using Bonferroni correction).

OPT: Optimal; CONF: Confirmation; CONT: Contradiction; CERT: Certainty

Table 9: Probit Regression of the Probability of Being Each Type

	OPT	CONF	CONT	CERT
Errors on comprehension questions	-0.489***	0.273***	-0.0497	-0.0525
	(0.0993)	(0.0651)	(0.0852)	(0.0883)
Constant	-0.382***	-0.936***	-1.140***	-1.299***
	(0.0909)	(0.0979)	(0.109)	(0.116)
Observations	344	344	344	344

Standard errors in parentheses. ***1%, **5%, *10% significance.

Statistical significance accounts for multiple hypothesis testing (adjusted using Bonferroni correction).

OPT: Optimal; CONF: Confirmation; CONT: Contradiction; CERT: Certainty

Table 10: Probit Regression of the Probability of Being Each Type

	OPT	CONF	CONT	CERT
Ideology	-0.129	0.127	-0.0385	-0.0838
	(0.129)	(0.128)	(0.153)	(0.165)
Partisanship	0.0808	0.156	0.00976	-0.109
	(0.103)	(0.102)	(0.122)	(0.132)
Political engagement	-0.0468	0.0430	0.0342	-0.0419
	(0.0506)	(0.0512)	(0.0580)	(0.0650)
Political Informedness	0.0558	-0.0697	0.0681	-0.0446
	(0.0715)	(0.0715)	(0.0869)	(0.0917)
Trust news	0.0760	-0.108	0.0790	-0.0292
	(0.0626)	(0.0650)	(0.0746)	(0.0831)
Attentive to news sources	0.0971	-0.257**	0.0761	0.185
	(0.0989)	(0.0965)	(0.120)	(0.130)
Constant	-1.181**	0.366	-2.070***	-1.519**
	(0.435)	(0.412)	(0.550)	(0.592)
Observations	344	344	344	344

Standard errors in parentheses. ***1%, **5%, *10% significance.

Statistical significance accounts for multiple hypothesis testing (adjusted using Bonferroni correction).

OPT: Optimal; CONF: Confirmation; CONT: Contradiction; CERT: Certainty

Table 11: Probit Regression of the Probability of Being Each Type

	OPT	CONF	CONT	CERT
Errors on comprehension questions	-0.433***	0.234***	-0.0221	-0.00256
	(0.115)	(0.0728)	(0.0938)	(0.102)
Wason Task	0.643**	-0.538	-0.470	-0.0794
	(0.227)	(0.294)	(0.344)	(0.317)
Belief Bias Score	0.179	0.0139	0.0646	-0.0264
	(0.0894)	(0.0789)	(0.0946)	(0.101)
Raven Score	0.107	-0.0349	0.0148	0.204
	(0.124)	(0.116)	(0.137)	(0.148)
CRT Score	0.0254	-0.142	-0.0355	0.0612
	(0.125)	(0.117)	(0.137)	(0.157)
Values Strong Opinions	0.0217	-0.0188	0.140	-0.196*
	(0.0710)	(0.0714)	(0.0859)	(0.0873)
Male	0.0323	0.417*	-0.394	0.0389
	(0.173)	(0.174)	(0.206)	(0.217)
Analytical Major	0.0341	-0.473**	0.487*	0.0855
	(0.171)	(0.169)	(0.198)	(0.208)
Confidence	0.832	-0.332	-0.365	0.153
	(0.504)	(0.488)	(0.562)	(0.614)
Ideology	-0.220	0.173	-0.0828	-0.0146
	(0.144)	(0.135)	(0.164)	(0.177)
Partisanship	0.109	0.219	-0.0407	-0.0969
	(0.114)	(0.110)	(0.130)	(0.141)
Political Informedness	-0.0356	-0.0686	0.149	-0.0818
	(0.0790)	(0.0758)	(0.0924)	(0.0953)
Trust news	0.0457	-0.0976	0.0580	-0.0534
	(0.0674)	(0.0664)	(0.0788)	(0.0865)
Attentive to news sources	0.0430	-0.265**	0.141	0.188
	(0.103)	(0.0984)	(0.123)	(0.128)
Constant	-1.966**	1.045	-2.649***	-1.717
	(0.749)	(0.685)	(0.851)	(0.884)
Observations	344	344	344	344

Standard errors in parentheses. ***1%, **5%, *10% significance.

Statistical significance accounts for multiple hypothesis testing (adjusted using Bonferroni correction).

OPT: Optimal; CONF: Confirmation; CONT: Contradiction; CERT: Certainty

The regressors included are those selected using LASSO.

Table 12: Probit Regression of the Probability of Being Each Type

E Further analysis on advice data

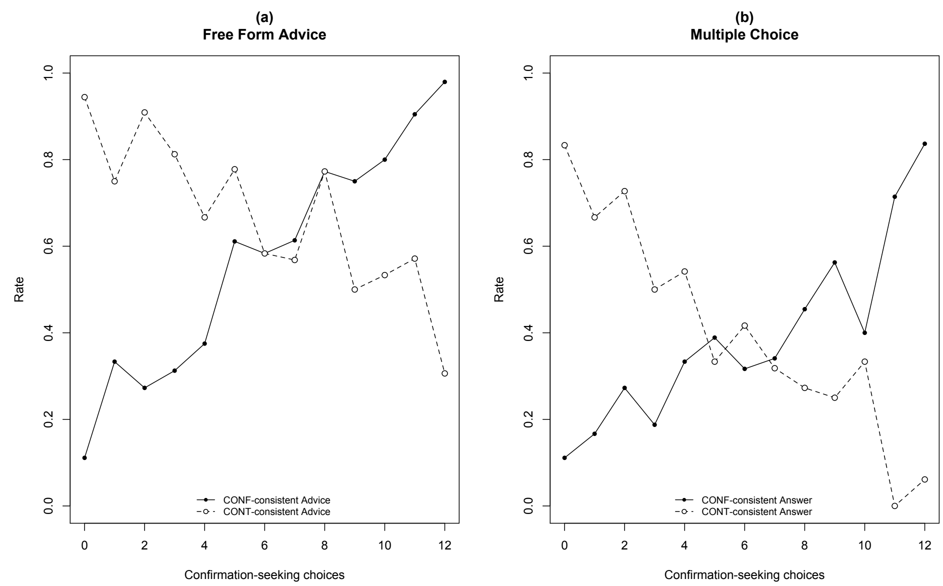


Figure 8: Relationship between confirmation-seeking choices and self-reported strategies. *Notes: CONF (CONT) -consistent advice/choice refers to free form advice (panel a) or multiple choice answer (panel b) that we have coded as consistent with confirmation (contradiction) seeking reasoning.*

We coded subjects' advices according to their consistency with the decision rules described in Section 5: CONF-consistent, CONT-consistent, OPT-consistent, CERT-consistent or NONE[94] (in ambiguous cases we coded advice as consistent with more than one rule). In Figure 8 (a), we plot the rate at which subjects submitted CONF-consistent and CONT-consistent advice as a function of the number of confirmation-seeking choices subjects made in the experiment. The results show a strong positive relationship between CONF-consistent advice and confirmation-seeking choices, with Confirmation types almost always giving advice coded as CONF-consistent. Likewise, there is a strong negative relationship between CONT-consistent advice and confirmation-seeking choices, with Contradiction types almost always giving advice coded as CONT-consistent.[95] (Optimal types also give advice coded as OPT-consistent nearly 75% of the time, though OPT advice is subtler and sometimes more difficult to differentiate from other types of advice).

As a robustness check, after the advice question we asked subjects to classify their information structure decisions by choosing one of the following from a multiple choice list: (1) "I mostly

[94]We coded subjects as NONE if they provided non-serious or indecipherable advice.

[95]Optimal and Certainty types cannot be distinguished as clearly on Figure 8 (both types make between 4 and 7 confirmation-seeking choices). Notably, in panel (b), where answers are most crisply coded, many subjects who make between 4 and 7 confirmation-seeking choices provide neither CONF or CONT-consistent answers to the survey (CONF-consistent and CONT-consistent sum to less than 1).

considered whether there were more orange or green balls in the basket and chose the advisor that gave this color advice most often."; (2) "I mostly considered whether there were more orange or green balls in the basket and chose the advisor that gave the opposite color advice most often."; (3) "Neither is a good description of how I chose an advisor."[96] In Figure 8 (b) we plot the rates at which subjects chose a CONF description (1) or CONT description (2) again as a function of the number of confirmation-seeking choices. Again, we find an extremely strong relationship, with confirmation-seeking subjects usually choosing the CONF description (and almost never choosing the CONT description) and contradiction-seeking subjects doing the reverse. These choices, too, are highly predictive of subject types as seen above.

	OPT	CONF	CONT	CERT
OPT-consistent advice	0.724***	-0.716***	-0.0207	-0.0382
	(0.180)	(0.219)	(0.217)	(0.247)
CONF-Consistent advice	-0.574**	1.983***	-0.944***	-0.290
	(0.211)	(0.362)	(0.266)	(0.259)
CONT-Consistent advice	-0.556*	0.642	0.575	-0.471
	(0.235)	(0.359)	(0.281)	(0.304)
CERT-Consistent advice	0.121	-0.721	0.0358	0.926***
	(0.227)	(0.334)	(0.262)	(0.272)
Constant	-0.505*	-1.852***	-1.160***	-1.250***
	(0.203)	(0.353)	(0.257)	(0.236)
Observations	344	344	344	344

Standard errors in parentheses. ***1%, **5%, *10% significance.

Statistical significance accounts for multiple hypothesis testing (adjusted using Bonferroni correction).

OPT: Optimal; CONF: Confirmation; CONT: Contradiction; CERT: Certainty

Table 13: Probit Regression of the Probability of Being Each Type

[96] This robustness check is useful because the advice task features a "beauty contest" structure that may conceivably push subjects to craft advice that they believe will be popular with other subjects. The multiple choice list cannot suffer from this problem. The similarity of the results across the two types of strategy elicitation suggests that subjects are likely honestly reporting strategies in both cases.

	OPT	CONF	CONT	CERT
Self-declared confirmation strategy	-0.0610	1.308***	-1.013***	-0.200
	(0.167)	(0.204)	(0.216)	(0.216)
Self-declared neutral strategy	-0.130	0.654**	-1.102***	-0.0876
	(0.201)	(0.241)	(0.290)	(0.252)
Constant	-0.630***	-1.505***	-0.655***	-1.240***
	(0.123)	(0.176)	(0.123)	(0.152)
Observations	344	344	344	344

Standard errors in parentheses. ***1%, **5%, *10% significance.

Statistical significance accounts for multiple hypothesis testing (adjusted using Bonferroni correction).

OPT: Optimal; CONF: Confirmation; CONT: Contradiction; CERT: Certainty

Table 14: Probit Regression of the Probability of Being Each Type

Examples from classification of advice

Classification	Advice
OPT-consistent	First, if the advisor only gives your advice on Orange or Green, you can choose the advisor who tends to choose the color of the ball that has more proportion. For example, there are more orange than green in the basket, you should choose the advice who are more likely to say orange because once he said green, it must be a green ball that will be chosen, otherwise, if he says orange, it will be more likely to be an orange ball because green ball has less chance to let him to say orange. / / Second, if the advisor also says nothing. Similar to the previous one, you may want the advisor to tends to keep quiet on when the color that has more proportion. If there is more orange than green, you may want him to keep more quiet on orange rather than green.
OPT-consistent	Choose the least ambiguous advice. If there are fewer green balls than orange balls, choose the adviser that can give you a response so that you know for certain the ball is green. Also, choose the adviser who will give the lower percent of the ambiguous advice to the less frequent ball colour. / / If the response "orange" from the adviser could mean either orange or green, choose the adviser who would is more likely to advise orange for the ball colour that is more frequent. The same goes for if saying nothing is an ambiguous response. / / The fewer of one colour there is, the less ambiguous you want the advice for that colour to be. The more of a colour there is, the more ambiguous it should be, as there is a greater chance of it being that coloured ball.
OPT-consistent	When picking an advisor, check the ratio of green to orange balls in the box above. If there is a majority colored ball, choose the advisor that will only say the minority color when it is picked. For example, when there is 10 orange balls vs 5 green balls, the orange balls are the majority. You should then chose the advisor that will say green only when green is pulled, that way you will know for sure that ball is green. If the advisor says orange and there is a possibility that the ball is green or orange, the risk is minimized since there is already a majority color of orange.
CONF-consistent	Look at the balls in the basket and see which one is highest, then look at which advisor leans more towards that color.
CONF-consistent	If there are more green ball than orange ball in the box, you should pick the one that more likely to say green; on the other hand, if there are more orange ball than green ball in the box, you should pick the one who will more likely to say orange.
CONF-consistent	Go with the advisor who is most certain about the more probabilistic outcome.
CONF-consistent	Run the numbers in your head. The ball's have probabilities that are multiples of five so they're easy, and the advisors are generally multiples of ten so they're easy too. Then see which advisors will give you the highest probability of telling the 'truth' and go with them.
CONF-consistent	You want to maximize the chance that the advisor will give you correct advice. To do this you should pick the advisor that gives you a 100% or at the very least gives you a higher chance of the correct color for the majority of balls. For example if there are 14 orange and 6 green. Pick the advisor that says 100% of saying orange if ball chosen is orange and like 70% chance of saying green if the ball chosen is green. This increases the chance of you getting the right answer.
CONF-consistent	When the amount of the orange balls is greater than the amount of the green balls, choose the advisor that gives more accurate information of the orange ball; / otherwise, choose the advisor that can give more accurate information of the green ball.

Classification	Advice
CONT-consistent	If there are more orange balls than green balls, pick the advisor that has a greater chance of saying green balls. Even if there are less green balls than orange balls which means there is a greater chance of an orange ball getting picked, if the advisor says there is a greater percentage of receiving a green ball, then it is most likely so since the advisor is more accurate with the ball being green than it is being orange.
CONT-consistent	Because many pairings of advisors had opposite likelihoods of truthfully disclosing the color of the ball, it makes sense to choose the advisor who is more likely to tell the truth about the color of ball which appears fewer times in the bag. For example, if there are 16 orange balls and 4 green balls, the advisor who tells the truth more often about the number of green balls is who you would be apt to choose, because regardless of possible deception about the number of orange balls, you are more likely to draw orange balls and thus answer more questions correctly.
CONT-consistent	Choose the advisor that has the greatest possibility of telling you the correct color of the ball. / / To do so, first determine how many of each color ball are in the basket. If there are a lot of orange, choose the advisor who is more likely to say green when the ball is green. If there are a lot of green, choose the advisor who is more likely to say orange when the ball is orange. This is because the ball is more likely to be the majority color and you would rather the advisor be more correct when the lesser probability color is chosen. / / Do a similar strategy when the advisors can say nothing. You would rather be more sure of when the lessor probable color occurs as it will occur less often.
CERT-consistent	1) Decide for each adviser if there is information that you can know with 100% certainty / / 2) Find the most likely color of the ball and pick the adviser that tells you that color with 100% certainty / / 3) If both advisers do this (ex: when they both say a color, they are certain its right, but they could also say nothing for both colors) look at which adviser will give you an answer for the most likely color more often and pick that advisor.

F Further analysis on oder effects

In this section, we study the extent to which there are order effects in the data. We focus on the endogenous block.

First we note that the rate at which the optimal information structure was chosen in the first block (first 6 rounds) is not significantly different from the rate corresponding to the second block.[97] These values are 56% and 60%, respectively.

Second, we investigate the extent to which the ordering of the blocks (whether it was omission or commission first) influences our estimates on the share of different types (Optimal, Confirmation, Contradiction and Certainty). If we focus on subjects who faced the commission problems first (allowing up to two deviations from a decision rule), we see that 69% of the population is typed as following one of these decision rules. The relative shares of each decision rule among those subjects who are typed are 27 % Optimal, 38% Confirmation, 23% Contradiction and 11% Certainty. If we focus on subjects who faced the omission problems first (allowing up to two deviations from a decision rule), we see that 71% of the population is typed as following one of these decision rules. The relative shares of each decision rule among those subjects who are typed are 43 % Optimal, 32% Confirmation, 11% Contradiction and 15% Certainty. With respect to these patterns, first we note that the share of Confirmation types remain above 30% with both orderings. The most striking pattern is the change in shares for the Optimal and Contradiction types (with total share of these types being between 50-55%). These results suggest that if subjects start thinking about optimality of an information structure within the context of the omission problems first, they have an easier time identifying optimal information structures in commission problems next, but that reasoning in the reverse order is harder. Namely, subjects who are prone to contradiction seeking cannot identify the optimal information structure when they are faced with the commission problems first.

Third, we look within a block to compare the rate at which the optimal information structure was chosen in the first three rounds vs. last three rounds. In aggregate, we find this rate to increase from 56% to 60% (from early to late rounds). This difference is significant ($p < 0.01$). When we break this up to look separately at omission and commission problems, we see that the effect is driven entirely by the omission problems. In the omission problems, the optimal information structure is chosen 64% chance in both early and late rounds. In the commission problems, the values are 48% and 56%, indicating some learning from introspection in these problems.

Next, we look at whether guessing patterns display any order effects. To this, we focus on

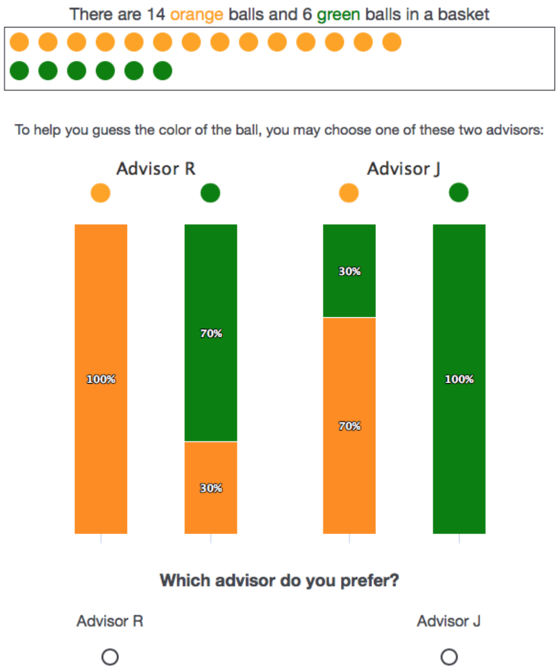
[97] All statistical statements are based on probit or OLS regressions with clustering at the subject level.

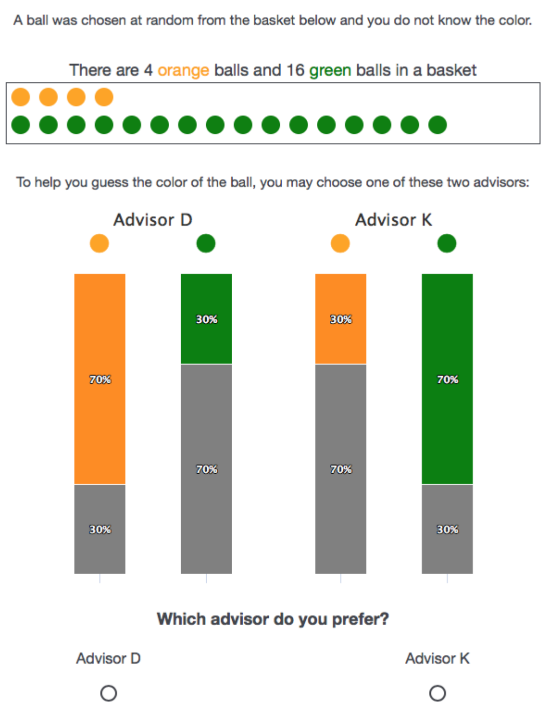
a measure of learning which captures increase in guessing accuracy relative to the prior in each question. We find some small but significant effects. For example, controlling for whether the optimal information structure was chosen, learning is 1% lower in the first block (first 6 rounds) relative to the second block ($p < 0.01$). In comparison, choosing the optimal information structure increases learning by 13%. Within a block, controlling for whether the optimal information structure was chosen, learning is 0.5% higher in the three rounds relative to the second three rounds ($p = 0.04$). Moreover, controlling for whether the optimal information structure was chosen, learning is is not significantly different between the Endogenous block and the Exogenous block.

In summary, while we do find some evidence of learning from introspection, these effects are relatively small. The most common types remain the same (either Optimal and Confirmation) and they in total make up about 65-70% of the observations that can be described by the four decisions rules considered in the paper. The order effects documented in this section suggest new directions for future work.

G Screenshots

G.1 Main experiment





G.2 Survey

Now we are going to ask you a set of 10 **Pay Questions**. In addition to your earnings so far, you will win 25 cents for each correct answer.

Pay Question 1

Below are four cards. Each has a symbol on one side and a number on the other side. You are given the following statement which may or may not be true of all four cards: "If a card has a star on one side, then it has a number larger than 10 on the other side."

Your task is to name those cards, and only those cards that need to be turned over in order to determine whether the statement is true or false.

- ☐ Star card
- ☐ Sun card
- ☐ 7 card
- ☐ 13 card

Pay Question 2

A farmer had 15 sheep and all but 8 died. How many are left?

- ○ none
- ○ 7
- ○ 8
- ○ 10

Pay Question 3

If you're running a race and you pass the person in second place, what place are you in?

- ○ first
- ○ second
- ○ third
- ○ fourth

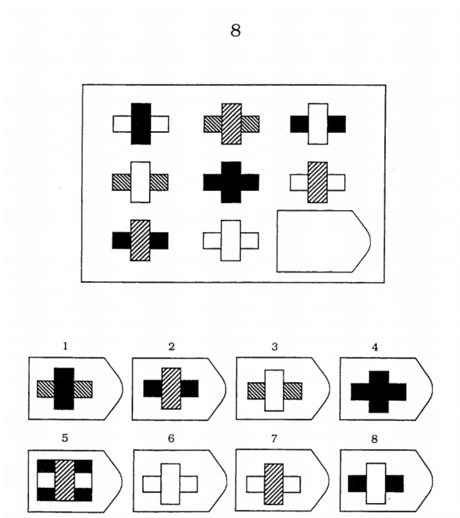
Pay Question 4

Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?

- ○ Can't know from information given
- ○ June
- ○ May
- ○ Emily

In each of the next three questions (5-7), there is a pattern with a piece missing and a number of pieces below the pattern. You have to choose which of the pieces below is the right one to complete the pattern.You will see 8 pieces that might complete the pattern. In every case, one and only one of these pieces is the right one to complete the pattern.

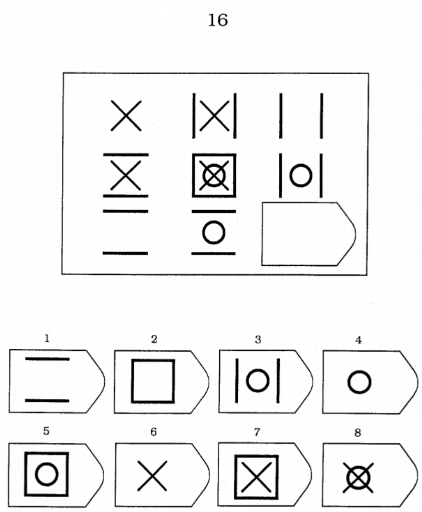
Pay Question 5



Which piece completes the pattern?

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8

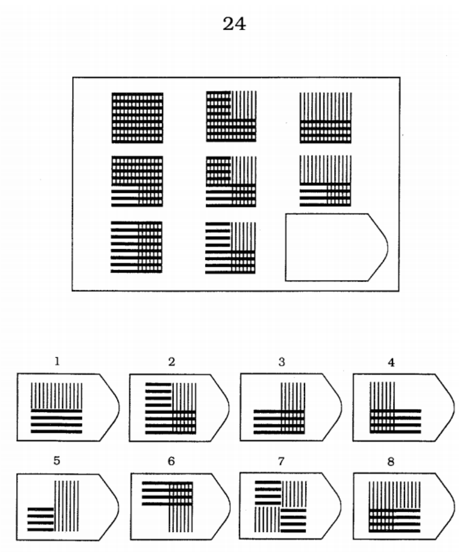
Pay Question 6



Which piece completes the pattern?

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

Pay Question 7



Which piece completes the pattern?

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8

For each of the following three problems (8-10), decide if the given conclusion follows logically from the premises. Select YES if, and only if, you judge that the conclusion can be derived unequivocally from the given premises. Otherwise select NO.

>>

Pay Question 8

Premise 1: All flowers have petals.
Premise 2: Roses have petals.
Conclusion: Roses are flowers.

Does the conclusion follow from the premises?

- ○ YES
- ○ NO

Pay Question 9

Premise 1: All mammals walk.
Premise 2: Whales are mammals.
Conclusion: Whales walk.

Does the conclusion follow from the premises?

- ○ YES
- ○ NO

Pay Question 10

Premise 1: All animals like water.
Premise 2: Cats do not like water.
Conclusion: Cats are not animals.

Does the conclusion follow from the premises?

- ○ YES
- ○ NO

How many of the preceding Pay Questions do you think you answered correctly?

How do you think you have done relative to others in the previous set of questions (the Pay Questions)?

- ○ Far above average
- ○ Moderately above average
- ○ Slightly above average
- ○ Average
- ○ Slightly below average
- ○ Moderately below average
- ○ Far below average

H Instructions

INSTRUCTIONS

You are about to participate in an experiment in the economics of decision-making. Follow these instructions carefully. In this experiment, you can earn a ***CONSIDERABLE AMOUNT OF MONEY****, which will be* ***PAID TO YOU IN CASH*** *at the end of the experiment.*

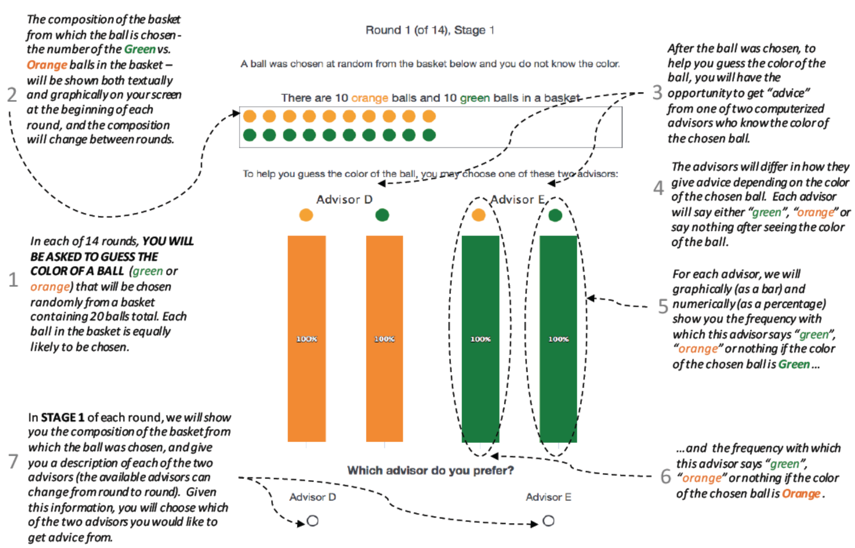
After the instruction period, ***we will ask you questions to check that you understand how the experiment works****. You should be able to answer all these questions correctly. You will not be able to participate in the experiment before you answer all questions correctly.*

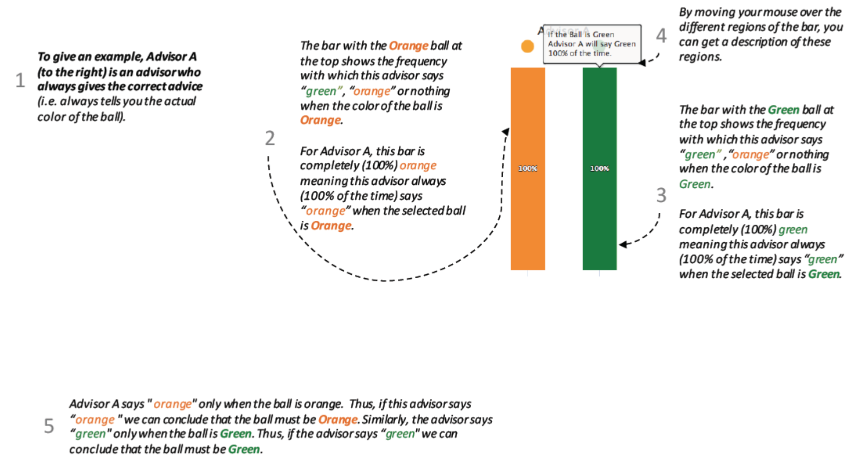
This experiment has two parts; these instructions are for the first part. Once this part is over, instructions for the second part will be given to you. Your decisions in this part have no influence on the other part.

The first part consists of ***14 rounds****. You will only be paid for* ***2 rounds*** *from the first part which will be randomly determined at the end of the experiment. Each round is equally likely to be selected.*

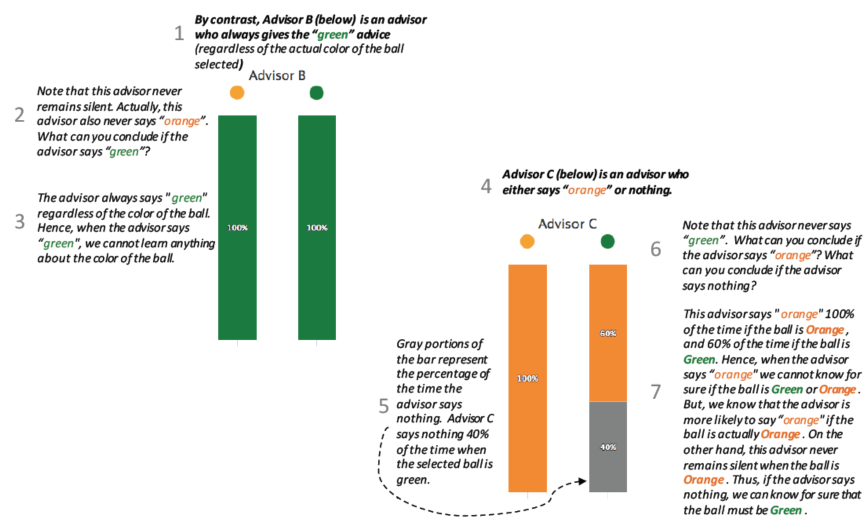
There is also a survey at the end of the experiment, and you will also receive additional payment for completing the survey.

1

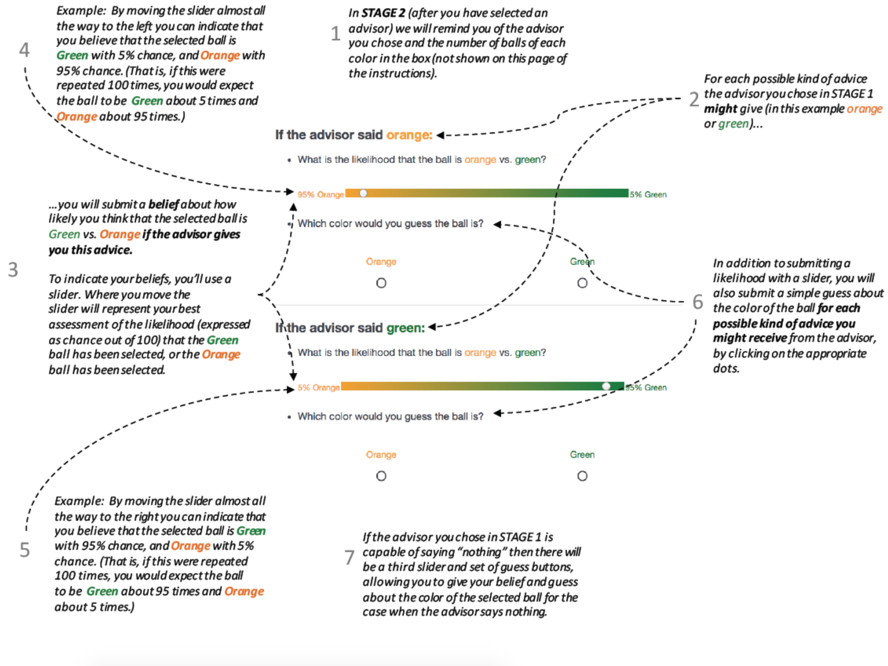


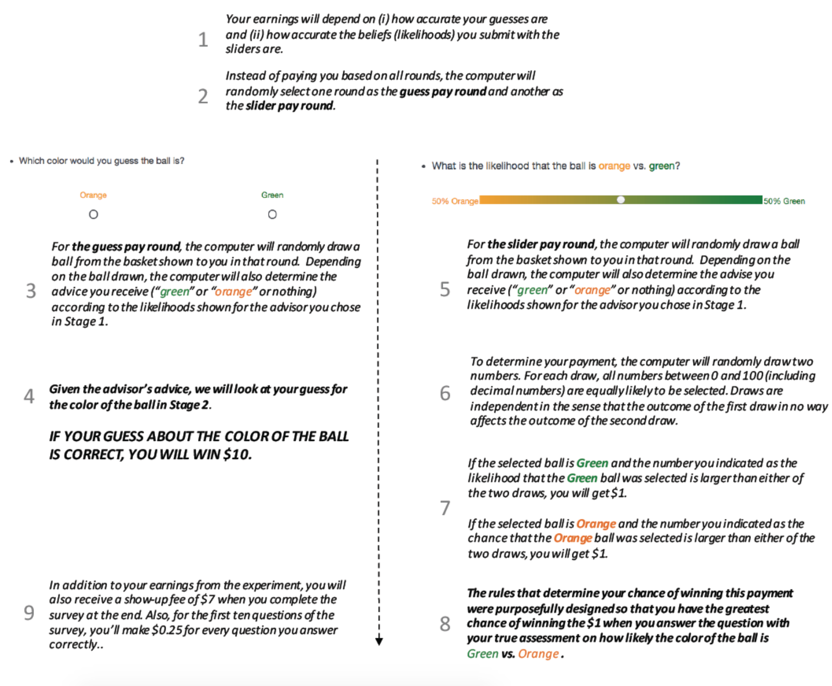


1 To give an example, Advisor A (to the right) is an advisor who always gives the correct advice (i.e. always tells you the actual color of the ball).
2 The bar with the Orange ball at the top shows the frequency with which this advisor says "green", "orange" or nothing when the color of the ball is Orange.
For Advisor A, this bar is completely (100%) orange meaning this advisor always (100% of the time) says "orange" when the selected ball is Orange.
100%
100%
4 By moving your mouse over the different regions of the bar, you can get a description of these regions.
3 The bar with the Green ball at the top shows the frequency with which this advisor says "green", "orange" or nothing when the color of the ball is Green.
For Advisor A, this bar is completely (100%) green meaning this advisor always (100% of the time) says "green" when the selected ball is Green.
5 Advisor A says "orange" only when the ball is orange. Thus, if this advisor says "orange" we can conclude that the ball must be Orange. Similarly, the advisor says "green" only when the ball is Green. Thus, if the advisor says "green" we can conclude that the ball must be Green.



1 By contrast, Advisor B (below) is an advisor who always gives the "green" advice (regardless of the actual color of the ball selected)
Advisor B
2 Note that this advisor never remains silent. Actually, this advisor also never says "orange". What can you conclude if the advisor says "green"?
3 The advisor always says "green" regardless of the color of the ball. Hence, when the advisor says "green", we cannot learn anything about the color of the ball.
100%
100%
4 Advisor C (below) is an advisor who either says "orange" or nothing.
Advisor C
100%
60%
40%
5 Gray portions of the bar represent the percentage of the time the advisor says nothing. Advisor C says nothing 40% of the time when the selected ball is green.
6 Note that this advisor never says "green". What can you conclude if the advisor says "orange"? What can you conclude if the advisor says nothing?
7 This advisor says "orange" 100% of the time if the ball is Orange, and 60% of the time if the ball is Green. Hence, when the advisor says "orange" we cannot know for sure if the ball is Green or Orange. But, we know that the advisor is more likely to say "orange" if the ball is actually Orange. On the other hand, this advisor never remains silent when the ball is Orange. Thus, if the advisor says nothing, we can know for sure that the ball must be Green.


4
Example: By moving the slider almost all the way to the left you can indicate that you believe that the selected ball is Green with 5% chance, and Orange with 95% chance. (That is, if this were repeated 100 times, you would expect the ball to be Green about 5 times and Orange about 95 times.)
1
In STAGE 2 (after you have selected an advisor) we will remind you of the advisor you chose and the number of balls of each color in the box (not shown on this page of the instructions).
2
For each possible kind of advice the advisor you chose in STAGE 1 might give (in this example orange or green)...
If the advisor said orange:
What is the likelihood that the ball is orange vs. green?
95% Orange
5% Green
Which color would you guess the ball is?
Orange
Green
3
...you will submit a belief about how likely you think that the selected ball is Green vs. Orange if the advisor gives you this advice.
To indicate your beliefs, you'll use a slider. Where you move the slider will represent your best assessment of the likelihood (expressed as chance out of 100) that the Green ball has been selected, or the Orange ball has been selected.
6
In addition to submitting a likelihood with a slider, you will also submit a simple guess about the color of the ball for each possible kind of advice you might receive from the advisor, by clicking on the appropriate dots.
If the advisor said green:
What is the likelihood that the ball is orange vs. green?
5% Orange
95% Green
Which color would you guess the ball is?
Orange
Green
5
Example: By moving the slider almost all the way to the right you can indicate that you believe that the selected ball is Green with 95% chance, and Orange with 5% chance. (That is, if this were repeated 100 times, you would expect the ball to be Green about 95 times and Orange about 5 times.)
7
If the advisor you chose in STAGE 1 is capable of saying "nothing" then there will be a third slider and set of guess buttons, allowing you to give your belief and guess about the color of the selected ball for the case when the advisor says nothing.


1
Your earnings will depend on (i) how accurate your guesses are and (ii) how accurate the beliefs (likelihoods) you submit with the sliders are.
2
Instead of paying you based on all rounds, the computer will randomly select one round as the guess pay round and another as the slider pay round.
Which color would you guess the ball is?
Orange
Green
3
For the guess pay round, the computer will randomly draw a ball from the basket shown to you in that round. Depending on the ball drawn, the computer will also determine the advice you receive ("green" or "orange" or nothing) according to the likelihoods shown for the advisor you chose in Stage 1.
4
Given the advisor's advice, we will look at your guess for the color of the ball in Stage 2.
IF YOUR GUESS ABOUT THE COLOR OF THE BALL IS CORRECT, YOU WILL WIN $10.
9
In addition to your earnings from the experiment, you will also receive a show-up fee of $7 when you complete the survey at the end. Also, for the first ten questions of the survey, you'll make $0.25 for every question you answer correctly..
What is the likelihood that the ball is orange vs. green?
50% Orange
50% Green
5
For the slider pay round, the computer will randomly draw a ball from the basket shown to you in that round. Depending on the ball drawn, the computer will also determine the advise you receive ("green" or "orange" or nothing) according to the likelihoods shown for the advisor you chose in Stage 1.
6
To determine your payment, the computer will 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.
7
If the selected ball is Green and the number you indicated as the likelihood that the Green ball was selected is larger than either of the two draws, you will get $1.
If the selected ball is Orange and the number you indicated as the chance that the Orange ball was selected is larger than either of the two draws, you will get $1.
8
The rules that determine your chance of winning this payment were purposefully designed so that you have the greatest chance of winning the $1 when you answer the question with your true assessment on how likely the color of the ball is Green vs. Orange.

INSTRUCTIONS FOR PART 2

1 *The second part of the experiment consists of 12 rounds and is very similar to the first part. In every round, you will be asked to guess the color of a ball that will be chosen randomly from a basket.*

2 *The only difference is that you will not be able to choose between advisors. Instead , the advisor will be given to you by the computer. Everything else will be the same.*

Similar to part 1, you will only be paid for two rounds. The computer will randomly select one round as the ***guess pay round*** *and another as the* ***slider pay round****.*

- Which color would you guess the ball is?

Orange ○ Green ○

3 *As before, for* ***the guess pay round****, the computer will randomly draw a ball from the basket shown to you in that round. Depending on the ball drawn, the computer will also determine the advice you receive ("green" or "orange" or nothing) according to the likelihoods shown for the advisor given to you in that round.*

4 ***Given the advisor's advice, we will look at your guess for the color of the ball.***

IF YOUR GUESS ABOUT THE COLOR OF THE BALL IS CORRECT, YOU WILL WIN $10.

- What is the likelihood that the ball is orange vs. green?

50% Orange — 50% Green

5 *For* ***the slider pay round****, the computer will randomly draw a ball from the basket shown to you in that round. Depending on the ball drawn, the computer will also determine the advise you receive ("green" or "orange" or nothing) according to the likelihoods shown for the advisor given to you in that round.*

6 *As before, you will have a chance to win $1 depending on your stated likelihood that the ball is Green* ***vs.*** *Orange for this advice. You can go back to the original instructions for how this payment is determined.*

7 ***To remind you, the rules that determine your chance of winning this payment were purposefully designed so that you have the greatest chance of winning the $1 when you answer the question with your true assessment on how likely the color of the ball is Green vs. Orange .***

8 *Finally, to remind you, in addition to your earnings from the experiment, you will also receive a show-up fee of $7 when you complete the survey at the end. Also, for the first ten questions of the survey, you'll make $0.25 for every question you answer correctly..*