Choosing Between Information Bundles (Job Market Paper)

Menglong Guan UC Santa Barbara

December 2023 (Check the latest version here.)

Abstract

This paper presents an experimental study on how people choose sets of information sources (referred to as information bundles). The findings reveal that subjects frequently fail to choose the more instrumentally valuable bundle in binary choices, largely due to the challenge of integrating the information sources within a bundle to identify their joint information content. The mistakes in choices can not be attributed to an inability to use information bundles. Instead, these mistakes are strongly explained by subjects' tendency to follow a simple but imperfect heuristic when valuing them, which I call "common source cancellation (CSC)". The heuristic causes subjects to mistakenly disregard the common information source in two bundles and focus solely on the comparison of the sources that the two bundles do not share. As a result, choices between information bundles are made without adequately considering the joint information content of each bundle. Notably, CSC emerges as a robust explanation for the information bundle choices for all subjects, including those who make perfect use of information bundles to make inferences.

Menglong Guan: Department of Economics, UC Santa Barbara, mguan@ucsb.edu. I thank Ryan Oprea, Sevgi Yuksel, and Erik Eyster for their invaluable guidance and support. This research was approved by UC Santa Barbara IRB. All errors are my own.

1 Introduction

In numerous contexts, people choose and make use of combined information sources (referred to as an information bundle) to form beliefs and facilitate judgments. For instance, doctors often choose multiple diagnostic tests to perform on patients, politicians assemble teams of consultants for advisory purposes, investors choose multiple financial market analysts to follow to seek investment advice, journal editors choose referees to review papers, and individuals decide which combinations of news sources to subscribe to. The optimal choice of information bundles hinges on a correct understanding of the joint information content of information sources within a bundle, and thus requires people to appropriately integrate multiple information sources. Information integration, which involves merging information from different sources in order to create a unified and comprehensive view, is potentially cognitively challenging. This is because it requires thinking through the possibility of receiving multiple pieces of information, the substitutability or complementarity of those pieces of information, and what they jointly imply.¹ For example, for a doctor to choose a proper set of diagnostic tests, she must understand the joint diagnosticity of different tests and know how to interpret possible combinations of test results. Similarly, an individual deciding between news sources must weigh their complementarity with existing sources and determine which combination yields the most comprehensive coverage.

In an age of abundant information, people can easily access many diverse information sources. Choosing which sources to use or pay attention to is thus an increasingly common decision problem people encounter in daily life. Understanding how people choose sets of information sources, i.e., information bundles, and what mistakes they make in those choices has therefore become increasingly relevant. Yet, to date, we know very little about these questions. To address the gap, this paper presents an experiment designed specifically to investigate people's choices of information bundles.

In the experiment, subjects face a simple guessing game in which they need to guess a binary state of the world. Before making a guess, subjects receive information from information sources that may improve the accuracy of their guesses. As illustrated in Figure 1, each information source is presented in an intuitive way that shows (i) the prior as a set of twenty objects (ten triangles and ten circles), one of which will be randomly drawn to determine the

¹Therefore, the challenges could lie in contingent reasoning, understanding and dealing with the correlation between information (sources), computational complexity involved, etc.



Figure 1: Examples of Representing Information Sources as Partitions Notes: σ , σ_0 and σ_b are three information sources. Subjects must guess the shape (triangle or circle) of a randomly drawn object among twenty objects. They are told which colored subset(s) contains the randomly drawn object under their chosen information source(s). $\sigma_0 \vee \sigma$, meaning the join of σ and σ_0 , is the integrated form of the information bundle { σ_0, σ }, which can be derived by finding out the intersections of signals (subsets) from σ and σ_0 .

true state (triangle or circle, i.e., T or C) and (ii) possible signals as subsets of the twenty objects (e.g., σ in Figure 1 has two subsets x and y).² When a subject receives a signal, she learns which subset contains the randomly drawn object before guessing the shape of the randomly drawn object. With multiple information sources, she receives information like this (i.e., which subset contains the true outcome) from each source.

This partition representation of information sources, which is built upon Guan, Oprea & Yuksel (2023) (GOY) and Brooks, Frankel & Kamenica (2023) (BFK), has two important features. First, it makes the characteristics of an information source visually transparent and can help remove classic mistakes in interpreting or using information (e.g., failures of Bayesian reasoning) that may bias the choice of information.³ Second, it pins down the joint information content between information sources and lays out a unique and seemingly straightforward way to correctly integrate information. For instance, the intersection of each possible pair of signals (subsets) of σ_0 and σ pins down their joint information content, described by $\sigma_0 \vee \sigma$, meaning the *join* of σ_0 and σ , as shown in Figure 1.⁴ The indicated

⁴Signal (subset) a in $\sigma_0 \vee \sigma$ is the intersection of x in σ and m in σ_0 ; b is the intersection of y and m; c is the intersection of y and n; and d is the intersection of x and n.

²Under this representation, an information source can be formally conceptualized as a partition of the extended state space $\Omega \times \{1, ..., 20\}$ (Green & Stokey 1978), where $\Omega = \{T, C\}$ is the set the payoff-relevant states. For example, the information source σ can be characterized as $\sigma = \{x, y\} = \{(T, \{1, ..., 5\}) \cup (C, \{11, ..., 18\}), (T, \{6, ..., 10\}) \cup (C, \{19, 20\})\}.$

 $^{^{3}}$ The results of the experiment strongly support this: conditional on receiving a signal from a given information source, subjects make optimal (from the Bayesian perspective) guesses about the shape of the randomly drawn object 98% of the time. In the experiment of GOY, which uses a different visualization of partition representation, subjects also use individual information sources optimally 98% of the time.

procedure of integrating information under this design captures how information integration is done in real-world scenarios: merging information from multiple sources to create a unified, cohesive, and comprehensive view.

The experiment consists in part of a sequence of binary choices (eight in total) between information bundles. In each of them, a subject chooses between a pair of information bundles, $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$, where σ_0 , σ and σ' are three distinct information sources. The subject will be provided with her chosen information bundle in a future payoff-relevant guessing game (in which she receives a signal from each source within the bundle before making guesses), making the choice elicitation incentive-compatible. According to standard economic theory, the subject should always choose the more instrumentally valuable bundle (i.e., the bundle that may induce a higher guessing accuracy).

My first main finding is that subjects' choices between information bundles are largely suboptimal: their likelihood of choosing the more instrumentally valuable bundles is only 56%. I further show that the suboptimal choices are strongly driven by subjects' failures to integrate information sources within a bundle and identify their joint information content. In a control setting, each pair of information bundles (denoted as $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$) are preintegrated into single information sources that contain the same information content as the bundles (e.g., $\sigma_0 \vee \sigma$ shown in Figure 1 contains the same information as σ_0 and σ together). Subjects then make choices between the two constructed *join* information sources (denoted as $\sigma_0 \vee \sigma$ and $\sigma_0 \vee \sigma'$). Removing the need to integrate sources and identify their joint information content, the optimality of information choices increases considerably to 77% (signed-rank test, p < 0.001).⁵ In addition, I find that subjects' choices between a pair of bundles barely correlate with their choices between the corresponding pair of *join* information sources (Kendall's $\tau = 0.148$, p = 0.62), indicating significant failures in identifying the joint information content of bundles and making choices accordingly.

The decrease in choice optimality from the control to the treatment (i.e., when facing information bundles) settings exhibits a pattern that aligns with a theory of difficulty in comparing information bundles suggested by BFK that I designed the experiment to test. BFK characterizes a set of comparison relationships between information sources (represented as partitions) and shows that for any σ_0 , { σ_0, σ } Blackwell dominates { σ_0, σ' } (meaning the

⁵At the subject level, 69% (85%) of subjects have a strictly (weakly) higher likelihood of choosing the more valuable information in binary choices between *join* information sources than those between (theoretically equivalent) information bundles.

former is weakly more instrumentally valuable) if and only if σ reveals-or-refines σ' . Revealor-refine means that each signal of σ either fully reveals the state or is a subset of some signal of σ' . A stronger relationship is refine, meaning that each signal of σ is a subset of some signal of σ' . The two relationships can be easily verified by "visual inspection" given the adopted partition representation of information sources.⁶ BFK's results suggest a theory of difficulty in comparing information bundles: When σ and σ' have a refine or reveal-or-refine relationship, the comparison of $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$ can be done easily without the need to integrate sources and identify the joint information content of each bundle. My experimental design incorporates the comparison relationships characterized by BFK. Findings show that the optimality gap between bundle choices and the corresponding join source choices is relatively smaller when σ refines or reveals-or-refines σ' , compared to other cases. This suggests that subjects' information bundle choices in those two cases are less distorted by the difficulties of information integration, supporting the theory of difficulty in comparing information bundles implied by BFK.

Next, I identify the source of the mistakes in information bundle choices. Making the optimal choice between information bundles generally requires subjects to think through the joint instrumental value of each bundle (i.e., how each bundle improves guessing accuracy) and then make choices accordingly. Suboptimal choices could arise from two plausible channels: (i) while subjects may intend to follow the optimal approach, they may be unable to properly interpret information bundles, leading to mistakes in valuing them; (ii) alternatively, subjects may entirely deviate from the optimal approach of valuing and comparing information bundles and make systematic mistakes in choices as a result.

To examine the first channel, I study how subjects make use of each bundle in the guessing game by eliciting their guesses about the randomly drawn object conditional on each possible pair of signals that the bundle might generate. Subjects make Bayesian optimal guesses 85% of the time, indicating fairly good use of the information bundles. Nonetheless, this was significantly lower than the 98% optimality rate when using join information sources. This suggests that the challenge of integrating information indeed leads to more errors in information usage.⁷ However, this reduction in guessing accuracy cannot explain mistakes in choices between bundles. Subjects only choose the bundle with a (weakly) higher "practical"

⁶Check Figure 17 in Appendix B for examples of σ reveals-or-refines σ' and σ refines σ' .

⁷I also find that 82 percent of the suboptimal guesses when using information bundles can be explained by subjects following a simple but incorrect way of combing signals (subsets). A more detailed discussion is provided in Section 4.2.

value conditional on their submitted guesses 62% of the time (significantly lower than the rate of 78% in choices between *join* sources). And the correlation between comparisons of the practical value of bundles and actual bundle choices is weak (Kendall's $\tau = 0.296$, p = 0.31). Following the elicitation of guesses, subjects are also asked to assess what level of guessing accuracy an information bundle induces conditional on how they use it.⁸ These subjective assessments cannot explain bundle choices either. Subjects choose the bundle to which they assign a (weakly) higher assessment only 61% of the time (significantly lower than the rate of 72% in choices between *join* sources). The correlation between assessments and actual choices is also minimal (Kendall's $\tau = 0.255$, p = 0.38). Taken together, these results suggest that subjects make information bundle choices without much consideration for how they would use the bundles, and therefore, mistakes in choices cannot be primarily attributed to errors or noise in the usage of information bundles. This is the second main finding of the paper.

Another possible channel driving mistakes in choices between information bundles is that subjects use some simpler decision rule that systematically deviates from the rational one. A potentially simple and intuitive decision rule is to reduce a choice between a pair of bundles $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$ to a choice between σ and σ' by "canceling" σ_0 , which I call the *common source cancellation (CSC)* heuristic. This heuristic is very appealing since it offers a way out of the difficulties associated with identifying the joint information content of information sources.^{9,10} To directly test if *CSC* drives information bundle choices, subjects are asked to make another sequence of binary choices between two single information sources (e.g., σ versus σ'). Each is designed to correspond to a binary choice between bundles $\{\sigma_0, \sigma\}$ and

⁸This belief elicitation is incentivized by the Binarized Scoring Rule (Hossain & Okui 2013) and implemented following the procedure proposed by Wilson & Vespa (2016).

⁹The heuristic is related to the "tendency to simplify decision problems" in human decision making emphasized by Rubinstein (1998) and a large literature on bounded rationality. Rubinstein (1998) hypothesizes that when comparing two choice alternatives, decision makers have the tendency to simplify the comparison by canceling the components of the two alternatives that are alike, which means canceling σ_0 when choosing between $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$ in my experiment.

¹⁰The heuristic might also be related to but can not be reduced to correlation neglect (Eyster & Weizsäcker 2011, Enke & Zimmermann 2019). Blackwell (1951, 1953) and Mu, Pomatto, Strack & Tamuz (2021) discuss that when σ_0 is independent (conditional on the true state) of σ and σ' , if σ Blackwell dominates σ' , then $\{\sigma_0, \sigma\}$ Blackwell dominates $\{\sigma_0, \sigma'\}$ as well (meaning the former is at least weakly more instrumentally valuable). However, this is not generally true when σ and σ' can not be Blackwell ordered. At least in that scenario, even a correlation-neglect subject still needs to think through the joint information content of each bundle to identify which one is more valuable.

 $\{\sigma_0, \sigma'\}$ from the experiment. If subjects follow the *CSC* heuristic, their choices between $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$ should align sharply with their choices between σ between σ' .

My third main finding is that CSC is the primary driver of information bundle choices. Subjects' likelihood of choosing σ over σ' strongly explains their likelihood of choosing $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$ (Kendall's $\tau = 0.764$, p < 0.01). Regression analysis further confirms that subjects' choices between bundles ($\{\sigma_0, \sigma\}$ versus $\{\sigma_0, \sigma'\}$) are significantly responsive to the difference in instrumental value (and informativeness) between σ and σ' rather than the value (informativeness) difference between the bundles.¹¹ When focusing on the mistakes in bundle choices, I find that the choices between σ and σ' can account for over 68 percent of all the suboptimal choices between information bundles. In addition, a heterogeneity analysis shows that CSC emerges as a primary explanation for information bundle choices and mistakes in those choices for all subjects, including those who make perfect use of information bundles in the guessing game. The heterogeneity analysis also reveals that subjects who are less able to integrate and interpret disaggregated information tend to rely more heavily on the heuristic in information bundle choices.

The common source cancellation heuristic that prevails in the data is very intuitive and is plausibly important in many choices of information sources. This heuristic means that people tend to compare information sources in isolation without considering their joint information content with other in-company sources (that they already have or choose together). One consequence of the heuristic is that it hinders people from diversifying their choices of information sources as they should. For instance, in the context of news consumption, this heuristic may potentially exacerbate polarization in news media choices. Imagine a scenario where a Republican is deciding between turning to either Fox News and The Blaze, or Fox News and CNN for political news. If influenced by the common source cancellation heuristic, the person would focus only on the comparison between The Blaze and CNN but not take into account the joint coverage of each combination of news sources. As a result, the person fails to recognize that the latter combination is likely to provide a more comprehensive coverage of political news (as Fox News and CNN are less overlapped). This oversight would lead to a missed opportunity for a more diverse and inclusive news consumption, resulting in less accurate beliefs.

This paper adds to a growing literature investigating how people choose or evaluate

¹¹In contrast, regression analysis shows that subjects' choices between information bundles are only slightly responsive to the practical value (conditional on guesses) or assessments of the bundles.

instrumentally valuable information sources (structures) (Ambuehl & Li 2018, Charness, Oprea & Yuksel 2021, Montanari & Nunnari 2022, Guan, Oprea & Yuksel 2023, Novak, Matveenko & Ravaioli 2023, Liang 2023).¹² Existing studies focus on choosing or evaluating single information sources in circumstances in which there is no need to consider the joint information content between sources.¹³ In contrast, this paper focuses on choices between information bundles, i.e., sets of information sources, in which the optimal choice requires correctly identifying the joint information content between sources.

This paper relates to recent theoretical works on the comparisons of information sources given some pre-existing information source (Brooks, Frankel & Kamenica 2023), and the dynamic acquisition of possibly complementary information sources (Liang & Mu 2020, Liang, Mu & Syrgkanis 2022).¹⁴ To my knowledge, the current paper is the first experimental study that examines whether and how people consider the joint information content between sources when it is a necessary step for the optimal choice of information. The experiment shows that people have limited ability to integrate information sources, and they do not take adequate account of the joint information content between sources when making choices.

This paper also relates to a strand of literature showing that people choose simple but imperfect decision rules as a way to avoid difficulties associated with developing or executing optimal strategies in cognitively challenging decision settings. The literature documents that System 1 thinking (i.e., the fast, automatic, intuitive, and effortless way of thinking) drives

¹⁴Blackwell (1951) and Mu, Pomatto, Strack & Tamuz (2021) discuss the comparison between sets of information sources but focus only on independent (conditional on the true state) information sources. Besides, some existing studies consider the settings that require thinking about the joint information content of multiple information sources or multiple pieces of information but do not focus on the choice of information. For example, Börgers, Hernando-Veciana & Krähmer (2013) characterize the complementarity and substitutability of two information sources (Blackwell experiments), Gentzkow & Kamenica (2017a,b) study information design games with multiple senders who provide potentially complementary information to influence a receiver, De Oliveira, Ishii & Lin (2021) focus on characterizing the optimal strategy of combining information sources that is robust to the correlation between information sources, Arieli, Babichenko & Smorodinsky (2018) study the robust aggregation of signals from information sources of which the decision maker may have limited knowledge, Levy & Razin (2021, 2022) study the optimal way of combining signals generated from multiple correlated information sources whose correlation structures are unknown or ambiguous, Enke & Zimmermann (2019), Hossain & Okui (2021) and Fedyk & Hodson (2023) experimentally study belief formation given signals from correlated information sources, etc.

 $^{^{12}}$ Some other work focuses on the demand for non-instrumental information or information source. The interested reader is referred to Nielsen (2020) or GOY for reviews of the literature.

 $^{^{13}}$ A recent work by Calford & Chakraborty (2023) studies the use, valuation and choice of multiple deterministic signals, rather than noisy information sources (structures).

human reasoning and decision making in many cases (Kahneman 2011), decision makers have a tendency to simplify decision problems (Rubinstein 1998), people narrowly frame choices by thinking about a choice in isolation without considering the broader context (Kahneman & Lovallo 1993, Barberis, Huang & Thaler 2006, Rabin & Weizsäcker 2009), decision makers often form a simplified model of the world and act using that simplified model (Gabaix 2014), etc. The current paper provides evidence of people following simplifying heuristics in a new and important context, the choices of sets of information sources.

The remainder of the paper is organized as follows. Section 2 introduces the conceptual framework. Section 3 describes the experimental design. Section 4 presents the main results. Section 5 discusses the possible reasons behind the emergence of the *common source cancellation* heuristic and other determinants of information choices. Section 6 concludes.

2 Conceptual Framework

2.1 Instrumental Value of Information

Let $\omega \in \Omega$ be the state of the world, where Ω is a finite state space. There is a prior distribution on Ω denoted by p. An information source (information structure) σ is a mapping from the state space Ω to a finite signal space S. Let σ_{ω}^{s} be the probability of the information source σ generating signal $s \in S$ conditional on state ω . Signal s induces a posterior distribution, denoted by q_{σ}^{s} , over the state space Ω . According to Bayes' Rule, $q_{\sigma}^{s}(\omega) = \frac{p(\omega)\sigma_{\omega}^{s}}{q_{\sigma}(s)}$, where $q_{\sigma}(s) = \sum_{\omega} p(\omega)\sigma_{\omega}^{s}$ is the probability of signal s being realized.

A decision problem D = (A, u) consists of a finite action set A and a utility function $u : A \times \Omega \to \mathbb{R}$. The decision maker (DM) chooses an action $a \in A$ after observing signal s generated by information source σ to maximize $\mathbb{E}[u(a, \omega)|s] = \sum_{\omega} q_{\sigma}^{s}(\omega)u(a, \omega)$. Following the standard definition in economics, the *instrumental value* of information source σ , in decision problem D, is the increase in expected utility due to the DM being able to condition her action choice on the realized signals. That is,

$$V_{\sigma} = \sum_{s \in S} q_{\sigma}(s) \max_{a \in A} \mathbb{E}[u(a, \omega)|s] - \max_{a \in A} \mathbb{E}[u(a, \omega)]$$

where $\mathbb{E}[u(a,\omega)] = \sum_{\omega} p(\omega)u(a,\omega).$

The decision problem D used in this paper is a simple guessing game. There is a binary

state of the world, i.e., $\Omega = \{T, C\}$, with a uniform prior p : p(T) = p(C) = 0.5. The DM makes a guess $a \in A = \{T, C\}$ with the objective of matching the underlying state. The DM earns a bonus of γ ($\gamma > 0$) if her guess matches the state and zero otherwise, i.e., $u(a, \omega = a) = \gamma$ and $u(a, \omega = -a) = 0$. A utility-maximizing DM always guesses the more likely state. With an information source, the DM guesses the underlying state to be the more likely state conditional on the realized signal. The guess will be correct, i.e., a = w, with a probability of max $\{q_{\sigma}^{s}, 1-q_{\sigma}^{s}\}$. Therefore, the *instrumental value* of σ can be simplified into:

$$V_{\sigma} = \left(\underbrace{\sum_{s \in S} q_{\sigma}(s) \underbrace{\max\{q_{\sigma}^{s}, 1 - q_{\sigma}^{s}\}}_{\text{Guessing accuracy conditional on } s}}_{\text{Expectation over } s} - \underbrace{p}_{\text{Guessing accuracy without information}}\right) \gamma \tag{1}$$

Expected improvement in guessing accuracy

which is the expected improvement of guessing accuracy induced by σ multiplying with the constant reward γ .

2.2 Information Bundles

An information bundle is a finite set of information sources. In this study, I focus on information bundles that consist of two distinct information sources, for example, an information bundle $b = \{\sigma, \sigma'\}$. With bundle b, the DM observes both a signal $s \in S$ from σ and a signal $s' \in S'$ from σ' before taking an action. Let $q_b(\{s, s'\}) = \sum_{\omega} p(\omega)p(\{s, s'\}|\omega)$ be the probability of observing s and s' at the same time, $S_b = \{\{s, s'\} : q_b(\{s, s'\}) > 0, s \in S, s' \in S'\}$ be the finite set of all possible signal combinations, and s_b be a realized signal combination.¹⁵ The information bundle b is then a mapping from state space Ω to S_b . It is convenient to think of each s_b as a re-defined signal such that s_b is equivalent to observing $\{s, s'\}$ and S_b as the set of the re-defined signals. Then the mapping characterized by b is just an information source, denoted as σ_b . Following BFK, the information source σ_b is referred to as the *join* of σ and σ' , denoted as $\sigma_b \equiv \sigma \vee \sigma'$, meaning σ_b is equivalent to observing both σ and σ' .

The *instrumental value* of information bundle b can be defined in the same way as above:

$$V_b = V_{\sigma_b} = \left(\sum_{s \in S_b} q_{\sigma_b}(s) \max\{q_{\sigma_b}^s, 1 - q_{\sigma_b}^s\} - p\right)\gamma$$

¹⁵With the partition representation of information sources, the correlation between two information sources is pinned down, and $p(\{s, s'\})$ are straightforward to identify.

Note that the *join* information source σ_b and the information bundle b are theoretically equivalent and equally valuable, but the use or evaluation of the latter requires a further step of integrating signals.

Given that γ is a constant, for simplicity, I will refer to the expected improvement in guessing accuracy induced by a certain information bundle (source) as its *instrumental value*. In standard economic theory, the choice between information bundles (sources) is assumed to rely only on the comparison of their instrumental value.

2.3 A Taxonomy of Comparisons of Information Bundles

When comparing and choosing between information bundles, in general, the DM needs to think through the joint instrumental value of each bundle (which necessarily involves information integration) and then make choices accordingly. An important recent paper by Brooks, Frankel & Kamenica (2023) studies the comparisons of information sources given some pre-existing information source. Their results provide a taxonomy of information bundle comparisons and characterize the scenarios in which the comparison can be done in an easy and intuitive way.

BFK adopts an alternative conceptualization of information sources (that was first formalized by Green & Stokey (1978)). Under that conceptualization, an information source is characterized as a *partition* of the extended state space $\Omega \times X$, where X is the set of "states" that govern the signal realization conditional on the payoff-relevant state (Ω), and a signal s is a subset of $\Omega \times X$, i.e., an element of the partition. Building upon the partition representation of information sources, BFK characterizes a list of comparison relationships between information sources, including (from strongest to weakest): (i) *Refine*, σ refines σ' , denoted as $\sigma R \sigma'$, if any signal of σ is a subset of some signal of σ' ; (ii) *Reveal-or-refine*, σ reveal-or-refine σ' , denoted as $\sigma O \sigma'$, if any signal of σ either fully reveals the state (i.e., $P(s|\omega) > 0$ for at most one ω) or is a subset of some signal of σ' ; (iii) *Sufficiency*, σ is sufficient for σ' , denoted as $\sigma S \sigma'$, if for any $s \in \sigma$ and any $s' \in \sigma'$, $P(s'|s, \omega) = P(s'|s)$, or equivalently, if for any decision problem D, $\sigma \vee \sigma'$ has the same value as σ ; (iv) *Blackwell*, σ Blackwell dominates σ' if σ is (weakly) more valuable than σ' for any decision problem D(Blackwell 1953).¹⁶ These relationships, especially the first two, are straightforward to check

 $^{^{16}}$ The interested reader is referred to BFK for a more detailed discussion of the listed comparison relationships (and an uncovered relationship *Martingale*, which is weaker than *Sufficiency* but stronger than

given the partition representation of information sources.¹⁷

What are the implications of these relationships between information sources on the comparison of information bundles? Consider any information source σ_0 . Its joint information content with σ (σ') can be characterized by the interactions of all possible signal combinations (each signal being a subset of $\Omega \times X$) of it and σ (σ'). By the definition of *refine*, if σ refines σ' , then any signal of $\sigma_0 \vee \sigma$ will be a subset of some signal of $\sigma_0 \vee \sigma'$, i.e., $\sigma_0 \vee \sigma$ refines $\sigma_0 \vee \sigma'$. Similarly, if σ reveals-or-refines σ' , then $\sigma_0 \vee \sigma$ reveals-or-refines $\sigma_0 \vee \sigma'$. So for any σ_0 , if $\sigma R \sigma'$ or $\sigma O \sigma'$, then $\sigma_0 \vee \sigma$ is (weakly) more instrumentally valuable than $\sigma_0 \vee \sigma'$ (as both *refine* and *reveal-or-refine* imply Blackwell), and equivalently, bundle { σ_0, σ } is (weakly) more valuable than { σ_0, σ' }. In fact, BFK proves that for any σ_0 , { σ_0, σ } Blackwell dominates { σ_0, σ' }, meaning the former is (weakly) more instrumentally valuable in any decision problem, if and only if $\sigma O \sigma'$.

BFK's results suggest a theory of difficulty in comparing (and choosing between) information bundles. When σ and σ' exhibit a *refine* or *reveal-or-refine* relationship, the comparison of $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$ becomes relatively intuitive and does not necessarily require the DM to integrate sources and recognize the joint information content (the joint instrumental value) of each bundle. In contrast, in other cases, the DM must carefully think through the joint information content to determine which bundle is more valuable and thus have to go through the difficulties associated with information integration and the computational burdens of identifying instrumental value. Or put differently, a choice (comparison) between $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$ can be simplified into a choice (comparison) between σ and σ' when the two sources exhibit a *refine* or *reveal-or-refine* relationship. However, such simplification is not correct and will lead to mistakes in other cases, as weaker relationships, such as *sufficiency* and *Blackwell*, between σ and σ' can not pin down the comparison relationship between bundles.

3 Experimental Design

The goal of the experiment is to study whether and under what circumstances people make optimal choices of information bundles, measure the impact that the challenge of information integration has on information choice, and explore the main forces driving choices of

Blackwell).

¹⁷Figure 17 in Appendix B presents examples of these comparison relationships.

information bundles, including why people make mistakes in these choices.

Subjects in the experiment face three types of decision tasks: (i) *Guessing Task*, eliciting subjects' guesses in the guessing game for all possible information that they might receive from a certain information bundle (i.e., measuring subjects' ability to use an information bundle); (ii) *Assessment Task*, following each Guessing task, eliciting subjects' assessments of the level of guessing accuracy an information bundle induces (i.e., measuring subjects' perceived usefulness of an bundle); (iii) *Information Choice Task*, eliciting subjects' choices between information bundles. Further details of the three types of tasks are described in Section 3.3 below. The Guessing and Assessment tasks study whether subjects make errors in using or evaluating the information content of information bundles. Recent experimental studies suggest that both the failures in evaluating information (e.g., Liang (2023) and GOY) and misuse of information (e.g., Ambuehl & Li (2018) and Guan, Lin, Zhou & Vora (2023)) can drive suboptimal demand for information.

The experiment employs a within-subjects design with two settings that turn on or off the requirement to integrate information from multiple sources (i.e., in order to identify the joint information content of a bundle):

- *Separated*, each information bundle is presented in its original form as a set of two information sources.
- **Joined**, each information bundle is replaced by its corresponding *join* information source.

This variation allows me to isolate the impact of the difficulties associated with information integration on the usage, assessment, and especially choices of information bundles.

3.1 Guessing Game and Visual Representation of Information

The guessing game used in the experiment is as follows: there is a set of twenty objects, including ten triangles and ten circles; one object is randomly drawn, and the subjects' task is to guess the shape of the randomly drawn object; subjects earn a bonus of \$12 if guessing correctly but zero otherwise.

Before making a guess, subjects receive information about the randomly drawn object from an information source or a bundle of sources. Each information source is represented as a *partition* of the twenty objects, i.e., grouping the twenty objects into non-empty subsets, referred to as groups in the experiment. An information source provides subjects with information about which group contains the randomly drawn object. Each group in the partition is thus a distinct signal that the information source might generate. The partition representation makes the characteristics of an information source visually transparent. The number of objects in a group visually shows the probability of the signal being realized; the composition of objects in a group intuitively reveals the posterior probability of the randomly drawn object being a triangle or circle. Note that posteriors inform optimal choices in the guessing game. Knowing posteriors and the probabilities of signal realizations is sufficient to identify the instrumental value (as defined in Equation (1)) of an information source. For example, σ in Figure 1 is a partition with two signals (groups), x and y, each visualized by the combination of a distinct color bar and a letter. With this information source, subjects learn which group (x or y) the randomly drawn object is in before they guess the shape of the randomly drawn object. It is intuitive to identify that the probability of signal x (y) being realized is $\frac{13}{20}$ ($\frac{7}{20}$) and the posterior of the randomly drawn object being triangle conditional on signal x (y) is $\frac{5}{13}$ ($\frac{5}{7}$).

Following Section 2.3, an information source represented in this way can be formally conceptualized as a finite partition of the extended state space $\Omega \times \{1, ..., 20\}$ (Green & Stokey 1978). For example, the information source σ in Figure 1 can be characterized as $\sigma = \{x, y\} = \{(T, \{1, ..., 5\}) \cup (C, \{11, ..., 18\}), (T, \{6, ..., 10\}) \cup (C, \{19, 20\})\}$. The interpretation of this conceptualization is that a random number is drawn uniformly from $\{1, ..., 20\}$ and determines the signal realization conditional on the state. This conceptualization highlights another important benefit of partition representation: it pins down the correlation between information sources and makes identifying the joint information content of multiple information sources straightforward. For instance, $\sigma_0 \vee \sigma$ shown in Figure 1 is the *join* of σ and σ' . Any signal realization from $\sigma_0 \vee \sigma$ is simply the intersection of *s* and *s'*, each being a subset of $\Omega \times \{1, ..., 20\}$, for some *s* from σ and some *s'* from σ' . Specifically, signal $a = (T, \{1, ..., 5\})$ from $\sigma_0 \vee \sigma$ is the intersection of signal $x = (T, \{1, ..., 5\}) \cup (C, \{11, ..., 18\})$ from σ and signal $m = (T, \{1, ..., 8\}) \cup (C, \{19, 20\})$ from σ' , denoted as $a = x \cap m$, and similarly, $b = y \cap m$, $c = y \cap n$ and $d = x \cap n$.

3.2 Information Bundles and Sources Studied in the Experiment

The experiment includes eight different pairs of information bundles, each pair being denoted as $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$. These pairs comprehensively encompass the comparison relationships between individual information sources σ and σ' introduced in Section 2.3, as well as cases in which σ and σ' can not be Blackwell ordered. This design incorporates the taxonomy of comparisons of information bundles characterized by BFK and enables me to test the implied theory of difficulty in comparing and choosing between information bundles.

| Isolated: | Joined: $\sigma_0 \lor \sigma$ vs. $\sigma_0 \lor \sigma'$ Separated: $\{\sigma_0, \sigma\}$ vs. $\{\sigma_0, \sigma'\}$ | |
|----------------------------|---|--------------------------|
| Comparison relationship | Difference in value: 0.05 | Difference in value: 0.1 |
| (1) Refine (R) | > | > |
| (2) Reveal-or-refine (O) | > | > |
| (3) Sufficiency (S) | > | > |
| (4) Blackwell (B) | > | > |
| (5) Not Blackwell (NB) | > | > |
| (6) Not Blackwell (-NB) | < | > |
| (7) - Blackwell (-B) | < | > |
| (8) - Sufficiency $(-S)$ | < | > |

Table 1: Studied Information Bundles and sources

Table 1 summarizes the studied information bundles, the corresponding individual information sources, and *join* information sources into eight cases. In each case, $\{\sigma_0, \sigma\}$ $(\sigma_0 \vee \sigma)$ is more valuable than $\{\sigma_0, \sigma'\}$ $(\sigma_0 \vee \sigma')$ by a 0.1 increment in guessing accuracy (\$1.2 increase in the expected payoff). I also manage to keep σ' the same or use its symmetric version in cases (1)-(6) to make these cases more comparable to each other. The difference in instrumental value between σ and σ' is fixed to be a 0.05 increment in guessing accuracy (\$0.6 increase in the expected payoff), but the sign is flipped in some cases, with > (<) denoting σ being more (less) valuable than σ' . This variation allows me to test whether subjects' information bundle choices might be misled by comparing σ and σ' individually, i.e., whether subjects incorrectly simplify choices between information bundles when the relationship between σ' and σ is weaker than *reveal-or-refine*. All of those information bundles and sources are presented in Figure 17 of Appendix B.

Notes: Each case of (1)-(8) corresponds to a pair of information bundles $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$, consisting of three distinct information sources, and a pair of join information sources corresponding to the bundles. Comparison relationships are introduced in Section 2.3. Value denotes the instrumental value, i.e., the expected improvement in guessing accuracy induced by an information bundle or source as defined in Section 2. > (<) denotes the left information bundle or source has a higher (lower) value than the right one in a comparison. -B(-S) denotes σ' Blackwell dominates (is sufficient for) σ .

There are twenty objects, including 10 triangles and 10 circles. One object will be **randomly** drawn by the computer. You will **earn \$12 if you correctly guess the shape** of the drawn object (triangle or circle).

You will learn which group the drawn object is in under the following information source before you guess its shape.



Figure 2: Screenshot of Guessing and Assessment Tasks in Part 1 Notes: The Assessment task appears below a Guessing task after subjects make all guessing choices and click a "Continue" button. The submitted guesses are shown on the screen but are no longer changeable when subjects work on the Assessment task.

3.3 Stages of the Experiment

The experiment consists of four parts.

Part 1 (Guessing and Assessment under the *Joined* setting, 16 rounds). This part contains 16 Guessing tasks. In each of them, an information source ($\sigma_0 \lor \sigma$ or $\sigma_0 \lor \sigma'$ from one of the eight cases listed in Table 1) as a partition is shown, and a subject submits her guesses about the shape of the randomly drawn object for each possible piece of information (i.e., each possible group containing the randomly drawn object) she might receive from the given information source. This elicits subjects' contingent plans about how to use an information source. Following each Guessing task, subjects are also asked to assess what level of guessing accuracy the information source induces, which reveals subjects' perceptions of the source's actual usefulness. The elicitation is incentivized by the Binarized Scoring Rule (Hossain & Okui 2013) and implemented following the procedure proposed by Wilson & Vespa (2016).

Figure 2 is a screenshot of Part 1. Note that the Assessment task appears right below a Guessing task, after subjects submit their guesses. The Guessing task and subjects' submitted guesses (which are no longer changeable) are shown on the screen when subjects work on the Assessment task.



Figure 3: Screenshot of Information Choice Task in Part 2

Part 2 (Choices between Information sources, 16 rounds). This part includes 16 Information Choice tasks. In each of them, subjects choose between two distinct information sources, i.e., $\sigma_0 \vee \sigma$ versus $\sigma_0 \vee \sigma$ (the *Joined* setting) or σ versus σ' (referred to as the *Isolated* setting afterward) from one of the eight cases. Figure 3 presents a screenshot of the task. To incentivize choices, subjects will be given their chosen information sources in a (potential) final Guessing task at the end of the experiment.

Part 3 (Guessing and Assessment under the *Separated* setting, 16 rounds). Subjects complete another 16 Guessing tasks and 16 follow-up Assessment tasks. The tasks are the same as those in Part 1 except that subjects now face information bundles, $\{\sigma_0, \sigma\}$ or $\{\sigma_0, \sigma'\}$ from the eight cases, instead of the *join* information sources, $\sigma_0 \vee \sigma$ or $\sigma_0 \vee \sigma'$. Figure 4 is a screenshot of the Guessing and Assessment tasks in Part 3.

Part 4 (Choices between Information Bundles, 8 rounds). This part consists of 8 binary choices between information bundles, i.e., $\{\sigma_0, \sigma\}$ versus $\{\sigma_0, \sigma'\}$. Each corresponds to one of the eight cases. Figure 5 is a screenshot of the task.

The four parts of the experiment are arrayed in ascending order of difficulty. Subjects start with relatively easy decision problems in Parts 1 and 2, become familiar with the three types of tasks and experiment interfaces, and then face relatively challenging problems in Parts 3 and 4. Having Guessing and Assessment tasks before Information Choice tasks also helps to mitigate the potential influence of failures in contingent thinking (Esponda & Vespa 2014, Martinez-Marquina, Niederle & Vespa 2019), i.e., subjects failing to foresee how they will use the information when making information choices. Additionally, the order of tasks within each part is randomized for each subject. In each Information choice task, the position of the two options (i.e., two bundles or sources) is also randomized.

There are twenty objects, including 10 triangles and 10 circles. One object will be **randomly** drawn by the computer. You will **earn \$12 if you correctly guess the shape** of the drawn object (triangle or circle).

Below is a pair of information sources (I and II). You will learn which group the drawn object is in under each of them before you guess its shape.



Figure 4: Screenshot of Guessing and Assessment Tasks in Part 3

3.4 Incentives and Implementation Details

The experiment was conducted at the LITE laboratory at the University of California, Santa Barbara, in June 2023. 100 subjects were recruited to participate in 7 sessions using the ORSEE recruitment system (Greiner 2015). The experiment used the software programmed by the author in oTree (Chen, Schonger & Wickens 2016). Between 7 and 20 subjects participated in each session, which lasted 90 minutes.

All subjects received a show-up fee of \$8. The experiment instructions contain six comprehension questions, and subjects got \$0.2 for each question they answered correctly in one attempt.¹⁸ Subjects' earnings from the experiment were determined according to a randomly selected round. For a subject, if one of the rounds in Parts 1 or 3 was selected, the subject's submitted guesses in that round were used to determine whether she received a \$12 reward from the Guessing task, and her answer in the follow-up Assessment task was used to determine whether she received another \$5 reward. If one of the Information Choice tasks

 $^{^{18}}$ The comprehension questions can be found in the experiment instructions in Appendix C.



Figure 5: Screenshot of Information Choice Task in Part 4

in Parts 2 or 4 was selected, the subject completed a final Guessing task given her chosen information bundle or source in that selected Information Choice task. Her guesses in the final task determined whether she received a \$12 reward. The average (median) final payoff is around \$22 (\$21).

4 Results

The main findings of the experiment are organized as follows. Section 4.1 analyzes and compares choices of information under the Separated and Joined settings. Section 4.2 looks at subjects' usage and assessment of the actual usefulness (conditional on the usage) of information bundles and corresponding *join* information sources. The section also examines whether mistakes in the choices of information bundles can be attributed to errors or noise in the usage of bundles. Section 4.3 then investigates whether the mistakes are instead systematic, driven by a simple but imperfect heuristic in information bundle choices. Section 4.4 explores the heterogeneity in these results among subjects.

4.1 Choices of Information Bundles and Join Information Sources

I begin by looking at the optimality of subjects' choices between information bundles, as measured by their likelihood of choosing the more instrumentally valuable bundle (i.e., the high-value bundle) over the other in binary choices, and to what extent the optimality is constrained by the challenge of information integration. The left panel of Figure 6 shows that



Figure 6: Choices of Information Bundles and Join information Sources Notes: The optimality of information choices is measured by the likelihood of choosing the high-value information bundle or source (relative to the other) in binary choices. Short vertical lines in the left panel denote 95 percent confidence intervals. In the right panel, the y-axis (x-axis) plots the likelihood of choosing bundle (join information source) $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma\}$ ($\sigma_0 \lor \sigma$ over $\sigma_0 \lor \sigma'$) in Information Choice tasks under the Separated (Joined) setting.

in binary choices between information bundles, subjects choose the high-value bundles only 56 percent of the time. The mistakes of failing to choose the high-value information turn out to be largely driven by subjects' failures to integrate information sources within a bundle and thereby identify their joint information content. Under the Joined setting, in which there is no need for information integration, subjects' likelihood of choosing the high-value information increases considerably, to over 77 percent (signed-rank test, p < 0.001).

Given my experimental design, the optimal decisions in the information choice tasks (i.e., binary choices) under the Separated and Joined settings are theoretically the same. If subjects are able to integrate sources and identify the joint information content of each bundle, then their choices under the Separated setting ought to align with their choices under the Joined setting. The right panel of Figure 6 presents a direct comparison of choices under the two settings. In the graph, each data point represents one case (eight cases in total as summarized in Table 1), the y-axis plots subjects' likelihood of choosing the bundle $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$, and the x-axis plots their likelihood of choosing the *join* information source $\sigma_0 \vee \sigma$ over $\sigma_0 \vee \sigma'$. Choices between *join* information sources poorly explain choices between (theoretically equivalent) information bundles and the two dimensions of likelihoods are barely correlated (Kendall's $\tau = 0.148$, p = 0.62), suggesting subjects largely fail to integrate sources and do not base their choices between information bundles on the joint information content of each bundle. Moreover, examining choices under the Separated and Joined settings subject by subject, I find that 69% (85%) of subjects have a strictly (weakly) higher likelihood of choosing the high-value information in binary choices under the Joined setting than under the Separated setting.

Result 1. Subjects' choices between information bundles are largely suboptimal and substantially deviate from their choices between theoretically equivalent join information sources.

Section 2.3 argues that BFK's characterization of comparison relationships between information sources suggests a theory of difficulty in comparisons of information bundles. When σ and σ' exhibit a refine (R) or reveal-or-refine (O) relationship, identifying which bundle, $\{\sigma_0, \sigma\}$ or $\{\sigma_0, \sigma'\}$ (for any σ_0), is more valuable does not necessarily require the DM to integrate sources and recognize the joint information content (the joint instrumental value) of each bundle. Otherwise, the DM has to carefully think about the joint information content and engage in the difficult task of information integration. A testable hypothesis related to this theory is that subjects' choices between information bundles should be less constrained by the challenge of information integration in cases in which σ and σ' exhibit a refine or reveal-or-refine relationship compared to other cases.

To test this, I focus on the difference in the optimality of information choices between the Joined and Separated settings and study how the difference changes across cases that vary in the comparison relationship between σ and σ' (eight cases in total as summarized in Table 1). Figure 7 depicts these differences. The left panel covers all data while the right panel focuses on the subjects for whom the more instrumentally valuable information bundle or *join* information source of each case is indeed more helpful in Guessing tasks (i.e., practically induces a weakly higher guessing accuracy). Note that these subjects have a relatively clear incentive to choose the high-value information bundle or *join* information source. In both panels, the x-axis denotes the eight different cases, and the y-axis plots the difference between the likelihood of choosing the high-value *join* information source under the Joined setting and the likelihood of choosing the high-value bundle under the Separated setting of each case. When $\sigma R \sigma'$ or $\sigma O \sigma'$ holds, the decrease in choice optimality is relatively small. The decreases under the two cases are the lowest if focusing on subjects with a clear incentive to choose high-value information, as the right panel shows. I take these as suggestive evidence that supports the theory of difficulty in comparisons of information bundles implied by BFK.

The figure reveals another noticeable pattern: the decrease in choice optimality is much smaller in cases in which the value comparison between σ and σ' is ordinally consistent with the comparison between bundles $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$ than in cases where the two value



Figure 7: Decrease in the Optimality of Information Choices from Joined to Separated Settings Notes: The left panel covers all data while, in each case, the right panel focuses on the subjects for whom the more instrumentally valuable information bundle or join source is indeed more helpful in Guessing tasks. Each data point plots the difference between the likelihood of choosing the high-value join information source under the Joined setting and the likelihood of choosing the high-value bundle under the Separated of a case. Eight different cases are introduced in Table 1. On the x-axis, the cases are ordered regarding the strength of the comparison relationship between σ and σ' . R denotes refine, O denotes reveal-or-refine, S denotes sufficiency, B denotes Blackwell, and NB denotes that two sources can not be Blackwell ordered. Detailed descriptions of these comparison relationships are in Section 2.3. Short vertical lines denote 95 percent confidence intervals by Bootstrapping.

comparisons go to opposite directions. This can be seen in either panel when comparing the first five cases with cases -NB, -B, and -S. I will show in later sections that this pattern is an important clue to the primary mechanism driving subjects' information bundle choices.

Result 2. The information choices are less optimal under the Separated setting compared to the Joined setting in every case. However, the decrease in choice optimality is relatively smaller, meaning subjects are less constrained by the challenge of information integration, when $\sigma R\sigma'$ or $\sigma O\sigma'$ holds. This supports the theory of difficulty in comparing information bundles implied by BFK.

4.2 Usage and Assessment of Information and Choice

The above results show that subjects often fail to make optimal choices of information bundles and the mistakes are largely due to the challenge of information integration. But how does the challenge of information integration induce mistakes in choices? One possibility is that the difficulties associated with information integration cause errors or noise in the usage of information bundles (ex-post), leading to mistakes in information bundle choices (ex-ante). In this section, I examine whether this channel is the main source of mistakes.



Figure 8: Guessing Optimality Notes: The optimality of guesses is measured by the share of submitted guesses that are optimal from the Bayesian perspective. The decrease in optimality is the difference in guessing optimality between the Joined and Separated settings.

The optimality of subjects' usage of information can be measured by the rate at which guesses about the shape of a randomly drawn object, conditional on receiving the information, are consistent with the Bayesian predictions. The left panel of Figure 8 presents the distribution of the subject-level optimality rates of guesses. Under the Joined setting, where subjects face a *join* information source in each Guessing task, 76 out of 100 subjects always make optimal (from the Bayesian perspective) guesses, and the average optimality rate is 98 percent. This near-perfect guessing behavior confirms that the partition representation of information sources removes typical errors (such as failures in Bayesian reasoning) people might make in using (single pieces of) information. In contrast, under the Separated setting, where subjects face an information bundle and have to integrate a pair of signals by themselves, the average optimality rate decreases to 85 percent, and only 32 subjects make optimal guesses all of the time. On the one hand, the guessing optimality under the Separated setting is still impressive, suggesting subjects are highly sensitive to joint information content when using a bundle of information sources. On the other hand, the reduction in guessing optimality due to the challenge of information integration is considerable (signed rank test, p < 0.001). Information integration seems to be challenging for most of the subjects. The right panel of Figure 8 presents the distribution of the subject-level decrease in guessing optimality rate from the Joined to the Separated settings. 66 (90) subjects have strictly (weakly) lower optimality rates when they have to integrate two pieces of information by themselves in Guessing tasks under the Separated setting.



Figure 9: Example of Information Bundle

I also explore what guessing errors subjects typically make in the presence of the challenge of information integration. The scenario in which the largest proportion of subjects guess suboptimally is when they learn groups b and q of the information bundle shown in Figure 9 contain the randomly drawn object. The Bayesian optimal guess is *Triangle*, but 54 subjects guessed *Circle*. This guessing error can be explained by the subjects integrating signals (groups) in a simple but incorrect way: count the total numbers of triangles and circles, respectively, that the two groups contain, then guess *Triangle* if the total number of triangles is higher and guess *Circle* otherwise. In the mentioned scenario, the decision rule predicts guessing *Circle* because groups b and q together contain more circles than triangles, i.e., 11 circles versus 7 triangles. Strikingly, this incorrect way of integrating signals can explain around 82 percent (770/942) of errors in the Guessing tasks under the Separated setting.^{19,20}

Result 3. The challenge of information integration leads to more errors in the usage of information: guesses are optimal 98% of the time when using join information sources; the optimality rate significantly decreases to 85% when using theoretically equivalent information bundles; and most (about 82%) of the guessing errors in the latter case can be attributed to an incorrect but simple way of integrating signals.

²⁰The decision rule is also highly correlated (though may not be reduced to) several documented rules of signal integration in the literature: (i) *correlation neglect*, perceiving the two signals to be independent and using the two signals separately to update beliefs; (ii) *DeGroot rule*, take a simple average of the posterior beliefs induced by two signals; (iii) *Not-To-Integrate*, focusing on only one signal (the more revealing one) but not the join of signals. These three alternative rules generate the same predictions of guesses given the studied information bundles in the experiment. These predictions deviate from the aforementioned decision rule in only 5 out of 63 scenarios and can explain 71% of guessing errors (67% if excluding one scenario in which the three rules give uniform predictions).

¹⁹Possible interpretation of the decision rule is that people do not cross-check information but simply pool information together and then make judgments based on the "quantity" comparison of "for" and "against" clues without thinking about the actual implication of the combination of multiple pieces of information.



Figure 10: Subjective Optimality of Information Choices Notes: In the left (right) panel, the choice optimality is measured by the likelihood of choosing the information bundle or source with a weakly higher value given guesses (assessments) in binary choices. Short vertical lines denote 95 percent confidence intervals.

Can subjects' choices between information bundles be explained by their (imperfect) usage of the bundles? To understand this, I compute the "practical" instrumental value of each information bundle conditional on how the bundle is used (i.e., conditional on subjects' submitted guesses in the Guessing task with the bundle), which I refer to as *value given guesses*, and examine whether it can explain choices between information bundles. As shown in the left panel of Figure 10, overall, subjects choose the bundle with a weakly higher value given guesses in binary choices only 62% of the time (significantly lower than the rate of 78% in choices between *join* information sources, p < 0.001). In addition, I compare the indicated likelihood of choosing one bundle over the other based on value given guesses with the actually observed choice likelihood across the eight binary choices between information bundles (Figure 15 in Appendix A depicts the comparison). I find that the two likelihoods are barely correlated (Kendall's $\tau = 0.296$, p = 0.31). These findings suggest that subjects do not take adequate account of their future usage of information bundles when they make choices.

It is also possible that subjects do think about their future usage of information bundles but in a noisy way. The Assessment task in the experiment directly elicits subjects' assessments of the practical usefulness of each information bundle (and corresponding *join* information source). Do the elicited assessments explain choices between information bundles? Results suggest that this is not the case, either. The right panel of Figure 10 shows that overall, subjects choose the bundle to which they assign a weakly higher assessment only 61% of the time (significantly lower than the rate of 72% in choices between *join* information sources, p < 0.001). Across the eight binary choices, the indicated likelihood of choosing a bundle over the other based on assessments barely correlates (Kendall's $\tau = 0.255$, p = 0.38) with the actually observed choice likelihood. These results once again indicate that subjects make choices between information bundles without much consideration for how they would use the bundles to make inferences.

Result 4. Subjects make information bundle choices without much consideration for how they would use the bundles, and therefore, mistakes in those choices cannot be primarily attributed to errors or noise in the usage of information bundles.

4.3 Common Source Cancellation in Information Bundle Choices

The previous section shows that subjects' choices between information bundles are only weakly related to their ability to use the bundles. This suggests that the mistakes subjects make in choosing between information bundles are likely driven by the use of a decision rule other than the optimal one – one that does not attempt to fully integrate the information contained in the bundles. The analysis in Section 4.1 shows that failures of integration (i.e. the difference between the optimality of information choice in the Joined vs. Separated settings) are much more severe when the bundle that contains a more valuable source (considered in isolation) is not the more valuable bundle. This finding indicates that subjects' choices between information bundles are sensitive to the direct comparison of the information sources the two bundles being compared do not share. This suggests a hypothesis: when choosing between information bundles { σ_0, σ } and { σ_0, σ' }, subjects might heuristically simplify their decision-making by "canceling" σ_0 and reducing a choice between bundles to a choice between individual sources σ and σ' . This simplifying heuristic, which I call *common source cancellation (CSC)*, is very intuitive and appealing as it circumvents the difficult task of integrating information and identifying the joint information content of each bundle.

Figure 11 provides evidence supporting that subjects follow the CSC heuristic. The left panel of the figure looks into the optimality of information bundle choices in two scenarios: (i) where σ is less valuable than σ' but $\{\sigma_0, \sigma\}$ is more valuable than $\{\sigma_0, \sigma'\}$ (cases (6)-(8) listed in Table 1), categorized as Individually Worse; and (ii) where the value comparison between σ and σ' aligns with the comparison between the two corresponding bundles (cases (1)-(5) in Table 1), categorized as Individually Better. In the first scenario, subjects make optimal information bundle choices only 45 percent of the time. In contrast, the optimality



Figure 11: Common Source Cancellation in Information Bundle Choices Notes: The left panel plots the likelihood of choosing bundle $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$, with the former being more instrumentally valuable than the latter by a 0.1 increment in guessing accuracy (i.e., \$1.2 increase in the expected payoff). "Individually Worse" refers to cases (6)-(8) listed in Table 1 and "Individually Better" refers to cases (1)-(5). Short vertical lines denote 95 percent confidence intervals by Bootstrapping. In the right panel, the y-axis plots the likelihood of choosing bundle $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$ under the Separated setting, and the x-axis plots the likelihood of choosing σ over σ' under the Isolated setting. Red dashed lines are the best linear fits, and the grey regions are 95 percent confidence intervals for predictions of the linear fits.

rate increases substantially to 63 percent in the second scenario. This pattern confirms that subjects are influenced by the direct comparison between σ and σ' when choosing between two corresponding bundles. The right panel of Figure 11 then directly compares subjects' choices between isolated information sources σ and σ' and their choices between corresponding bundles { σ_0, σ } and { σ_0, σ' }. As the graph shows, the likelihood of choosing σ over σ' strongly explains the likelihood of choosing { σ_0, σ } over { σ_0, σ' } across the eight cases and overall, the two likelihoods are highly correlated (Kendall's $\tau = 0.764, p < 0.01$). Moreover, when focusing on the suboptimal choices between information bundles, I find that subjects' choices between σ and σ' can account for over 68 percent of the mistakes in information bundle choices. These findings, aligning with the *CSC* heuristic, strongly suggest that when choosing between information bundles, subjects tend to focus solely on the comparison between σ and σ' without thinking about the joint information content of each bundle.

Table 2 offers additional statistical evidence for these findings. Regression model (1) in the table regresses the choice of bundle $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$ on the difference in instrumental value (measured with respect to guessing accuracy) between the two corresponding isolated information sources σ and σ' , with the constant term capturing the difference in instrumental value (i.e., 0.1 increment in guessing accuracy) between the two bundles. Results show that subjects' choices between information bundles strongly respond to the value comparison of the two isolated sources but barely respond to the value comparison of the two bundles. Regression model (2) additionally includes the difference in assessments and the difference in value given guesses between two bundles as independent variables, both being measured with respect to guessing accuracy as well. Subjects' choices also seem to be (slightly) responsive to subjective assessments and value given guesses of bundles. However, the effect size of both is much smaller than that of the value comparison of the two corresponding isolated sources, suggesting the CSC heuristic is the primary driver of information bundle choices.

| | Logit Regression | | |
|---|--------------------------|--|--|
| | (choose $\{\sigma_0,$ | σ } over { σ_0, σ' }) | |
| | (1) | (2) | |
| Difference in Value (Isolated, σ vs. σ') | 7.158^{***} (1.634) | 7.116^{***} (1.671) | |
| Difference in Assessment | | 1.663^{*} (0.871) | |
| Difference in Value Given Guesses | | 1.998^{**} (0.877) | |
| Constant | 0.157^{*} (0.092) | $0.065 \\ (0.085)$ | |
| No. of subjects | 100 | 100 | |
| Ν | 800 | 800 | |

Table 2: Choices Between Information Bundles

Notes: Logit regressions with the dependent variable being whether bundle $\{\sigma_0, \sigma\}$ is chosen in a binary choice. The difference in (theoretical) instrumental value between two bundles is always 0.1 increment in guessing accuracy and is captured by the constant term. Assessment refers to the elicited assessment of the instrumental value of an information bundle. Value given guesses denotes the "empirical" instrumental value of an information bundle accounting for how the bundle is used. Value, assessment, and value given guesses are all measured regarding guessing accuracy. For the Optimal group, value given guesses equals the theoretical instrumental value. Therefore, the difference in value given guesses is always 0.1 between a pair of bundles, making its coefficient to be 0. Clustered standard errors in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

Result 5. Subjects primarily follow the common source cancellation (CSC) heuristic when choosing information bundles. The choices between two isolated information sources σ and σ' strongly explain choices between two corresponding bundles { σ_0, σ } and { σ_0, σ' } and can account for most of the mistakes in the latter.

4.4 Heterogeneity

The results so far establish that, at the aggregate level, the *CSC* heuristic is the primary driving force behind subjects' choices between information bundles and can account for most of the mistakes in choices. Is this true for all subjects? Do subjects who seem to understand the joint information content of bundles (i.e., being able to interpret and use information bundles in an optimal way) still follow this heuristic? More broadly, is the tendency of *common source cancellation* associated with subjects' ability to integrate information (which is necessary for making optimal choices between information bundles)? Answering these questions will help us to understand the significance and prevalence of the *CSC* heuristic in the context of choosing information bundles and shed light on the determinant of the heuristic.

I examine heterogeneity by classifying subjects into three groups with respect to how well they can make use of information bundles (i.e., a proxy of the ability to properly integrate information): (i) Naive, subjects follow exactly the incorrect way of integrating signals as discussed in Section 4.2 (10 subjects) or worse (i.e., those with a guessing optimality rate lower than 0.746) in the Guessing tasks under the Separated setting; (ii) In-Between, subjects make better use of information bundles than the Naive group but are not fully optimally; (iii) Optimal, subjects make perfect use of information bundles. The three groups include 37, 31, and 32 subjects, repectively. Table 5 in Appendix A compares the three groups in terms of the optimality of their usage, assessment, and choices of information under both the Joined and Separated settings. The optimality rates of the Optimal group are always the highest, and the rates of the Naive group are almost always the lowest. The Optimal group also has the lowest decreases in optimality rates from the Joined setting to the Separated setting, indicating this group of subjects is less constrained by the difficulties associated with information integration relative to other groups.

Figure 12 studies whether and to what extent each group follows the *CSC* heuristic when choosing between information bundles. The figure replicates the right panel of Figure 11 with the data from each group of subjects separately. As the figure shows, choices between isolated information sources strongly explain the choices between information bundles in each group. Even for the *Optimal* group, who use information bundles optimally 100% of the time, their choices of information under the Isolated and Separated settings are qualitatively aligned. Moreover, choices between isolated information sources can account for 72 percent, 65 percent, and 67 percent of suboptimal choices between bundles of the three groups, respectively.



Figure 12: Common Source Cancellation in Information Bundle Choices – By Group Notes: In each panel, the y-axis plots the likelihood of choosing bundle $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$ under the Separated setting, and the x-axis plots the likelihood of choosing σ over σ' under the Isolated setting. Naive, In-Between, and Optimal denote three groups of subjects that are classified according to their guessing optimality under the Separated setting (details can be found above). Red dashed lines are the best linear fits, and the grey regions are 95 percent confidence intervals for predictions of the linear fits.

These results suggest that the CSC heuristic plays a vital role in explaining choices between information bundles of each group. Figure 12 also indicates that the tendency of *common source cancellation* is stronger among subjects who make worse use of information bundles. The heuristic near-perfectly explains the choices between information bundles of the *Naive* group, while its influence is relatively weaker (though still considerable) among the other two groups. This indicates people are more likely to follow the CSC heuristic if they are less able to integrate information and interpret and use the joint information content correctly.

Table 3 replicates regression (2) in Table 2 with the data of each group separately. Regression results show that choices between information bundles of each group are significantly responsive to the value comparison of the two corresponding isolated information sources, confirming that each group has the tendency of *common source cancellation* when choosing between information bundles. The effect size is the largest for the *Naive* group and becomes relatively smaller for the other two groups. In addition, the regression analysis reveals that the choices of the *In-Between* group are also significantly responsive to subjective assessments of information bundles, though the effect size is substantially smaller than that of the difference in instrumental value between two isolated sources. The choices of the *Optimal* group are also strongly responsive to the difference in instrumental value of two bundles, suggesting this group of subjects is sensitive to the joint information content of each bundle when making binary choices. Figure 16 in Appendix A further shows that combining *CSC* with the mechanism of following subjective assessments explains the bundle choices of the *In-Between* group quantitatively well, and combining *CSC* with the mechanism of basing information bundle choices on the joint information content of each bundle explains the choices of the *Optimal* group quantitatively well.

| | Logit Regression (choose $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$) | | |
|---|---|---|---|
| | Naive | In-Between | Optimal |
| Difference in Value (Isolated, σ vs. σ') | 8.200^{***} (2.842) | 7.758^{**} (3.246) | 6.227^{**} (3.049) |
| Difference in Assessment | -1.244 (1.126) | $\frac{4.192^{***}}{(1.603)}$ | $2.558 \\ (1.923)$ |
| Difference in Value Given Guesses | $1.024 \\ (0.912)$ | $\begin{array}{c} 0.723 \\ (1.431) \end{array}$ | 0.000 (.) |
| Constant | -0.140 (0.097) | -0.144 (0.149) | $\begin{array}{c} 0.632^{***} \\ (0.226) \end{array}$ |
| No. of subjects | 37 | 31 | 32 |
| Ν | 296 | 248 | 256 |

Table 3: Choices Between Information Bundles – By Group

Notes: Logit regressions with the dependent variable being whether bundle $\{\sigma_0, \sigma\}$ is chosen in a binary choice. The difference in (theoretical) instrumental value between two bundles is always 0.1 increment in guessing accuracy and is captured by the constant term. Assessment refers to the elicited assessment of the instrumental value of an information bundle. Value given guesses denotes the "empirical" instrumental value of an information bundle accounting for how the bundle is used. Value, assessment, and value given guesses are all measured regarding guessing accuracy. For the Optimal group, value given guesses equals the theoretical instrumental value. Therefore, the difference in value given guesses is always 0.1 between a pair of bundles, making its coefficient to be 0. Clustered standard errors in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

Result 6. There is heterogeneity in the ability to integrate information among subjects. But the common source cancellation heuristic emerges as a primary driver of the choices between information bundles of each group of subjects, including those who make perfect use of each information bundle in the guessing game.

5 Discussion

Why Common Source Cancellation?

What are the reasons behind the emergence of the *common source cancellation* (CSC) heuristic? First, following the heuristic in information bundle choices may be due to subjects approaching the choice problem in a wrong way from the beginning. For instance, they believe that the individually better source always constitutes a better bundle and think that

the common component σ_0 can be canceled out when comparing two bundles $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$. Second, it is also possible that subjects know that the heuristic is not the optimal approach but still choose to use it as a way out of the difficulties associated with information integration and to save cognitive efforts.

While examining or distinguishing the two possible reasons is beyond the scope of the current experiment, there is suggestive evidence that both might be at play. The finding that the impact of the CSC heuristic is more pronounced among subjects who struggle with information integration and the effective use of information bundles (arguably more likely to approach the choice problem incorrectly or have a more limited ability to approach the problem) seem to align with the first explanation. On the other hand, the finding that the *Optimal* group, who make perfect use of each bundle and demonstrate sensitivity to the joint information content of bundles, also largely follow the CSC heuristic supports the second explanation. However, it should be noted that these arguments are only suggestive but not conclusive.

Other Determinants of Information Choices

A growing literature shows many factors other than instrumental value may influence information choices (see Nielsen (2020) or GOY for a review). The most related to the current paper is GOY, which finds that the demand for single information sources is influenced by *informativeness*, the fundamental characteristic of information sources, in addition to being responsive to instrumental value.²¹ Aligning with GOY, subjects' information choices in the current experiment also exhibit a sharp aversion to non-instrumental informativeness.

The left panel of Figure 13 presents the likelihood of choosing the high-value information source in binary choices. Under both the Isolated and Joined settings, on average, high-value sources are more likely to be chosen (i.e., the likelihoods are significantly larger than 0.5). Besides, the likelihood significantly increases (signed rank test, *p*-value < 0.001) as the value difference between a pair of information sources increases from 0.05 increment in guessing accuracy under the Isolated setting to 0.1 under the Joined setting. The right panel of Figure 13 examines the impact of excess informativeness on the choice of individual information sources. Each data point represents a binary choice, and the y-axis plots the likelihood of choosing the high-value source in each binary choice. Grey and blue dots denote the data of Isolated and Joined settings, respectively, and the dashed lines are the best linear

²¹Informativness is measured by the mutual information (Shannon 1948) between prior and posterior beliefs induced by given information (Cabrales, Gossner & Serrano 2013).



Figure 13: Choices between single Information sources Notes: In both panels, choice likelihood denotes the likelihood of choosing the high-value information source in binary choices. Short vertical lines in the left panel denote 95 percent confidence intervals. Δv denotes the difference in instrumental value between a pair of information sources. Informativeness is measured by the mutual information between prior and posterior beliefs that the certain information source induces. Dashed lines in the right panel are the best linear fits. Data of Isolated and Joined settings are distinguished by color, grey versus blue.

fits. The graph shows that subjects are averse to non-instrumental informativeness: as the high-value information source becomes more informative (relative to the low-value source in the binary choice), the likelihood of choosing it decreases.

The above results are confirmed by regression analyses shown in Table 4. With whether to choose the high-value information source as the dependent variable, the difference in informativeness between a pair of sources is included as the independent variable, and the constant term captures the effect of the difference in instrumental value (being 0.05 increment in guessing accuracy under Isolated and 0.1 under Joined). The difference in informativeness has a significantly negative impact under either setting, suggesting subjects are averse to informativeness. The constant term is significantly positive and substantially larger under the Joined setting than under the Isolated setting, reflecting that subjects are responsive to instrumental value when choosing single sources.

Additionally, in line with the *common source cancellation* heuristic, subjects' choices between information bundles are not influenced by the difference in value or informativeness between bundles. Instead, those choices are significantly responsive to the differences in value and informativeness between the corresponding isolated sources contained in the bundles. This responsiveness is also similar to that in choices under the Isolated setting.

| | Isolated | Joined | Separated | |
|--|---------------------------------|---|--|------------------------------|
| | $(\sigma \text{ vs. } \sigma')$ | $(\sigma_0 \lor \sigma \text{ vs. } \sigma_0 \lor \sigma')$ | $(\{\sigma_0,\sigma\})$ | vs. $\{\sigma_0, \sigma'\})$ |
| Diff in Informativeness | -4.869^{***} (0.737) | -4.265^{***} (1.173) | -0.002 (1.034) | 1.185 (1.588) |
| Diff in Value (σ vs. σ') | | | | $18.662^{***} \\ (3.904)$ |
| Diff in Informativeness (σ vs. σ') | | | | -3.473^{***} (1.179) |
| Constant | $\frac{1.528^{***}}{(0.163)}$ | $2.385^{***} \\ (0.374)$ | $\begin{array}{c} 0.242\\ (0.282) \end{array}$ | -0.083 (0.414) |
| No. of Subjects | 100 | 100 | 100 | 100 |
| Ν | 800 | 800 | 800 | 800 |

Table 4: Informativeness Aversion in Information Choices

Notes: Logit regressions with the dependent variable being whether to choose the high-value information bundle or source in a binary choice. Under Isolated, the difference in instrumental value between a pair of information sources is always a 0.05 increment in guessing accuracy; the difference is always 0.1 under Joined and Separated. Informativeness is the mutual information between prior and posterior beliefs that a certain information bundle or source induces. Clustered standard errors in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

6 Conclusion

This paper investigates experimentally how people choose information bundles (i.e., sets of information sources), whether and under what circumstances they make mistakes, and where those mistakes mainly come from. The study shows that subjects often fail to choose the more instrumentally valuable bundle because of difficulties in integrating sources within a bundle to identify their joint information content. Mistakes in information bundle choices are systematic and can be primarily attributed to subjects following an intuitive bit imperfect heuristic I call *common source cancellation (CSC)*. This heuristic causes subjects to fail to consider the joint information content of each bundle and to mistakenly reduce a choice between bundles to a choice between the non-shared information sources in the two bundles. A heterogeneity analysis reveals the wide prevalence of this heuristic among subjects and shows that those with a more limited ability to integrate information tend to rely more heavily on the heuristic in information bundle choices. Given that information integration is likely to be more challenging (and thus people are probably less able to do it) in real-world settings than in the simplified setting of my experiment, it is plausible that the heuristic exerts an even more pronounced influence in many real-world contexts.

This study has several implications. The results suggest that information integration is challenging and leads to errors in information usage and choice (even in a simplified experimental setting). To facilitate people taking up valuable information and using it to improve decision making, information should better not be provided in a disaggregated way whenever possible. Besides, the prevalence of the *common source cancellation* heuristic highlights that people tend to compare information sources in isolation without considering their joint information content with other available sources. Influenced by the heuristic, people are unlikely to diversify their information choices and consumption as they should. This calls for interventions aimed at directing individuals to think about the joint information content of multiple sources and enhancing their ability to integrate information.

References

- Ambuehl, S. & Li, S. (2018), 'Belief updating and the demand for information', Games and Economic Behavior 109, 21–39.
- Arieli, I., Babichenko, Y. & Smorodinsky, R. (2018), 'Robust forecast aggregation', Proceedings of the National Academy of Sciences 115(52), E12135–E12143.
- Barberis, N., Huang, M. & Thaler, R. H. (2006), 'Individual preferences, monetary gambles, and stock market participation: A case for narrow framing', *American Economic Review* 96(4), 1069–1090.
- Blackwell, D. (1951), Comparison of experiments, in 'Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability', University of California Press, Berkeley, California, pp. 93–102.
- Blackwell, D. (1953), 'Equivalent comparisons of experiments', The Annals of Mathematical Statistics 24(2), 265–272.
- Börgers, T., Hernando-Veciana, A. & Krähmer, D. (2013), 'When are signals complements or substitutes?', Journal of Economic Theory 148(1), 165–195.
- Brooks, B., Frankel, A. & Kamenica, E. (2023), Comparisons of signals, working paper.
- Cabrales, A., Gossner, O. & Serrano, R. (2013), 'Entropy and the value of information for investors', *American Economic Review* **103**(1), 360–77.
- Calford, E. & Chakraborty, A. (2023), The value of and demand for diverse news sources, Working Papers 355, University of California, Davis, Department of Economics.
- Charness, G., Oprea, R. & Yuksel, S. (2021), 'How do people choose between biased information sources? evidence from a laboratory experiment', *Journal of the European Economic* Association 19(3), 1656–1691.
- Chen, D. L., Schonger, M. & Wickens, C. (2016), 'otree an open-source platform for laboratory, online, and field experiments', *Journal of Behavioral and Experimental Finance* 9, 88–97.
- De Oliveira, H., Ishii, Y. & Lin, X. (2021), Robust merging of information, working paper.
- Enke, B. & Zimmermann, F. (2019), 'Correlation neglect in belief formation', *The Review* of Economic Studies 86(1), 313–332.

- Esponda, I. & Vespa, E. (2014), 'Hypothetical thinking and information extraction in the laboratory', American Economic Journal: Microeconomics 6(4), 180–202.
- Eyster, E. & Weizsäcker, G. (2011), Correlation neglect in financial decision-making, Discussion Papers of DIW Berlin 1104, DIW Berlin, German Institute for Economic Research.
- Fedyk, A. & Hodson, J. (2023), 'When can the market identify old news?', Journal of Financial Economics 149(1), 92–113.
- Gabaix, X. (2014), 'A sparsity-based model of bounded rationality', *The Quarterly Journal* of *Economics* **129**(4), 1661–1710.
- Gentzkow, M. & Kamenica, E. (2017*a*), 'Bayesian persuasion with multiple senders and rich signal spaces', *Games and Economic Behavior* **104**, 411–429.
- Gentzkow, M. & Kamenica, E. (2017b), 'Competition in persuasion', The Review of Economic Studies 84(1), 300–322.
- Green, J. R. & Stokey, N. L. (1978), Two representations of information structures and their comparisons, working paper.
- Greiner, B. (2015), 'Subject pool recruitment procedures: organizing experiments with orsee', Journal of the Economic Science Association 1(1), 114–125.
- Guan, M., Lin, C., Zhou, J. & Vora, R. (2023), Preference for sample features and belief updating, unpublished.
- Guan, M., Oprea, R. & Yuksel, S. (2023), Too much information, working paper.
- Hossain, T. & Okui, R. (2013), 'The binarized scoring rule', *The Review of Economic Studies* **80**(3 (284)), 984–1001.
- Hossain, T. & Okui, R. (2021), Belief formation under signal correlation, working paper.
- Kahneman, D. (2011), *Thinking Fast and Slow*, 1 edn, New York: Farrar Straus and Giroux.
- Kahneman, D. & Lovallo, D. (1993), 'Timid choices and bold forecasts: A cognitive perspective on risk taking', *Management Science* **39**(1), 17–31.
- Levy, G. & Razin, R. (2021), 'A maximum likelihood approach to combining forecasts', *Theoretical Economics* **16**(1), 49–71.
- Levy, G. & Razin, R. (2022), 'Combining forecasts in the presence of ambiguity over correlation structures', *Journal of Economic Theory* 199, 105075. Symposium Issue on Ambiguity, Robustness, and Model Uncertainty.

- Liang, A. & Mu, X. (2020), 'Complementary Information and Learning Traps', The Quarterly Journal of Economics 135(1), 389–448.
- Liang, A., Mu, X. & Syrgkanis, V. (2022), 'Dynamically aggregating diverse information', *Econometrica* 90(1), 47–80.
- Liang, Y. (2023), Boundedly rational information demand, working paper.
- Martinez-Marquina, A., Niederle, M. & Vespa, E. (2019), 'Failures in contingent reasoning: the role of uncertainty', *American Economic Review* **109**(10), 3437–3474.
- Montanari, G. & Nunnari, S. (2022), Audi alteram partem: an experiment on selective exposure to information, working paper.
- Mu, X., Pomatto, L., Strack, P. & Tamuz, O. (2021), 'From blackwell dominance in large samples to rnyi divergences and back again', *Econometrica* **89**(1), 475–506.
- Nielsen, K. (2020), 'Preferences for the resolution of uncertainty and the timing of information', Journal of Economic Theory 189.
- Novak, V., Matveenko, A. & Ravaioli, S. (2023), The Status Quo and Belief Polarization of Inattentive Agents: Theory and Experiment, CRC TR 224 Discussion Paper Series crctr-224-2023-385, University of Bonn and University of Mannheim, Germany.
- Rabin, M. & Weizsäcker, G. (2009), 'Narrow bracketing and dominated choices', American Economic Review 99(4), 1508–43.
- Rubinstein, A. (1998), *Modeling Bounded Rationality*, Vol. 1 of *MIT Press Books*, 1 edn, The MIT Press.
- Shannon, C. E. (1948), 'A mathematical theory of communication', Bell System Technical Journal 27(3), 379–423.
- Wilson, A. & Vespa, E. (2016), Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language, working paper.

Online Appendix for

Choosing Between Information Bundles

Menglong Guan

CONTENTS:

- **A.** Additional Plots and Tables
- **B.** Information Bundles and Sources Studied in the Experiment
- **C.** Experiment Instructions

Additional Plots and Tables Α



Figure 14: Likelihood of Choosing High-Value Information – All Three Settings Notes: The figure plots the likelihood of choosing σ over σ' in the Isolated setting, the likelihood of choosing $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$ in the Separated setting, and the likelihood of choosing $\sigma_0 \lor \sigma$ over $\sigma_0 \lor \sigma'$ in the Joined setting. R denotes refine, O denotes reveal-or-refine, S denotes sufficiency, B denotes Blackwell, and NB denotes that two sources can not be Blackwell ordered. Detailed descriptions of these comparison relationships are in Section 2.3. The cases are ordered in terms of the likelihood of choosing σ over σ' under them in the Isolated setting. Short vertical lines denote 95 percent confidence intervals.



Figure 15: Value given Guesses or Assessments versus Choices between Information Bundles Notes: Value given guess denotes the actually realized instrumental value of an information bundle conditional on how it is used. Assessment denotes the elicited assessments of the instrumental value of information bundles. In both panels, the y-axis plots the likelihoods of subjects choosing a bundle over another in binary choices under the Separated setting. In the left (right) panel, the x-axis plots the likelihoods indicated by the value given guesses (subjective assessments) of the pairs of bundles. If a pair of information bundles have the equal value given guesses (are assigned with equal assessments), then the choice of the corresponding subject between that pair of bundles is considered to be 0.5 when computing the choice likelihood indicated by value given guesses (assessments).

| Setting | Group | Guess | Choice (instrum | Assessment nental value) | Choice (value gi | Assessment iven guesses) |
|--------------------|------------|-------|--------------------|-----------------------------|---------------------|-----------------------------|
| | Naive | 0.71 | 0.49 | 0.36 | 0.61 | 0.53 |
| Separated | In-Between | 0.87 | 0.51 | 0.50 | 0.57 | 0.54 |
| | Optimal | 1 | 0.69 | 0.68 | 0.69 | 0.68 |
| | Naive | 0.95 | 0.75 | 0.63 | 0.75 | 0.64 |
| Joined | In-Between | 0.99 | 0.76 | 0.68 | 0.76 | 0.67 |
| | Optimal | 0.99 | 0.81 | 0.75 | 0.82 | 0.75 |
| | Naive | 0.24 | 0.26 | 0.27 | 0.14 | 0.11 |
| Joined - Separated | In-Between | 0.12 | 0.25 | 0.18 | 0.19 | 0.13 |
| | Optimal | -0.01 | 0.12 | 0.07 | 0.13 | 0.07 |

Table 5: Optimality of Decision Making – By Group

Notes: "Joined-Separated" presents the differences in optimality rates between the Joined and Separated settings.



Figure 16: Combining Mechanisms of Information Bundle Choices Notes: The y-axis plots the likelihood of choosing bundle $\{\sigma_0, \sigma\}$ in the binary choices under the Separated setting. I define a subject's "Isolated+Assessment" ("Isolated+Joined") choices as either her choices under the Isolated setting or choices indicated by her assessments of the instrumental value of bundles (or choices under the Joined setting), depending on which are more consistent with her choices between information bundles under the Separated setting. In the left panel, the x-axis plots the predicted likelihood of choosing bundle $\{\sigma_0, \sigma\}$ regarding the defined "Isolated+Assessment" choices. In the right panel, the x-axis plots the predicted likelihood regarding the defined "Isolated+Joined" choices. In each panel, the red dashed line is the best linear fit, the grey region shows the 95 percent confidence intervals for predictions of the linear fit, and the grey dashed line is the diagonal line y = x.

B Information Bundles and Sources



Figure 17: Studied Information Bundles and Sources Notes: To be continue on next pages.



Case (6): Not Blackwell, $\sigma < \sigma'$

Figure 17-2: Studied Information Bundles and Sources



Cuse (0). Sumereney, 0.00

Figure 17-3: Studied Information Bundles and Sources

C Instructions

Introduction • The study includes 4 parts. If you complete the study, you will receive a guaranteed payment of \$8. In addition, you will receive a sizable performance-based payment that is based on your choices in the study. • In each part, we will start by providing you with instructions. Please read and follow the instructions closely and carefully. We will ask you comprehension questions to check that you understand the instructions. You will receive a \$0.2 bonus for each comprehension question that you answer correctly. · Your choices in one part will NOT impact another part. At the end of the study, your performance-based payment will be determined based on your choices in a randomly selected part. Continue Part 1: Guessing Questions • This part includes 16 Guessing questions . In each question, we will show you twenty objects, including 10 triangles and 10 circles, like below: 3 4 5 6 7 8 9 10 11 12 13 14 15 16(17) (18) 19 (20) • In a Guessing question, one of the twenty objects will be randomly drawn by the computer. Each object is equally likely to be drawn. Your task is to guess the shape (triangle or circle) of the randomly drawn object. • You will earn a \$12 bonus payment if you guess the shape correctly and \$0 otherwise. You will submit your guess on the shape of the randomly drawn object by clicking radio buttons like below: • I guess the shape of the randomly drawn object to be: ○ Triangle ○ Circle Comprehension Question How do you earn the \$12 bonus payment? \bigcirc By correctly guessing the shape of the randomly drawn object. \bigcirc By randomly clicking on things. Continue

Part 1: Information

- Before you guess the shape of the randomly drawn object, you will receive information about which object has been drawn by the computer from an information source.
- An information source will divide the twenty objects into Groups, which are denoted by color bars and letters, like this:



(This is just an example. You may see many different groupings across questions.)

- In each Guessing question, we will tell you which group the randomly drawn object is in BEFORE you guess its shape.
- · For instance, in the example above:
 - If the drawn object is one of triangles 1-6 or circles 17-20, we will tell you that the drawn object is in Group a before you guess its shape;
 - If the drawn object is one of triangles 7-10 or circles 11-16, we will tell you that the drawn object is in Group b before you guess its shape;
- In fact, we will ask you to submit your guess for each possible information (about which group the drawn object is in) you might receive.
- For instance, in the example above, we will ask you to guess the shape of the randomly drawn object if you learn it is in Group a AND guess the shape if you learn it is in Group b.
- You will submit your guesses by finishing a list of choices, each corresponding to a piece of information you might receive, like below:

| 0 | If I learn the drawn object is in Group | a, I will guess its shape to be: | ○ Triangle | ○ Circle |
|---|---|----------------------------------|------------|----------|
| 0 | If I learn the drawn object is in Group | b, I will guess its shape to be: | ○ Triangle | O Circle |

- At the end of the experiment, if one of the Guessing questions is selected for payment: The computer will randomly draw an object. We will then
 look at which Group the randomly drawn object is in, and whether your guess for that Group matches the shape of the drawn object. If it
 matches, that means your guess is correct and you earn the \$12 bonus payment.
- · This means you should always submit your BEST GUESS to maximize your expected earnings.

Comprehension Questions

What information will an information source provide to you?

- \bigcirc The shape of the randomly drawn object.
- \odot The numbering of the randomly drawn object.
- \bigcirc Which group the randomly drawn object is in.

How do you maximize your chances of winning the \$12 bonus payment?

- \bigcirc Always guess the shape of the randomly drawn object to be Triangle.
- \bigcirc Always guess the shape of the randomly drawn object to be Circle.
- \odot Always make the same guess about the shape of the randomly drawn object for each possible piece of information you might receive.
- Always make the best guess about the shape of the randomly drawn object for each possible piece of information you might receive.

Part 1: Follow-up Assessment Task

• In each round of Part I, after you complete a Guessing question and click "Continue", you will see a follow-up Assessment task on your screen, like below:

Example

If this Guessing question is selected for payment, what is the likelihood do you think that your guess is correct? (Reminder: Your answer to this question will not impact your chance of winning the \$12 bonus from the Guessing question. You have the greatest chance of earning an extra \$5 bonus by submitting your TRUE assessment.)

- As shown in the example, you can drag the slider to pick a number (between 50 and 100) to indicate how likely do you think that your guess
 is correct. Your answer in the Assessment task will NOT impact your chance of winning the \$12 bonus from the Guessing question.
- At the end of the experiment, if a Guessing question is selected for payment, aside from the \$12 bonus you can earn if you guess the shape correctly, you also have chance to get **an extra \$5 bonus** from the Assessment task.
- Given your picked number (that indicates your assessment on the likelihood) in the Assessment task, whether you can get the extra \$5 bonus will be determined in the following way:
 - 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 is no way affects the outcome of the second draw.
 - If your guess is correct and your picked number is larger than either of the two draws, you will get \$5.
 - If your guess is incorrect and your picked number is smaller than either of the two draws, you will get \$5.
 - The procedure is purposefully designed so that you have the greatest chance of winning the \$5 when you answer truthfully in the assessment task.

Comprehension Question

Which of the following statement is NOT correct?

- \odot I have the greatest chance of winning the extra \$5 bonus when I answer truthfully in the Assessment task.
- \odot My answer in the Assessment task will not impact my chance of winning the \$12 bonus from the Guessing question.
- \bigcirc I should answer randomly in the Assessment task.

Practice Round

Practice Guessing Question

There are twenty objects, including 10 triangles and 10 circles. One object will be **randomly** drawn by the computer. You will **earn \$12 if you correctly guess the shape** of the drawn object (triangle or circle).

You will learn which group the drawn object is in under the following information source before you guess its shape.



Please indicate your guess for each possible piece of information (i.e., which group the drawn object is in under the information source) you might receive:

| • | If I learn the drawn object is in Group $\mathbf{m},$ I will guess its shape to be: | Triangle | Circle |
|---|--|------------------------------|----------|
| • | If I learn the drawn object is in Group \mathbf{n} , I will guess its shape to be: | Triangle | O Circle |

Please finish all the choices and then click Continue.

(Remember: If this were not a practice question, these choices would determine your actual guess and therefore your bonus payment!)

Practice Feedback

If the randomly drawn object is one of triangles 1-2 or circles 16-20, we will implement your guess for Group m (the first guess above).

If the randomly drawn object is one of triangles 3-10 or circles 11-15, we will implement your guess for Group n (the second guess above).

If your implemented guess matches the shape of the drawn object, that means your guess is correct and you can earn \$12.

Practice Assessment Task

What is the likelihood do you think that your guess is correct in the above practice Guessing question?

I think the likelihood of my guess being correct is: 65%

Practice Feedback

If this were not a practice round, given the payoff determination procedure explained in the instructions, you have the greatest chance of winning an extra \$5 when you answer **truthfully** in the Assessment Task.

Click "Continue" to start Part 1.



Part 2: Choose Information Source

- This part includes **16 Choose-Information-Source questions**. In each question, we will show you two different information sources (each divides the twenty objects into groups) and ask you to **choose which one you would rather receive information from** before you guess the shape of the randomly drawn object.
- · You will make your choice of information source in an interface like below:

Example

Below are two information sources. Please choose which one you would rather receive information from before you guess the shape of the randomly drawn object.



• At the end of the experiment, if one of the Choose-Information-Source questions is selected for payment, you will receive information (i.e., which group the drawn object is in) from the information source you choose before you are asked to guess the shape of the drawn object. You will earn a \$12 bonus payment if your guess is correct.

Comprehension Question

If one of the Choose-Information-Source questions is selected for payment, which of the following statement is CORRECT?

- O Before making a guess, I will receive the same information no matter which information source I have chosen.
- \odot Before making a guess, I will receive information from the information source I have chosen.
- O Before making a guess, I will be equally likely to receive information from any of the two shown information sources.

Part 3: More Guessing Questions

- This part includes another 16 Guessing questions. In each question, we will show you a pair of information sources. And you will receive information from both of them before you guess the shape of the drawn object.
- For instance, below is a pair of information sources (I and II). You will learn which group the randomly drawn object is in under each of them before you guess the shape. The information will help you in the Guessing question:
 - If you learn that the drawn object is in Group a of source I and Group c of source II, that means the drawn object must be one of triangles 1-3 or circles 17-20;
 - If you learn that the drawn object is in Group a of source I and Group d of source II, that means the drawn object must be one of triangles 4-6;
 - If you learn that the drawn object is in Group b of source I and Group c of source II, that means the drawn object must be one of circles 13-16;
 - If you learn that the drawn object is in Group b of source I and Group d of source II, that means the drawn object must be one of triangles 7-10 or circles 11-12;



(This is just an example. You may see many different pairs of information sources across questions.)

• Similarly, we will ask you to submit your guess for each possible pair of information (i.e., which group the drawn object is in under each of the two information sources) you might receive, like below:

| • | If I learn the drawn object is in Group | of source I a | nd Group C | of source II, I will guess its shape to be: | ○ Triangle | ○ Circle |
|---|---|----------------------|------------|---|------------|-------------------|
| > | If I learn the drawn object is in Group | of source I a | nd Group d | of source II , I will guess its shape to be: | ○ Triangle | ○ Circle |
| • | If I learn the drawn object is in Group | of source I a | nd Group c | of source II , I will guess its shape to be: | ○ Triangle | \bigcirc Circle |
| • | If I learn the drawn object is in Group | of source I a | nd Group d | of source II , I will guess its shape to be: | ○ Triangle | ○ Circle |

- At the end of the experiment, if one of these Guessing questions is selected for payment: The computer will randomly draw an object. We will then look at which Group the drawn object is in under source I and which group it is in under source II, and check whether your guess for that combination of information matches the shape of the drawn object. If it matches, that means your guess is correct and you earn the \$12 bonus payment.
- This means you should always submit your BEST GUESS to maximize your expected earnings.
- Besides, as in the Part I, after you complete a Guessing question, we will also ask you to finish an Assessment Task. Your answer in the Assessment task will NOT impact your chance of winning the \$12 bonus from the Guessing question. You should always answer your true assessment on the likelihood of your guess being correct so to maximize your chance of getting an extra \$5 bonus.

Click "Continue" to start a Practice Round.

Practice Round

Practice Guessing Question

There are twenty objects, including 10 triangles and 10 circles. One object will be **randomly** drawn by the computer. You will **earn \$12 if you correctly guess the shape** of the drawn object (triangle or circle).

Below is a pair of information sources (I and II). You will learn which group the drawn object is in under each of them before you guess the shape.



Please indicate your guess for each possible pair of information (i.e., which group the drawn object is in under each of the two information sources) you might receive:

| • | If I learn the drawn object is in Group | а | of source I and Group | с | of source II , I will guess its shape to be: | Triangle | Oircle |
|---|---|---|------------------------------|---|---|------------|--------|
| • | If I learn the drawn object is in Group | a | of source I and Group | d | of source II , I will guess its shape to be: | Triangle | Circle |
| • | If I learn the drawn object is in Group | b | of source I and Group | с | of source II , I will guess its shape to be: | O Triangle | Circle |
| • | If I learn the drawn object is in Group | b | of source I and Group | d | of source II , I will guess its shape to be: | O Triangle | Circle |

Please finish all the choices and then click Continue.

(Remember: If this were not a practice question, these choices would determine your actual guess and therefore your bonus payment!)

Practice Feedback

If the randomly drawn object is one of triangles 1-3 or circles 17-20, we will implement your guess for the case when you learn the drawn object is in Group a of source I and Group c of source II (the first guess above).

If the randomly drawn object is one of triangles 4-6, we will implement your guess for the case when you learn the drawn object is in Group a of source I and Group d of source II (the second guess above).

If the randomly drawn object is one of circles 13-16, we will implement your guess for the case when you learn the drawn object is in Group b of source I and Group c of source II (the third guess above).

If the randomly drawn object is one of triangles 7-10 or circles 11-12, we will implement your guess for the case when you learn the drawn object is in Group b of source I and Group d of source II (the fourth guess above).

If your implemented guess matches the shape of the drawn object, that means your guess is correct and you can earn \$12.

Practice Assessment Task

What is the likelihood do you think that your guess is correct in the above practice Guessing question?



If this were not a practice round, given the payoff determination procedure explained in the instructions, you have the greatest chance of winning an extra \$5 when you answer **truthfully** in the Assessment Task.

Click "Continue" to start Part 3.

Part 4: Choose Information Source

- This part includes another 8 Choose-Information-Source questions. In each question, we will show you two pairs of information sources and ask you to choose which pair you would rather receive information from before you guess the shape of the drawn object.
- You will make your choice in an interface like below:

Example

Below are two pairs of information sources. Please choose which pair of information sources you would rather receive information from before you guess the shape of the drawn object.



• At the end of the experiment, if one of these questions is selected for payment, you will receive two pieces of information, **one from each information source in the pair you choose** (i.e., you will learn which group the drawn object is in under each of the two information sources), before you are asked to guess the shape of the randomly drawn object. You will earn a \$12 bonus payment if your guess is correct.

Comprehension Question

If one of these Choose-Information-Source questions is selected for payment, which of the following statement is CORRECT? O Before making a guess, you will learn which group the drawn object is in under each information source in the pair you have chosen. O Before making a guess, you will learn which group the drawn object is in under only one information source in the pair you have chosen.

The Final Question

One of the Choose-Information-Source questions is randomly selected to determine your performance-based payment.

Below is the pair of information sources you chose in that question.



Please indicate your guess for each possible pair of information (i.e., which group the drawn object is in under each of the two information sources) you might receive:

| • If I learn the drawn object is in Group a of source I and Group e of source II , I will guess its shape to be: | \bigcirc Triangle | ○ Circle |
|--|---------------------|----------|
| • If I learn the drawn object is in Group b of source I and Group e of source II, I will guess its shape to be: | \bigcirc Triangle | ○ Circle |
| • If I learn the drawn object is in Group b of source I and Group f of source II , I will guess its shape to be: | \bigcirc Triangle | ○ Circle |

Please finish all THREE choices and then click Submit.

(Remember: these choices determine your actual guess and therefore whether you can earn the \$12 bonus.)

Submit

Experiment Results

Your Final Payoff

In the Final Question, the randomly drawn object is a circle, and your guess of its shape is triangle. You earn the \$8 guaranteed payment and \$0.8 for answering 4 comprehension questions correctly. So your total earnings are \$8.80. We will send you the payment via Venmo by the end of the day.