University of California
Santa Barbara

Experimental Studies on Information Economics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Economics

by

Menglong Guan

Committee in charge:

Professor Ryan Oprea, Chair
Professor Erik Eyster
Professor Sevgi Yuksel

June 2024

The Dissertation of Menglong Guan is approved.

Professor Erik Eyster

Professor Sevgi Yuksel

Professor Ryan Oprea, Committee Chair

April 2024

Experimental Studies on Information Economics

by

Menglong Guan

Acknowledgements

I would like to express my sincere gratitude to my advisor, Ryan Oprea, and my committee members, Sevgi Yuksel and Erik Eyster. I thank Ryan and Sevgi for their guidance and encouragement throughout my PhD journey. Their insightful feedback and unwavering dedication to my growth as a researcher have been instrumental in shaping my academic path. I thank Erik for his thoughtful comments, constructive suggestions, and the time he has devoted to helping me with my research. His expertise and insights have been extremely valuable.

I want to thank Daniel Martin, Cheng-Zhong Qin, Gary Charness, and other faculty from the UCSB Economics Department and my fellow graduate students for providing helpful feedback on my research projects. I am very fortunate to be in the Economics PhD program at UCSB, and I am thankful to Mark Patterson for his diligent and extraordinary work in supporting me and the program. Additionally, I would like to thank my undergraduate advisor and coauthor, Sen Geng, for igniting my passion for academic research when I was an undergraduate and for his continued support and collaboration during my graduate studies.

Finally, I am deeply grateful to my family for their unconditional love, understanding, and encouragement throughout this challenging yet rewarding journey. I could not have achieved this milestone without their support. I also want to thank my partner, Han Xiao, for standing by my side and supporting me, particularly during the most stressful moments of this journey. Your presence has been invaluable to me.

Curriculum Vitæ

Menglong Guan

Contact

Address	North Hall 2053, Department of Economics, University of California, Santa Barbara, Santa Barbara, CA 93106
Email	mguan@ucsb.edu
Website	https://www.menglongguan.com/

Education

2024	Ph.D. in Economics (Expected), University of California, Santa Barbara.
2019	M.A. in Economics, University of California, Santa Barbara.
2017	B.A. in Economics, Xiamen University

Fields of Study

Experimental Economics, Behavioral Economics, Microeconomic Theory

Advised by Ryan Oprea (Chair), Erik Eyster and Sevgi Yuksel

Research Papers

1. Choosing Between Information Bundles

2. Too Much Information (with Ryan Oprea and Sevgi Yuksel)

3. Trustworthy by Design (with Sen Geng) - *Games and Economic Behavior - 141 (2023): 70-87*

4. Three Faces of Complexity in Strategic Choice (with Ryan Oprea)

5. Preference for Sample Features and Belief Updating (with ChienHsun Lin, Jing Zhou and Ravi Vora)

Permissions and Attributions

The content of Chapter 3 and Appendix C is the accepted manuscript of an article published by Elsevier in Games and Economic Behavior, Vol 141, Sen Geng and Menglong Guan, 'Trustworthy by design', Page 70-87, 2023. The article is available online: https://doi.org/10.1016/j.geb.2023.05.009. The reuse of the content in this dissertation is permitted by the publisher.

Abstract

Experimental Studies on Information Economics

by

Menglong Guan

This dissertation consists of three experimental studies on information economics, exploring the topics of the demand for information, the choice and use of information, and information design within strategic contexts.

Chapter 1 studies 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 we 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.

Chapter 2, based on a joint work with Ryan Oprea and Sevgi Yuksel, studies how people's demand for information structures is shaped by their informativeness—the reduction in uncertainty they produce. To do this, we introduce new methods that remove confounds for information demand like failures of Bayesian reasoning. We show that

people (i) strongly demand informativeness when it has instrumental value but also (ii) display a sharp aversion to informativeness when it cannot be used to improve choice, sometimes leading to costly errors in information choice. Several strands of evidence suggest that this aversion is driven by subjective information processing costs that rise with informativeness.

Chapter 3, based on a joint work with Sen Geng, explores theoretically and experimentally whether information design can be used by trustees as a signaling device to boost trusting acts. In our main setting, a trustee partially or fully decides a binary payoff allocation and designs an information structure; then a trustor decides whether to invest. In the control setting, information design is not available. In line with the standard equilibrium analysis, we find that introducing information design increases trustworthiness and trusting acts, and some trustees choose full trustworthiness with the most informative structure. We also find systematic behavioral deviations, including some trustees' choosing zero trustworthiness with the least informative structure and trustors' overtrusting in low informative structures. We finally provide a model of heterogeneity in prosociality and strategic sophistication, which rationalizes the experimental findings.

JEL: D01, D80, D91

Contents

Chapter 1

Choosing Between Information Bundles

1.1 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 1.1, 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 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 1.1 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

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

[2] Under 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 in Figure 1.1 can be characterized as $\sigma = \{x, y\} = \{(T, \{1, ..., 5\}) \cup (C, \{11, ..., 18\}), (T, \{6, ..., 10\}) \cup (C, \{19, 20\})\}$.

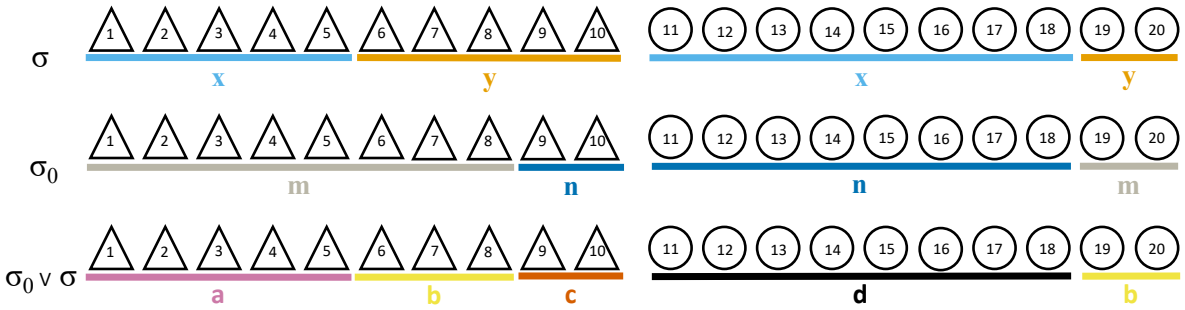


Figure 1.1: Examples of Representing Information Sources as Partitions *Notes: σ, σ_0 and σ_b are three information sources. Subjects must guess the shape (triangle or circle) of a randomly drawn object among twenty objects. They are told which colored subset(s) contains the randomly drawn object under 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 $\{\sigma_0, \sigma\}$, which can be derived by finding out the intersections of signals (subsets) from σ and σ_0.*

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 1.1.[4] The indicated procedure of integrating information under this design captures how information integration is done in real-world scenarios: merging

[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.

[4]Signal (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.

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 pre-integrated into single information sources that contain the same information content as the bundles (e.g., $\sigma_0 \vee \sigma$ shown in Figure 1.1 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.

[5]At 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.

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, $\{\sigma_0, \sigma\}$ Blackwell dominates $\{\sigma_0, \sigma'\}$ (meaning the former is weakly more instrumentally valuable) if and only if σ *reveals-or-refines* σ'. *Reveal-or-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

[6] Check Figure A.4 in Appendix A.2 for examples of σ *reveals-or-refines* σ' and σ *refines* σ'.

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" 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 of how they would

[7] I 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 1.4.2.

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

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 $\{\sigma_0, \sigma'\}$ 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

[9]The 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.

[10]The 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.

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 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.

[11] In 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.

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

[12]Some other work focuses on the demand for non-instrumental information or information sources. 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).

[14]Blackwell (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 (2017*a*,*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.

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 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 1.2 introduces the conceptual framework. Section 1.3 describes the experimental design. Section 1.4 presents the main results. Section 1.5 discusses the possible reasons behind the emergence of the *common source cancellation* heuristic and other determinants of information choices. Section 1.6 concludes.

1.2 Conceptual Framework

1.2.1 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^s_ω 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^s_σ, over the state space Ω. According to Bayes' Rule, $q^s_\sigma(\omega) = \frac{p(\omega)\sigma^s_\omega}{q_\sigma(s)}$, where $q_\sigma(s) = \sum_\omega p(\omega)\sigma^s_\omega$ 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 $\mathbb{E}[u(a,\omega)|s] = \sum_\omega q^s_\sigma(\omega)u(a,\omega)$. 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 $\max\{q^s_\sigma, 1 - q^s_\sigma\}$. Therefore, the *instrumental*

value of σ can be simplified into:

$$V_\sigma = \Big(\underbrace{\sum_{s\in S} q_\sigma(s) \underbrace{\max\{q_\sigma^s, 1-q_\sigma^s\}}_{\substack{\text{Guessing accuracy} \\ \text{conditional on } s}}}_{\text{Expectation over } s} - \underbrace{p}_{\substack{\text{Guessing accuracy} \\ \text{without information}}} \Big) \gamma \tag{1.1}$$

$$\text{Expected improvement in guessing accuracy}$$

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

1.2.2 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

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

above:

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

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.

1.2.3 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 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 $\{\sigma_0, \sigma\}$ is (weakly) more valuable than $\{\sigma_0, \sigma'\}$. In fact, BFK proves that for any σ_0, $\{\sigma_0, \sigma\}$ Blackwell dominates $\{\sigma_0, \sigma'\}$, 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

[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 *Blackwell*).

[17] Figure A.4 in Appendix A.2 presents examples of these comparison relationships.

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.

1.3 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 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 a bundle); (iii) *Information Choice Task*, eliciting subjects' choices between information bundles. Further details of the three types of tasks are described in Section 1.3.3 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.

1.3.1 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 (1.1)) of an information source. For example, σ in Figure 1.1 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 1.2.3, 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 1.1 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 1.1 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 $d = x \cap n$.

1.3.2 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 1.2.3, 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.

Table 1.1: Studied Information Bundles and sources

	Isolated: σ vs. σ'	*Joined*: $\sigma_0 \vee \sigma$ vs. $\sigma_0 \vee \sigma'$ *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)	$<$	$>$

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 1.2.3. Value denotes the instrumental value, i.e., the expected improvement in guessing accuracy induced by an information bundle or source as defined in Section 1.2. $>$ ($<$) 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.

Table 1.1 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 A.4 of Appendix A.2.

1.3.3 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 ($\sigma_0 \vee \sigma$ or $\sigma_0 \vee \sigma'$ from one of the eight cases listed in Table 1.1) 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 1.2 is a screenshot of Part 1. Note that the Assessment task appears right

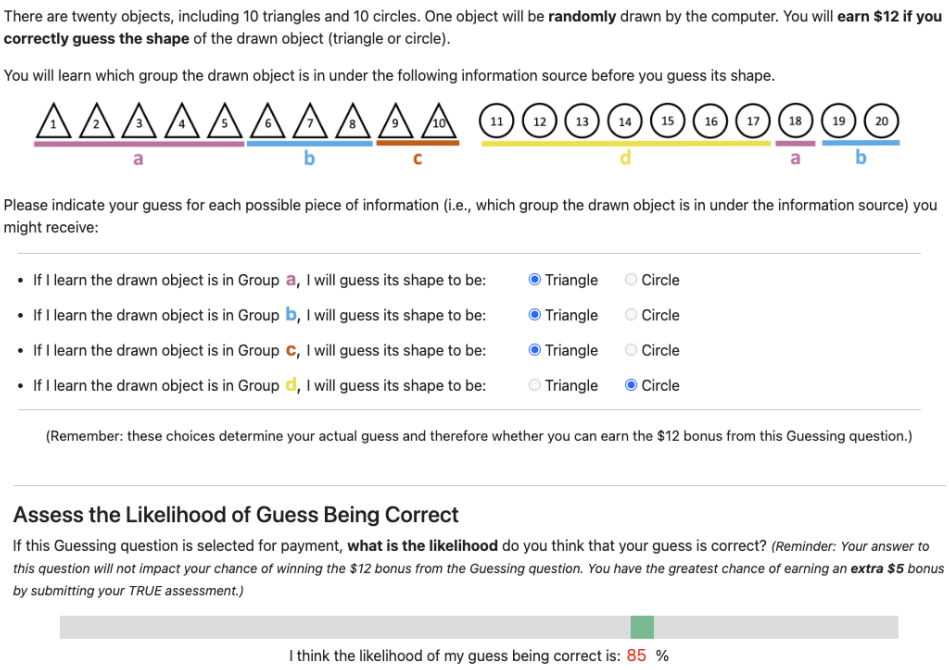


Figure 1.2: Screenshot of Guessing and Assessment Tasks in Part 1 *Notes: The Assessment task appears below a Guessing task after subjects make all guessing choices and click a "Continue" button. The submitted guesses are shown on the screen but are no longer changeable when subjects work on the Assessment task.*

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.

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 1.3 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, $\{\sigma_0, \sigma\}$

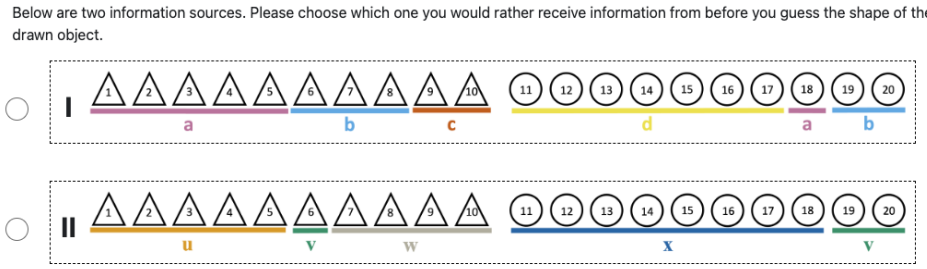


Figure 1.3: Screenshot of Information Choice Task in Part 2

or $\{\sigma_0, \sigma'\}$ from the eight cases, instead of the *join* information sources, $\sigma_0 \vee \sigma$ or $\sigma_0 \vee \sigma'$. Figure 1.4 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 1.5 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.

1.3.4 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

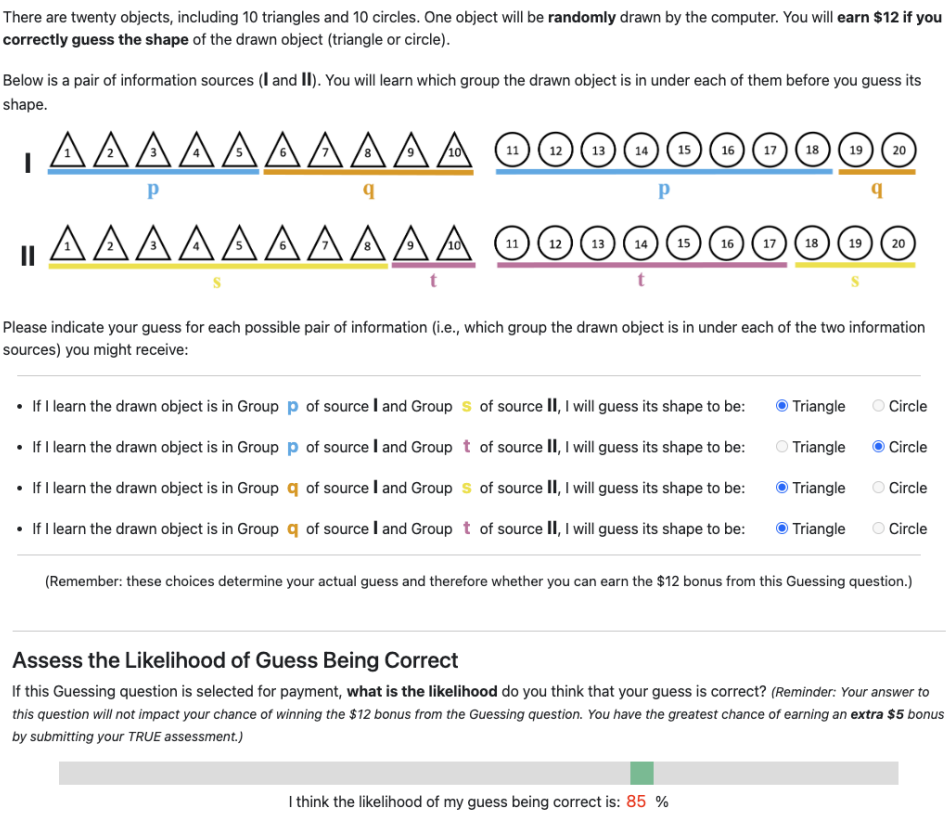


Figure 1.4: Screenshot of Guessing and Assessment Tasks in Part 3

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. 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 in Parts 2 or 4 was selected, the subject completed a final Guessing task

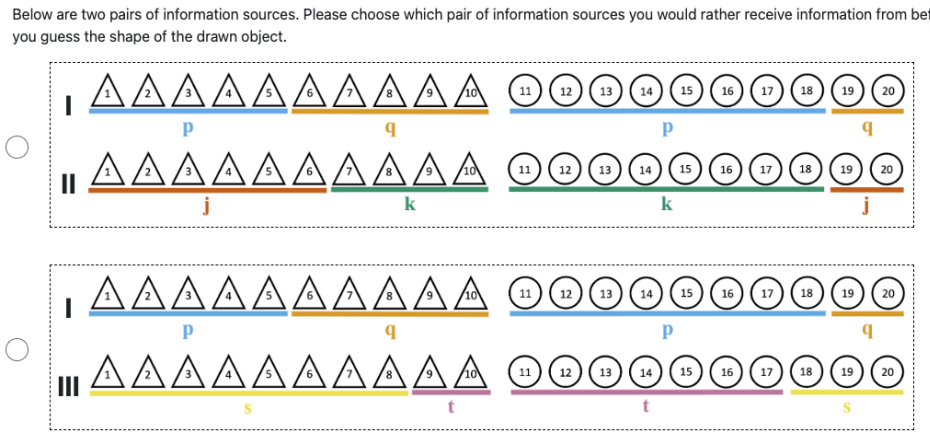


Figure 1.5: Screenshot of Information Choice Task in Part 4

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).

1.4 Results

The main findings of the experiment are organized as follows. Section 1.4.1 analyzes and compares choices of information under the Separated and Joined settings. Section 1.4.2 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 1.4.3 then investigates whether the mistakes are instead systematic, driven by a simple but imperfect heuristic in information bundle choices. Section 1.4.4 explores the heterogeneity in these results among subjects.

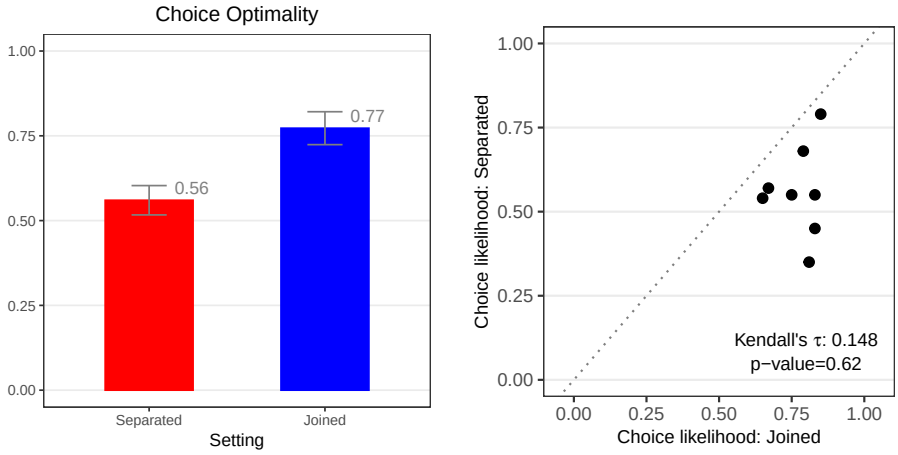


Figure 1.6: Choices of Information Bundles and Join information Sources *Notes: The optimality of information choices is measured by the likelihood of choosing the high-value information bundle or source (relative to the other) in binary choices. Short vertical lines in the left panel denote 95 percent confidence intervals. In the right panel, the y-axis (x-axis) plots the likelihood of choosing bundle (join information source) $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma\}$ ($\sigma_0 \vee \sigma$ over $\sigma_0 \vee \sigma'$) in Information Choice tasks under the Separated (Joined) setting.*

1.4.1 Choices of 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 1.6 shows that 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 1.6 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 1.1), 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 1.2.3 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 1.1). Figure 1.7 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 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*

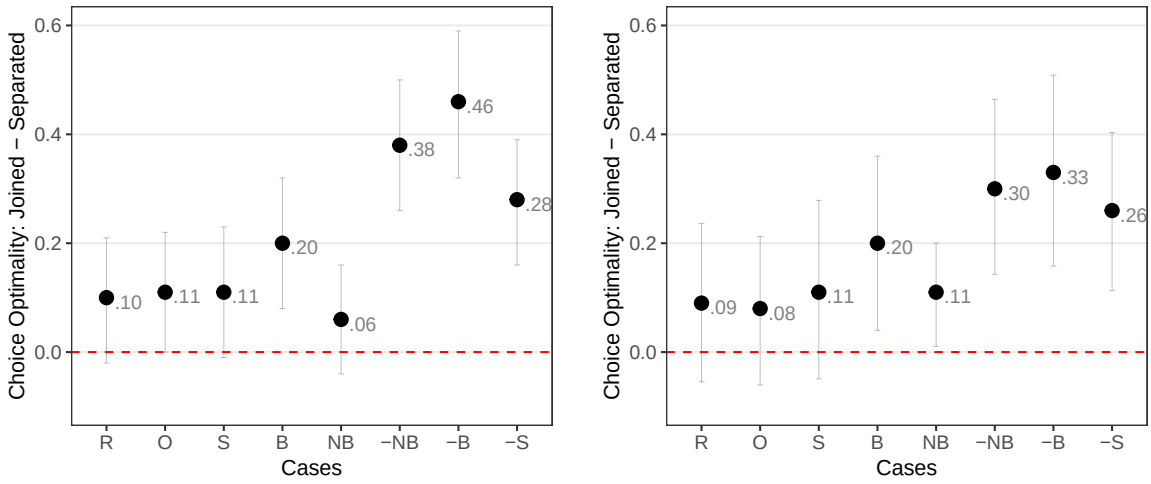


Figure 1.7: Decrease in the Optimality of Information Choices from Joined to Separated Settings *Notes: The left panel covers all data while, in each case, the right panel focuses on the subjects for whom the more instrumentally valuable information bundle or join source is indeed more helpful in Guessing 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 1.1. 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 1.2.3. Short vertical lines denote 95 percent confidence intervals by Bootstrapping.*

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.

1.4.2 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.

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 1.8 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 1.8 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.

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 1.9 contain the randomly drawn object. The Bayesian optimal guess is *Triangle*, but 54 subjects guessed *Circle*. This guessing error can be explained by

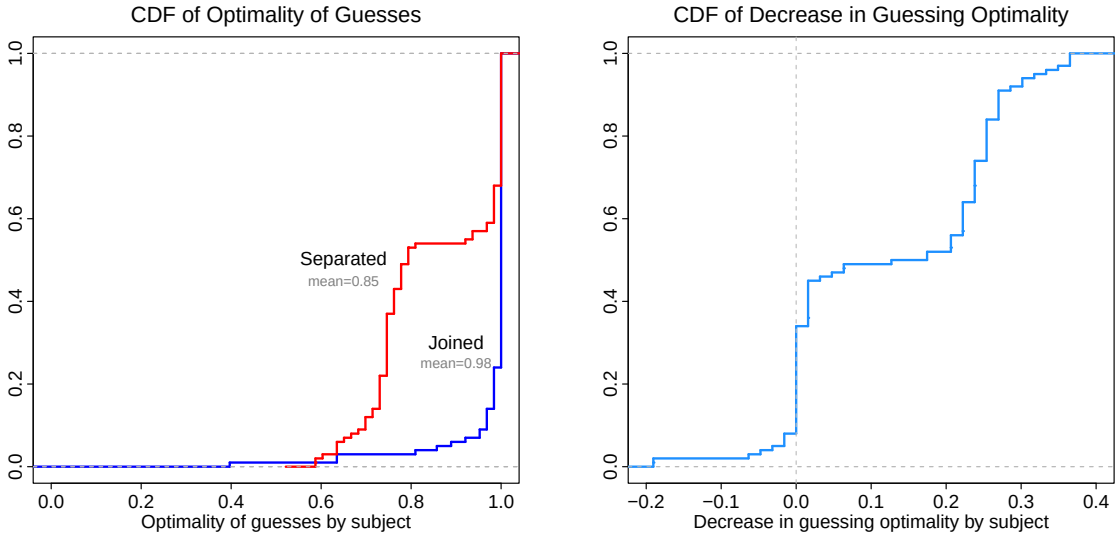


Figure 1.8: Guessing Optimality *Notes: The optimality of guesses is measured by the share of submitted guesses that are optimal from the Bayesian perspective. The decrease in optimality is the difference in guessing optimality between the Joined and Separated settings.*

the 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.[18,19]

[18]Possible 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.

[19]The 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).

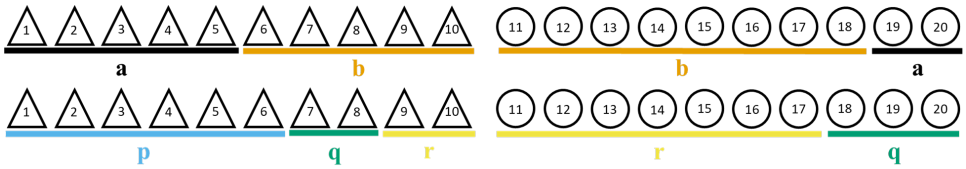


Figure 1.9: Example of Information Bundle

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.*

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 1.10, 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 A.2 in Appendix A.1 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'

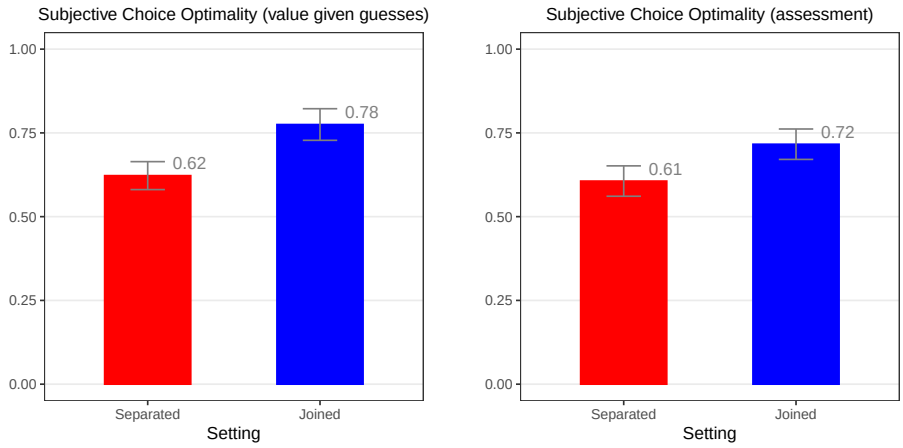


Figure 1.10: Subjective Optimality of Information Choices *Notes: In the left (right) panel, the choice optimality is measured by the likelihood of choosing the information bundle or source with a weakly higher value given guesses (assessments) in binary choices. Short vertical lines denote 95 percent confidence intervals.*

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 1.10 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 one 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 of how they would use the bundles to make inferences.

Result 4 *Subjects make information bundle choices without much consideration of 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.*

1.4.3 Common Source Cancellation in 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 1.4.1 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 $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$, 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 1.11 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 1.1), 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 1.1), categorized as Individually Better. In the first scenario, subjects make optimal information bundle choices only 45 percent of the time. In contrast, the optimality rate increases substantially to 63 percent in the second

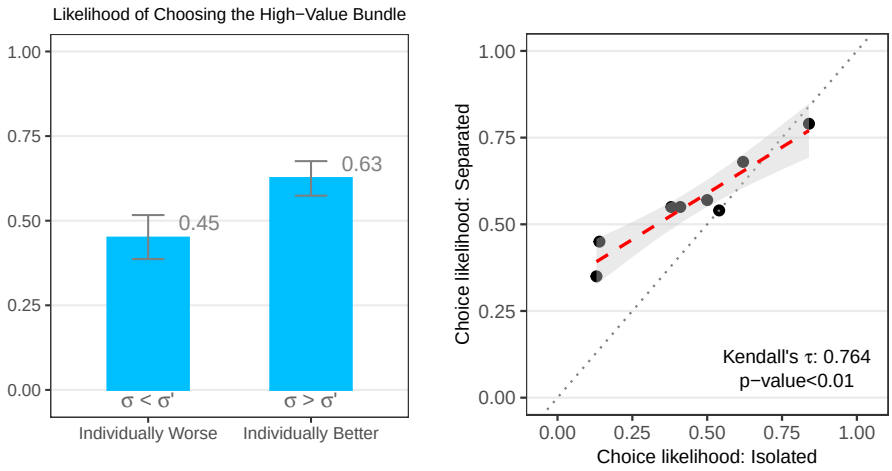


Figure 1.11: Common Source Cancellation in Information Bundle Choices *Notes: The left panel plots the likelihood of choosing bundle* $\{\sigma_0, \sigma\}$ *over* $\{\sigma_0, \sigma'\}$*, with the former being more instrumentally valuable than the latter by a 0.1 increment in guessing accuracy (i.e., \$1.2 increase in the expected payoff). "Individually Worse" refers to cases (6)-(8) listed in Table 1.1 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. The red dashed line is the best linear fit, and the grey region is the 95 percent confidence interval for predictions of the linear fits.*

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 1.11 then directly compares subjects' choices between isolated information sources σ and σ' and their choices between corresponding bundles $\{\sigma_0, \sigma\}$ and $\{\sigma_0, \sigma'\}$. As the graph shows, the likelihood of choosing σ over σ' strongly explains the likelihood of choosing $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$ 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 1.2 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.

Table 1.2: Choices Between Information Bundles

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

*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* $\{\sigma_0, \sigma\}$ *and* $\{\sigma_0, \sigma'\}$ *and can account for most of the mistakes in the latter.*

1.4.4 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 1.4.2 (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, respectively. Table A.1 in Appendix A.1 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 1.12 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 1.11 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. These results suggest that the *CSC* heuristic plays a vital role in explaining choices between information bundles of each group. Figure 1.12 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 1.3 replicates regression (2) in Table 1.2 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*

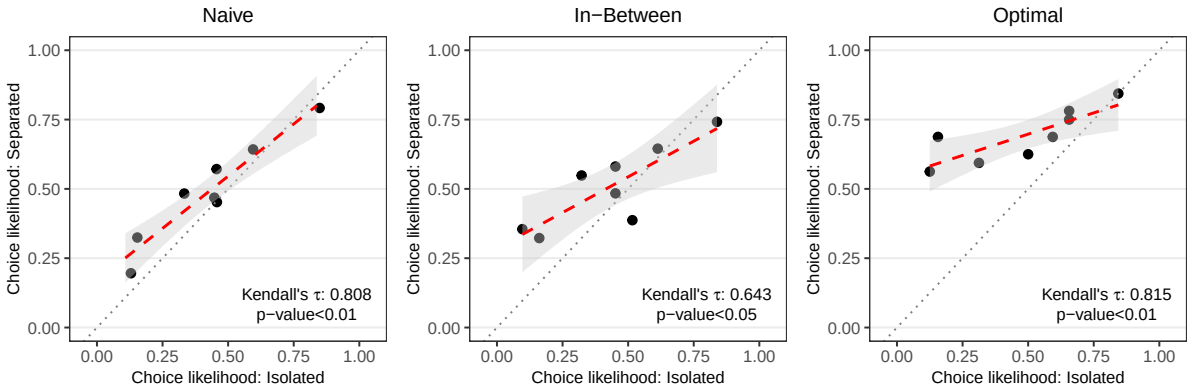


Figure 1.12: Common Source Cancellation in Information Bundle Choices – By Group *Notes: In each panel, the y-axis plots the likelihood of choosing bundle $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$ under the Separated setting, and the x-axis plots the likelihood of choosing σ over σ' under the 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.*

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 A.3 in Appendix A.1 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.

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*

Table 1.3: Choices Between Information Bundles – By Group

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

*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.*

between information bundles of each group of subjects, including those who make perfect use of each information bundle in the guessing game.

1.5 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.[20] Aligning with GOY, subjects' information choices in the current experiment also exhibit a sharp aversion to non-instrumental informativeness.

The left panel of Figure 1.13 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,

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

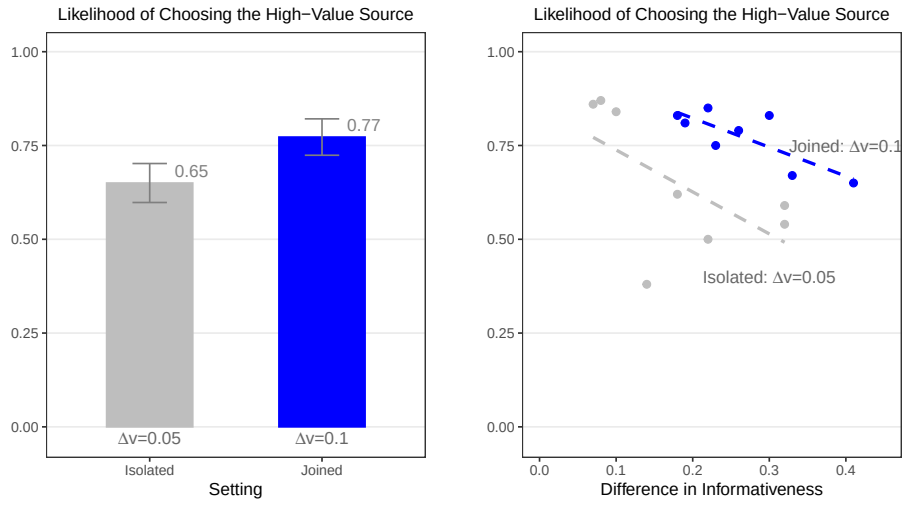


Figure 1.13: Choices between single Information sources *Notes: In both panels, choice likelihood denotes the likelihood of choosing the high-value information source in binary choices. Short vertical lines in the left panel denote 95 percent confidence intervals. Δv denotes the difference in instrumental 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.*

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 1.13 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 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 1.4. 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.

Table 1.4: Informativeness Aversion in Information Choices

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

*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.*

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.

1.6 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.

Chapter 2

Too Much Information

with Ryan Oprea and Sevgi Yuksel

2.1 Introduction

In this paper we experimentally study how people's demand for information structures is shaped by *informativeness*. Informativeness is the basic descriptive characteristic of information structures, measuring the reduction in uncertainty an information source is expected to induce (Frankel & Kamenica 2019). In standard economic theory informativeness influences information demand only to the extent that it produces instrumental value – i.e., to the extent that it is expected to improve decision making in relevant decision tasks. To the extent this is true, conditional on instrumental value, decision makers should be indifferent to informativeness and it should therefore have no direct impact on information demand. The goal of our paper is to treat this standard theoretical assumption as a null hypothesis, and compare it to two natural alternatives.

First, decision makers may directly value informativeness, above and beyond its contribution to instrumental value. That is, perhaps due to natural human curiosity, distaste

for residual uncertainty, caution or deep-seated information-seeking heuristics that arise from the free disposal nature of information, people strictly prefer more informative information structures to less, even when holding instrumental value fixed. Second, it is possible that people instead display an *aversion* to informativeness that does not contribute directly to instrumental value. Richer (more informative) information structures are, after all, more complex in the sense that they require more intensive information processing to properly evaluate. If decision makers are unable (or unwilling) to bear the costs of fully understanding these structures, they may put a smaller premium on them relative to simpler (less informative) structures. Thus, an alternative hypothesis is that, conditional on instrumental value, demand for information falls with informativeness.

Studying people's demand for informativeness is difficult because it is easily confounded with other forces that shape and distort information demand. First, in typical *experimental* paradigms, the demand for information is confounded by well-known mistakes people make in *interpreting* and making use of information. Perhaps most importantly, most prior research on information in experimental economics is conducted in prior-signal updating settings in which subjects must apply Bayes' rule properly before they can *even understand* how a piece of information will influence their beliefs and actions. Because people have systematic tendencies to violate Bayes' rule, typical methods therefore run the risk of confounding systematic *confusion* about how information informs choice with *preferences* for information. Second, in typical *naturally occurring* observational settings, informativeness is easily confounded with drivers of taste for information that are difficult to theoretically operationalize and are therefore difficult to measure and control. For instance, sources of information may vary in how *entertaining* or *worrisome* they are, producing or inhibiting demand for information for reasons that have little to do with informativeness. These "affective characteristics" of information structures may be easily confounded with informativeness in ways that are difficult to

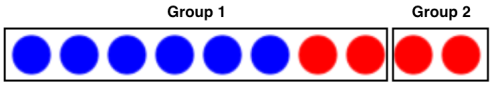


Figure 2.1: Example on the Representation of Information Structures in the Experiment *Notes: The subjects' task is to guess the color of a randomly selected ball. The information structure reveals whether the randomly selected ball is in Group 1 or Group 2.*

formally account for in measurement exercises.

Our contribution is to propose and implement a method for eliciting preferences for information that is free from these confounds, allowing us to cleanly measure how informativeness shapes information demand. First, like most experiments in this literature, we study a simple setting in which subjects (i) must guess a binary state of the world (red or blue) but (ii) first receive a signal that may improve the accuracy of that guess. Because this experimental paradigm is completely abstract, variation in informativeness and instrumental value are unlikely to be confounded with affective characteristics of information (e.g., variation in how "entertaining" a source of information is) that might influence demand in uncontrolled or framed settings. Second, unlike most experiments, we present information in a way that doesn't require difficult applications of Bayes' rule. Instead, we present information as in Figure 2.1 by showing subjects (i) the prior as a set of ten balls (six blue, four red), one of which will be randomly selected to determine the true state and (ii) signals as subsets of these balls that, together, partition the ten balls (signals are drawn as boxes around balls – in the example the ten balls are partitioned into two subsets, so there are two possible signals). Subjects are told which subset the actually-selected ball is from before guessing the state. We find that this way of presenting signals entirely removes classical biases like over-under inference, motivated reasoning and confirmation bias: upon receiving a signal, subjects make the rational Bayesian decision (i.e., make an optimal guess about the color of the ball) 98% of the time.

We use these techniques to elicit subjects' preferences over sixteen distinct informa-

tion structures, presented in this debiased way, with each structure corresponding to a unique partition. This variation across structures independently varies informativeness and instrumental value, allowing us to judge between our motivating hypotheses. We measure preferences in two ways. First, we elicit weak ordinal preferences by having subjects rank structures in order of preferences –higher-ranked information structures are more likely to be assigned to them for a payoff-relevent future choice, making the weak ranking incentive compatible. Second, we elicit strict cardinal preferences by eliciting subjects' willingness to pay to receive this information structure in a future choice, using an incentive compatible BDM mechanism (Becker, Degroot & Marschak 1964).

Our first main finding is that subjects' information preferences are strikingly sensitive to instrumental value. Subjects, on average, strictly prefer more instrumentally valuable information structures to less and sometimes reveal median valuations that are reasonably close to true instrumental value. However, we also find significant failures to rank information structures in terms of instrumental value and clear evidence that some information structures are significantly mis-valued.

Our second main finding is that conditional on instrumental value, subjects display a strong *aversion* to informativeness. When comparing two information structures with the same instrumental value, subjects tend to strictly prefer the less informative of the two. Indeed, our elicitations show that subjects are often willing to pay strictly less for information structures that are more informative. Using agnostic clustering techniques, we show that 2/3 of subjects display this strict aversion to informativeness (conditional on instrumental value) while a smaller cluster covering 1/4 of subjects displays behavior that suggests a preference *for* informativeness. We show that the dominant aversion to informativeness in the population is sometimes severe enough to make subjects prefer less instrumentally valuable information to more valuable alternatives. Indeed, our results suggest that aversion to informativeness is an important driver of failures to rank

information structures according to instrumental value.

In the final part of the paper we examine *why* subjects display an aversion to informativeness that can't be used to improve choice. Our design allows us to rule out a number of salient hypotheses. First, we show that the results cannot be rationalized as an outgrowth of subjects' preferences for the timing of information: our results go in the opposite direction of recent measures of such preferences (Nielsen 2020). Indeed, our results suggest that this behavior is likely to be unrelated to uncertainty or risk preferences of any sort: our design includes a treatment in which we remove uncertainty (and with it preferences related to uncertainty) from these tasks while maintaining identical information processing in the valuation task, and document broadly similar results. Second, as discussed above, our design rules out classical inferential errors in assessing the information produced by these structures: subjects in our tasks show no signs of Bayesian errors like over-/under-inference, confirmation bias or motivated reasoning when making use of information. This means that aversion to informativeness can't be driven by rational anticipation of systematic mistakes in the use of information. Third, our design rules out the possibility that aversion to informativeness is due to inability to reason about how informative structures will be used to inform choice. In a diagnostic treatment, we have subjects make choices for every possible signal from every information structure *before* evaluating any of them, and remind subjects of these choices at the evaluation stage. Aversion to informativeness is no less strong in this treatment, suggesting that failures to contingently reason about the use of information does not underlie this result.

Instead, our results suggest that aversion to informativeness is a consequence of the fact that more informative structures are more costly to precisely evaluate and are therefore less well-understood by subjects, making them less attractive. Several pieces of evidence point towards this "complexity" interpretation, rooted in the costs and difficulties of evaluation. First, more informative information structures require significantly

more time to evaluate than less informative structures: decision time or runtime is a direct resource cost of information processing that is often used in the literature (and throughout computer science) as a measure of complexity and effort.[1] Second, this is likely due to the fact that more informative information structures tend to consist of a larger number of strongly differentiated pieces of information that have to be *aggregated* in order to properly value them (e.g., they tend to contain more possible signals and induce more heavily differentiated distinct posteriors). Thus, in a very direct sense, properly valuing informative structures requires more work on the part of decision makers. Finally, we ran a treatment that removes uncertainty from information structures, leaving only the complexity of aggregating their features as a potential driver of misvaluations. We find similar aversion to "informative" structures in this data, strongly suggesting that the costs and difficulties of evaluation are the primary driver of this aversion to informativeness.

Taken together, our findings suggest that more informative information structures are less desirable, ceteris paribus, because they are more costly or difficult to evaluate, leading subjects to undervalue them. It is important to be clear that this is a ceteris paribus conclusion, made on the basis of a tightly controlled experiment designed to deliberately isolate important primitives of interest to information economics. Clearly, in applications many other characteristics of information that are harder to account for in economic theory compete with informativeness (and instrumental value) to shape information demand. For instance, people often pursue information that is informative but not instrumentally valuable because it is *entertaining* or *interesting*, leading them to demand informative but instrumentally useless trivia – characteristics that we do not

[1]Decision time is controversial as a complexity measure in some settings because subjects may choose to spend less time on more difficult problems (i.e., problems that seem too difficult to correctly solve). This is less of a problem in our setting because subjects virtually always make optimal decisions, conditional on information.

yet know how to model, measure or control. Because of this, our experiment (like most experiments and indeed most models) deliberately brackets off these kinds of affective drivers of information demand in order to study the influence of primitives of information structures that we know how to measure and interpret. Doing this, we find that aversion to informativeness is substantial and sometimes strong enough to cause subjects to prefer less instrumentally valuable information to more, leaving accuracy and earnings "on the table".

Our paper contributes to several literatures.

First, we make a methodological contribution to the growing experimental literature studying information demand. Our design allows us to study demand for information in a setting where making optimal use of information is very easy and does not require complex Bayesian reasoning. Thus, our visual representation of information structures successfully excludes non-Bayesian reasoning or misinterpretation of information structures as confounds for studying the demand for information. While we use these techniques to study informativeness, they can be easily applied to study many other questions about people's taste for information.

Second, our paper adds to a growing literature studying how factors other than instrumental value influence information demand. Prior studies have experimentally or theoretically examined the role of confirmation seeking (Charness, Oprea & Yuksel 2021, Montanari & Nunnari 2022), preference for certainty (Ambuehl & Li 2018, Novak, Matveenko & Ravaioli 2023), timing of resolution of uncertainty (Grant, Kajii & Polak 1998, Eliaz & Schotter 2007, 2010, Nielsen 2020, Falk & Zimmermann 2022, Je 2023), skewness of information (Masatlioglu, Orhun & Raymond 2023), anticipatory feelings (Caplin & Leahy 2001), motivated attention (Falk & Zimmermann 2022, Golman & Loewenstein 2018, Golman, Loewenstein, Molnar & Saccardo 2022), and behavioral motivations stimulated by changes in beliefs like disappointment aversion (Palacios-Huerta

1999, Dillenberger 2010, Andries & Haddad 2020), loss aversion (Koszegi & Rabin 2009), dissonance avoidance (Festinger 1957), suspense and surprise (Ely, Frankel & Kamenica 2015*a*), etc., play in information demand. Our contribution to this literature is to study how demand is influenced by "informativeness", the basic descriptive characteristic of information structures. We show that informativeness has a powerful influence above and beyond its contribution to instrumental value. Because informativeness is a fundamental and universal characteristic of information structures, our findings have particularly wide-spread normative implications for information design and positive implications for predicting and interpreting information demand.

Most closely related to our study in this literature is Liang (2023), a concurrent paper that studies suboptimal valuation of information structures and provides evidence that is broadly supportive of the mechanism underlying our main results. In particular, his results suggest that subjects have difficulty foreseeing and integrating payoffs from multiple information-contingent choices, but it is mostly difficulties with integration which get in the way of optimal valuation. Specifically, in diagnostic treatments in which information-contingent choices are predetermined and presented as such, subjects behave more optimally. We also find suboptimal information demand that seems to derive from similar difficulties in evaluating information structures. Our results also link these types of difficulties in identifying instrumental value to informativeness.

Third our study provides a new kind of evidence in support of the central trade-off at the heart of rational inattention models: people acquire information to maximize utility net of information costs (Sims 2003, Matějka & McKay 2015, Caplin & Dean 2013). These models assume that decision makers face information costs that are (in typical parameterizations) increasing in Shannon mutual information between prior and posterior beliefs – the same measure of informativeness we use in most of our empirical work. Our paper contributes to this literature by expanding our understanding of when and

why agents act as if information is costly. The rational inattention literature, strictly speaking, assumes that information costs are costs of information *gathering* or *extraction* (this is why they are called "inattention" models). For instance, state-of-the-art experiments on rational inattention typically ask subjects to extract information from complex visual images (Dewan & Neligh 2020, Caplin, Csaba, Leahy & Nov 2020, Dean & Neligh 2023) or from a series of equations (Ambuehl, Ockenfels & Stewart 2022). By contrast, we deliberately minimize information gathering costs by giving subjects direct and easily interpreted information on the state of the world. The fact that we nonetheless observe an aversion to informativeness suggests that information costs are not driven only by the cognitive effort required to *gather* or *extract* information, but are also driven by the cognitive effort required to *evaluate* the ex-ante value of information. Our results therefore suggest that rational inattention models may also be effective models of complexity (information processing) aversion, and may therefore have a much wider scope of application than is typically supposed.

Finally, our study relates to a growing literature showing the role complexity plays in a wide-range of economically important settings. Recent work suggests that people dislike engaging in complex behaviors (Oprea 2020), that this distaste has a strong distorting effect on choice (Banovetz & Oprea 2023), and that complexity limits and distorts the kinds of beliefs people form (Kendall & Oprea 2023). As a result, complexity (broadly defined as a cognitive processing costs) has been shown in recent work to be a major driver of behavioral anomalies in a number of canonical choice settings including, e.g., lottery anomalies (Enke & Graeber 2023, Oprea 2023), intertemporal choice anomalies (Enke, Graeber & Oprea 2023) and failures of Bayesian reasoning (Ba, Bohren & Imas 2023). Our work complements and extends this literature by providing evidence that valuations for a very different (but no less canonical) choice object (information structures) are also fundamentally shaped by complexity, leading to systematic anomalies in information

demand.

The remainder of the paper is organized as follows. Section 2.2 presents our theoretical framework. Section 2.3 describes the experimental design. Section 2.4 presents results and Section 2.5 examines the mechanism driving these results. Section 2.6 concludes by discussing the implications of our results.

2.2 Theoretical Framework and Behavioral Hypotheses

2.2.1 Informativeness and Instrumental Value

Consider a finite state space Ω, with a typical state denoted by ω. The prior distribution on Ω is denoted by p. An information structure σ is a stochastic mapping from the state space Ω to a finite set of signals S. It is useful to think of σ as inducing a distribution over posteriors.[2] That is, given p, an information structure σ induces (i) a distribution q_σ over S and (ii) conditional on each signal s, a posterior distribution p^s_σ over the state space.

The amount of information generated by an information structure is described by a metric we will call its *informativeness*: the expected *reduction in uncertainty* induced by the information structure (see Frankel & Kamenica (2019) for an in-depth discussion). Several measures of informativeness can be defined because several metrics of "uncertainty" (and thereby "uncertainty reduction") can be selected for the purpose. For instance, the most prominent measure in the literature is based on Shannon entropy (Shannon 1948), a canonical measure of uncertainty in beliefs defined as

[2]For each ω, let $\sigma_\omega(s) \in \Delta(S)$ denote the probability that signal s is realized. The probability of observing signal s is $q_\sigma(s) := \sum_\omega p(\omega)\sigma_\omega(s)$. For each signal, the posterior distribution on Ω can be computed using Bayes' rule. Conditional on each signal s, $p^s_\sigma(\omega) = \frac{p(\omega)\sigma_\omega(s)}{q_\sigma(s)}$.

$H(p) = -\sum_{\omega\in\Omega} p(\omega)\ln p(\omega)$. As characterized in Cabrales, Gossner & Serrano (2013), the *entropy informativeness* of information structure σ is the expected reduction of entropy of the decision-maker's beliefs as a result of the information conveyed by σ, that is, the Shannon mutual information between prior and posterior beliefs:

$$I_\sigma = H(p) - \sum_{s\in S} q(s)H(p^s). \tag{2.1}$$

This measure of informativeness is equal to zero when σ carries no information (posterior always equal prior) and is maximized at $H(p)$ when σ fully reveals the state. In summary, entropy informativeness provides a numeric measure of informativeness independent of the decision problem, allowing complete ordering of information structures. When studying "informativeness" we will focus on this entropy-based metric throughout the paper, but in Appendix B.6 we show that little depends on this choice: our results are robust to varying the specific definition of informativeness we use.[3]

In standard economic theory, the value of an information structure to a decision-maker (DM) depends on the decision problem the information structure is meant to inform. Suppose the DM faces a decision problem in which she observes signal s from information structure σ and takes action $a \in A$ to maximize $\mathbb{E}[u(a,\omega)|s] := \sum p^s(\omega)u(a,\omega)$, where $u(a,\omega)$ describes the decision-maker's state-dependent utility function. The *instrumental value* (or simply *value*) of σ, given the set of actions A available and utility function u, is the expected increase in utility made possible by the DM being able to condition her

[3]For instance, an alternative ordering of informativeness across information structures is provided by Blackwell (1953). According to Blackwell's ordering, an information structure is more informative than another whenever the latter is a garbling of the former, i.e., signals form the less informative structure can be interpreted as observing those from the more informative one with noise. Blackwell requires a strong condition for the comparison between information structures. By Blackwell's Theorem, a more informative structure (according to Blackwell ordering) generates higher instrumental value (as defined later in this section) in *any* decision problem. Thus, Blackwell provides only a partial order of informativeness, making it less useful for our purposes than entropy informativeness.

action on the realized signals:

$$V_\sigma = \underbrace{\sum_{s\in S} q(s) \underbrace{\max_{a\in A} \mathbb{E}[u(a,\omega)|s]}_{\substack{\text{Expected utility} \\ \text{conditional on } s}} - \underbrace{\max_{a\in A} \mathbb{E}[u(a,\omega)]}_{\substack{\text{Expected utility} \\ \text{without } s}}}_{\text{Expectation over } s}.$$

Note that, when u is denominated in money, V_σ can further be interpreted as the the greatest price a rational decision-maker would be willing to pay for information from σ before facing a specific decision problem. Instrumental value is the key characteristic shaping information demand (i.e., preferences for information) in standard information economics.

Although informativeness and instrumental value are related, they are not the same thing. When comparing information structures σ and σ', it is possible for σ to be as (or even more) valuable than σ' while being less entropy informative. The reason for this is intuitive: in the context of any given decision problem it is possible for an information structure to reduce uncertainty in ways that are not useful for informing choice. This observation is what motivates our experiment.

2.2.2 Question and Hypotheses

Our question is how informativeness shapes people's preferences for (or demand for) information structures. As suggested above, economic theory gives a clear answer to this question: informativeness influences the demand for information *only* to the extent that it improves expected utility in a decision problem by allowing the decision maker to make better choices..

H0. Conditional on instrumental value, demand is not influenced by informativeness.

H0 hypothesizes that people evaluate information structures exclusively through the

lens of the relevant decision problem that the information will be used to inform. An alternative possibility is that evaluation of information might be at least partially divorced from the specifics of the decision problem at hand. That is, it may be that decision makers' demand for information is influenced not only by the decision but also (at least in part) by characteristics of the information structure itself.

For instance, decision makers might prefer *more informative structures*, even when that informativeness is not useful for improving decision making. One reason for this might be that decision makers are drawn to more information as a hedge against the possibility that they've misunderstood how they would use that information. Or taste for informativeness might arise from rules of thumb, culled from natural life, that often instruct us to cautiously seek out information even when it is not immediately obvious how to use it. After all, information can always be disregarded if it doesn't prove to be useful. A final possibility is *curiosity* – a direct preference for more over less information in information sources (Golman & Loewenstein 2018, Golman, Loewenstein, Molnar & Saccardo 2022)– which might cause people to, ceteris paribus, prefer more informative information structures to less. We state this broad possibility as a second hypothesis:

H1. Conditional on instrumental value, demand for information is increasing with informativeness.

A final possibility is that people instead display an aversion to informativeness, particularly when additional information is not useful. Why might this be the case? Perhaps the most salient possibility is that information structures with high informativeness require more *information processing* and are therefore more costly to properly interpret and evaluate. That is, more informative structures may be more *complex* to process, using the definition of the term from computer science. Indeed, the idea that entropy reduction is costly to decision makers is an assumption often made in models of bounded rationality like rational inattention models (Sims 2003, Matějka & McKay 2015, Caplin

& Dean 2013), though typically in settings somewhat different from ours. If decision makers are unable (or unwilling) to bear the costs of fully understanding these structures, they may put a smaller premium on them relative to simpler (less informative) structures, ceteris paribus. We state this possibility as a final hypothesis.

H2. Conditional on instrumental value, demand for information is decreasing with informativeness.

Our experiment, discussed below, is designed to distinguish between these hypotheses. In particular we designed our experiment to study (i) to what degree instrumental value predicts how people rank and bid on information structures and (ii) to what degree informativeness acts as an additional driver of information demand once its contribution to instrumental value is accounted for.

2.2.3 A Guessing Task

To study the questions posed above, we focus on a simple guessing task in our experiment. The state of the world is binary, i.e. $\Omega = \{b, r\}$, where b (as will be clear in the next section) can be interpreted as referring to the color *blue* and r referring to the color *red*. The decision-maker's prior is fixed at $p := p(b) = 0.6$. The decision-maker takes a binary action, $A = \{b, r\}$, with the goal of matching the state such that $u(a, \omega) = B$ (where B is a bonus) if $a = \omega$ and zero otherwise.

With this simple specification, the instrumental value of an information structure reduces to the following:

$$V_\sigma = \Big(\underbrace{\sum_{s \in S} q(s) \underbrace{\max\{p^s, 1 - p^s\}}_{\substack{\text{Guessing accuracy} \\ \text{conditional on } s}}}_{\text{Expectation over } s} - \underbrace{p}_{\substack{\text{Guessing accuracy} \\ \text{without information}}} \Big) B. \tag{2.2}$$

Note that the expected utility of the decision-maker in this problem is equal to their guessing accuracy times the bonus associated with guessing correctly.[4] Thus, the value of an information structure for a decision maker-facing such a guessing task is directly linked to the expected improvement in their guessing accuracy. Although this setting is simple, as we will show in the next section, it is rich enough to allow us to construct a set of information structures that independently vary in instrumental value and informativeness, allowing us to test our hypotheses.

2.3 Experimental Design

2.3.1 Guessing Task and Representation of Information

We built our experimental design around the simple guessing task introduced in the last section. A random ball is drawn from a set of 10 which always consists of six blue balls and four red balls. The subjects' task is to correctly guess the color of that randomly selected ball. If they correctly guess the ball's color, they earn a bonus payment of $10.

Before making their guess, subjects receive partial information about the randomly selected ball from an information structure. Each information structure is represented as a partition of the 10 balls. The information structure provides information about the randomly selected ball by revealing to the subject *which cell of the partition* the randomly selected ball belongs to. Thus, each cell of the partition is a distinct signal that the structure might generate. The size of any cell visually represents the probability with which that signal will be realized, and the composition of the balls within each cell visually represents the posterior probability that the ball is blue or red conditional

[4]For example, without additional information, the decision-maker guesses the state to be b (since $p > 0.5$). This guess is correct with probability p. Similarly, conditional on signal realization s, the decision-maker guesses the state to be the most likely state. Such a guess is correct with probability $\max\{p^s, 1-p^s\}$.

on that signal being realized. Presenting information in this way therefore makes the characteristics of an information structure particularly transparent to subjects.

Figure 2.1 provides an example of an information structure that partitions the 10 balls into two cells (two possible signals). The first cell (labeled "Group 1") consists of six blue balls and two red balls; the second cell (labeled "Group 2") consists of the remaining two red balls. This information source provides information about the randomly selected ball by revealing which cell (Group 1 vs. Group 2) the randomly selected ball belongs to. Formally, the information structure generates a binary signal. The first (second) signal—corresponding to Group 1 (Group 2)—is realized with 80 (20) percent probability. Conditional on the first (second) signal, the posterior probability that the color of the randomly selected ball is blue is 75 (zero) percent.

We study a total of 16 information structures, depicted in Figure 2.2, which were selected to independently vary instrumental value and informativeness (as well as other characteristics). Table B.10 in Appendix B.5 provides a comparison of these information structures on both of these (and other) measures. For each of these information structures we (i) study how people make use of these information structures (by eliciting guesses for each information structure and signal) and (ii) study how people value these information structures (by eliciting rank preferences over and willingness-to-pay for these structures).

2.3.2 Eliciting Preferences for Information Structures

The main section of the experiment elicits subjects' preferences for the 16 information structures depicted in Figure 2.2 (we will call this the "Demand" section). We do this in two distinct ways, each of which has advantages and disadvantages. Figure 2.3 shows screenshots of each.

Ranking: Subjects are asked to rank the 16 information structures from most preferred

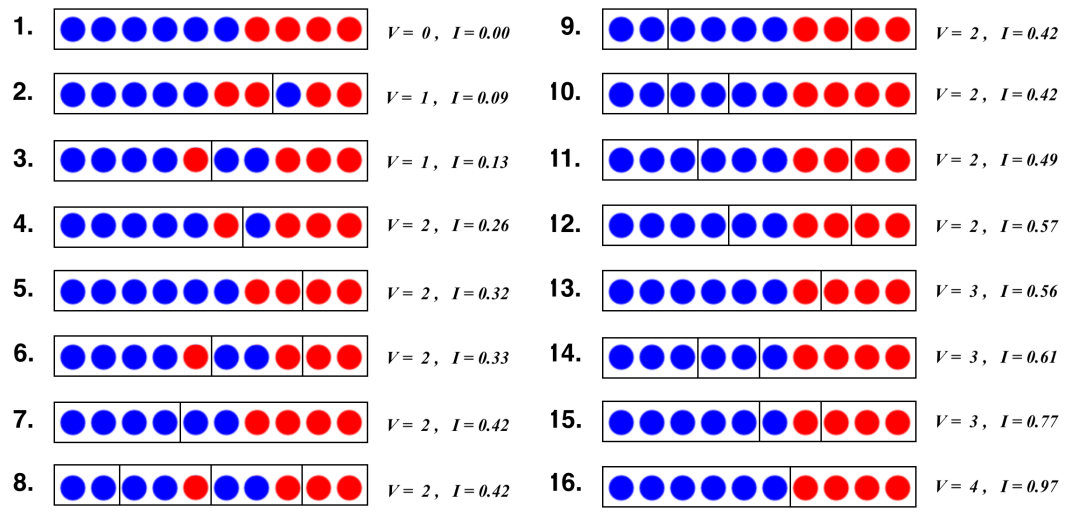
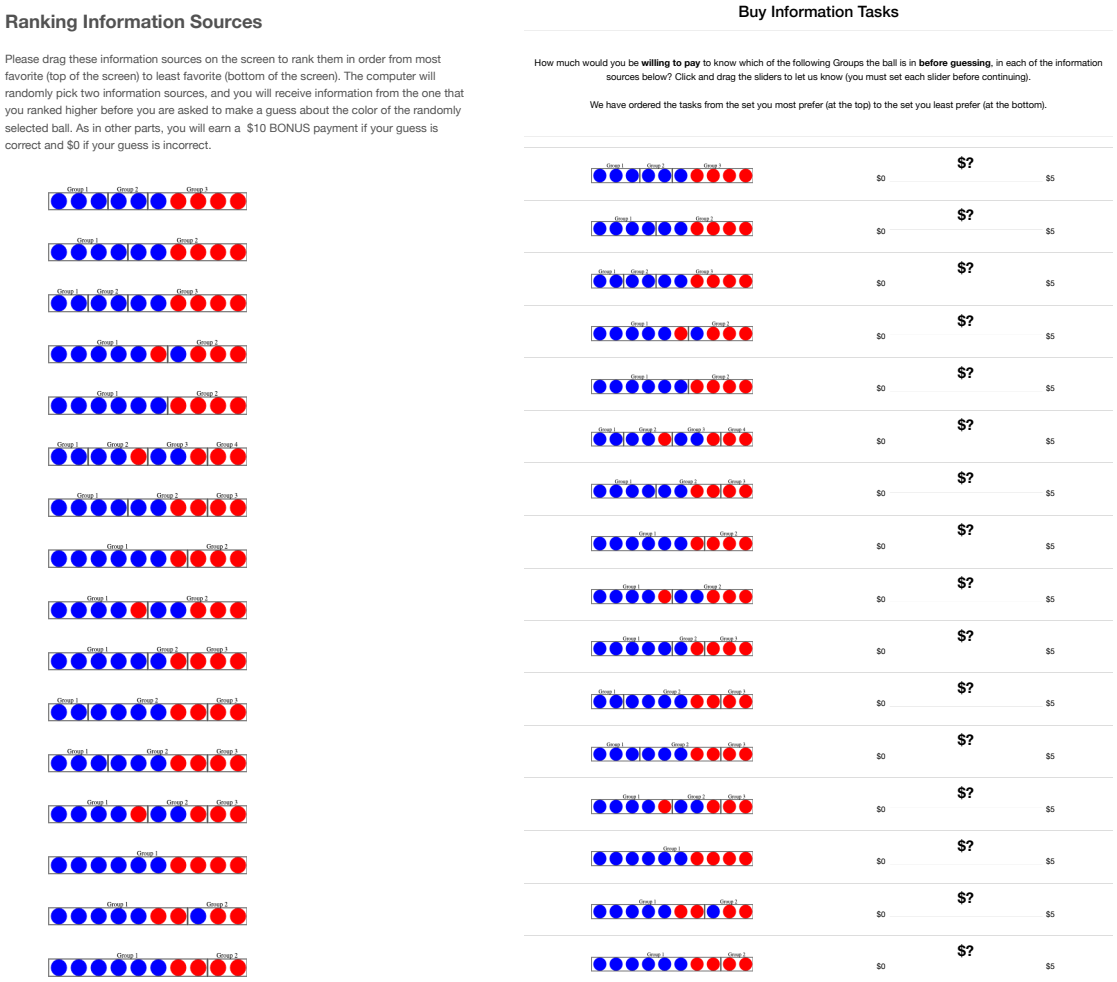


Figure 2.2: The 16 Information Structures used in the Experiment *Notes: The information structures are listed lexicographical in order of instrumental value (V as defined in equation 2.2) and entropy informativeness (I as defined in equation 2.1) . To make the partitions more salient in the experiment, each cell of the partition was labeled as "Group 1", "Group 2," etc. See Appendix B.5 for details on other characteristics of these information structures.*

to least preferred. Specifically, the 16 structures are presented to subjects (in an order randomized on the individual level) and subjects are tasked with reordering them using a drag-and-drop interface. Subjects are incentivized to place more preferred information structures above (higher in the list than) less preferred structures: structures that are placed higher on the list are more likely to be assigned to subjects for a paid guessing task that occurs at the end of the experiment. The advantage of the Ranking elicitation is that it is extremely simple and intuitive for subjects, likely providing cleaner evidence of rank preferences. The disadvantage is that, strictly speaking, this method measures only a weak preference ordering: subjects who are indifferent between two structures nonetheless must rank one higher than another.

WTP: After ranking information structures, subjects are given an endowment of $5 and shown the information structures in the order they ranked them. For each of the 16 information structures, subjects are then asked to express (using a slider) their (maximum) willingness to pay (WTP) to receive information from that structure in a guessing task



(a) Ranking (b) WTP

Figure 2.3: Eliciting Demand for Information

that will occur at the end of the experiment. Incentives are provided using the Becker-Degroot-Marschak (BDM) method (Becker, Degroot & Marschak 1964) and if subjects do not purchase the information structure, they will receive no information to inform their guess. The advantage of the WTP elicitation is that it measures strict preferences. The disadvantage is that elicited WTP (using BDM and related mechanisms) is much more complicated than Ranking, and is therefore known to be noisy and subject to biases

(in particular, the pull-to-the-center effect in Danz, Vesterlund & Wilson (2022)).[5]

We will make use of each of these elicitations in our analysis for robustness, allowing each method to compensate for weaknesses of the other. In some of our analysis we will do this in an explicit way. Because of known noisiness in WTP elicitations, when comparing WTP between different information structures, in addition to reporting aggregate results, we will also report results for the subset of observations in which WTP comparisons agree with the ordering elicited in the simpler (and arguably therefore more reliable) Ranking elicitations.

2.3.3 Eliciting Guesses

In the Guesses section of the experiment, we ask subjects to make a guess of the ball's color for each possible signal in each of the information structures depicted in Figure 2.2. Figure 2.4 shows a screenshot from this task. We remind subjects of the task and incentives and then show subjects a partition of the blue and red balls, labeling each element as a "group." At the bottom, subjects are asked to give their guess of the ball's color for each possible group the ball might be in (for each possible signal they might receive).

The purpose of the Guesses section of the experiment is to study how people make use of information from information structures and whether subjects use that information in a suboptimal way. This information is important for interpreting subjects' demand for information structures.

[5]WTP elicitation also gives us quantiative measures of value that can be compared to theoretical benchmarks. For a risk neutral agent, WTP should be \$10 times the increase in guessing accuracy enabled by each information structure (as captured in Equation 2.2). In Appendix B.1, Figure B.1 shows that the WTP of a reasonably risk averse or loving agent does not deviate much from the WTP of the risk neutral agent. In much of our analysis we focus on relative comparison of WTP amounts (whether a subject is willing to pay more for one information structure relative to another). These comparisons should be determined entirely by instrumental value (as defined by Equation 2.2) independent of risk preferences, under standard theory.

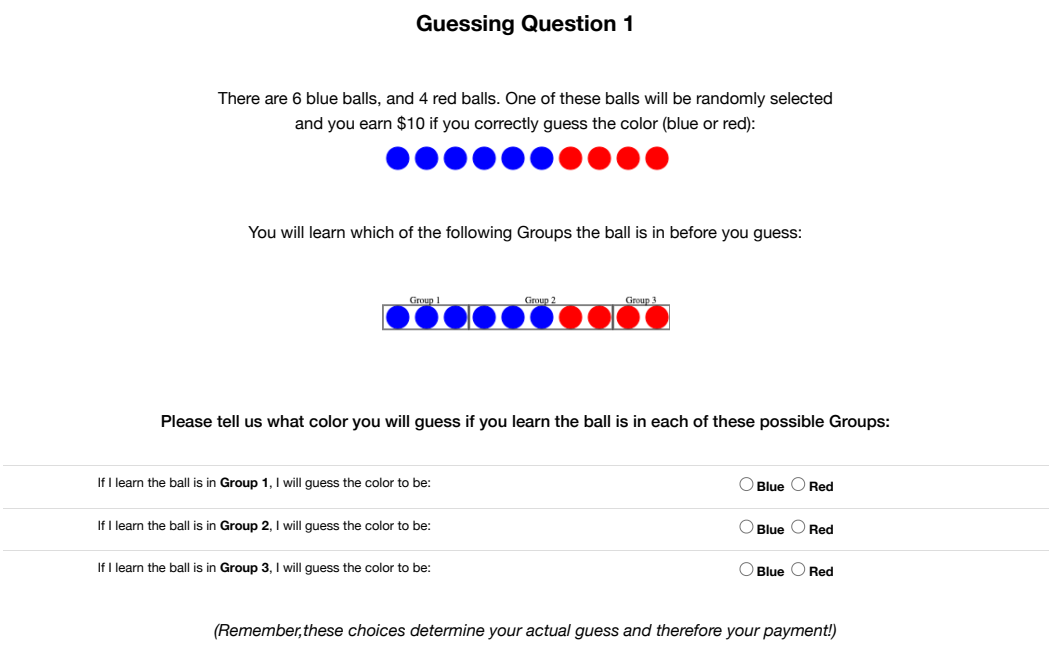


Figure 2.4: Elicit Guesses

2.3.4 Treatment Variations

In our **Baseline** treatment ($N = 109$ subjects), we ask subjects to perform the Demand section of the experiment (the Ranking and WTP elicitations) first, and the Guesses section afterwards. To this we add two diagnostic treatments that will help us to interpret our results.

First, in our **Reverse** treatment ($N = 54$ subjects), we reversed the order: subjects were assigned Guesses first and Demand afterwards. In these sessions, during the Demand section, subjects were actually shown the guesses they had made earlier, conditional on each possible signal for that information structure.[6] This information was not binding, but was designed to remind subjects of how different signal realizations are likely to generate different guessing patterns. The purpose of this treatment was to study whether having already made use of information structures (and being reminded of how they are

[6]See Figure B.7 in Appendix B.6 for a screenshot how this was displayed to subjects.

used) improves subjects' identification of instrumental value in guiding their demand. This will be useful for understanding the mechanism behind our results.

Second, in our **No Uncertainty** treatment ($N = 61$ subjects), we removed uncertainty from the design using techniques employed by Martinez-Marquina, Niederle & Vespa (2019) and Oprea (2023). Specifically, instead of drawing only one ball, we informed subjects that we would draw *all balls* and pay subjects based on the accuracy of their guess for each of the balls.[7,8] Everything else in the experiment remains identical. Doing this retains much of the information processing involved in evaluating and interpreting information structures, but removes scope for risk, uncertainty or timing preferences (e.g., preferences for the timing of the resolution of information). Again, this data will be useful for understanding the mechanism behind our results.

2.3.5 Incentives and Implementation Details

All sessions were conducted at the LITE laboratory at the University of California, Santa Barbara between December 2021 and March 2023. We recruited subjects from across the curriculum to participate in 15 sessions using the ORSEE recruiting software (Greiner 2015). The experiment was conducted using software programmed by the authors in Qualtrics. Between 8 and 21 subjects participated in each session and sessions lasted for 30-40 minutes.

All subjects received a show up fee of $7. Subjects' earnings depended on a randomly selected section of the experiment. If the Demand section was selected for payment, we randomized between Ranking and WTP. If Ranking was selected, two information

[7] In this case, the partition associated with an "information" structure constrains the types of guesses subjects can make by requiring them to make the same guess for all balls in the same cell of the partition.

[8] To achieve a clean comparison to the other treatments, subjects received a prize of $1 for each of the balls for which their guesses were correct. This implies, for example, an information structure that enables a guessing accuracy of 90 percent will generate the same expected bonus payment in all three treatments.

structures (from the set of 16) were randomly selected and the subject was assigned to receive information from the one that was ranked higher. If WTP was selected, one information structure (from the set of 16) was randomly selected and whether or not the subject received information from this source was determined according to the BDM mechanism. In either case (regardless of whether Ranking or WTP was selected), at the end of the experiment, the subject was presented with one additional guessing task with information from the selected information structure and subjects were paid based on this guess.[9] If the Guesses section was selected for payment, a random information structure was picked and the subject's guesses for that case were used to determine payment for a randomly drawn ball. Note that in all cases, whether or not subjects received a bonus payment of $10 ultimately depended on the accuracy of their guess about the color of a randomly selected ball.

2.4 Results

2.4.1 Optimality of Guesses

We begin by confirming that our experimental design successfully removes inferential errors like failures of Bayesian reasoning when subjects use signals from information structures.[10] To do this, we study how subjects make use of the information provided to them. We focus on two measures. The *accuracy* of subjects' guesses is the likelihood that those guesses match the state (the true color of the randomly selected ball). The *optimality* of subjects' guesses is the likelihood with which subjects' guesses are optimal

[9]Note that this could mean the subject receives no information depending on the subject's WTP if WTP is selected.

[10]As we show in Section 2.5, none of our results are impacted by the order with which subjects are assigned the Guess and Demand tasks. For this reason, throughout this section, we will pool the Baseline and Reverse treatments; all of our results continue to hold if we instead focus on the Baseline treatment alone.

given the signal realization.[11]

Figure 2.5a presents data on guessing accuracy, and shows how this varies with the instrumental value of the information structure. Note that subjects start with a prior of 0.6; hence, in the absence of any information, we would expect guessing accuracy to be equal to this value if guesses are optimal (i.e. if they guess the color of the ball to blue). We find, indeed, guessing accuracy to be 0.59 in this case. With information structures that are capable of improving guessing accuracy by 10, 20 or 30 percent, we find that guessing accuracy increases to 0.69, 0.79 and 0.89. When the information structure is fully revealing, guessing accuracy goes up to 1. Thus, subjects make near-perfect use of information structures to inform guessing accuracy.

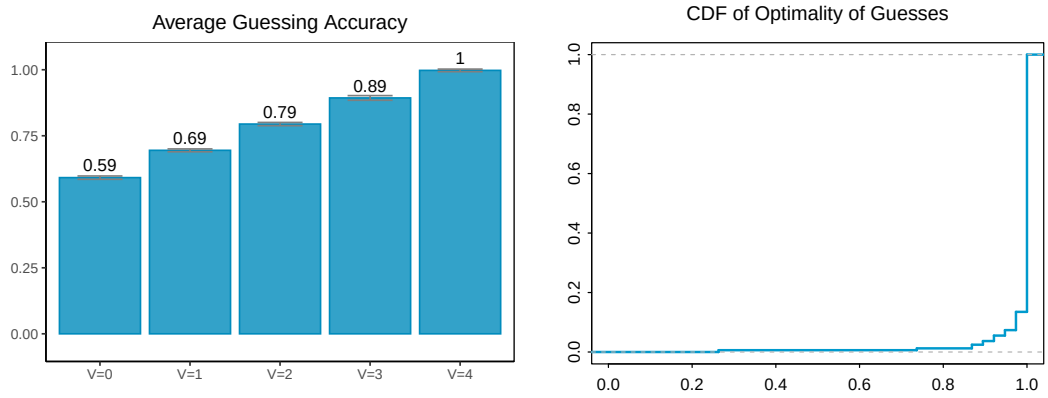
Figure 2.5b shows the optimality rate of guesses, and depicts the distribution of this measure— the share of signals for which the subject's guess is optimal—computed at the individual level. A vast majority of subjects (86 percent) *always* make optimal choices. Overall 98 percent of guesses are optimal conditional on the information available to the subjects. This near-perfect optimality strongly suggests that our methods for providing information avoid Bayesian errors (and other errors like confirmation bias), removing a major barrier to measuring information preferences.

Result 7 *Subjects make near optimal use of information. Guesses conditional on signal are optimal 98% of the time. 86% of subjects make optimal guesses 100% of the time.*

2.4.2 Demand for Information

Figure 2.6 provides a first look at how demand for information is shaped by instrumental value and informativeness, by examining data from our elicitations. Panel (a) of the Figure examines all pairwise ranking comparisons between information structures in

[11] To allow for clean interpretation of this measure, we restrict attention to signals (with Bayesian posterior different from 0.5) for which there is a unique optimal guess.



(a) Guessing Accuracy by Value (b) Distribution of Optimality of Guesses

Figure 2.5: Optimality of the Use of Signals *Notes: In panel (a) V denotes instrumental value (see Equation 2.2 for formal definition). Gray lines denote 95 percent confidence intervals. In panel (b) optimality of guesses is computed on the individual level and denotes the share of signals for which guess was optimal.*

which the one structure is strictly more instrumentally valuable than the other. The bars show the likelihood with which the more instrumentally valuable structure was ranked higher than (as more preferred to) the second. We plot a separate bar for cases in which the value difference is weakly below the median for the dataset ($\Delta_v = 1$) or strictly above the median ($\Delta_v > 1$). When the value difference is relatively low, the optimal structure is preferred 72.7 percent of the time. This increases to 85.3 percent when the value difference is relatively high.

Panel (b) of the same Figure studies how often subjects display a preference for the more *informative* information structure, among all of the pairwise comparisons in which structures can be ranked by informativeness. As in panel (a), the "low" bar represents pairs of information structures in which the difference in informativeness is relatively low (weakly below the median difference of 0.24), while high represents the cases where the difference is high (strictly above the median difference of 0.24). When the difference in informativeness is low, the more informative structure is preferred 60.2 percent of the time. This increases to 75.7 percent when difference in entropy informativeness is

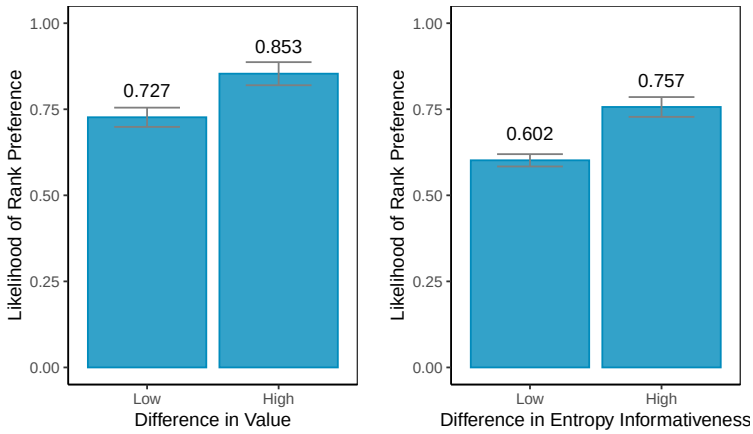


Figure 2.6: Demand for Information by Value and Informativeness *Notes: The figures condition on all pairwise comparisons between information structures where there there is a strict positive difference—on value for (a) or entropy informativeness for (b)—between the first and the second structure. See Section 2.2 for formal definitions of value and informativeness. The bars depict the likelihood with which the first structure was ranked as more preferred to the second.* Low *(*High*) represents all pairwise comparisons where the difference is weakly lower (strictly higher) than the median difference: 1 for value, and 0.24 for entropy informativeness. Gray lines denote 95 percent confidence intervals.*

high.[12] Thus, in uncontrolled comparisons, subjects tend to prefer more informative to less informative information structures.

Figure 2.6 confirms that demand for information is strongly predicted by its instrumental value. We show the same thing in a more disaggregated way in Figure 2.7 by plotting the distributions of responses (for both Ranking and WTP) by the instrumental value of the information structure. The plot shows that 58 percent of subjects treat information structure 1 (see Figure 2.2), which has zero instrumental value, as the least preferred information structure and 83 percent of subjects treat information structure 16, which has the highest possible instrumental value (by fully revealing the state), as the most preferred information structure. Overall, there is first order stochastic dominance between distributions whenever we compare information structures with low value to high value. Similar patterns are also observed in the distribution of WTP amounts.[13]

[12]Here and throughout the results we will focus on measures of "entropy informativeness." However in Appendix B.6, we show that these results also hold with alternative ways of comparing informativeness, including Blackwell ordering and the variance of posterior.

[13]There are some deviations from this pattern in WTP for information structure 16 with value of 4. As expected, the WTP data also displays clear compression often seen with BDM elicitations: subjects on average overpay for information structures of low value and underpay for information structures of

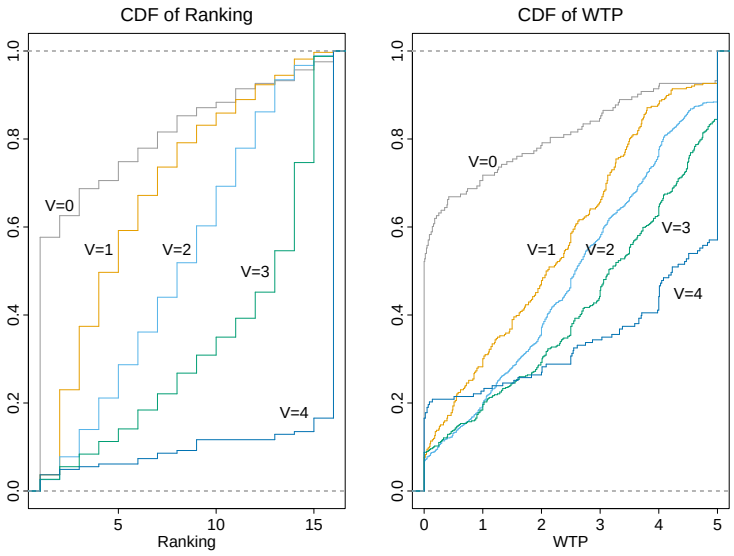


Figure 2.7: Distribution of Rank and WTP by Value *Notes: When subjects order information structures, each structure is assigned a ranking from 1 (least preferred) to 16 (most preferred). V denotes the value of an information structure as defined in equation 2.2.*

Although these results show that (as expected) instrumental value is a major driver of information demand, they also reveal significant deviation from payoff maximizing behavior. As Figure 2.6 shows, in 27.3 percent of relevant cases, subjects rank a less intramentally valuable information structure higher (i.e., as more preferred) than a more valuable information structure when the value difference is relatively low. Even in such cases these mistakes come at significant cost to subjects. By ranking less valuable structures as more preferred, conditional on the pairwise comparison being relevant for payment, subjects reduce their guessing accuracy by 10 percent (22 percent) in *Low* (*High*) cases, leaving approximately \$1 (\$2.2) on the table.[14] Such mistakes are less frequently observed (14.7 percent) when the value difference is higher, but in these cases the mistakes are also twice as costly.

Result 8 *Demand for information is strongly influenced by its instrumental value, but*

high value.

[14]To compute this, we look at subjects' expected bonus payment conditional on each information structure for each of these violations.

there are serious deviations from payoff maximizing behavior. Subjects also tend to prefer more informative structures to less informative structures in the raw data.

2.4.3 Preferences for Informativeness

We now turn to our main questions, by evaluating the hypotheses posed in Section 2.2.2. In the raw data (e.g., Figure 2.6) subjects show a clear preference for more informative information structures, but this measure is confounded with instrumental value. To separate the two notions, we study "excess informativeness": informativeness that does not improve instrumental value, measured by examining the effect of informativeness on information demand in comparisons that hold instrumental value fixed.

Figure 2.8 gives us a first view of the effects of excess informativeness on information demand. Each data point in the Figure represents a pair of information structures. Due to symmetry, we focus on pairs in which the "first" information structure in the pair has a weakly higher value than the "second" one. On the x-axis of both panels we plot the difference in informativeness between the two structures. On the y-axis we plot (i) the likelihood that the weakly higher-value structure is ranked higher in the Ranking elicitation (in the left hand panel) and (ii) the difference in WTP in between the weakly higher and lower value structures (in the right hand panel).[15] We separate the data points (by color and shape) based on the value difference, Δ_v between the two information structures. We also separate out (and show in faded colors) pairs in which at least one information structure is visually disordered (i.e. blue and red balls

[15]WTP data is inconsistent with ranking data in 25 percent of pairwise comparisons. These are cases where one structure is ranked as more preferred to another, but WTP for the former is strictly lower than the other. Many features of the data suggest that Ranking data is a better representation of subjects' preferences than WTP: (i) Subjects on average spend 50 percent more time on Ranking than WTP; (ii) in cases where Ranking and WTP data are inconsistent, value difference is predictive of relative ranking ($p < 0.01$), but not differences in WTP. This is not surprising given the typical noisiness of WTP elicitations in the literature. Thus, to facilitate stronger interpretation of the results, in Appendix B.6 Figure B.6 we re-conduct the analysis of WTP restricting attention to WTPs that are ranked consistently with the Ranking data. We find very similar results in this analysis.

are not shown in orderly clusters) in order to account for the higher visual complexity of these tasks.[16] The graphs also depict linear fits: gray lines include all pairs within a value-difference class, darker lines in the corresponding color restrict attention to pairs within each class where there is no visual disorder, hence providing us with the cleanest comparison.

The results show striking evidence in support of Hypothesis 2: demand for information decreases with informativeness when we control for instrumental value. Focusing first on pairs with identical instrumental values ($\Delta_v = 0$) plotted with green squares, we find that as an information structure becomes more informative (relative to an alternate structure) it (i) becomes less likely to be ranked more highly than the alternative information structure (left panel) and (ii) has a smaller WTP assigned to it relative to the alternative structure (right panel). What is even more striking is that the same pattern is also observed when we look at pairs in which the instrumental value difference between the two information structures rises to 1 (blue circles). Payoff maximizing behavior here requires the high value information structure to be preferred to the low value one with 100 percent probability (left panel). While this instrumental value difference clearly increases the likelihood of preference for the higher value structure (blue circles are almost always above the green squares, conditional on informativeness difference), the likelihood is substantially below one in most of these cases. Furthermore, the likelihood decreases (generating more violations of payoff-maximizing behavior) as the informativeness of the more valuable information structure increases. Similar patterns are observed when we focus on relative WTP (right panel). Eventually, once the instrumental value difference between the two information structures increases to 2 or more (purple or gray triangles,

[16]While this was not anticipated at the design stage, the results clearly reveal subjects to have an aversion towards information structures that are visually disordered (as seen in regressions reported in Appendix B.6 Table B.14). This seems consistent with findings on order-preference that cannot be explained by consumption utility in Evers, Inbar, Loewenstein & Zeelenberg (2014).

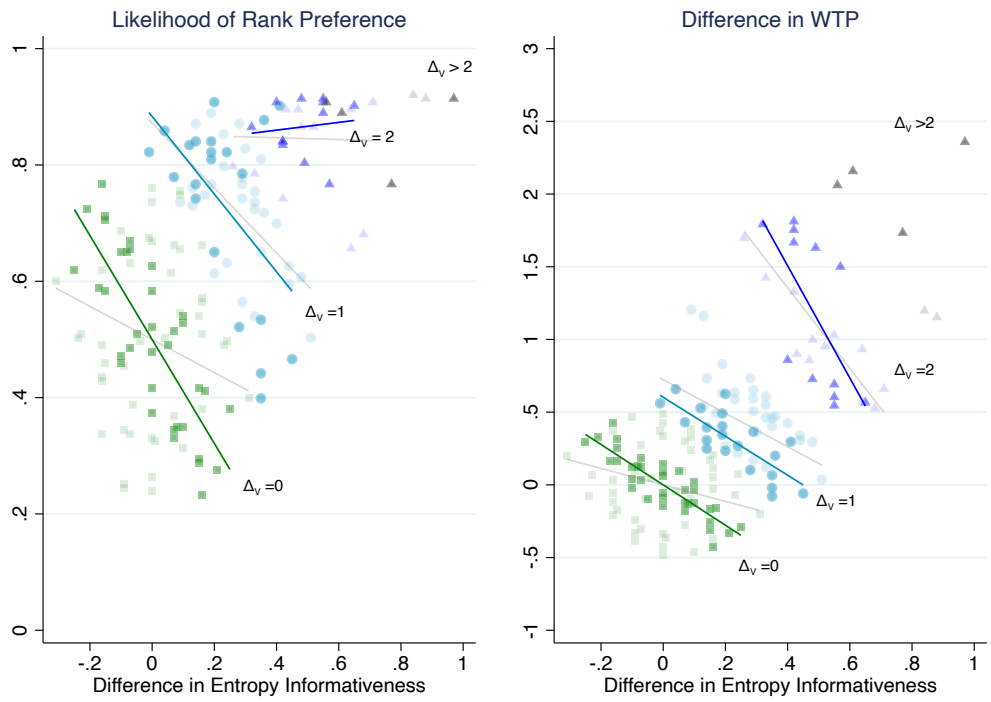


Figure 2.8: Preference for Information Structure by Value and Informativeness *Notes: Each dot denotes a pair of information structures. Green squares denote pairs where there both information structures are of the same value. Blue dots denote pairs where the value difference between the first and second information structure is 1. The darker blue (gray) triangles denote pairs where the value difference is equal to (larger than) 2. To account for the potential impact of visual complexity, pairs with at least one information structure where the blue and red balls are not displayed in order are depicted in a lighter color. Gray lines depict the best linear fits for each of the first three categories. Darker lines in the corresponding colors denote the best linear fits where the pairs depicted in a lighter color are not included.*

respectively), the pattern disappears for rank preference (left panel) suggesting that instrumental value eventually overtakes the negative effect of informativeness in rankings. While differences in WTP between the two information structures increase substantially when the value difference increases to 2 or more, we still observe the pattern that difference in WTP is decreasing as the informativeness of the more valuable information structure increases (right panel).[17]

[17] Another important observation on Figure 2.8 is that the WTP data is much noisier than the ranking data (although it produces the same qualitative patterns). For example, when the value difference between two information structures is 2 or higher, subjects rank the more valuable one as more preferred about 85 percent of the time. While the average WTP for the former one is \$1.2 higher than the other one (p-value<0.001), the aggregate likelihood that the WTP for the former is higher than the other one (by more than 10 cents when it should be around \$2) is only 67 percent.

Table 2.1: Determinants of Demand for Information

	Ranking (Logit)			**Difference in WTP (OLS)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Difference in Value	0.800***		1.607***	0.527***		0.924***
	(0.063)		(0.150)	(0.046)		(0.104)
Difference in Informativeness		2.453***	-3.436***		1.812***	-1.716***
		(0.193)	(0.487)		(0.167)	(0.338)
Clusters	163	163	163	163	163	163

*Notes: See Section 2.2 for formal definitions of value and entropy informativeness. Regressions control for differences in visual disorder (whether the blue and the red balls were presented out of order). Detailed results are presented in Appendix B.6 Table B.14. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

In Table 2.1 we provide statistical support for these results by examining pairs of information structures and regressing Rankings (using Logit) and WTP differences (using OLS) on (i) the difference in instrumental value between the two structures and (ii) the difference in informativeness. When we include these as independent variables individually, we find exactly what we documented in the previous subsection: that both relative Ranking and differences in WTP are significantly increasing in both differences in instrumental value and informativeness. However, when we include both variables, controlling for instrumental value, the coefficient on informativeness becomes negative and highly significant, indicating demand that instead decreases in informativeness. Thus, when we control for instrumental value, we find strong evidence that informativeness decreases demand for information as visualized in Figure 2.8. Detailed results are presented in Appendix B.6 Table B.14.[18]

To better understand these results, Figure 2.9 shows some especially diagnostically

[18]Since informativeness and value are necessarily correlated, in Table B.15 of this Appendix, we also include separate regressions with only differences in entropy informativeness and visual disorder as right-hand-side variables for different classes of information structure pairs that are separated by value difference. The regressions further support the main conclusions of this section. In Tables B.16-B.17 of this Appendix, we also show that these results are not driven by the specific measure of informativeness used in the analysis: aversion to more information controlling for value is also seen when we compare informativeness of information structures using the Blackwell order or the variance of the induced posteriors (see Frankel & Kamenica (2019) for further discussions on these alternative comparisons of informativeness).

valuable examples from the dataset. Information structures 14 and 15 have the same instrumental values, inducing guessing accuracy of 80 percent, but 15 is more entropy informative (it also dominates 14 according to the Blackwell ordering). This is easy to verify visually. 15 generates the same posterior as 14 for the first five blue balls, but breaks down the posterior of 20 percent conditional on the remaining balls into two distinct posteriors: 50 percent with 40 percent probability and zero percent with 60 percent probability. Subjects, in the aggregate, prefer 14 to 15: 77 percent of subjects rank the former as more preferred than the latter one.[19] The comparison of these information structures to information structure 10, which is less instrumentally valuable (inducing guessing accuracy of only 70 percent) reveals that the negative impact of informativeness on information demand can be large enough to distort valuations relative to instrumental value. 81 percent of subjects rank 14 as more preferred than 10, making an earnings-maximizing choice. However, when the more informative 15 is compared to 10, only 40 percent of subject rank 15 higher. In other words, 60 percent of subjects behave suboptimally when comparing 15 to 10. This suggests that informativeness can distort subjects' evaluation of information structures to the degree that it generates violations in how information structures are ranked relative to instrumental value – a finding that is mirrored in the aggregate statistics.

We summarize the main result of this section as a third result:

Result 9 *Controlling for instrumental value, subjects display a preference for structures that are* less *informative. High informativeness decreases demand, and increases the likelihood of suboptimal demand.*

The negative impact of informativeness on information demand is important, be-

[19]The example suggests that the negative impact of informativeness on information demand might be due to an aversion to signals that are maximally uncertain (associated with a posterior of 0.5). We discuss this possibility in relation to other alternative explanations in Section 2.5.1.

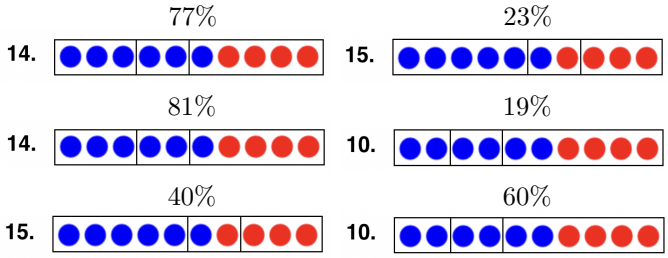


Figure 2.9: Example of Aversion to Informativeness *Notes: The percentage of subjects who rank an information structure as more preferred to the other (shown on the same row) is displayed above the information structure.*

cause it generates "mistakes" in valuation relative to instrumental value. Indeed, our results suggest that informativeness that does not contribute to instrumental value is an important driver of suboptimal demand in our data.

2.4.4 Heterogeneity

Our results so far have been based on an aggregate analysis. Does demand for information vary with informativeness and instrumental value similarly for all subjects? To find out, we examine heterogeneity by classifying subjects into types using clustering analysis. Since there are many possible factors driving preferences over information structures in our setting, we take an agnostic approach, using clustering to seek out "natural" groupings in the data without reliance on labeled examples of what prominent types might be in the data.[20]

Specifically, we transform subjects' rank preference to 120 pairwise comparisons, recording which information structure was preferred in each possible pair. We then fit a Bernoulli Mixture Model (BMM) pre-specifying the number of clusters $k \in \{2, 3, 4, 5, 6\}$.[21]

[20]Recently, Charness, Oprea & Yuksel (2021) used a similar approach in an experiment on information demand (focused on how people interpret biased information structures). They found that heterogeneity was important for understanding information preferences in their data. In particular, this type of analysis revealed an important minority cluster of subjects who behaved in a way that was systematically different than aggregate behavior.

[21]Choice of the BMM model is natural here as it is developed specifically for clustering of multidimensional binary data. This clustering method first estimates parameters of the subpopulations (mixture components), each being a multidimensional Bernoulli distribution, and weights of each subpopulation

Then we select the number of clusters k associated with the lowest BIC (Bayesian information criterion). This method partitions our subjects into 4 clusters. In order of size, the first cluster covers 74 subjects (45% of the data), the second cluster 40 subjects (25% of the data), the third cluster 35 subjects (21% of the data), and the fourth cluster 14 subjects (9% of the data).

Table B.4 in Appendix B.4 compare subjects in these different clusters in terms of the optimality of their guesses and their rankings of information structures. The largest two clusters are very similar, displaying high optimality on both dimensions, matching central tendencies in our aggregate analysis. While guesses are highly optimal for the third cluster, optimality of the ranking of information structures weakens significantly for this group. Finally, the smallest fourth cluster is particularly noisy on both dimensions. For this reason, we focus our analysis in the main text to the biggest three clusters (91 percent of our subjects) and relegate corresponding analysis of the smallest cluster to Appendix B.4.

Figure B.2 in Appendix B.4 reproduces Figure 2.8 separately for each of the main clusters in our data. The most striking observation is the contrast in behavior between the first two clusters. In the largest cluster, demand for information is decreasing in informativeness once we control for value, matching our aggregate findings. This is reflected in both Ranking and WTP data. The second cluster displays the opposite behavior (at least when focusing on Ranking data): controlling for value, subjects in this cluster display a preference for more informativeness, however there is weaker evidence for this in WTP data. Patterns observed in the third cluster are reminiscent of those from the first cluster (at least with respect to sensitivity to informativeness), but behavior is noisier and as noted above, there are more deviations from optimal behavior. Table 2.2

(mixture weights). Then, using these estimates, clustering simply becomes a matter of using Bayes' rule to classify each observation as belonging to the mixture component most likely to have produced it.

Table 2.2: Determinants of Demand for Information By Cluster

	Ranking (Logit)			**Difference in WTP (OLS)**		
	C1	C2	C3	C1	C2	C3
Difference in Value	3.744***	0.809***	1.085***	1.425***	0.548***	0.617**
	(0.164)	(0.182)	(0.170)	(0.162)	(0.115)	(0.237)
Difference in Informativeness	-9.855***	2.942***	-2.843***	-3.189***	-0.148	-1.332*
	(0.638)	(0.755)	(0.651)	(0.523)	(0.439)	(0.786)
Clusters	74	40	35	74	40	35

*Notes: C1, C2 and C3 refer to Clusters 1, 2 and 3 and represent 45%, 25% and 21% of the data, respectively. See Section 2.2 for formal definitions of value and entropy informativeness. Regressions control for differences in visual disorder (whether the blue and the red balls were presented out of order). Detailed results are presented in Tables B.5 - B.7 in Appendix B.4. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

provides further support for these observations using regression analysis (reproducing Table 2.1 separately for the largest three clusters). We find that relative Ranking and differences in WTP are both significantly *decreasing* in informativeness when we control for instrumental value for the first and the third clusters, but the opposite result is observed (at least when focusing on relative ranking) for cluster 2.[22]

Overall, in the first and the third clusters (covering 67 percent of our data) demand for information is decreasing in informativeness when we control for instrumental value, while we find no evidence for such a tendency in the second cluster (25 percent of the data) which shows some evidence of preferences for more informativeness. It is worth emphasizing that the clustering method does not assume in any way that informativeness should play a role in how subjects are classified into different types. Thus, these results strongly reinforce our finding that the influence of informativeness on information demand plays a key role in organizing heterogeneity in our data.

Result 10 *For a majority of subjects, demand for information is* decreasing *in informativeness when controlling for its instrumental value.*

[22]Using different clustering methods (such as K-Modes clustering) generates similar results. See Appendix B.4 for further details.

2.5 Mechanism

Why is demand for information decreasing in informativeness, ceteris paribus, for a majority of subjects in our data? In this section we use data from diagnostic treatments and analysis of auxilary data to answer this question. In Sections 2.5.2 and 2.5.1 we provide evidence that rules out explanations rooted in subjects' preferences over timing of the resolution of uncertainty or risk or misunderstandings of how to make use of information structures. In Section 2.5.3 we instead present evidence that aversion to informativeness is likely a consequence of the fact that informative information structures are more costly to evaluate properly (i.e., are more "complex" to value). As a result, subjects avoid these information structures and are instead drawn to less informative ones, sometimes leaving instrumental value "on the table" in order to avoid these information processing costs.

2.5.1 Timing and Risk Preferences

One natural class of explanation for aversion to informativeness is non-standard preferences over risk, loss or the timing of the resolution of uncertainty. We find that preference-based mechanisms of this sort cannot explain our data, for two reasons. First, our data does not seem to fit with the most promising such explanations, given prior empirical findings. Second, we designed our No Uncertainty treatment to shut down scope for such preferences altogether and we find results that are broadly similar to those from our main treatments.

Perhaps the most relevant preference-based mechanism given our experiment is that subjects might have preferences over the timing of the resolution of uncertainty (e.g., Kreps & Porteus (1978), Grant, Kajii & Polak (1998)). Prior experimental results have shown that subjects have seemingly strict preferences for earlier revelation of non-

instrumental information from information structures (Eliaz & Schotter 2007, 2010, Nielsen 2020, Falk & Zimmermann 2022). However, our results seem to show the exact opposite: subjects display a preference to *avoid* non-instrumental information (i.e., informative structures that don't improve choice), which means subjects reveal (if anything) a preference to delay learning about the true payoff state.

Other preference-based explanations seem similarly unlikely to fully account for our results. For instance, in our setting, loss aversion (Koszegi & Rabin 2009) could possibly manifest as an aversion towards information structures which generate maximally uncertain signals (with associated posterior of 0.5) as such signals imply a lower likelihood of winning the prize than the prior.[23] Our main results on informativeness aversion (controlling for instrumental value) remain when we remove such information structures from the analysis.[24] Our results also don't seem consistent with suspense/surprise preferences, a'la Ely, Frankel & Kamenica (2015*a*): we find no evidence that ranking or valuation of information structures can be explained by the amount of suspense or surprise they generate when we control for their instrumental value. Likewise, the results don't seem to be driven by preferences over the skewness of information provided by our information structures: although there is some suggestion in our data (as in some prior experiments like Masatlioglu, Orhun & Raymond (2023)) that subjects could be drawn to positively skewed structures, controlling for this does not change our findings on informativeness attitudes. Mechanisms like motivated attention and dissonance avoidance (Festinger 1957) likewise can't account for our findings. Curiosity preferences as in Gol-

[23] Although this seems unlikely given that, by definition, all information structures increase expected likelihood of winning the prize.

[24] However, as seen in Appendix B.6 Table B.18, removing these information structures can influence the degree to which informativeness impacts demand for information controlling for instrumental value. This suggests that some subjects might indeed particularly dislike information structures generating maximally uncertain signals. Our experiment is not designed to identify the relative magnitude of these effects, and this may be difficult in general because increasing informativeness without increasing instrumental value necessarily involves changing the distribution of induced posterior beliefs to move closer to extremes of 0.5 and 0 or 1 (as observed in the example of Figure 2.9).

man & Loewenstein (2018) and Golman, Loewenstein, Molnar & Saccardo (2022) predict informativeness loving behavior that goes in the opposite direction of our main findings.

Our No Uncertainty treatment allows us to assess the relevance of preference-based mechanisms for our results in a more comprehensive way. In this treatment, instead of paying subjects based on their guess for a randomly drawn ball, we paid them based on the accuracy of their guesses for *all* ten balls. That is, subjects were told that all ten balls would be selected, but guesses needed to be the same for all balls in the same cell of a partition associated with an "information" structure. Thus, the objective (and certain) value of each "information" structure in this treatment is exactly equal to the instrumental value of that information structure. Furthermore, information processing required for guesses and evaluation of each "information" structure in this treatment is very similar to that required in the other two treatments. But (i) since guesses are known to apply to all balls, there is no actual information provided to subjects in the experiment and (ii) there is no objective uncertainty anywhere in the experiment. Because of (i), preferences related to the timing of information cannot apply in this treatment. Because of (ii) there is likewise no scope for explanations related to risk preferences since there is no uncertainty in these tasks. In Appendix B.3 we show that the No Uncertainty treatment produces broadly similar results to our Baseline treatment. We continue to find that subjects display an aversion to "informativeness" controlling for value even though there is no actual learning occurring about an unknown state in this task.[25] This suggests that preferences for timing or risk likely are not a primary driver of our findings.

Result 11 *Our results are not consistent with recent findings (from settings where in-*

[25] As documented in Appendix B.3, behavior (particularly WTP) is more noisy in this treatment, likely attenuating the effect we find clearly in the Ranking data. This noise might be driven by the somewhat unnatural framing of the problem relative to the other two treatments. Probably due to this attenuation, the WTP data in aggregate does not display aversion to informativeness controlling for instrumental value. However, there is evidence suggestive of this pattern when we restrict analysis to WTP data that is ordinally consistent with ranking data.

formation has no instrumental value) on preferences for the timing of resolution of uncertainty. Similar patterns are observed when uncertainty is removed from the task, suggesting that our results are not an outgrowth of risk or timing preferences.

2.5.2 Biases and Mistakes

A second natural class of explanations for aversion to informativeness is that it derives from mistakes in how people make use of information.

First, subject might make mistakes in interpreting and optimally making use of information, and reveal these mistakes when valuing information structures. There is, after all, a great deal of evidence in the prior literature that subjects suffer from a range of judgement errors when evaluating signals in standard prior-signal updating tasks, including for example, over- and under-inference, confirmation bias and motivated reasoning. Thus, one natural hypothesis is that subjects suffer from one or more of these biases when assessing the information contained in information structures, leading them to fail to properly value them in ways that spuriously resemble distaste for informativeness.

We can rule out such mistakes by studying behavior in the Guessing section of the experiment which was specifically designed to evaluate this hypothesis. As we detail in Section 2.4.1 above, subjects show no evidence of systematic biases in interpreting signals from information structures, and in fact show little evidence of noise either. Indeed, our subjects make nearly perfect use of information structures to inform Guessing behavior: 98% of Guessing choices are optimal (expected earnings maximizing) in our dataset and the vast majority of our subjects *always* make optimal use of information. This strongly suggests that subjects are overwhelmingly capable of understanding how to use the information contained in information structures to inform choice.

A more subtle version of the same class of explanation is that, although subjects are

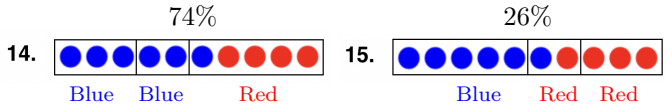


Figure 2.10: Example of Aversion to Informativeness with Information on Signal-Contingent Guesses *Notes: The percentage of subjects who rank an information structure as more preferred to the other is displayed above the information structure. Text below cells of the partition represent the signal-contingent guesses made in an identical problem earlier. We restrict attention to subjects in the Reverse treatment whose guesses with the two information structures are same for each ball (as seen above).*

capable of interpreting each piece of information they receive from information structures, they fail to properly account for this from an ex-ante perspective when valuing and comparing unfamiliar information structures before they receive information. Perhaps subjects can make optimal use of information, but fail to foresee how they will do so ex ante when evaluating information structures. There is now a growing literature showing that subjects often neglect to "unfold" decision trees, failing to contingently reason about events in the future (Esponda & Vespa 2014, Martinez-Marquina, Niederle & Vespa 2019). Perhaps subjects make a similar error here and fail to accurately think through the payoff-relevant consequences (i.e., guesses induced by each signal) of receiving information from each information structure producing an apparent aversion to informativeness.

To test this kind of explanation, we ran the Reverse treatment, which has subjects make guesses for all signals under all 16 information structures *before* they are asked to value them. The purpose of this treatment was to minimize the scope for this kind of error by giving subjects a precise sense of how each information structure impacts choice. To further strengthen the effect of this treatment, we also reminded subjects of the guesses they made (contingent on every signal) in these previous guessing tasks on their elicitation screens, allowing subjects to recall their prior engagement with each information structure. To provide a concrete example of how subjects experienced this treatment, Figure 2.10 displays how a subject might have seen information structures 14 and 15 (the example from Figure 2.9).[26] In particular, subjects are shown their

[26]Figure 2.10 displays how these information structures were presented to a subset of the subjects

signal-contingent guess of the state (red or blue) underneath each possible signal when evaluating the structure. Under the hypothesis that our main findings are driven by subjects' failure to contingently reason about how different signals induce different guesses, we would expect this treatment to eliminate or at least reduce the severity of our results. Instead, in Appendix B.2, we show that the Reverse treatment has no effect on our results at all. Even among these subjects, we find a strong preference for less informativeness.

Result 12 *Auxiliary tasks and a diagnostic treatment suggest that failures in optimal use of information or failures in foreseeing how signals will induce guesses cannot explain why demand for information is decreasing in informativeness when we control for instrumental value.*

2.5.3 Valuation Complexity

A third possibility is that subjects attach less value to informative structures, because these structures are more costly and difficult (i.e., more complex) to evaluate. Properly valuing an information structure, after all, requires the decision maker to aggregate a number of pieces of information about the structure: she has to consider each of the signals that could be realized, determine the optimal action and compute payoff consequence for each case, and then aggregate over all of these possibilities. There are strong ex ante reason to think there is more information to aggregate in more informative relative to less informative structures. In particular, more informative information structures tend to have more signals, tend to generate more distinct posteriors and tend to include more extreme posteriors, all of which plausibly make the aggregation problem more difficult. Indeed, such aggregation costs have been shown to severely impact valuations in the domain of risk (Enke & Graeber 2023, Oprea 2023) and intertemporal choice (Enke, Graeber

(50%) in the Reverse treatment for whom induced guesses (for each ball in the set of ten) is exactly the same for the two information structures.

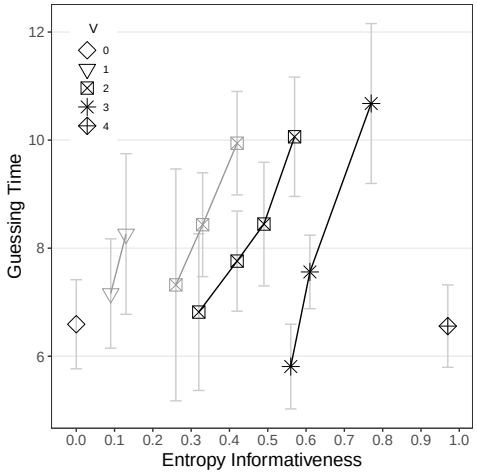


Figure 2.11: Cost of Informativeness: Guessing Time *Notes: Figure depicts average guessing time spent on each of the 16 information structures that are arranged by value and entropy informativeness. The information structures where the blue and red balls are not displayed in order are depicted in gray rather than black. Vertical lines denote 95 percent confidence intervals.*

& Oprea 2023) in recent work, implicating them in some of the most famous anomalies in behavioral economics such as probability weighting and hyperbolic discounting. Could aggregation difficulties be similarly driving anomalous valuations here?

Our data and design provide several strands of evidence that seem to directly support the idea that aggregation costs of this sort drive subjects' apparent distaste for informativeness.

First, our data produces direct evidence that evaluating more informative structures requires more effort for subjects. In particular, controlling for instrumental value, subjects spend more time making signal-by-signal decisions in Guessing tasks involving more informative information structures. Figure 2.11 shows clear evidence that guessing time—the number of seconds it takes subjects to make guesses conditional on each possible signal realization from an information structure— increases with informativeness. For example, among the non-disordered information structures that have an instrumental value of 2 (the largest class in our design), the average time it takes subjects to make the full set of signal-contingent guesses is 47% percent higher (p-value<0.001) with the highest entropy

informative information structure than with the lowest entropy informative structure.[27] Decision time ("runtime") is the most commonly used metric of complexity (information processing costs) in computer science, supporting the hypothesis that more informative structures are more complex to evaluate.[28]

Second, our finding of distaste for informativeness remains even after we remove uncertainty from the task in our No Uncertainty treatment. With risk and timing removed, the only possible reason to fail to maximize value in this problem is the difficulty of properly valuing the descriptive components of the information structure. Indeed, arguably the only substantive connection between the Baseline treatment and the No Uncertainty treatment is the similar cognitive processing required to evaluate the structure. Broadly similar aversion to informativeness in the two settings therefore seems to suggest that this shared cognitive burden is the driver of this result.

Third, the evident difficulty of valuing more informative structures, seems to be rooted in the costs of aggregating its components and not in other costs in the choice problem. In particular (as already discussed) our design rules out the hypothesis that valuation difficulties stem from difficulties in assessing signal-contingent choices: having subjects make these choices ahead of time, and showing them to subjects directly during valuation (as we do in the Reverse treatment) has no impact on measured informativeness aversion. The only other possibility would seem to be that the difficulty (cognitive costs) of optimally *using* information is higher for more informative structures, and that subjects anticipate this difficulty when valuing those structures. However, this hypothesis doesn't bear scrutiny. To decide how to optimally respond to a signal in our experiment,

[27] The same pattern in also observed in the Reverse and No Uncertainty treatments.

[28] The use of decision time as a metric of complexity is controversial in the literature, because subjects may choose not to expend effort at all on particularly difficult tasks. This will produce a non-monotonic relationship between complexity and decision time. This is arguably not a concern in our data since we have clear evidence that subjects virtually never "give up" – almost all subjects make rational choices in our Guessing tasks.

a subject simply needs to identify whether there are more red or blue balls in a single partition subset clearly indicated by the experimental software. This is identically true for all signals and all information structures. If anything, we might expect this task to be somewhat *easier* for more informative structures since partition subsets (signals) tend to contain fewer balls in these structures. This seems to leave only the difficulty of aggregation itself as the source of the difficulty of evaluating information structures.

Why do informativeness-driven aggregation difficulties lead to a systematic aversion to informativeness? The most likely explanation is that decision makers tend to economize on the costs of precisely calculating instrumental value by imprecisely estimating value instead – and do so increasingly as information structures grow more informative and the aggregation task grows more burdensome. Subjects respond to the imperfect or incomplete understanding of value that results from this decision by cautiously undervaluing informative structures, shading their valuations towards the experiment's default of no information.[29] This kind of cautious attenuation has recently been formalized by "noisy coding" models (Woodford 2020) which have been empirically successful, accounting for a wide range of anomalous behaviors including small stakes risk aversion (Khaw, Li & Woodford 2021), probability weighting (Enke & Graeber 2023, Vieider 2023, Frydman & Jin 2023), Bayesian reasoning failures (Ba, Bohren & Imas 2023) and hyperbolic discounting (Gabaix & Laibson 2022, Vieider 2021, Enke, Graeber & Oprea 2023).[30]

Result 13 *More informative structures require more cognitive effort to evaluate. This and several other strands of evidence suggest that distaste for informativeness is driven by information processing costs of valuing structures, which increase in their informa-*

[29] In our elicitations, especially our WTP elicitations receiving no information is quite literally the default outcome that the subject is paying to avoid

[30] An alternative possibility is that subjects simply have a primitive distaste for complex information structures (independent of difficulties in evaluating them) that leads them to undervalue them. This seems less plausible to us, but is also consistent with our data.

tiveness.

2.6 Discussion

In standard economic theory informativeness influences information demand only to the extent that it produces instrumental value – i.e., to the extent that it is expected to improve decision making in relevant decision tasks. The results reported in this paper suggest that this is not the case: informativeness has an independent, first order influence on information demand. In particular, demand for information sharply decreases in informativeness, conditional on instrumental value. We find that this aversion to informativeness is often severe enough to make subjects prefer less instrumentally valuable information to more valuable alternatives, leading them to leave earnings "on the table."

Additional evidence from our experiment suggest that this aversion to informativeness arises because informative structures are more 'complex' (i.e. more costly or difficult to evaluate), leading subjects to undervalue them. Our design allows us to rule out several alternative explanations for this pattern: they cannot be (i) rationalized as an outgrowth of subjects' preferences for the timing of information, (ii) driven by mistakes in optimal use of information; (iii) arise from failures to reason about information-contingent actions.

These results suggest that even in the simplest possible settings (e.g., our experiment), decision makers face a difficult aggregation problem when evaluating information structures: they have to consider and weigh all possible signal realizations, determine the optimal action and compute payoff consequence for each case, and then aggregate over all these possibilities, condensing this information into a value. The costs and difficulties of doing this correctly rise with the informativeness of the structure, meaning the distortions it produces follow a predictable pattern that can be used in modeling and information design.

Indeed, our results have policy implications on how information should be provided to decision makers to most effectively influence behavior. Our results suggest especially that information should be narrowly targeted to the decision problem the information is meant to inform. Extraneous or irrelevant information is likely to be treated as an economic bad in the market for information and makes information less effective as a policy instrument. Even in contexts (such as that of our experiment) in which decision makers are able to easily make optimal use of the signals informative structures produce, they nonetheless should be expected to undervalue and avoid such structures because of the costs of evaluating their value, ex ante.

These findings reveal a connection between information demand and a growing number of contexts in which complexity of valuation (which typically involves aggregating different pieces of information) has been shown to produce first order distortions in choice. For instance, the difficulty of correctly valuing an information structure is formally similar to the problem of evaluating a compound lottery, which decision makers have well-documented difficulties with (Halevy 2007, Chew, Miao & Zhong 2017). Similarly, recent research suggests that the complexity of aggregation is a key driver of classical anomalies in risky choice, such as probability weighting (Enke & Graeber 2023, Oprea 2023) and classical anomalies in intertemporal choice, such as hyperbolic discounting (Enke, Graeber & Oprea 2023). Thus our findings suggest a parsimonious connection between anomalies in information demand and anomalies in a number of other canonical choice settings in economics.

Much of what we know about information demand so far comes from settings in which information has no instrumental value. A key methodological lesson from our experiment is that behavior in such settings is likely shaped by a very different reasoning process and hence influenced by different characteristics of information structures than settings in which information is required to inform choice. In particular, previous literature (as

discussed in 2.5.1) often documents information loving behavior when information has no instrumental value. Our results, which go in the opposite direction, suggest that the aversion to informativeness documented in our paper is a byproduct of the difficulty decision makers encounter in assessing the instrumental value of an information structure – a problem that is lifted in non-instrumental settings. For this reason, our results suggest that we cannot easily counterfactually project findings from non-instrumental settings to instrumental settings (or vice versa).

Finally, our results provide a new kind of evidence in support of the central trade-off at the heart of rational inattention models, one of the most influential and often used formal theories of bounded rationality. In these models, people acquire information to maximize utility but suffer information costs that influence this choice (Sims 2003, Matějka & McKay 2015, Caplin & Dean 2013). Our paper contributes to this literature by expanding our understanding of when and why agents act as if information is costly. We show that these information costs need not be limited to costs associated with generating this information or making use of this information, but can also arise due to the cognitive effort required to *evaluate* the ex-ante value of information. Our results therefore suggest that rational inattention models may also be effective models of complexity (information processing) aversion, and may therefore have a much wider scope of application than is typically supposed.

Chapter 3

Trustworthy by Design

with Sen Geng[1]

3.1 Introduction

Trust is a crucial ingredient for many social interactions and economic transactions.[2] However, foreseeing the risk of being exploited by trustees often discourages trustors from placing trust. Addressing this social dilemma is a continuing focus of the literature on trust. Recent work (Brown et al. 2004; Charness & Dufwenberg 2006; Bracht & Feltovich 2009; Andreoni 2018) has established mechanisms that enhance trusting acts through reputation, cheap talk, or a marketing strategy with a satisfaction guarantee.[3] A novel aspect is to explore voluntary trust-enhancing mechanisms that are applicable

[1]The content of Chapter 3 and Appendix C was the accepted manuscript of an article published in Games and Economic Behavior, Vol 141, Sen Geng and Menglong Guan, 'Trustworthy by design', Page 70-87, Copyright Elsevier (2023). The article is available at: https://doi.org/10.1016/j.geb.2023.05.009. The reuse of the content in this dissertation is permitted by the publisher.

[2]For the positive impact of trust on economic outcomes, see Knack & Keefer (1997), Zak & Knack (2001), Guiso et al. (2004, 2009), and Gennaioli et al. (2021), among others.

[3]The implementation of such mechanisms sometimes relies on recurring interactions (Brown et al. 2004), third-party involvement, such as a form of competition (Huck et al. 2012), or binding guarantees (Andreoni 2018).

to one-shot interactions and have rational foundations.

In addition, information design, pioneered by Kamenica & Gentzkow (2011), investigates how senders influence receivers' behavior via provision of information when senders can commit to an information disclosure policy. The models have been widely applied in various contexts.[4] A novel exploration is to apply information design in trade settings, in which an allocation of gains from trade may be endogenously determined and enhancing aggregate welfare is usually a prerequisite for improving individual welfare. For this application, the key question is whether information design can be used by self-interested market players to increase their own benefits through fostering social welfare.

Fitting into the intersection of the two strands of literature, our study explores whether information design can be used by trustees as a signaling device to boost trust and consequently realize gains from trade. Empowering trustees to signal their trustworthiness makes it easier for trustors to infer typically invisible behavior; in this sense, it seems intuitive that outcomes can be improved. However, the endogeneity of providing information and choosing to be trustworthy complicates the problem, and a key issue is how much this capability is used.[5] Thus, in this article, we ask the following questions: To what extent and in which way does trustees' capability to design information improve trustworthiness and trusting acts? What is the theory underlying such effect or null effect?

These questions are difficult to address in the field due to the lack of exogenous variation in trustees' capability to design information. It is also challenging to induce such variation in the field. We thus propose two games that characterize two distinct settings: no capability versus full capability to design information, and conduct a laboratory

[4]See, for example, applications to grading in schools (Boleslavsky & Cotton 2015), entertainment (Ely et al. 2015*b*) and financial sector stress tests (Goldstein & Leitner 2018).

[5]The problem is further complicated when social preferences are involved: trustors may place trust when signaling trustworthiness is not feasible (See Berg et al. (1995)), but introduction of a signaling device may backfire when trustees choose an opaque information disclosure policy.

study.[6]

In Game 1, a trustee, who partially or fully decides a binary payoff allocation (equivalently a binary state) and has no capability to design information, precedes a trustor, who decides whether to invest. Game 2 resembles Game 1 except that the trustee is capable of designing an information structure, which is characterized by two state-dependent distributions of signal realizations. A binary signal about the underlying state is then generated according to the trustee's choice of the state and the information structure. We explore the changes in trusting acts and trustworthiness between games based on the equilibrium model and a model of heterogeneity in prosociality and strategic sophistication. We also employ a within-subjects experiment to explore treatment effects and systematic behavioral patterns.

We find theoretically and experimentally that introducing information design increases trustworthiness and trusting acts. The experimental results show that, compared to Game 1, trustworthiness and trusting acts increase substantially in Game 2. We also observe that trustees signal high trustworthiness via high informativeness and obscure low trustworthiness via low informativeness. Almost all trustors correctly infer the underlying state (and trustworthiness) from the realized signal when observing a high informative structure, but some fail to understand that the information structure per se also conveys a message about trustworthiness and overtrust in low informativeness. Both models predict that introducing information design, together with its bringing about preliminarity of the trustee's choice and an increase in the trustee's power to commit to a certain level of trustworthiness, fosters trusting acts. Nonetheless, only the second model predicts all major patterns. Especially, it predicts that some trustees optimally choose zero trustworthiness with the least Blackwell informative structure, and trustors with

[6]Given the focus of this study, our games abstract features about trust, trustworthiness and information design from the multifaceted features of practices. Karlan (2005) shows that experimental trust games are directly linked to real-life decisions including financial decisions.

a lower level of strategic sophistication are more likely to place trust than the most sophisticated trustors.

In terms of implications, our work proposes rational foundations of market practices involving trust and information design. In business practices involving a seller and buyer or an investee and investor, trustees often take the initiative, including information design, to build trust. Our findings show that information design is indeed an effective mechanism to enhance trusting acts; thus, trustees have a good reason to take such actions. We also show that signaling high trustworthiness with high informativeness and obscuring low trustworthiness with low informativeness are optimal approaches for some trustees given trustors' strategic sophistication. These results explain why, in practice, the relevant information is designed at different levels of informativeness. In terms of policy implications, our results show that there is a potential gain in alerting trustors to the possible association between low trustworthiness and low informativeness. Since trustees' optimal trustworthiness and information structure are responsive to trustors' strategic sophistication, this intervention is likely to motivate trustees to endogenously enhance trustworthiness and informativeness.

Our study is related to the literature involving trust: it provides models and experimental evidence showing that information design is a useful signaling device to enhance trust.[7] Existing studies of trust games (mostly variants of Berg et al. (1995)), including the aforementioned trust-enhancing mechanisms, are based on the paradigm of trustors' decision to place trust preceding trustees' decision to allocate payoff. By contrast, we set up a new paradigm in which trustees take the lead in the strategic interaction. By applying this innovative perspective, we model the market practice of gaining trust through information design as a combination of a reverse trust game and an information design game.

[7]See Fehr (2009) for a survey of the literature on trust.

Our study is also related to the literature concerning information design led by Kamenica & Gentzkow (2011).[8] In terms of applying information design, our study shows that information design can be used by self-interested market players to foster trust in trade settings and in turn fulfill welfare-increasing transactions. In terms of extending the basic model, the information designer in our setting endogenously chooses the state of the world, which deviates from the common assumption about the exogenously determined state. This deviation extends the original scope of information design to settings where the designer decides both a payoff-relevant state and an information structure.[9]

Our study is additionally related to recent experimental work on information design. For instance, in Fréchette et al. (2022), senders first commit to an information structure and then may revise the signal that concerns the realized state depending on the specified degree of commitment. They focus on the role of commitment and information verifiability and find that overall, subjects react to commitment in the direction predicted by the theory. Besides different focuses and designs, an important difference is that the payoff-relevant state is endogenously determined in our setting while it is exogenous in theirs.

The remainder of this paper proceeds as follows. Section 3.2 introduces the games and presents the equilibrium outcomes. Section 3.3 presents the experimental design and procedure, followed by experimental results in Section 3.4. Section 3.5 proposes a behavioral model that rationalizes the experimental findings. Section 3.6 discusses an alternative experimental design, the role of risk preference, how our games are related to classic trust games, and potential causal mechanisms. Section 3.7 concludes. Appendix

[8] Kamenica (2019) categorizes the existing research about information design into two directions and reviews the strand of literature that extends the basic model.

[9] In this sense, our study is related to Asriyan et al. (2023) and Szydlowski (2021). Asriyan et al. (2023) consider a setting where a designer chooses an action and a factor that affects a receiver's choice of information structure. Szydlowski (2021) investigates a setting where an entrepreneur jointly chooses information disclosure and financing policies. Unlike ours, the state of the world is exogenously determined in his setting.

C provides details about the behavioral model, proofs, additional data analysis, and experimental instructions.

3.2 Games and Equilibrium Analysis

3.2.1 Two games

We introduce two games to study whether trustees' being capable of designing information resolves the social dilemma involving trust. In both games, a trustee first decides payoff allocation and then a trustor decides whether to invest. This is different from the decision order in classic trust games, where the trustor typically moves first and the trustee moves second.[10] Trustees' moving first reflects the fact that trustees take the lead in some contexts. For example, a seller chooses target customers and makes efforts to seek customers' trust in a product. An entrepreneur takes the initiative to gain investors' trust in a funding proposal. We use a feature of whether the trustee is capable of designing information to differentiate a reverse trust game (Game 1) from a trustworthiness design game (Game 2).

In Game 1, a trustee (player A) moves first to decide a payoff-relevant state: he selects a probability $p \in [0, 1]$, with which state 1 is to be realized and, correspondingly, the probability of state 2 being realized is $1 - p$. Knowing neither the value of p nor the realized state, a trustor (player B) decides $z \in \{1, 0\}$, that is, to invest or not invest. If player B does not invest, player A and player B receive a payoff of v_0^A and v_0^B, respectively. If player B invests and the realized state is state 1, player A and player B receive a payoff of $\rho_1 v$ and $(1 - \rho_1)v$, respectively. If player B invests and the realized state is state 2, player A and player B receive a payoff of $\rho_2 v$ and $(1 - \rho_2)v$, respectively.

[10] Section 3.6 discusses the effect of the reversed decision order: it is likely to make the trust motive weaker than in classic trust games but still remarkable.

Game 2 resembles Game 1 except that a stage of information design is introduced. Specifically, player A moves first to decide a probability p of state 1 being realized and also decides an information structure, i.e., the conditional likelihoods of signals $(q_1, q_2) \in [0,1] \times [0,1]$ specifying the probability of generating a binary signal $s \in \{b, w\}$ given the realized state. When the realized state is state 1 (state 2), signal $s = b$ is realized with the probability q_1 (q_2) and signal $s = w$ is realized with the probability $1 - q_1$ $(1 - q_2)$. Player B observes neither the value of p nor the realized state. After observing the conditional likelihoods and a realized signal (q_1, q_2, s) , player B decides $z \in \{1, 0\}$.

We assume that $v > v_0^A + v_0^B$, $v_0^A < \rho_1 v < \rho_2 v$ and $(1 - \rho_2)v < v_0^B < (1 - \rho_1)v$. That is, investing increases the aggregate payoff, state 1 refers to a payoff allocation that is beneficial to both players, and state 2 refers to a payoff allocation that is beneficial to player A but harmful to player B. Let $\rho_0^A = \frac{v_0^A}{v}$ and $\rho_0^B = \frac{v_0^B}{v}$. The assumptions can be rewritten as: $\rho_0^A + \rho_0^B < 1$, $\rho_0^A < \rho_1 < \rho_2$ and $1 - \rho_2 < \rho_0^B < 1 - \rho_1$. These assumptions capture a main feature of trust behavior: trust is socially desirable and potentially beneficial to both parties, but placing trust in others exposes one to the risk of being betrayed.[11]

Discussion of Game 2

We discuss the characteristics in Game 2 that player A chooses (p, q_1, q_2) and player B observes (q_1, q_2, s) from two perspectives: its interpretation in two motivating examples and its usefulness.

Consider first the example of a seller trying to identify a target customer and sell to the customer a complex financial product such as annuity. There are two states of

[11] We interpret p as a measure of the level of trustworthiness. p can also be interpreted alternatively: it reflects player A's determination on choosing state 1 or reflects player A's applying a mixed strategy to choose the payoff-relevant state. In classic trust games, the trustee moves second and decides the amount to return back to the trustor after being invested, and the ratio of the back-transfer serves as a measure of the level of trustworthiness.

the world: The product is suitable or unsuitable for a customer. That is, the state of the world is customer specific. The customer gets utility v_0^B for not buying the product. When buying the product, she gets utility $(1-\rho_1)v$ $(> v_0^B)$ from a suitable product and gets utility $(1-\rho_2)v$ $(< v_0^B)$ from an unsuitable product. The seller gets utility v_0^A for not selling the product. He gets utility $\rho_1 v$ $(> v_0^A)$ if he sells to a customer for whom the product is suitable, and gets utility $\rho_2 v$ $(\rho_2 > \rho_1)$ if he sells to a customer for whom the product is unsuitable.[12]

We can think of the seller's decision process as consisting of two interim stages. In stage one, the seller selects a target customer to whom he markets the product. To do that, the seller collects information about potential customers' profiles (e.g., age, gender, income, likes, and dislikes, etc), becomes knowledgeable of the chances of the product being suitable for these customers given his expertise, and markets the product to a customer for whom the product is suitable with probability p.[13] In this setting, the seller chooses a trustworthy, untrustworthy, or intermediate trustworthy action if he markets the product to a customer for whom the product is suitable with certainty, unsuitable with certainty, or suitable with a probability strictly between zero and one.

In stage two, the seller helps the target customer to learn about product suitability by providing product information and employing an appropriateness test, which can be a set of questions constructed by the seller. It is required by law (e.g., European Union MiFID II Article (25)) that the seller must truthfully report the test outcome to the customer. In practice, the set of questions the seller chooses to ask may be fully, partially or little diagnostic. We thus formalize the seller's constructing test questions as choosing the

[12]For example, an insurance product may be unsuitable for a customer who is unlikely to make an insurance claim and suitable for a customer who is likely to make a claim. Selling the product to the former seems more profitable than selling it to the latter.

[13]The seller in this example does not *physically* choose p as player A does in Game 2, but instead selects a customer to persuade. Nevertheless, the essence of their such actions is arguably the same: they choose to be trustworthy, untrustworthy or intermediate trustworthy.

conditional likelihoods of signals, (q_1, q_2), and formalize the test outcome as a realized signal (s) indicating the underlying state.

It is implicitly assumed that in Game 2, player A publicly commits to an information structure and his commitment to choosing a certain level of trustworthiness relies on the presence of information design. The first assumption holds naturally in this example because the customer observes the test questions and the test outcome must be truthfully reported. The second assumption is also plausible in our example for the reasons below. How the seller chooses a target customer is typically unobservable and the seller has an incentive to market the product to a customer for whom the product is suitable with a lower probability (i.e., undercut p). These make it difficult for the seller to commit to choosing a certain level of trustworthiness. Supplying product information and conducting an appropriateness test help a customer learn about product suitability, which in turn increases the seller's power to commit to choosing a certain level of trustworthiness.

As another illustrative example, consider the scenario in which a producer decides on product quality and selects a customer review site that discloses quality information through product reviews. Specifically, the producer determines the percentage of products that have no defect. Depending on the design features of the review system, including incentives for buyers to post reviews and visibility of posted reviews of different attitudes, customer review sites differ in their levels of informativeness about revealing product quality. The producer chooses a customer review site that maximizes the probability of selling the product to a consumer. Knowing the level of informativeness of the producer's selected review site, the consumer makes an informed purchase decision after reading reviews.[14] In this example, the provision of a product without defect incurs more costs to the producer than the provision of a defective product; purchasing the former

[14]According to a study by the Spiegel Research Center of Northwestern University, nearly 95% of shoppers read online reviews before making a purchase.

is more desirable for the consumer than purchasing the latter. The producer's choice of the percentage of products without defect and the specific review site correspond to player A's choice of p and (q_1, q_2). The consumer's knowledge of the informativeness of the review site and reading reviews correspond to player B's observing (q_1, q_2, s).

Regarding the usefulness of the characteristics, a setting of $p \in [0, 1]$ provides a more flexible measure of trustworthiness than a setting of binary $p \in \{0, 1\}$ does, which is particularly important given that a main variable of interest in this article is trustworthiness. This approach also helps to generate both an equilibrium with intermediate trustworthiness and an equilibrium with full trustworthiness. A setting of $q_1 \in [0, 1]$ and $q_2 \in [0, 1]$ is general and enables us to seamlessly incorporate existing knowledge of informativeness ordering and information design without any distortion.[15]

3.2.2 Equilibrium Analysis

Game 1's Nash equilibria include: player A chooses $p \leq \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}$ and player B does not invest. Note that player A has a unique best response, $p = 0$, once player B invests with any non-zero probability. Following Simon & Stinchcombe (1995), which extends the trembling-hand perfection in finite games to perfection in infinite games, we refine the Nash equilibria of Game 1. This refinement guarantees a unique equilibrium: player A chooses $p = 0$ and player B does not invest.

We now explore Game 2's equilibrium outcomes. Our solution concept is the perfect Bayesian equilibrium (PBE).[16] Note that for information sets $(q_1 = 1, q_2 = 0, s = b)$ and

[15] We also investigate two other natural cases of the information structure. In one case, (q_1, q_2) is taken from a subset of $[0, 1] \times [0, 1]$ such that $q_1 \geq q_2$ (equivalently $\frac{q_1}{q_2} \geq \frac{1-q_1}{1-q_2}$), according to which signal $s = b$ and signal $s = w$ can be interpreted as favoring state 1 and state 2, respectively. In another case, (q_1, q_2) is taken from a subset of $[0, 1] \times [0, 1]$ such that $q_1 \geq q_2$ and $q_1 + q_2 = 1$, according to which the conditional probability of generating signal $s = b$ in state 1 is identical to that of generating signal $s = w$ in state 2. It will be clear from Figure 3.1 that the sets of equilibria are reduced in these two cases, yet the main predictions that introducing information design increases trustworthiness and trusting acts remain unchanged.

[16] Simon & Stinchcombe (1995) state that their trembling-hand perfection does not apply to continuum

$(q_1 = 0, q_2 = 1, s = w)$, player B's posterior belief about state 1 is always one and player B optimally invests. Then, sequential rationality dictated by PBE implies that player A can treat player B's always placing no trust as a non-credible threat. Accordingly, PBE rules out the no trust equilibria.

We characterize players' PBE actions in Proposition 1. Players' actions off the equilibrium path and player B's PBE belief system are specified in Appendix C.2.1.[17]

Proposition 1 *In PBEa of Game 2, player A has two types of equilibrium actions. (1) Full trustworthiness:* $(q_1 = 1, q_2 \leq \frac{\rho_1 - \rho_0^A}{\rho_2 - \rho_0^A}, p = 1)$ *or* $(q_1 = 0, q_2 \geq \frac{\rho_2 - \rho_1}{\rho_2 - \rho_0^A}, p = 1)$. *(2) Intermediate trustworthiness:* $(q_1 = 1, q_2 = \frac{\rho_1 - \rho_0^A}{\rho_2 - \rho_0^A}, \underline{p} \leq p < 1)$ *or* $(q_1 = 0, q_2 = \frac{\rho_2 - \rho_1}{\rho_2 - \rho_0^A}, \underline{p} \leq p < 1)$, *where* $\underline{p} \equiv \frac{\rho_0^B - (1-\rho_2)}{(1-\rho_1-\rho_0^B)\frac{\rho_2 - \rho_0^A}{\rho_1 - \rho_0^A} + \rho_0^B - (1-\rho_2)} \in (0, 1)$. *Player B's equilibrium action is:* $z = 1(0)$ *if* $q_1 = 1, q_2 \leq \frac{\rho_1 - \rho_0^A}{\rho_2 - \rho_0^A}$ *and* $s = b(w)$, *and* $z = 0(1)$ *if* $q_1 = 0, q_2 \geq \frac{\rho_2 - \rho_1}{\rho_2 - \rho_0^A}$ *and* $s = b(w)$.

We provide an intuition of Proposition 1 as follows. Knowing that player B invests in state 1 and does not invest in state 2, player A uses an information structure (q_1, q_2) to "signal" that he does not choose state 2. To make his signaling credible, his choice of (q_1, q_2) must be such that given player B's strategy, his expected payoff from choosing $p = 1$ is no less than that from choosing $p = 0$. Also, player A's expected payoff from choosing $p = 1$ must be at least $\rho_1 v$ in any PBE because choosing $(q_1 = 1, q_2 = 0, p = 1)$ can give him this amount. These necessitate the following conditions: player B invests in one signal and does not invest in another signal; the likelihood of generating the first signal in state 1 is sufficiently large and the likelihood of generating the first signal in state 2 is sufficiently small. It turns out that player B's following strategy meets the

extensive form games like Game 2 and the extension of their analysis to these games remains an open issue. So we use a different solution concept for Game 2.

[17] In Appendix C.2.1, we employ a refinement of PBE in the spirit of In & Wright (2018) and Nguyen & Tan (2021) to restrict players' beliefs and actions off the equilibrium path and show that the set of equilibrium actions according to this refinement equals to the set of PBE actions.

above requirement: invests only when she observes a certain signal and the chances of such signal realization in the two states are 100% and not exceeding a threshold value, respectively. In addition, when player A chooses 100% and the threshold value as the two likelihoods, his expected payoff from choosing $p = 1$ equals that from choosing $p = 0$. Yet, he still needs to choose a reasonably large p so that player B has no incentive to deviate from the above strategy (i.e., $\underline{p} \leq p \leq 1$).

We illustrate both players' equilibrium strategies and expected payoffs on the equilibrium paths in Figure 3.1 and Figure 3.2, respectively. Player A's expected payoff is a constant value, $\rho_1 v$, in all PBEa. The intuition is that player A's expected payoff in any PBE cannot exceed $\rho_1 v$ because his expected payoff function is linear in p: a choice of $p = 0$ will lead player B to never invest and in turn player A's expected payoff is v_0^A ($< \rho_1 v$); the highest expected payoff for player A is $\rho_1 v$ when he optimally chooses a certain p greater than zero. Recall that player A has an expected payoff of at least $\rho_1 v$ in any PBE, so he must have an expected payoff of $\rho_1 v$ in any PBE. Player B's expected payoff is an increasing function of player A's choice of $p \in [\underline{p}, 1]$, with a minimum of v_0^B and a maximum of $(1 - \rho_1)v$.[18]

Equilibria insights. Proposition 1 suggests that introducing information design increases trustworthiness and trusting acts in two ways. First, player A commits to state 1 and picks a cutoff information structure so that player B invests with certainty. Second, player A does not commit to a state and constructs a delicate information structure to ensure that player B invests with positive probability. Player A employs the delicate information structure to assure his opponent that it is in his best interest to choose a level of trustworthiness that sustains player B's optimal investing in a favorable signal. Compared to Game 1, the insights show that introducing information design, together

[18]Player B's expected payoff function on the equilibrium paths: $EV_B(p) = p[(1-\rho_1)v - (1-\rho_2)\frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}v - \frac{\rho_2-\rho_1}{\rho_2-\rho_0^A}v_0^B] + (1-\rho_2)\frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}v + \frac{\rho_2-\rho_1}{\rho_2-\rho_0^A}v_0^B$.

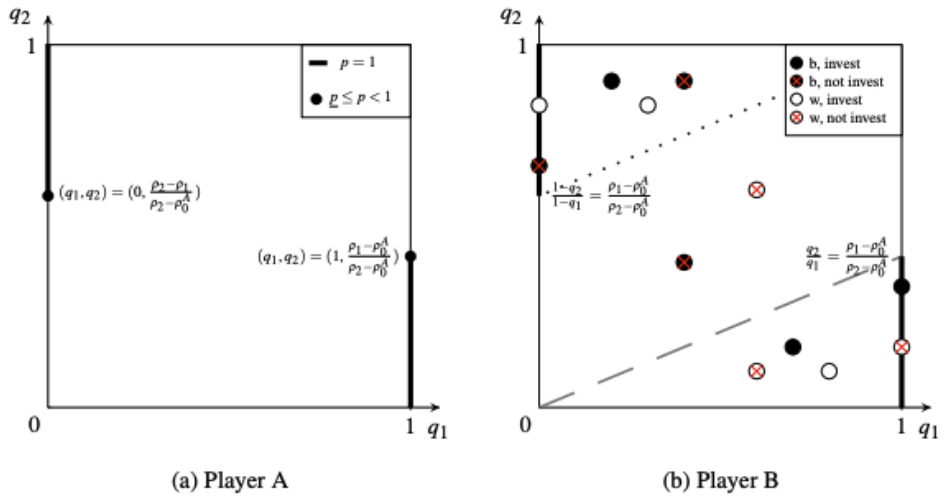
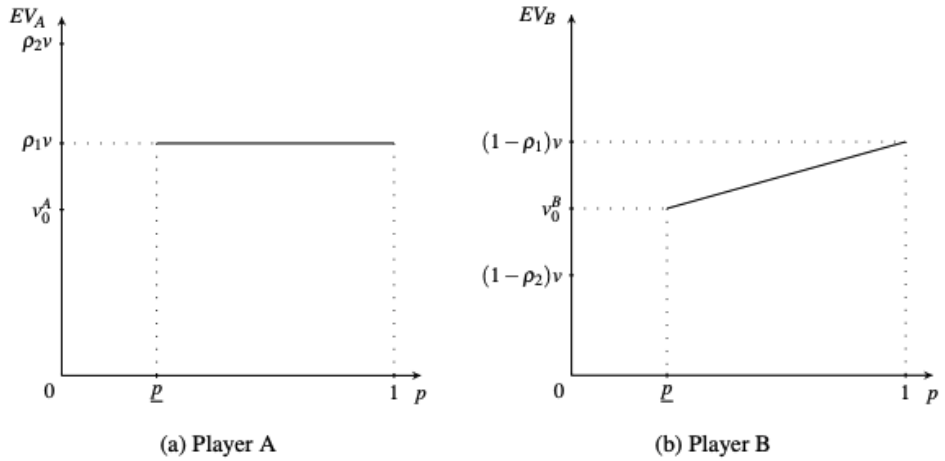


Figure 3.1: Equilibrium Strategy in Game 2 *Notes: In panel (a), player A's equilibrium actions of full trustworthiness and intermediate trustworthiness are highlighted in thick lines and circles, respectively. In panel (b), the dashed and dotted lines are* $\frac{q_2}{q_1} = \frac{\rho_1 - \rho_0^A}{\rho_2 - \rho_0^A}$ *and* $\frac{1-q_2}{1-q_1} = \frac{\rho_1 - \rho_0^A}{\rho_2 - \rho_0^A}$*, respectively. Player B's equilibrium strategy can be characterized by three parts: (1) only invest in a favorable signal when* (q_1, q_2) *is on the equilibrium path (the area of thick lines), (2) always invest in a favorable signal and invest in an unfavorable signal with positive probability when* (q_1, q_2) *is off the equilibrium path and the likelihood ratio* $\frac{q_2}{q_1}$ *or* $\frac{1-q_2}{1-q_1}$ *falls below a cutoff value (the area above the dotted line or below the dashed line), and (3) never invest otherwise (the area between the dashed line and the dotted line), where signal* $s = b(s = w)$ *is favorable if* $\frac{q_1}{q_2} > 1(\frac{1-q_1}{1-q_2} > 1)$ *and is unfavorable otherwise.*

with its bringing about preliminarity of player A's choice and an increase in his power to commit to a certain level of trustworthiness, fosters trustworthiness and trusting acts.

Extension from Kamenica & Gentzkow (2011). Their model implies that when p is exogenous and observable, player A optimally chooses $q_1 = 1$ and q_2 such that player B is indifferent to either investing or not investing given a favorable signal. Therefore, the optimal q_2 is a strictly increasing function of p. However, Proposition 1 reveals that this implication does not hold when p is endogenous and unobservable. In Game 2, if $q_1 = 1$ and q_2 is excessively large, player A will choose $p = 0$ given player B's investing in a favorable signal; consequently, investing in a favorable signal is no longer player B's optimal strategy. To ensure player B's optimal investing in a favorable signal, player A's choice of q_2 cannot exceed an upper bound that exactly makes player A indifferent to either choosing $p = 0$ or choosing the equilibrium p. Given the current payoff structure,



(a) Player A (b) Player B

Figure 3.2: Expected Payoffs on the Equilibrium Paths in Game 2 *Note: On the equilibrium paths,* $p \in [\underline{p}, 1]$*, where* $\underline{p} \equiv \frac{\rho_0^B - (1-\rho_2)}{(1-\rho_1-\rho_0^B)\frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}+\rho_0^B-(1-\rho_2)} \in (0,1)$.

the upper bound is $\frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}$ and when $q_2 = \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}$, $p = \underline{p}$ exactly makes player B indifferent to either investing or not investing in a favorable signal. Thus, whichever of $p \in [\underline{p}, 1]$ player A chooses, the optimal q_2 for him remains constant.

Equilibria outcome changes across games. Player A chooses to be trustworthy with zero probability in Game 1, but he chooses to be trustworthy with up to full probability in Game 2, that is, $p \in [\underline{p}, 1]$. Player B does not invest in Game 1, but in Game 2, she invests on the equilibrium paths $(q_1 = 1, q_2 \leq \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}, s = b)$ and $(q_1 = 0, q_2 \geq \frac{\rho_2-\rho_1}{\rho_2-\rho_0^A}, s = w)$. In terms of ex ante expected payoffs, in Game 1, player A and B get v_0^A and v_0^B, respectively. In Game 2, player A's payoff is $\rho_1 v$ $(> v_0^A)$ and player B's payoff generally exceeds v_0^B; thus, the aggregate payoff increases from Game 1 to Game 2.

3.3 Experimental Design and Procedure

Our experiment tests the effect of introducing information design on trustworthiness and trusting acts, and also the theory predictions on players' behavioral patterns. To

isolate the effect of introducing information design from the effect of experience or learning, this experiment collects subjects' initial responses to the games. We discuss an alternative experimental design in Section 3.6.

3.3.1 Experimental Design

We employ a within-subjects design according to which each subject plays both Game 1 and Game 2. We set $v_0^A = v_0^B = 10$, $(\rho_1 v, (1-\rho_1)v) = (15, 15)$ and $(\rho_2 v, (1-\rho_2)v) = (22, 8)$. Experimental instructions are provided in Appendix C.4.

The task of deciding the probability p is described to subjects intuitively: (1) player A picks an integer number P from 0 to 100; (2) if $P = 100$, then payoff allocation $(15, 15)$ is realized, and if $P = 0$, then payoff allocation $(22, 8)$ is realized; (3) if $0 < P < 100$, then payoff allocation $(15, 15)$ is realized with a probability of $P\%$ and payoff allocation $(22, 8)$ is realized with a probability of $(100 - P)\%$. The task of deciding the conditional likelihood q_i $(i = 1, 2)$ is also described intuitively: (1) when payoff allocation $(15, 15)$ $((22, 8))$ is realized, a ball from urn 1 (urn 2) with 100 balls in total is randomly drawn; (2) player A picks an integer number Q_i from 0 to 100 so that urn i contains Q_i black balls and $100 - Q_i$ white balls.

In Game 1, payer A first decides a number P. Then, player B decides whether to invest in the project without knowing player A's choice of P and which of the two payoff allocations is realized. In Game 2, player A first decides the numbers P, Q_1, and Q_2. Then, a payoff allocation is realized by a computer server according to the number P, and a ball is randomly drawn from the corresponding urn. Finally, player B observes only the numbers Q_1, Q_2 and the randomly drawn ball, and decides whether to invest in the project. In both games, the realized payoff allocation is implemented when player B invests in the project, and each player keeps the endowment of 10 when player B does

not invest.

Each subject is randomly assigned a constant role throughout the experiment, either player A or player B. Each subject plays five rounds of each game,[19] and the order of the two games is varied across sessions to control the game order effect. Each subject is paired with a new anonymous opponent in each round and receives no feedback.[20] The payoff unit in the experiment is tokens and each token can be redeemed for two Chinese yuan. Subjects are paid according to the tokens they earn in one randomly selected round, so they have no incentive to play different strategies across rounds to hedge against risks. Each subject also receives an additional show-up fee of five Chinese yuan. After the paid rounds, subjects are asked to complete a six-question survey, which includes belief questions about whether they think player A (B) on average receives a higher, lower, or identical payoff in Game 2 compared to that in Game 1.

To ensure that subjects understand the two games, we design a practice stage consisting of four rounds before the paid rounds. In the practice stage, a subject plays each role in each game only once against a computer opponent who always makes random choices.[21] Subjects are informed of the computer opponent's strategy and receive feedback at the end of each practice round.

3.3.2 Discussion of the Design

We discuss below a few aspects of the design.

[19]Playing multiple rounds of Game 2 is useful for identifying player B's strategy. The number of rounds played for Game 1 is set to be the same for comparability.

[20]This design is standard in the literature on eliciting initial responses. See, for instance, Costa-Gomes & Crawford (2006) and Alaoui & Penta (2015), among others. This design excludes or suppresses confounds such as reputation consideration, learning about the opponents' strategies, and the effect of past experience.

[21]The setting of making random choices by a computer opponent ensures that the practice experience will not produce confounds in the paid experiment. To prevent subjects from misunderstanding their opponents' choices in the paid rounds, we make it clear in experimental instructions that their opponents in practice rounds and paid rounds are different in nature.

Payoff allocations. The payoff allocations we select satisfy the assumptions in the theoretical setting. Specifically, $\rho_0^A + \rho_0^B = \frac{20}{30} < 1$, $\rho_0^A = \frac{10}{30} < \rho_1 = \frac{15}{30} < \rho_2 = \frac{22}{30}$, and $1 - \rho_2 = \frac{8}{30} < \rho_0^B = \frac{10}{30} < 1 - \rho_1 = \frac{15}{30}$. Our selection is also for the practical purpose. Bohnet et al. (2008) also use the payoff allocations $(10, 10)$, $(15, 15)$, and $(22, 8)$ in their trust game and conduct experiments in a highly diverse set of countries (Brazil, China, Oman, Switzerland, Turkey, and the United States).

Treatments. First, player A in Game 2 has full flexibility to decide an information structure: $Q_1 \in \{0, 1, \cdots, 99, 100\}$ and $Q_2 \in \{0, 1, \cdots, 99, 100\}$, which closely matches the theoretical setting. Alternatively, one may want to make some restrictions on sets of information structure in lab implementation such as: (1) $(Q_1, Q_2) \in \{(100, 0), (50, 50)\}$, (2) $Q_1 + Q_2 = 100$, or (3) $Q_1 \geq Q_2$. Each of the first two restrictions will rule out the possibility of subjects playing according to equilibria with intermediate trustworthiness. The third restriction designates a black ball as a signal favoring state 1 and is the most innocent restriction. However, a considerable fraction of A players in our experiment still prefer using a white ball as a signal favoring state 1. Second, the equilibria values of p in Games 1 and 2 equal 0 and an interval of $[\frac{1}{7}, 1]$ given the values of payoff parameters, respectively. We thus allow P to take any value from $\{0, 1, \cdots, 99, 100\}$, which corresponds to the set of $\{0, 0.01, \cdots, 0.99, 1\}$ for values of p. Third, we employ a within-subjects design instead of a between-subjects design. While both designs can gauge treatment effects, the former additionally makes it possible for us to investigate subject heterogeneity in the effect of introducing information design.

Constant roles. Burks et al. (2003) show that a subject's playing both roles in trust games brings about confounds. Thus, we let each subject play a constant role throughout the experiment.

3.3.3 Experimental Procedure

There were twenty participants in each session of our experiment. Ten subjects were randomly chosen to be player A and the other ten were selected to be player B. Each subject was paired with an opponent subject exactly once. Each participant was provided with a hard copy of the experimental instructions. When all participants were seated and assigned identification numbers, pre-recorded audio of the experimental instructions was played through loudspeakers to maintain uniformity across sessions. After the audio was played, an experimental investigator answered any questions about understanding the experimental instructions. The experiment started when participants' questions had all been addressed. The experimental interface was programmed through z-Tree (Fischbacher 2007). At the end of the experiment, each subject's earnings were displayed only on her computer screen. Subjects were paid privately and then left the laboratory.

We conducted sixteen sessions at the Finance and Economics Experimental Laboratory of Xiamen University, with five sessions on December 6th, eight sessions on December 7th and three sessions on December 8th of 2019. Game 2 preceded Game 1 in the eight sessions on December 7th and Game 1 preceded Game 2 in the other eight sessions. A total of 320 student subjects at Xiamen University from different majors were recruited , each of whom participated in only one session. Each session lasted approximately 35 minutes. Subjects playing role A (B) on average earned 32.6 (26.05) Chinese yuan (about 4.66 and 3.72 US dollars at the contemporaneous exchange rate), compared to a local minimum hourly wage of 18 Chinese yuan in that year.

3.4 Experimental Results

An observation in our sample refers to the values of variables for a pair of subjects in one round. The main variables of interest include trustworthiness and information

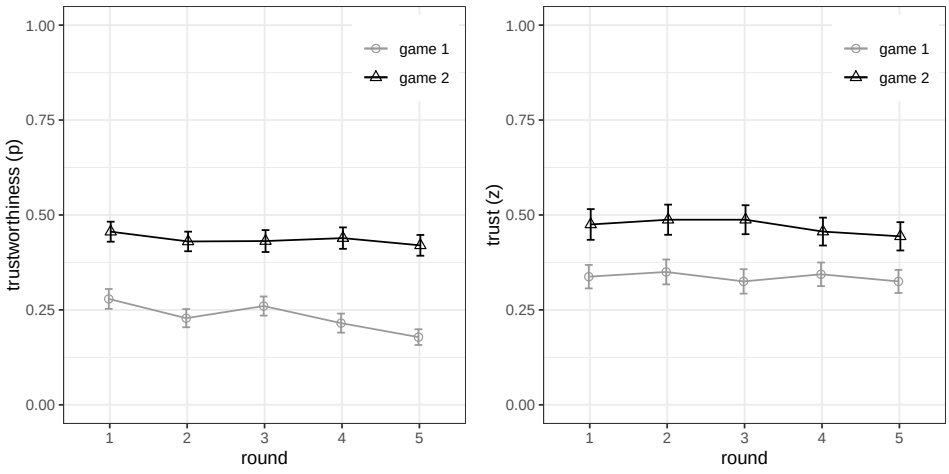


Figure 3.3: Choices Across Rounds *Notes: Circles or triangles represent sample means; line segments represent ± standard errors that have been adjusted for within-subject correlations.*

structure, measured by $(p = \frac{P}{100}, q_1 = \frac{Q_1}{100}, q_2 = \frac{Q_2}{100})$, the signal indicated by the color of the drawn ball $(s \in \{b, w\})$, trust indicated by the action of investing or not investing $(z \in \{1, 0\})$, and a dummy indicating whether the observation is from Game 1 or Game 2. Overall, the experiment collected 1600 observations from 1600 rounds, including 800 observations of each game. In the following analysis, we further cluster the observations at the subject or session level whenever it is needed.

3.4.1 Treatment effects

Figure 3.3 demonstrates the round-by-round data of trustworthiness (p) and trust (z). It is clear that trustworthiness and trusting acts increase from Game 1 to Game 2 irrespective of round.

Table 3.1 shows that when we consider the full sample, subsample 1, in which Game 1 is played first, or subsample 2, in which Game 2 is played first, the treatment effects are qualitatively consistent with the equilibrium analysis. The average of player A's chosen probability of state 1, p, nearly doubles from 0.23 to 0.44 (p-value$<$0.001, Wilcoxon signed rank test on the averages at the subject level). The fraction of rounds in which player

B invests increases by 40% from 0.34 to 0.47 (p-value<0.001, Wilcoxon signed rank test on the averages at the subject level). Since the aggregate payoff for a pair of subjects is 20 when player B does not invest and is 30 when player B invests, the increase in the frequency of player B's investing implies that the aggregate payoff increases.

Table 3.1: Treatment Effects

	Game 1	Game 2	Difference
Trustworthiness (p)	0.23	0.44	0.20***
	0	$\geq \frac{1}{7}$	
	0.26	0.42	0.15**
	0.2	0.45	0.25***
Trust (z)	0.34	0.47	0.13***
	0	$\geq \frac{1}{2}$	
	0.39	0.45	0.06*
	0.28	0.49	0.21***

*Notes: For each variable, rows 1-4 represent the full sample, equilibrium prediction, subsample 1 and subsample 2, in which Game 1 is played first and second, respectively. The used test is Wilcoxon signed rank test on the averages at the subject level. *, **and *** indicate significance at the 10%, 5% and 1% levels, respectively.*

As to each player role's payoff change, its compliance with the equilibrium analysis is mixed. No significant difference is observed in player A's average payoff between games.[22] Player B's average payoff increases significantly from 9.91 to 11.35 (p-value<0.001, Wilcoxon signed rank test on the averages at the subject level).[23]

Figure 3.4 presents the data in a disaggregated form: the distributions of each subject's average p, z and payoff. The figure clearly shows that the distributions of player A's level of trustworthiness and player B's average payoff in Game 2 first-order stochastically

[22]In Appendix C.3.1, we also analyze player A's average payoff change between games under each realized state: the average payoff increases significantly under state 1 and decreases significantly under state 2. This observation indicates that introducing information design benefits only those A players who choose to be trustworthy.

[23]According to the post-experiment questionnaire, among three belief options of being higher, lower and equal, over half of subjects believe that, on average, player A receives a higher payoff in Game 2 than that in Game 1, and more than three-quarters of the subjects believe that player B receives a higher payoff in Game 2 than that in Game 1.

dominate those in Game 1, respectively. It also shows that player B trusts more often in Game 2 than in Game 1, and the distributions of player A's average payoff in the two games are not clearly distinguishable.

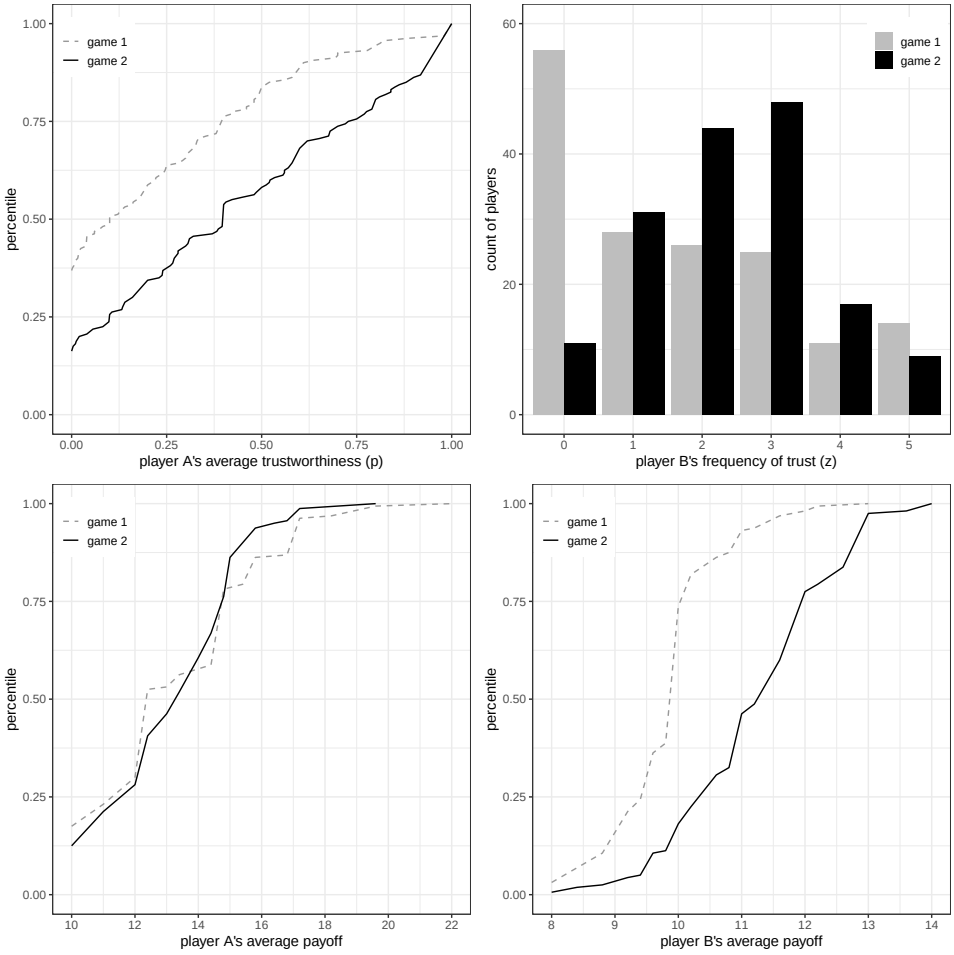


Figure 3.4: Distributions of p, z and payoff at the subject level

In Appendix C.3.2, we supplement the above analyses by considering only the first round play of each game. We still find trustworthiness, trusting act and player B's payoff increase significantly from Game 1 to Game 2. In addition, the data in a disaggregated form exhibits a similar pattern as Figure 3.4 does.

3.4.2 Measure of Informativeness

Before looking into the observed behavioral patterns, we now introduce a measure of informativeness of an information structure $(q_1, q_2) \equiv Q$, which itself is based on Blackwell's informativeness order (Blackwell 1953). Information structure Q is (weakly) more Blackwell informative than information structure Q' if we can replicate the second information structure from the first by randomly drawing a signal after each observation of a signal in the first information structure. Being strictly more informative, less informative, identically informative, the most informative, and the least informative can then be defined.

Blackwell's order is not complete, so there is no numerical representation in the standard sense; that is, there does not exist a mapping from information structures to the real number set that satisfies two properties: (1) if Q is strictly more informative than (identically informative as) Q', then the former is assigned a greater number (the same number); and (2) the converse statement of (1). Nevertheless, an index that satisfies only the first property is useful for differentiating information structures.[24] For such an index, Q is no less informative than Q' if the former has an index value no less than that of the second. Below, we propose a natural definition of such an index.

Definition 1 *We define index u_Q for Blackwell's informativeness of Q as the probability that Q is more informative than any Q' randomly drawn from $[0,1]\times[0,1]$ according to a uniform distribution, that is, $u_Q \equiv Pr(Q' \sim U[0,1]\times[0,1] : Q$ is more informative than $Q')$.*[25]

[24] Notably, an index that satisfies only the second property is less useful, as can be illustrated by an example. Consider three information structures $(0.3, 0.1)$, $(0.8, 0.6)$, and $(0.2, 0.1)$. Based on Claim 1 in the proof of Proposition 2 in Appendix C.2.2, the first two are not comparable and the last two are not comparable, but $(0.3, 0.1)$ is strictly more informative than $(0.2, 0.1)$. In this example, there is no index that satisfies the second property and in turn $(0.3, 0.1)$ and $(0.2, 0.1)$ cannot be differentiated.

[25] One may consider an alternative definition of the index based on a dual perspective: an index of Q is defined as one minus the probability that any Q' randomly drawn from $[0,1]\times[0,1]$ according to a uniform distribution is more informative than Q. This alternative index satisfies only a weak version of the first property.

We show that the index can be characterized by the difference between the two likelihoods of an information structure, and it indeed satisfies the first property.

Proposition 2 *For any Q, $u_Q = |q_1 - q_2|$. If Q is strictly more Blackwell informative than (identically informative as) Q', then $u_Q > u_{Q'}$ ($u_Q = u_{Q'}$).*

Note that $Q = (1, 0)$ or $(0, 1)$ is the most informative structure and $Q = (q, q)$ is the least informative structure. Since $u_Q = 1$ if and only if $Q = (1, 0)$ or $(0, 1)$, and $u_Q = 0$ if and only if $Q = (q, q)$, we can strengthen the interpretation of $u_Q = 1(0)$: Q is the most (least) informative structure if $u_Q = 1(0)$.

3.4.3 Player A's patterns

We first examine the extent to which player A's choice at the aggregate level is in compliance with the equilibrium prediction. The percentages of rounds in which player A chooses $p = 0$ (equilibrium strategy), $0 < p \leq 0.5$, $0.5 < p < 1$, and $p = 1$ in Game 1 are 57.75%, 22.25%, 9.5% and 10.5%, respectively. In Game 2, 24.25% of observations comply with the equilibrium predictions of player A's actions.[26] All belong to full trustworthiness with the most informative structure: $(p = 1, q_1 = 1, q_2 = 0)$ (163 observations) and $(p = 1, q_1 = 0, q_2 = 1)$ (31 observations).[27]

We now consider player A's choice of information structure at the aggregate level. According to the observed distribution of the index u_Q, the quartiles of u_Q are 0, 0.2, 1 and 1, with a mean value of 0.40. In addition, the fractions of $u_Q = 0$ and $u_Q = 1$ are 28% and 26.38%, respectively. Note that the equilibrium model predicts only the pattern

[26] We also consider allowing 10% perturbation in the equilibrium predictions and find only slight improvement: 25.88% of observations comply with the predictions.

[27] In the subsample in which (q_1, q_2) is on the equilibrium path, almost all observations (211 out of 213) have the most informative structure. In the subsample in which (q_1, q_2) is on (off) the equilibrium path, 91.08% (57.75%) of observations comply with the equilibrium prediction of p.

of $u_Q = 1$. Therefore, one may attribute the pattern of $u_Q = 0$ to errors or propose an alternative model to rationalize it.

Pattern 1 (Information structure) *More than half of the information structures that subjects choose are the most (26%) or least (28%) informative.*[28] *Moreover, in almost all (92%) the observations with the most informative structure,* $p = 1$*; in approximately 57% of the observations with the least informative structure,* $p = 0$*.*

How a player A's choice of p is associated with his choice of information structure Q in Game 2 is also of our interest. A correlation test between p and u_Q on the averages at the subject level shows that they are positively related (Pearson's $r = 0.764$, p-value < 0.001).[29] We then investigate the association at the subject level. Let $\bar{p}$ be the average value of p chosen by a subject in five rounds of a game. We divide $\bar{p}$ into two categories: low $\bar{p}$ if $\bar{p} \leq 0.5$ and high $\bar{p}$ if $\bar{p} > 0.5$. Similarly, let $\overline{u_Q}$ be the average value of u_Q chosen by a subject in five rounds of a game. We label $\overline{u_Q}$ as low (high) informativeness if $\overline{u_Q} \leq 0.5$ ($\overline{u_Q} > 0.5$). We consider whether a subject with a low or high $\bar{p}$ chooses a more informative structure.

Table 3.2 suggests that some subjects signal their high trustworthiness by choosing high informativeness and other subjects obscure their low trustworthiness by choosing low informativeness.[30] The first case is predicted by the equilibrium model, and the second case likely reflects subjects' non-equilibrium strategic reasoning. Additionally, those who choose low $\bar{p}$ and high informativeness appear strategically naive and selfish, and those who choose high $\bar{p}$ and low informativeness appear strategically naive and prosocial.

Pattern 2 (Association of trustworthiness and information structure) *Among*

[28]The fractions of choosing ($p = 0, u_Q = 0$) and ($p = 1, u_Q = 1$) are 16% and 24.25%. When 10% perturbation of the two choices is allowed, the fractions are 27.25% and 24.88%.

[29]The positive correlation still holds even if we exclude those A players such that $(\bar{p}, \overline{u_Q}) = (1, 1)$ or $(0, 0)$ (Pearson's $r = 0.547$, p-value < 0.001).

[30]We also investigate the association at the round level and find a similar result.

Table 3.2: Association of p and u_Q at the Subject Level in Game 2

	Low informativeness	High informativeness	Total
Low $\bar{p}$	53.75%(86)	4.38%(7)	58.13%(93)
High $\bar{p}$	13.13%(21)	28.75%(46)	41.88%(67)
Total	66.88%(107)	33.13%(53)	100%(160)

Notes: The number of subjects in each category is reported in parentheses. Low (high) informativeness: $\overline{u_Q} \leq 0.5 (> 0.5)$; low (high) $\bar{p}$: $\bar{p} \leq 0.5 (> 0.5)$. Pearson's chi-square test: p-value<0.001; Fisher's exact test: p-value<0.001.

those subjects who choose a low $\bar{p}$, almost all (92%) choose low informativeness. Among those who choose a high $\bar{p}$, the majority choose high informativeness (69%). A small fraction of subjects choose a low $\bar{p}$ with high informativeness or a high $\bar{p}$ with low informativeness.

We finally explore how player A's choice of p in Game 2 is associated with his choice of p in Game 1. A correlation test shows that player A's $\bar{p}$ in Game 1 and Game 2 are positively related (Pearson's $r = 0.25$, p-value<0.01). We then consider the fractions of subjects whose $\bar{p}$ increases, remains unchanged or decreases from Game 1 to Game 2. In principle, the first group is predicted by the equilibrium model, the second group may be attributed to a null effect of introducing information design, and the third group may be attributed to errors or an adverse effect of introducing information design. We find the fractions are 60.63%, 13.75%, and 25.63% respectively, and the first two groups account for 74.38% of A players. Moreover, the fraction of subjects whose $\bar{p}$ exceeds 0.5 in both games is 13.75%, which may be attributed to social preferences.

Pattern 3 (Change in trustworthiness) *For the majority of subjects, $\bar{p}$ in Game 2 is no less than that in Game 1. A considerable fraction of subjects choose high $\bar{p}$ in both games.*

3.4.4 Player B's patterns

We first examine player B's behavioral compliance with the equilibrium prediction. While the equilibrium model predicts that player B does not invest in Game 1, the percentage of rounds in which player B invests in Game 1 is 33.63%. Compared to the equilibrium predictions in Game 2, the consistency ratio of player B's actions exceeds 96% on the equilibrium path (q_1, q_2), which is expected given that player A chooses the most informative structure on the equilibrium path. Player B's actions are consistent with predictions approximately 75% of the time off the equilibrium path (q_1, q_2), partly due to the equivocal prediction of actions in many off-equilibrium paths.

Note that in Game 2, player B is off the equilibrium path (q_1, q_2) for most of time given player A's low compliance with the equilibrium strategy. Thus, the above simple description of player B's binary action off the equilibrium path may not be sufficiently informative. Below, we investigate how player B's investing decision hinges on her observed information structure and signal.

We categorize a signal generated from an information structure, $s|Q$, based on whether it recommends investing (favorable) or not investing (unfavorable). Note that player B's optimal state-dependent action is investing if and only if the underlying state is state 1. Thus, we say $s|Q$ is a favorable signal if the likelihood of generating the signal in state 1 is higher than that in state 2, that is, $(s = b, \frac{q_1}{q_2} > 1)$ or $(s = w, \frac{1-q_1}{1-q_2} > 1)$; otherwise, $s|Q$ is an unfavorable signal.

We run a probit regression of player B's investing choice on a dummy indicating a favorable signal, the interaction between u_Q and the dummy, the interaction between u_Q and another dummy indicating an unfavorable signal, the frequency of investing in Game 1 and a dummy indicating Game 1 being played first, with standard errors clustered at the subject level. The estimated coefficients (standard errors) are 0.170 (0.161), 2.852

(0.299), -0.462 (0.281), 1.532 (0.281) and -0.304 (0.154), respectively. The results suggest that the effect of informativeness is positive (negative) for favorable (unfavorable) signals, and the effect is much larger for favorable signals.

We then consider player B's investing choice at the subject level. Based on a dichotomy of information structure, i.e., low (high) index if $u_Q \leq 0.5$ ($u_Q > 0.5$), and a dichotomy of signal (favorable or not), player B's information set can be classified into four categories: H_1 (H_2), unfavorable signal and information structure of low (high) index; and H_3 (H_4), favorable signal and information structure of low (high) index. Given such classification, player B's strategy space consists of 16 strategies, each of which specifies a binary action of investing or not investing for each H_i. One can look at the exact distribution of player B's strategy, but this information is not very useful.[31] We thus consider the distribution at a coarse level: we investigate actions under each H_i in the subsample of subjects whose action in that H_i is self-consistent. Take H_1 for example, we consider the subsample in which a subject takes the same action in all rounds in which she observes H_1.

Overall, 158 of 160 subjects have self-consistent actions for at least one H_i. Among them, 117, 44, 100 and 121 subjects have self-consistent actions for information sets H_1, H_2, H_3 and H_4, respectively, and the corresponding fractions of subjects who invest are 10.26%, 9.09%, 43% and 98.35%.[32] Consistent with the equilibrium prediction, the

[31] First, a strategy space consisting of 16 strategies is still substantially scattered given the sample of 160 subjects, which leads to a small sample size of subjects who choose each specific strategy. Second, a subject may observe only some of the four categories of information sets in five rounds and, in turn, she can be classified as multiple strategy types. For example, a subject who observes only two (one) categories of information sets can be classified as four (eight) strategy types. Third, when a subject observes a specific category of information set in more than one round, she may invest in one round and not invest in another round; consequently, we cannot classify her as any of the 16 strategy types. In particular, this scenario is likely to occur when observing $(q_1, q_2 = q_1, b/w)$.

[32] We also investigate the distribution of subject's action in the subsample in which a subject has self-consistent actions for *any* H_i she observes. Overall, 103 of 160 subjects belong to this subsample. Among them, 101, 27, 73 and 85 subjects choose self-consistent actions in information sets H_1, H_2, H_3 and H_4, respectively, and the corresponding fractions of subjects who invest are 10.89%, 11.11%, 39.73% and 98.82%.

fraction of subjects who invest is very high (low) when observing a favorable (unfavorable) signal and an information structure with a high index.

Interestingly, the fraction of subjects who invest is considerable when observing a favorable signal and an information structure with a low index. Given that an information structure with a low index is associated with a low level of trustworthiness, strategically sophisticated subjects would be better not to invest when observing such information structures. This observation indicates that some subjects may fail to understand that the information structure per se conveys a message about trustworthiness, which is not captured by the equilibrium prediction.[33] [34]

Pattern 4 (Distribution of player B's strategy in Game 2) *The fraction of subjects who invest is the highest (approximately* 98%*) when observing a favorable signal and an information structure with a high index. The fraction is the lowest (approximately* 9%*) when observing an unfavorable signal and an information structure with a high index. In addition, a considerable fraction of subjects (approximately* 43%*) invest when observing a favorable signal and an information structure with a low index.*

We finally investigate the association of player B's strategies in the two games. A correlation test shows that the numbers of investing rounds in Game 1 and Game 2 of each player B are positively related (Kendall's $\tau = 0.300$, p-value<0.001).[35] We classify a subject into the "trusting group" if she invests in at least three rounds of Game 1 and into the "no trusting group" otherwise: there are 50 subjects in the former group and

[33] An analysis of subjects' action patterns contingent on the information set at the aggregate level conveys a similar message. The fractions of rounds in which subjects invest for H_1 to H_4 are 21.70%($\frac{79}{364}$), 10.20%($\frac{5}{49}$), 45.22%($\frac{71}{157}$) and 96.09%($\frac{221}{230}$).

[34] We rule out the possibility that these subjects mistakenly believe that their opponents make random choices in both paid rounds and practice rounds: the correlation coefficient between a vector of these subjects' actual choices and a vector of their predicted choices under the assumption of their holding the wrong belief is 0.196, which is typically interpreted as a very weak correlation.

[35] At the subject level, the fractions of subjects whose frequency of investing in Game 2 is larger, equal to, and lower than that in Game 1 are 53.1%, 24.4%, and 22.5% respectively. The first two groups account for 87.5% of B players.

110 subjects in the latter group. As expected, the distributions of observed information sets are not statistically different between the two groups (see the detailed analysis in Appendix C.3.3). However, the frequency of investing in Game 2 of the "trusting group" is always higher than that of the "no trusting group", as shown in Table 3.3, indicating that some subjects seem to be inherently more trusting than others, regardless of the game setting.

Table 3.3: Player B's Strategies Across Games

	Game 1	Game 2			
		H_1	H_2	H_3	H_4
"Trusting group"	75.60%($\frac{189}{250}$)	39.84%($\frac{49}{123}$)	25%($\frac{3}{12}$)	68.63%($\frac{35}{51}$)	98.44%($\frac{63}{64}$)
"No trusting group"	14.55%($\frac{80}{550}$)	12.45%($\frac{30}{241}$)	5.41%($\frac{2}{37}$)	33.96%($\frac{36}{106}$)	95.18%($\frac{158}{166}$)
Difference	61.05%	27.39%	19.59%	34.67%	3.26%

Notes: "Trusting/no trusting group" refers to subjects who invest in no less than/less than three rounds of Game 1. The percentage is the frequency of investing. H_1 (H_2), unfavorable signal and information structure with a low (high) index; H_3 (H_4), favorable signal and information structure with a low (high) index. Chi-square test with Donner's adjustment on clustered data at the subject level for Game 1 (p-value<0.001), H_1 (p-value<0.001), H_2 (p-value=0.060), H_3 (p-value<0.001) and H_4 (p-value=0.298).

Pattern 5 (Association of player B's strategies across games) *In Game 2, the "trusting group" is always more likely to invest than the "no trusting group", especially when observing an information structure with a low index (H_1 or H_3).*

3.5 Prosociality and Strategic Sophistication

We show that the treatment effects and some behavioral patterns are in line with the equilibrium analysis. However, there are also a few notable deviations including some A players' choosing zero trustworthiness with the least informative structure and some B players' seemingly failing to understand that the information structure per se conveys a

message about trustworthiness. In this section, we propose a model of heterogeneity in prosociality and strategic sophistication to rationalize the treatment effects and all major patterns.

We first discuss two alternative modeling approaches: a level-k analysis with the standard preference and an equilibrium analysis with social preferences. The two approaches are somewhat appealing in the sense that they deviate from the standard framework only in one dimension. However, they have limitations in rationalizing the aforementioned deviations. For example, the level-k model with the standard preference cannot make a *sharp* prediction about player A's optimally choosing $(p = 0, q_1 = q_2)$: while the strategy can be optimal if player B of a lower level believes that player A chooses $p = 1$, a continuum of other strategies (e.g., $p = 0, 0 < q_1, q_2 < 1$) are also optimal in this situation. The equilibrium model with social preferences cannot make a sharp prediction about player A's optimally choosing $(p = 0, q_1 = q_2)$ either, and requires that player B correctly understands that the information structure per se conveys a message about trustworthiness. We thus conduct a level-k analysis with social preferences: this behavioral model generates sharp predictions about the deviations when retaining the main predictions based on the standard equilibrium model.

3.5.1 The behavioral model

We provide detailed assumptions and intuitions of the behavioral model in Appendix C.1. The optimal strategy for each behavioral type is characterized in Proposition 3.

Proposition 3 *Given the assumptions about players' heterogeneity in prosociality, viewpoints about prosociality and strategic sophistication, and the assumption about conditionally pessimistic posterior belief in zero-probability information sets, the optimal strategy for each type is summarized in Table 3.4.*

Table 3.4: Optimal Strategy for Each Behavioral Type

Role	Type	Game 1	Game 2
A	$(P, L_0/L_1)$	$p = 1$	$p = 1,\ q_1 \sim U[0,1],\ q_2 \sim U[0,1]$
	$(P, L_{k\geq 2})$	$p = 1$	$p = 1, (q_1, q_2) = (1,0)$ or $(0,1)$
	$(S, L_0/L_1)$	$p = 0$	$p = 0,\ q_1 \sim U[0,1],\ q_2 \sim U[0,1]$
	$(S, L_2/L_3)$	$p = 0$	$p = 0, q_1 = q_2$
	$(S, L_{k\geq 4})$	$p = 0$	$p = 1, (q_1, q_2) = (1,0)$ or $(0,1)$
B	$(P, L_{k\geq 0}, \pi)$	$z = 1$	$z = 1$
	(S, L_0, π)	$z = 0$	$z = 0$
	$(S, L_1/L_2, \pi)$	$z = 1$ if $\pi \in (\frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}, \overline{\pi})$; $z = 0$ if $\pi \in (0, \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}]$	$z = 1$ if $(\frac{q_1}{q_2} > \frac{\rho_0^B - (1-\rho_2)}{1-\rho_1-\rho_0^B}\frac{1-\pi}{\pi}, b)$ or $(\frac{1-q_1}{1-q_2} > \frac{\rho_0^B - (1-\rho_2)}{1-\rho_1-\rho_0^B}\frac{1-\pi}{\pi}, w)$; $z = 0$ otherwise
	$(S, L_{k\geq 3}, \pi)$	same as above	$z = 1$ if $(0 < q_1 \leq 1, q_2 = 0, b)$ or $(0 \leq q_1 < 1, q_2 = 1, w)$; $z = 0$ otherwise

Notes: P and S index prosocial and selfish players, respectively. L_k refers to a level-k player. π refers to player B's belief about the fraction of prosocial A players. The description of $\overline{\pi}$ can be found in Appendix C.1.

Similar to the equilibrium model, the behavioral model explains the treatment effects, the details of which are discussed in Appendix C.1.3. Below, we highlight how the behavioral model rationalizes those patterns that are not predicted by the equilibrium model.

Association of trustworthiness and information structure. Type $(S, L_2/L_3)$ chooses $(p = 0, q_1 = q_2)$ and type $(S, L_{k\geq 4})$ chooses $(p = 1, q_1 = 1, q_2 = 0)$ or $(p = 1, q_1 = 0, q_2 = 1)$. $(q_1 = q_2)$ and $(q_1, q_2) = (1,0)/(0,1)$ are the least and the most informative structures, respectively. This suggests that some A players optimally obscure their zero trustworthiness through the least informative structure and other A players optimally signal their full trustworthiness through the most informative structure, which rationalizes Patterns 1 and 2.

Trust based on lenient or stringent conditions. L_1/L_2 selfish B players are more likely

to be persuaded to invest than $L_{k\geq 3}$ selfish B players. The former can be persuaded to invest when observing a favorable signal and an information structure with a finite likelihood ratio. By contrast, the latter invest only when observing a favorable signal and an information structure with an infinite likelihood ratio. That is, the latter invest only when the observed signal and information structure perfectly reveal that the underlying state is state 1. In other words, type $(S, L_1/L_2, \pi)$ behaves like an unsophisticated player who treats the information structure as exogenously given and consequently fails to understand its association with player A's choice of p. Type $(S, L_{k\geq 3}, \pi)$ behaves like a sophisticated player who anticipates the association between player A's choice of p and choice of information structure. This result explains Pattern 4 about the distribution of player B's strategy in Game 2.

Association of player B's strategies across games. We categorize player B's types into two groups depending on whether the type invests in Game 1. Selfish players with viewpoint $\pi > \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$ and prosocial players belong to the first group, who invest in Game 1, and selfish players with viewpoint $\pi \leq \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$ belong to the second group, who do not invest in Game 1. Note that in Game 2, type $(S, L_1/L_2, \pi > \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1})$ is more likely to invest than type $(S, L_1/L_2, \pi \leq \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1})$, and the likelihood that type $(S, L_{k\geq 3}, \pi)$ invests is a constant value irrespective of π, ceteris paribus. This implies that the first group also invest more frequently in Game 2 than the second group under mild conditions,[36] which matches Pattern 5.

[36]For example, one sufficient condition is that the proportion of type $(S, L_{k\geq 3}, \pi > \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1})$ of those with $\pi > \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$ is not greatly larger than the proportion of type $(S, L_{k\geq 3}, \pi \leq \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1})$ of those with $\pi \leq \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$.

3.5.2 Estimation of behavioral types

To further understand subjects' behavior in a disaggregated form, we classify subjects' types based on complying exactly with the behavioral model prediction (Exact type). That is, a subject is assigned a certain Exact type if his/her action in each of ten rounds complies *exactly* with the prediction and is otherwise treated as unclassified. We also classify subjects' types based on the maximum likelihood estimation (MLE type).[37] Our MLE type estimation procedure is similar to that of Crawford & Iriberri (2007*b*), and the details are provided in Appendix C.3.4. Table 3.5 summarizes subjects' type distributions based on the criteria of Exact type, MLE type, a refinement of MLE type, which treats subjects for whom the explanation of random choice can not be rejected at the five percent level as unclassified, and a further refinement, which additionally treats subjects with assignment of multiple types as unclassified.

One may notice a sharp increase in the number of classified subjects under the MLE classification compared to under the Exact classification. It is due to the fact that the Exact classification is very stringent while MLE allows inconsiderable deviations. For example, while choice variables can take values from intervals, player A is assigned Exact type $(S, L_2/L_3)$ only when he chooses $p = 0$ and $(p = 0, q_1 = q_2)$ in each of five rounds of Game 1 and Game 2, respectively. Similarly, player A is assigned Exact type $(S, L_{k\geq 4})$ only when he chooses $p = 0$ and $(p = 1, |q_1 - q_2| = 1)$ in each of five rounds of Game 1 and Game 2, respectively. In contrast, the MLE classification allows errors in action choices. The assignment of MLE type is reasonable as long as the choices of subjects who are assigned a MLE type are close to the optimal choices of the corresponding Exact

[37] We omit type L_0 in these classifications because L_0 players are assumed to exist only in the mind of L_1 players. Additionally, type L_0 is generally not separated from type L_1 based on the Exact type classification. Moreover, given that type L_0 chooses an action with certainty in our specification, the likelihood of this type's choosing any action is difficult to specify when we apply the MLE method. In fact, it is typical in the literature that type L_0 is econometricially classified only if type L_0 is assumed to make a random action, when its likelihood function can be appropriately determined.

type. We find that the average and median choices of subjects with a MLE type are indeed in line with the predicted choices of the corresponding Exact type. For example, for those subjects who are assigned type $(S, L_2/L_3)$ under MLE, the averages (medians) of p in Game 1 and Game 2, and $|q_1 - q_2|$ are 0.15 (0.07), 0.16 (0.14) and 0.09 (0.07), respectively; these numbers are 0.14 (0), 0.67 (0.7) and 0.68 (0.69) for those subjects who are assigned type $(S, L_{k\geq 4})$.

Table 3.5: Summary of Subjects' Type Distributions

Role	Type	Exact	MLE	MLE, excluding random	MLE, excluding random and multiple types
A	(P, L_1)	2	10	4	4
	$(P, L_2/L_3)$	2	14	10	10
	$(P, L_{k\geq 4})$		12	8	8
	(S, L_1)	15	25	22	22
	$(S, L_2/L_3)$	3	51	47	47
	$(S, L_{k\geq 4})$	13	55	53	53
	Unclassified	130	0	16	16
B	$(P, L_1/L_2, \pi \leq \frac{2}{7})$	4	11	8	8
	$(P, L_1/L_2, \pi > \frac{2}{7})$		0	0	0
	$(P, L_{k\geq 3}, \pi \leq \frac{2}{7})$		5	0	0
	$(P, L_{k\geq 3}, \pi > \frac{2}{7})$		0	0	0
	$(S, L_1/L_2, \pi \leq \frac{2}{7})$	47	98	83	50
	$(S, L_1/L_2, \pi > \frac{2}{7})$	7	18	11	11
	$(S, L_{k\geq 3}, \pi \leq \frac{2}{7})$	33	49	43	10
	$(S, L_{k\geq 3}, \pi > \frac{2}{7})$	3	12	8	8
	Unclassified	103	0	40	73

Notes: Columns 3-6 report, respectively, the number of subjects for each Exact type, MLE type, and refined MLE type by treating those subjects for whom the explanation of random choice can not be rejected at the five percent level or/and those subjects being assigned multiple types as unclassified. For prosocial players, more types are differentiated through the MLE than through the Exact classification because some types share the same optimal strategy but have different expected payoff functions. In columns 2-5, the total number of subjects for each role over types may exceed 160 because some subjects can be assigned multiple types. A comparison between columns 5 and 6 shows that when we focus on those subjects who are assigned types according to "MLE, excluding random", each player A is assigned a unique type, and 33 B players can be classified as both types $(S, L_1/L_2, \pi \leq \frac{2}{7})$ and $(S, L_{k\geq 3}, \pi \leq \frac{2}{7})$ due to a lack of sufficient variation of their information sets.

Below, we show from a few perspectives that the MLE results lend empirical support to the behavioral model.

First, we perform likelihood ratio tests at the subject level on the null hypothesis of a random choice model. The null hypothesis is rejected at the significance level of five percent for 90% of A players and for 75% of B players, which suggests that for most subjects, the proposed model is favored against a random choice model.

Second, the proportions of prosocial players and less sophisticated players are considerable.[38] Based on the MLE type classification, the proportion of A players (B players) who are assigned to be prosocial is 20.63% (10%), and the proportion of A players (B players) who are less sophisticated is 60.63% (79.38%).

Third, 98 out of 160 A players never choose ($p = 1, |q_1 - q_2| = 1$) even though this choice can give them a payoff of 15 for sure. The MLE type estimates reveal that about half of them are likely to do so intentionally: 50 out of the 98 subjects are assigned type $(S, L_2/L_3)$, who optimally chooses zero trustworthiness with the least informative structure (i.e., $p = 0$ and $|q_1 - q_2| = 0$)) according to the behavioral model; the averages (medians) of their choices of p in Game 1 and Game 2 and $|q_1 - q_2|$ are 0.15 (0.09), 0.16 (0.14) and 0.09 (0.07), respectively.

Fourth, recall that two model predictions rely on mild assumptions about the size of a certain type's proportion. The MLE type estimates indeed validate these assumptions. Specifically, the treatment effect of increasing trusting acts is partly attributed to a small proportion of selfish B players with $\pi > \frac{2}{7}$ out of all selfish B players. According to the MLE type estimates, selfish B players with $\pi > \frac{2}{7}$ constitute only one-fifth of selfish B players. In addition, to predict that the group who invest in Game 1 invest more frequently in Game 2 than the group who do not invest in Game 1, a mild sufficient condition is that the proportion of type $(S, L_{k\geq 3}, \pi > \frac{2}{7})$ of those with $\pi > \frac{2}{7}$ is not greatly larger than the proportion of type $(S, L_{k\geq 3}, \pi \leq \frac{2}{7})$ of those with $\pi \leq \frac{2}{7}$. We find that the proportions are 40% and 37.69%, respectively, which verifies the sufficient

[38] We call $L_{k<4}$ A players and $L_{k<3}$ B players less sophisticated players.

condition.

3.6 Discussion

Alternative experimental design

We discuss an alternative experimental design, in which feedback is provided at the end of each round. The merit of such design is that subjects can learn from past experience and may converge to a stable strategy.[39] However, the design has its weakness: subjects' learning to play the game and learning about opponents' trustworthiness or trust occur simultaneously. The effects of the latter type of learning will entangle with the effects of introducing information design, making it hard to identify which force is driving subjects' behavior. For these reasons, we take a compromised approach: subjects receive feedback in practice rounds, in which they play against a computer opponent who always makes random choices, but get no feedback in paid rounds.

The role of risk preference

Subjects are implicitly assumed to be risk neutral in our equilibrium model and behavioral model. While it is a typical assumption in the literature, we discuss how subjects' risk preferences might affect the interpretation of the experimental findings. First, the treatment effects on trustworthiness and trusting acts are not affected because we employ a within-subjects design and these effects largely hold at the subject level. Second, although we can not completely rule out the effect of risk preference, there is suggestive evidence that our interpretation of behavioral patterns is at least not largely confounded by it. Subjects' answers to two post-experiment survey questions reveal

[39]Note that even in the design with feedback, players' behavior may still substantially deviate from the standard equilibrium prediction. For instance, in the experiment of Fréchette et al. (2022), which features information design, each subject played 25 rounds with complete feedback at the end of each round. Their analysis of the last ten rounds still finds substantial deviations from the equilibrium predictions.

their risk premiums.[40] Focusing on subjects who are assigned types according to "MLE, excluding random and multiple types" in Table 3.5, we find: (1) for A players, the risk premiums of any two different types are not significantly different at the 10% level (t-test with Benjamini-Hochberg multiple testing correction); (2) for B players, except that type $(S, L_1/L_2, \pi > \frac{2}{7})$ has a significantly lower risk premium than type $(S, L_1/L_2, \pi \leq \frac{2}{7})$ or type $(S, L_{k\geq 3}, \pi \leq \frac{2}{7})$, the risk premiums of any other two types are not significantly different at the 10% level.[41]

Relationship with classic trust games

Game 1 resembles classic trust games except for the flipped decision order. The motive of reciprocating trust is arguably less likely to be triggered in Game 1 than in classic trust games because, in the former setting, the trustee does not observe the trusting act when making decisions. The level of trustworthiness and the frequency of trust in Game 1 are 0.23 and 0.34, respectively. We compare them with the results of two previous trust game experiments. Bohnet et al. (2008) study a trust game with the same payoff structure as Game 1. In their game, an investor chooses her minimum acceptable probability of pairing with a trustworthy trustee that makes her willing to invest, and a trustee determines final payoffs of either $(15, 15)$ or $(8, 22)$ if being invested (strategy method). The proportion of trustee subjects who choose $(15, 15)$ is 0.29 in their sessions run in China. In Bracht & Feltovich (2009), given a payoff allocation of $(2, 0)$ for not investing, a trustor first decides whether to invest, and conditional on investing, a trustee decides a binary allocation between $(4, 4)$ and $(0, 8)$. They find that the frequency

[40] The two questions are: (i) Suppose Plan I pays you 22 with X% chance and 10 with $(100 - X)$% chance, and Plan II pays you 15, at least how large X should be so that you would be willing to choose Plan I? (ii) Suppose Plan I pays you 15 with Y% chance and 8 with $(100 - Y)$% chance, and Plan II pays you 10, at least how large Y should be so that you would be willing to choose Plan I? The two questions resemble player A and B's decision problems in the experiment and we compute player A and B's risk premium based on the player's answer to question (i) and (ii), respectively.

[41] Given that subjects with type $(S, L_1/L_2, \pi > \frac{2}{7})$ seem to be less risk-averse than subjects with type $(S, L_1/L_2, \pi \leq \frac{2}{7})$ or type $(S, L_{k\geq 3}, \pi \leq \frac{2}{7})$, risk preference is also a possible explanation of Pattern 5.

of investing is 0.567, and the conditional frequency of choosing $(4, 4)$ is 0.376 in the first five rounds. The comparison shows that trustworthiness and trusting acts in Game 1 are relatively lower but still remarkable, suggesting that the flipped decision order does not substantially change subjects' perception of Game 1 as a trust game.

Direct and indirect effect of information design on trust

The change in trusting acts between Game 1 and Game 2 reflects the total effect of information design on trust. A natural next step is to assess potential causal mechanisms: the direct effect and the indirect effect (Robins & Greenland 1992; Pearl 2001). In our context, the direct effect is represented by the change in trusting acts caused only by the change in treatment status when fixing the level of trustworthiness in either treatment status, and the indirect effect is represented by the change in trusting acts caused only by the change of trustworthiness when keeping the same treatment status. We conduct the causal mediation analysis following Tingley et al. (2014) and find that the proportions of the direct and indirect effect of information design are about 60% and 40%, respectively (see details of the analysis in Appendix C.3.5).

3.7 Conclusion

This article shows that information design can be used as a signaling device to improve trustworthiness and trusting acts and, consequently, increase social welfare. More precisely, the presence of information design, together with its bringing about preliminarity of the trustee's choice and an increase in the trustee's power to commit to a certain level of trustworthiness, fosters trustworthiness and trusting acts. We create a laboratory environment that induces exogenous variation in trustees' capability to design information. We find that trustworthiness and trusting acts substantially increase from a control condition of no capability to a treatment condition of full capability. We

also observe that trustees associate the choice of information structure with the choice of trustworthiness, but some trustors fail to understand the association. To understand the underlying mechanism, we explore an equilibrium model and an alternative model that allows for heterogeneity in prosociality and strategic sophistication. The alternative model provides a better explanation for the findings; especially, it rationalizes some trustees' choice of zero trustworthiness with the least informative structure.

Our study points to several promising topics for future research. First, it proposes an innovative perspective of letting trustees take the lead in the strategic interaction. By following this perspective, one can investigate other approaches that trustees may employ to win trust. Second, no capability and full capability are two boundary cases of trustees' capability to design information. One can naturally imagine the cases of intermediate capability in the field. Varying levels of limited capability and exploring how trustees use limited capability is an important task for future research. Third, limited capability in some field settings may be trustees' endogenous choice. This endogeneity is related to, but conceptually different from, the endogeneity of how to use a certain capability. A comprehensive analysis of such settings will involve explicit modeling of costs and benefits in both dimensions.

Appendix A

Appendix for "Choosing Between Information Bundles"

A.1 Additional Analysis

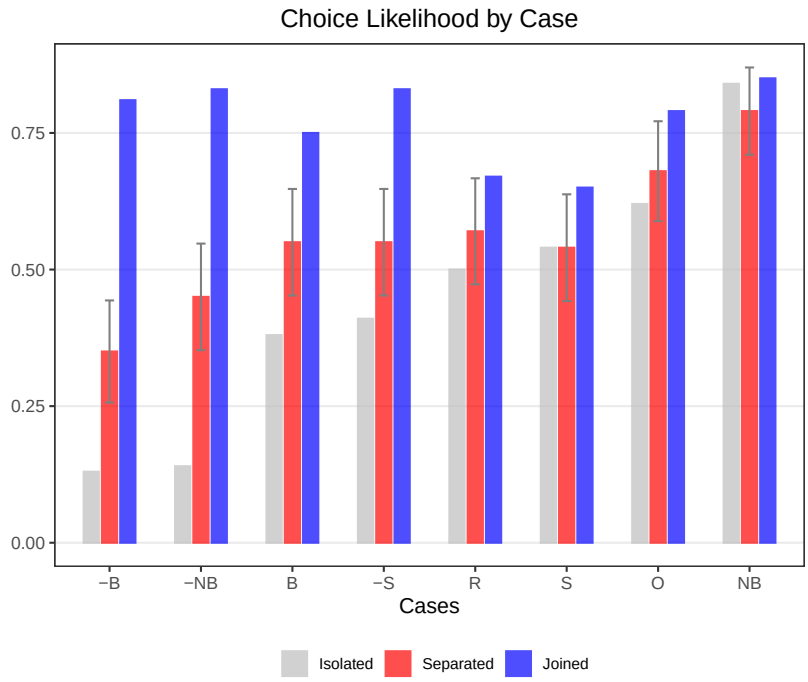


Figure A.1: Likelihood of Choosing High-Value Information – All Three Settings *Notes: The figure plots the likelihood of choosing σ over σ' in the Isolated setting, the likelihood of choosing $\{\sigma_0, \sigma\}$ over $\{\sigma_0, \sigma'\}$ in the Separated setting, and the likelihood of choosing $\sigma_0 \vee \sigma$ over $\sigma_0 \vee \sigma'$ 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 1.2.3. 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.*

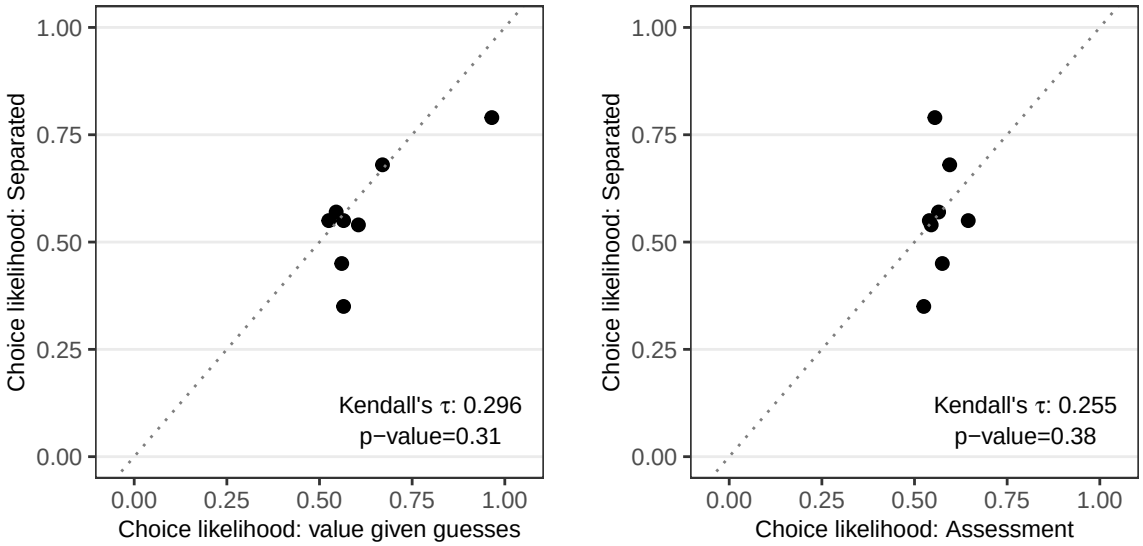


Figure A.2: Value given Guesses or Assessments versus Choices between Information Bundles

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).

Table A.1: Optimality of Decision Making – By Group

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

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

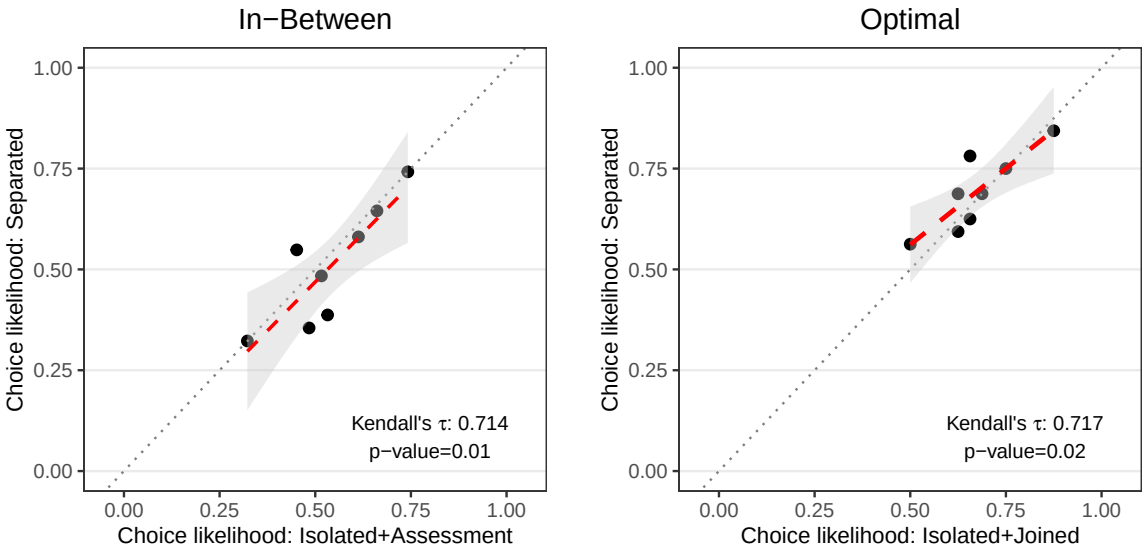
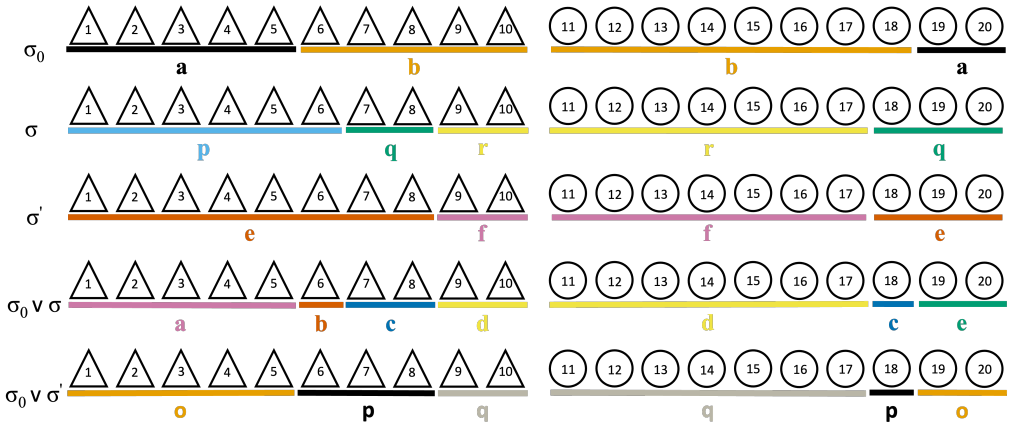
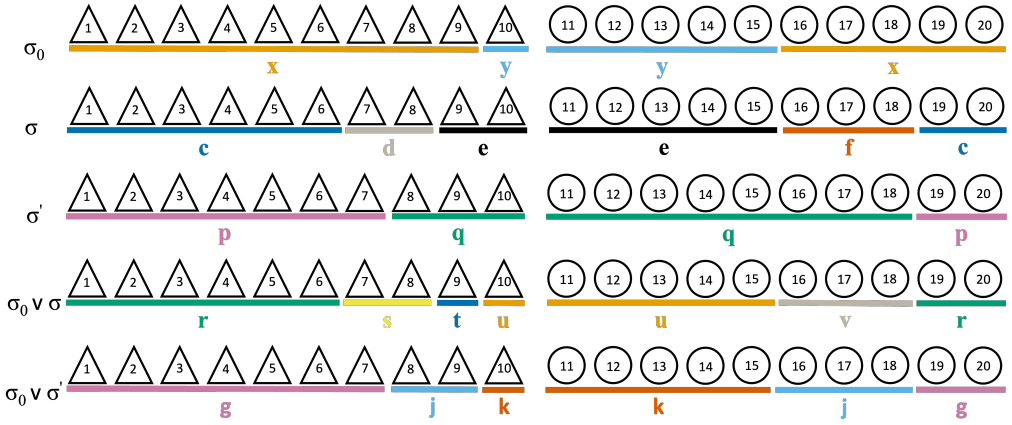


Figure A.3: Combining Mechanisms of Information Bundle Choices *Notes: The y-axis plots the likelihood of choosing bundle $\{\sigma_0, \sigma\}$ in the binary choices under the Separated setting. I define a subject's "Isolated+Assessment" ("Isolated+Joined") choices as either her choices under the Isolated 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 $y = x$.*

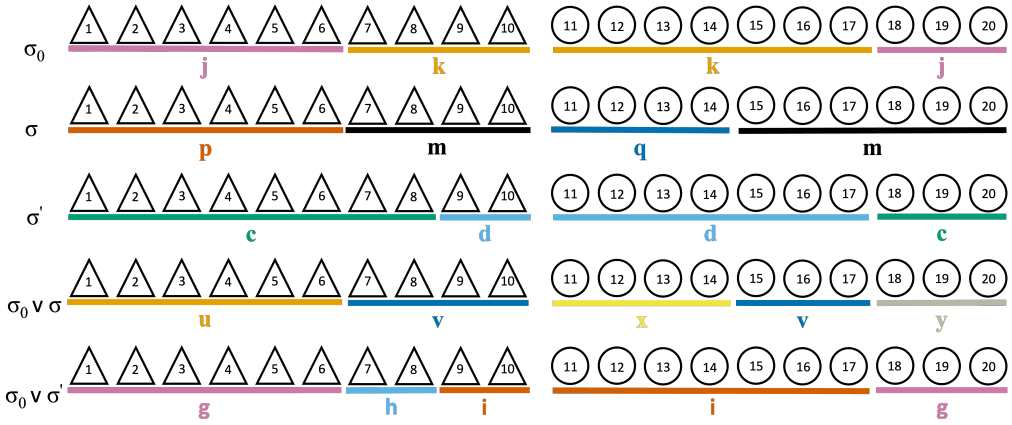
A.2 Information Bundles and Sources



Case (1): Refine, $\sigma R \sigma'$

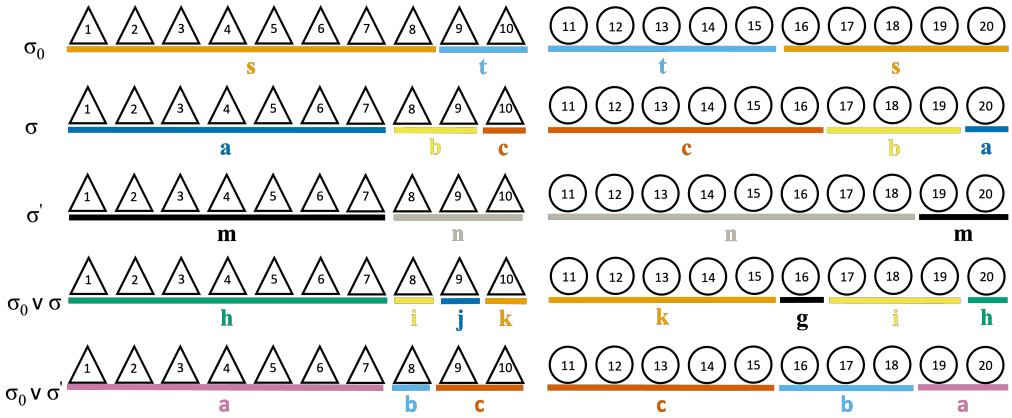


Case (2): Reveal-or-Refine, $\sigma O \sigma'$

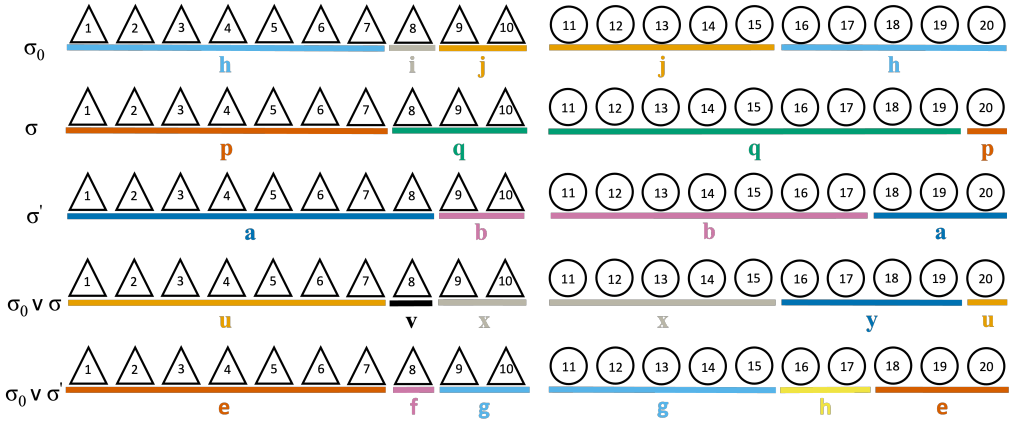


Case (3): Sufficiency, $\sigma S \sigma'$

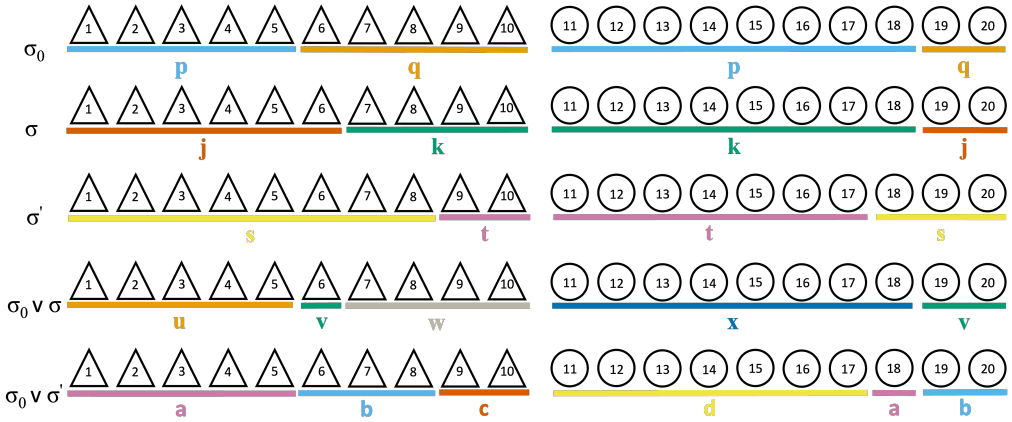
Figure A.4: Studied Information Bundles and Sources *Notes: To be continue on next pages.*



Case (4): Blackwell, $\sigma B \sigma'$

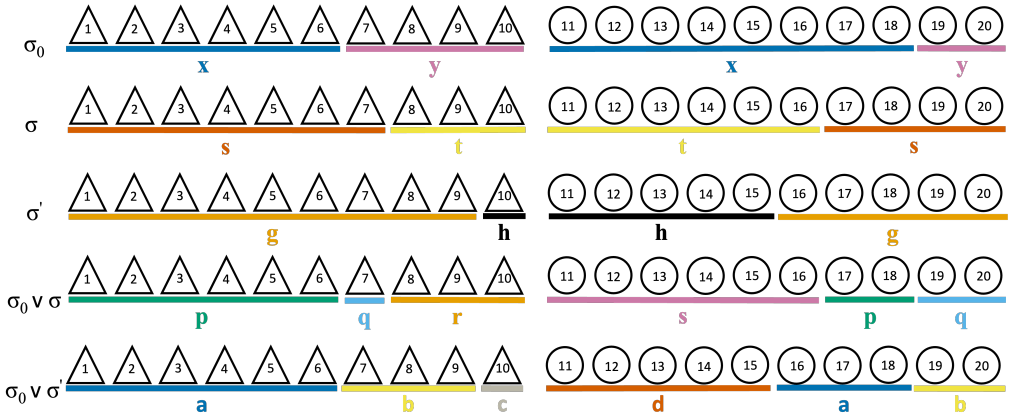


Case (5): Not Blackwell, $\sigma > \sigma'$

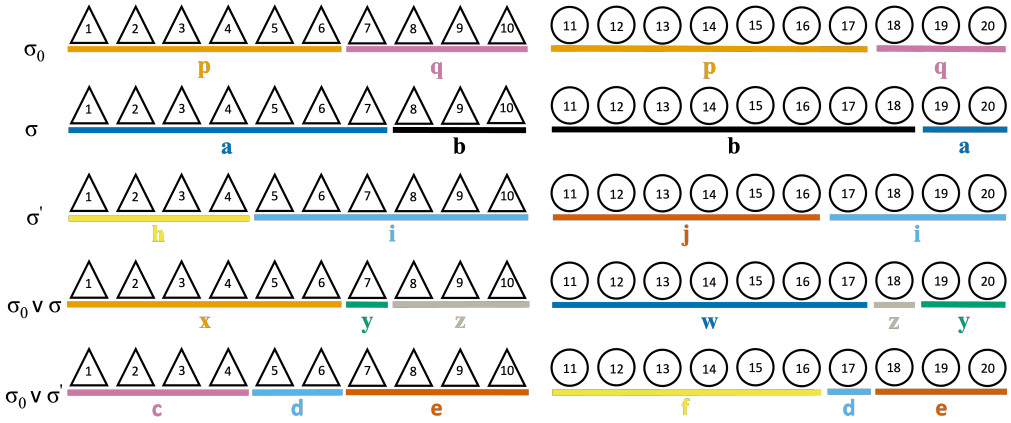


Case (6): Not Blackwell, $\sigma < \sigma'$

Figure A.4-2: Studied Information Bundles and Sources



Case (7): - Blackwell, $\sigma' B \sigma$



Case (8): - Sufficiency, $\sigma' S \sigma$

Figure A.4-3: Studied Information Bundles and Sources

Appendix B

Appendix for "Too Much Information"

B.1 Optimal WTP for an Information Structure

The goal of this section is to study how WTP for an information structure changes with risk preferences. Given the CRRA (constant relative risk aversion) utility function

$$u(x) = \begin{cases} \dfrac{x^{1-r}-1}{1-r}, & \text{if } r \neq 1 \\ log(x), & \text{if } r = 1 \end{cases}$$

in which r is the coefficient of relative risk aversion, we consider a relatively wide range of $r \in [-1, 3]$ and compute the optimal WTPs for information structures with value V (as defined in equation 2.2).[1,2]

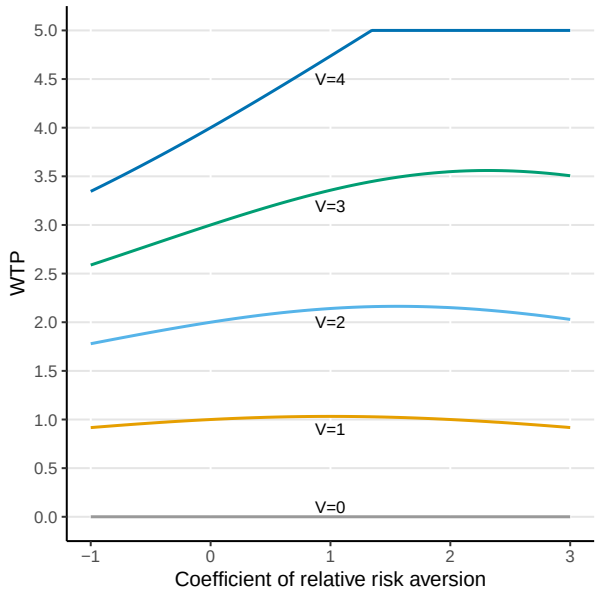


Figure B.1: Optimal WTP for Information Structures under CRRA Utility *Notes: The optimal WTP of a risk neutral agent is V, that is, the WTP at the risk aversion coefficient being 0 in the figure. Given the budget of $5, theoretical WTPs (for the information structure with V = 4) that are higher than 5 are recorded as 5.*

[1]Gandelman & Hernández-Murillo (2015) estimate the coefficient of relative risk aversion of 75 countries and find the values are between 0 to 3. We add the range [-1,0] so to take into account risk loving behavior as well.

[2]The optimal WTP is solved by finding out y that makes $pu(22) + (1-p)u(12) = p^s u(22-y) + (1-p^s)u(12-y)$, where p and p^s are guessing accuracies without or with information, respectively. Note that a subject who guesses correctly (incorrectly) and pays \$y for an information structure receives $\$22 - y$ ($\$12 - y$) at the end of the experiment.

B.2 Varying the order of Guess versus Elicitation of Demand for Information

B.2.1 Baseline Treatment

Table B.1: Determinants of Demand for Information (Elicitation of Demand First)

	Ranking (Logit)			**Difference in WTP (OLS)**			
	(1)	(2)	(3)	(1)	(2)	(3)	(4)
Difference in Value	0.880***		1.678***	0.565***		0.975***	1.289***
	(0.075)		(0.183)	(0.056)		(0.119)	(0.110)
Difference in Informativeness		2.710***	-3.395***		1.954***	-1.766***	-2.400***
		(0.219)	(0.587)		(0.204)	(0.386)	(0.390)
Difference in Disorder	-0.304***	-0.155*	-0.658***	0.057	0.182**	-0.124*	-0.242***
	(0.087)	(0.081)	(0.105)	(0.066)	(0.071)	(0.075)	(0.077)
Clusters	109	109	109	109	109	109	108

*Notes: Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. OLS regression (4) includes only pairwise comparisons in which WTP data is ordinally consistent with Ranking data. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

B.2.2 Reverse Treatment

Table B.2: Determinants of Demand for Information (Elicitation of Guesses First)

	Ranking (Logit)			**Difference in WTP (OLS)**			
	(1)	(2)	(3)	(1)	(2)	(3)	(4)
Difference in Value	0.658***		1.486***	0.448***		0.823***	1.169***
	(0.111)		(0.262)	(0.082)		(0.204)	(0.210)
Difference in Informativeness		1.981***	-3.528***		1.525***	-1.615**	-2.378***
		(0.358)	(0.881)		(0.287)	(0.663)	(0.691)
Difference in Disorder	-0.340***	-0.245**	-0.711***	0.002	0.095	-0.164	-0.315**
	(0.129)	(0.123)	(0.163)	(0.095)	(0.096)	(0.120)	(0.133)
Clusters	54	54	54	54	54	54	54

*Notes: Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. OLS regression (4) includes only pairwise comparisons in which WTP data is ordinally consistent with Ranking data. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

B.3 The No Uncertainty Treatment

Table B.3: Determinants of Demand for Information (No Uncertainty)

	Ranking (Logit)			**Difference in WTP (OLS)**			
	(1)	(2)	(3)	(1)	(2)	(3)	(4)
Difference in Value	0.735***		1.143***	0.387***		0.362***	0.704***
	(0.109)		(0.206)	(0.069)		(0.134)	(0.134)
Difference in Informativeness		2.478***	-1.759**		1.489***	0.106	-0.760*
		(0.381)	(0.719)		(0.251)	(0.399)	(0.422)
Difference in Disorder	-0.541***	-0.368***	-0.725***	-0.033	0.092	-0.022	-0.244**
	(0.094)	(0.088)	(0.111)	(0.085)	(0.091)	(0.107)	(0.100)
Clusters	61	61	61	61	61	61	61

*Notes: Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. OLS regression (4) includes only pairwise comparisons in which WTP data is ordinally consistent with Ranking data. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

B.4 Further Analysis on Clusters

B.4.1 Comparison of Clusters

Table B.4: Optimality of Guesses and Demand by Cluster

	Share	Opt. of Guesses	Opt. of Demand (Rank)	Opt. of Demand (WTP)
Cluster 1	45	99	85	63
Cluster 2	25	99	89	61
Cluster 3	21	98	64	48
Cluster 4	9	94	36	42

Notes: Numbers denote percentages. Here we focus on cases with strict value difference. Demand is coded as optimal given WTP data if subjects pay more (by more than 10 cents) for the more optimal information structure.

B.4.2 Analysis on the Largest Three Clusters

Table B.5: Determinants of Demand for Information (Cluster 1)

	Ranking (Logit)			**Difference in WTP (OLS)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Difference in Value	1.292***		3.744***	0.686***		1.425***
	(0.066)		(0.164)	(0.065)		(0.162)
Difference in Informativeness		3.099***	-9.855***		2.252***	-3.189***
		(0.182)	(0.638)		(0.229)	(0.523)
Difference in Disorder	-0.266***	-0.194**	-1.328***	0.053	0.173**	-0.275***
	(0.096)	(0.082)	(0.132)	(0.077)	(0.084)	(0.097)
Clusters	74	74	74	74	74	74

*Notes: Only Cluster 1 (45% of the subjects) are included. Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

Table B.6: Determinants of Demand for Information (Cluster 2)

	Ranking (Logit)			**Difference in WTP (OLS)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Difference in Value	1.454***		0.809***	0.514***		0.548***
	(0.136)		(0.182)	(0.093)		(0.115)
Difference in Informativeness		5.855***	2.942***		1.945***	-0.148
		(0.538)	(0.755)		(0.372)	(0.439)
Difference in Disorder	-1.123***	-0.606***	-0.837***	-0.090	0.067	-0.105
	(0.158)	(0.165)	(0.171)	(0.118)	(0.118)	(0.115)
Clusters	40	40	40	40	40	40

*Notes: Only Cluster 2 (25% of the subjects) are included. Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

Table B.7: Determinants of Demand for Information (Cluster 3)

	Ranking (Logit)			**Difference in WTP (OLS)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Difference in Value	0.421***		1.085***	0.309***		0.617**
	(0.059)		(0.170)	(0.090)		(0.237)
Difference in Informativeness		1.237***	-2.843***		1.025***	-1.332*
		(0.213)	(0.651)		(0.311)	(0.786)
Difference in Disorder	-0.081	-0.039	-0.374**	0.041	0.098	-0.096
	(0.156)	(0.161)	(0.167)	(0.119)	(0.124)	(0.136)
Clusters	35	35	35	35	35	35

*Notes: Only Cluster 3 (21% of the subjects) are included. Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

B.4.3 Analysis on the Smallest Cluster

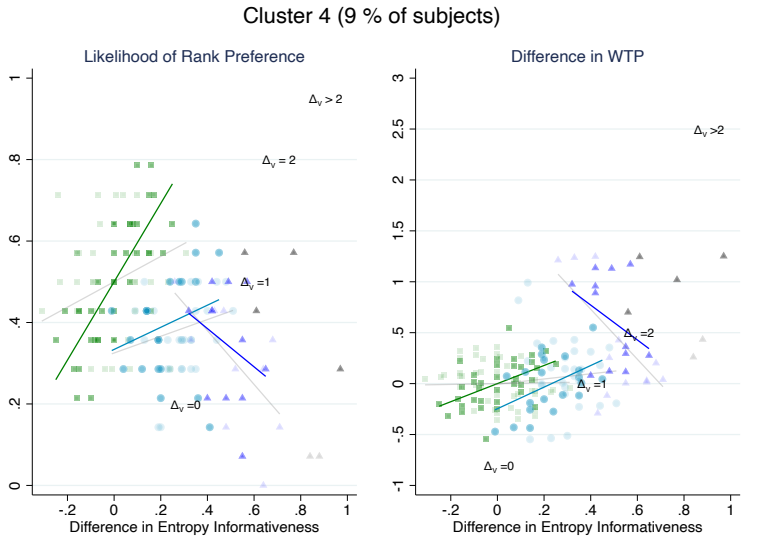


Figure B.3: Preference for Information by Value and Informativeness (Cluster 4) *Notes: Each dot denotes a pair of information structures. Green squares denote pairs where both information structures are of the same value. Blue dots denote pairs where the value difference between the first and second information structure is 1. The darker blue (gray) triangles denote pairs where the value difference is equal to (larger than) 2. To account for the potential impact of visual complexity, pairs with at least one information structure where the blue and red balls are not displayed in order are depicted in a lighter color. Gray lines depict the best linear fits for each of the first three categories. Darker lines in the corresponding colors denote the best linear fits where the pairs depicted in a lighter color are not included.*

Table B.8: Determinants of Demand for Information (Cluster 4)

	Ranking (Logit)			**Difference in WTP (OLS)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Difference in Value	-0.309**		-0.633	0.264		0.117
	(0.146)		(0.459)	(0.180)		(0.339)
Difference in Informativeness		-1.005*	1.397		1.080	0.632
		(0.528)	(1.725)		(0.642)	(0.919)
Difference in Disorder	0.349	0.292	0.493	0.326*	0.428**	0.391*
	(0.295)	(0.322)	(0.372)	(0.180)	(0.194)	(0.215)
Clusters	14	14	14	14	14	14

*Notes: Only Cluster 4 (9% of the subjects) are included. Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

B.4.4 Alternative Clustering method: K-Modes

Table B.9: Determinants of Demand for Information (By K-Modes Cluster)

	Ranking (Logit)			**Difference in WTP (OLS)**		
	C1	C2	C3	C1	C2	C3
Difference in Value	2.746***	0.956***	2.591***	1.254***	0.567***	1.115***
	(0.217)	(0.169)	(0.310)	(0.157)	(0.112)	(0.305)
Difference in Informativeness	-7.337***	1.670***	-8.109***	-2.655***	-0.212	-2.981***
	(0.633)	(0.564)	(1.071)	(0.469)	(0.401)	(1.035)
Difference in Disorder	-1.394***	-0.664***	-0.385**	-0.288**	0.004	-0.126
	(0.137)	(0.132)	(0.193)	(0.110)	(0.098)	(0.144)
Clusters	64	54	35	64	54	35

*Notes: C1, C2 and C3 refer to Clusters 1, 2, and 3 under the K-Modes clustering and represent 39%, 33%, and 21% of the data, respectively. Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

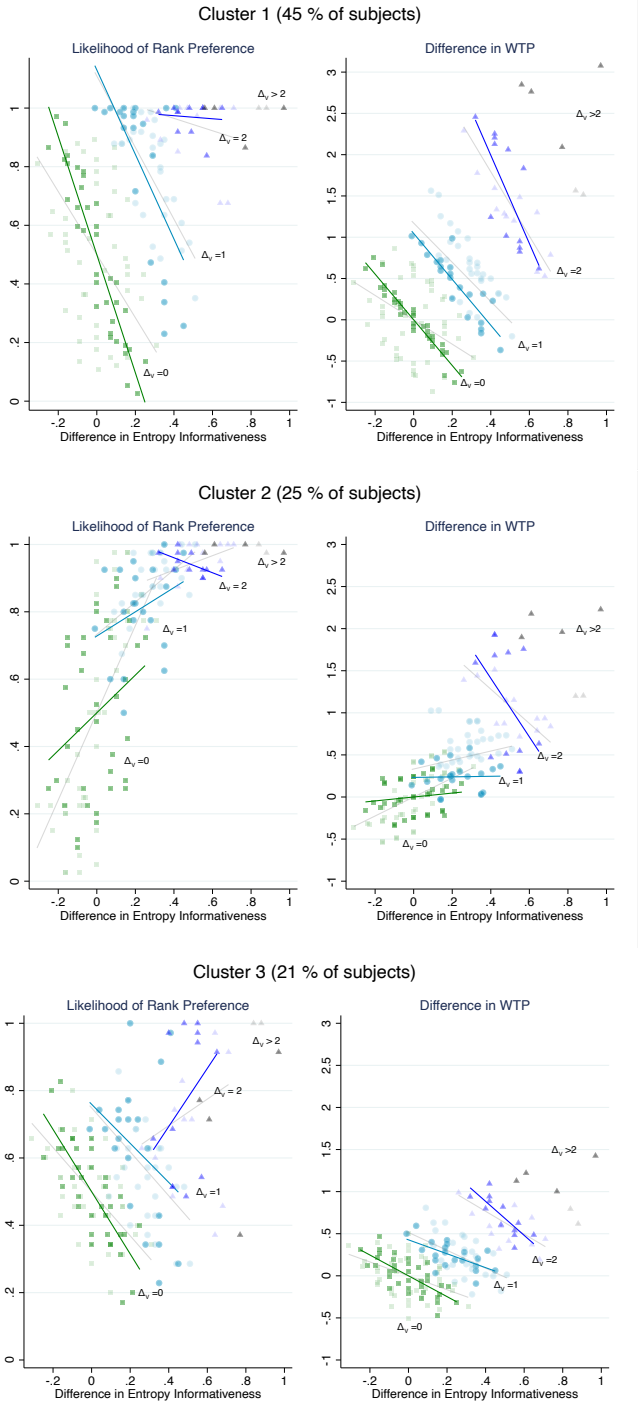


Figure B.2: Preference for Information Structure by Value and Informativeness Separated by Cluster *Notes: Each dot denotes a pair of information structures. Green squares denote pairs where both information structures are of the same value. Blue dots denote pairs where the value difference between the first and second information structure is 1. The darker blue (gray) triangles denote pairs where the value difference is equal to (larger than) 2. To account for the potential impact of visual complexity, pairs with at least one information structure where the blue and red balls are not displayed in order are depicted in a lighter color. Gray lines depict the best linear fits for each of the first three categories. Darker lines in the corresponding colors denote the best linear fits where the pairs depicted in a lighter color are not included.*

B.5 Characteristics of Information Structures

Table B.10: Characteristics of Information Structures

Information	value	informativeness	disorder	varpost	nsignal	ndistinctpost	uncertain	certain	skewness
1	0	0	0	0	1	1	0	0	
2	1	0.09	1	0.03	2	2	0	0	-0.873
3	1	0.13	1	0.04	2	2	0	0	0
4	2	0.26	1	0.082	2	2	0	0	-0.408
5	2	0.32	0	0.09	2	2	0	0.2	-1.5
6	2	0.33	1	0.093	3	3	0	0.2	-1.372
7	2	0.42	0	0.107	2	2	0	0.4	0.408
8	2	0.42	1	0.107	4	3	0	0.4	-0.868
9	2	0.42	0	0.107	3	3	0	0.4	-0.868
10	2	0.42	0	0.107	3	2	0	0.4	0.408
11	2	0.49	0	0.12	3	3	0	0.5	-0.577
12	2	0.57	0	0.14	3	3	0.4	0.6	-0.344
13	3	0.56	0	0.154	2	2	0	0.3	-0.873
14	3	0.61	0	0.16	3	2	0	0.5	0
15	3	0.77	0	0.19	3	3	0.2	0.8	-0.398
16	4	0.97	0	0.24	2	2	0	1	-0.408

Notes: Information denotes the information structures shown in Figure 2.2; value is instrumental value as defined in equation 2.2; informativeness is entropy informativeness as defined in equation 2.1; disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls are presented out of order; varpost denotes the variance of $P(b|s)$, i.e., the variance of Bayesian posterior of the drawn ball being blue given signal s; nsignal denotes the number of distinct signals that an information structure can generate; ndistinctpost denotes the number of distinct posteriors that an information structure can induce; uncertain denotes the probability of generating maximally certain signals (i.e., signals induce posterior of 1 or 0); certain denotes the probability of generating maximally uncertain signals (i.e., signals induce posterior of 0.5); skewness denotes the third normalized moment of Bayesian posterior $P(b|s)$.

Table B.11: Blackwell Ordering

Information	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	0		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
3		0	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
4			0								-1	-1	-1	-1	-1
5				0	-1		-1	-1		-1	-1	-1		-1	-1
6					0		-1	-1		-1	-1	-1		-1	-1
7						0			0		-1		-1	-1	-1
8							0	0		-1	-1			-1	-1
9								0		-1	-1			-1	-1
10									0		-1		-1	-1	-1
11										0	-1			-1	-1
12											0			-1	-1
13												0		-1	-1
14													0	-1	-1
15														0	-1
16															0

Notes: Information denotes the information structures shown in Figure 2.2. The Blackwell ordering takes values 1, 0, -1 when the row information structure is more, equally or less Blackwell informative than the column one. The value is missing if Blackwell comparison cannot be made.

Table B.12: Determinants of Demand for Information (Ranking)

	Ranking (Logit)					
	(1)	(2)	(3)	(4)	(5)	(6)
Difference in Value	1.607***	1.210***	1.532***	1.367***	1.674***	1.107***
	(0.150)	(0.137)	(0.144)	(0.136)	(0.150)	(0.126)
Difference in Informativeness	-3.436***	-1.577***	-3.000***	-2.142***	-3.896***	-1.064**
	(0.487)	(0.478)	(0.451)	(0.441)	(0.473)	(0.474)
Difference in Disorder	-0.677***	-0.589***	-0.615***	-0.488***	-0.674***	-0.422***
	(0.088)	(0.086)	(0.090)	(0.087)	(0.089)	(0.087)
Difference in Uncertain		-0.624***				-0.544***
		(0.096)				(0.101)
Difference in # Signals			-0.129**			0.018
			(0.055)			(0.062)
Difference in # Distinct Posteriors				-0.326***		-0.312***
				(0.066)		(0.075)
Difference in Certain					0.102**	0.098***
					(0.046)	(0.035)
Clusters	163	163	163	163	163	163

*Notes: Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Descriptions of the other characteristic measures can be found in Table B.10. Except for Value and Informativeness, the other differences in characteristic measures are defined as follows: 1 if the first information structure has a higher characteristic measure than the other in a pairwise comparison, -1 if the opposite, and 0 otherwise. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

Table B.13: Determinants of Demand for Information (WTP)

	Difference in WTP (OLS)					
	(1)	(2)	(3)	(4)	(5)	(6)
Difference in Value	0.924***	0.951***	0.987***	1.027***	0.999***	1.067***
	(0.104)	(0.105)	(0.104)	(0.106)	(0.105)	(0.110)
Difference in Informativeness	-1.716***	-1.840***	-2.103***	-2.282***	-2.246***	-2.591***
	(0.338)	(0.355)	(0.337)	(0.353)	(0.346)	(0.402)
Difference in Disorder	-0.137**	-0.143**	-0.197***	-0.229***	-0.132**	-0.216***
	(0.064)	(0.062)	(0.062)	(0.062)	(0.064)	(0.062)
Difference in Uncertain		0.043				0.007
		(0.060)				(0.057)
Difference in # Signals			0.128***			0.046
			(0.045)			(0.039)
Difference in # Distinct Posteriors				0.155***		0.102**
				(0.053)		(0.048)
Difference in Certain					0.128***	0.083***
					(0.027)	(0.013)
Clusters	163	163	163	163	163	163

*Notes: Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Descriptions of the other characteristic measures can be found in Table B.10. Except for Value and Informativeness, the other differences in characteristic measures are defined as follows: 1 if the first information structure has a higher characteristic measure than the other in a pairwise comparison, -1 if the opposite, and 0 otherwise. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

B.6 Additional Plots and Tables

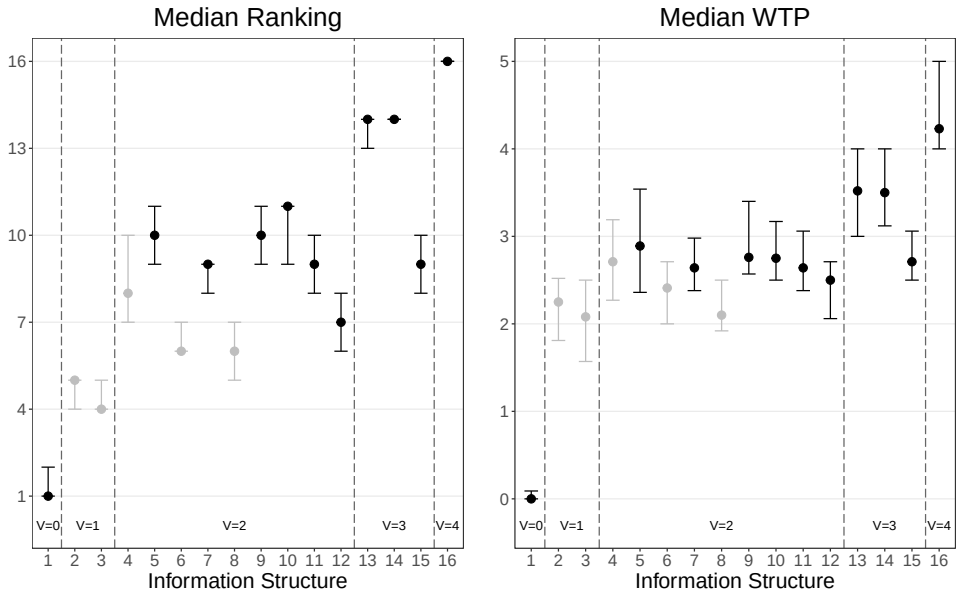


Figure B.4: Median Rank and WTP by Information Structure *Notes: Information structures are the ones as shown in Figure 2.2. V denotes the instrumental value of an information structure as defined in equation 2.2. Vertical lines are 95 percent confidence intervals that are derived from the exact sign test.*

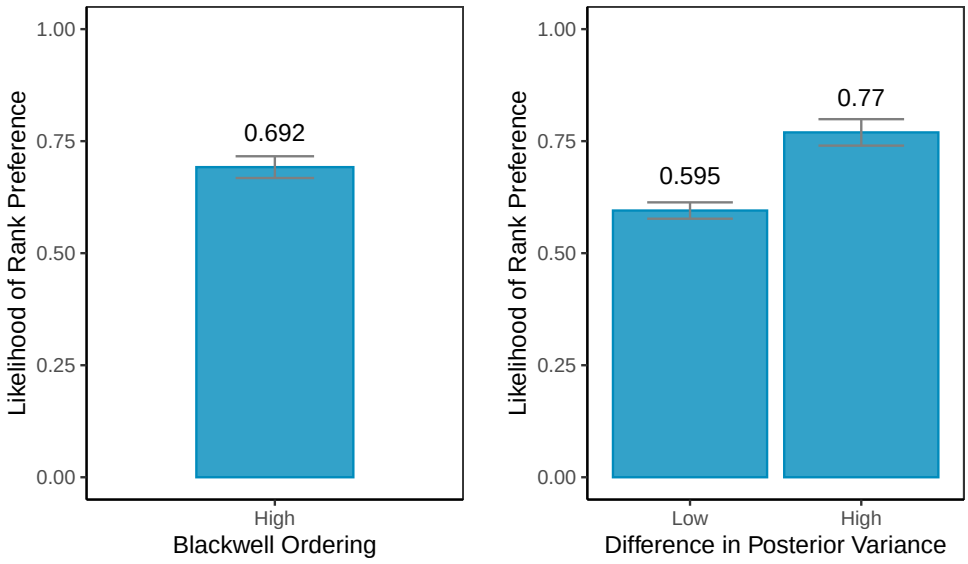


Figure B.5: Demand for Information by Blackwell Ordering and Posterior Variance *Notes: The figures condition on all pairwise comparisons between information structures where there is a strict positive difference—on Blackwell Ordering for (a) or variance of posterior for (b)—between the first and the second structure. The bars depict the likelihood with which the first structure was ranked as more preferred to the second. In (a),* High *represents the first structure is strictly more Blackwell informative. In (b),* Low *(*High*) represents all pairwise comparisons where the difference is weakly lower (strictly higher) than the median difference of variance of posterior (i.e., 0.0635).*

Table B.14: Determinants of Demand for Information

	Ranking (Logit)			**Difference in WTP (OLS)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Difference in Value	0.800***		1.607***	0.527***		0.924***
	(0.063)		(0.150)	(0.046)		(0.104)
Difference in Informativeness		2.453***	-3.436***		1.812***	-1.716***
		(0.193)	(0.487)		(0.167)	(0.338)
Difference in Disorder	-0.317***	-0.186***	-0.677***	0.039	0.153***	-0.137**
	(0.072)	(0.067)	(0.088)	(0.054)	(0.057)	(0.064)
Clusters	163	163	163	163	163	163

*Notes: Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Standard errors (clustered at the subject level) are in parentheses. *** 1%, ** 5%, * 10% significance.*

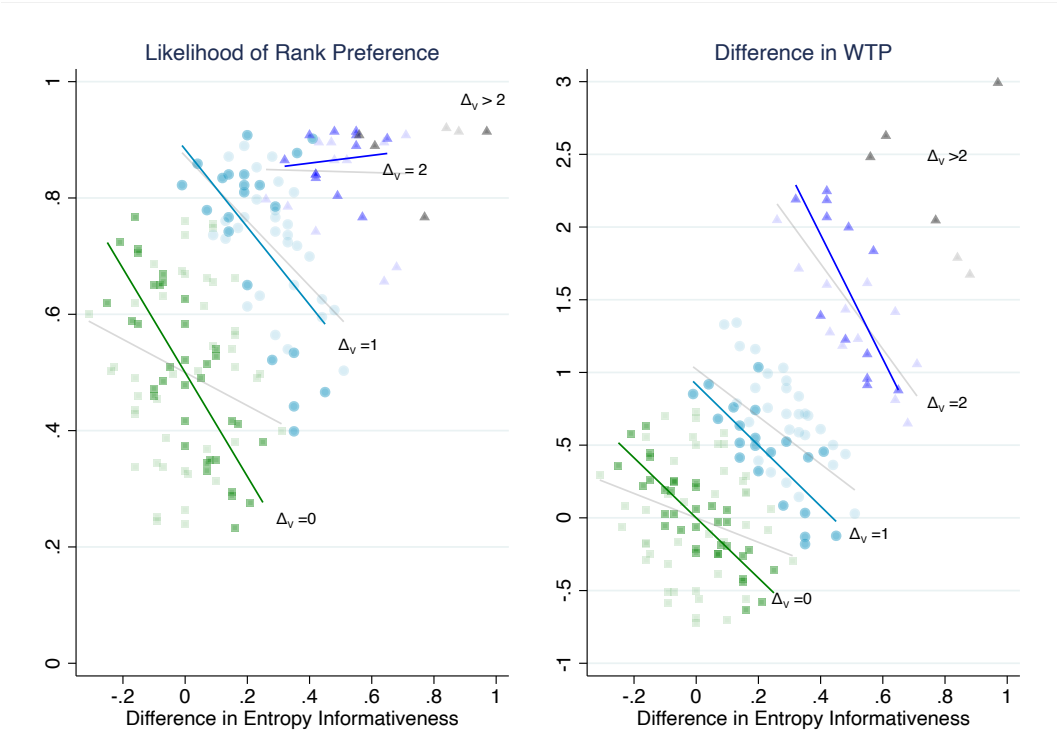


Figure B.6: Preference for Information Structure by Value and Informativeness *Notes: This is the same as Figure 2.8 except that pairwise comparisons with WTPs ranked inconsistently with the Ranking data are dropped in the right panel.*

Table B.15: Determinants of Demand for Information (By Value Difference)

	Ranking (Logit)				**Difference in WTP (OLS)**			
	$\Delta V = 0$	$\Delta V = 1$	$\Delta V = 2$	$\Delta V > 2$	$\Delta V = 0$	$\Delta V = 1$	$\Delta V = 2$	$\Delta V > 2$
Difference in Informativeness	-4.638***	-4.023***	-0.959**	-0.210	-1.845***	-1.543***	-2.541***	0.443**
	(0.573)	(0.536)	(0.468)	(0.395)	(0.400)	(0.363)	(0.391)	(0.200)
Difference in Disorder	-0.876***	-0.494***	-0.234**	-0.534**	-0.368***	-0.169***	0.064	0.961***
	(0.103)	(0.086)	(0.111)	(0.233)	(0.071)	(0.064)	(0.069)	(0.139)
Constant	0.178***	1.876***	2.123***	2.051***	0.048**	0.761***	2.366***	1.755***
	(0.051)	(0.166)	(0.281)	(0.353)	(0.019)	(0.105)	(0.241)	(0.202)
Clusters	163	163	163	163	163	163	163	163

*Notes: ΔV denotes the value (as defined in equation 2.2) difference between two structures in a pairwise comparison. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

Table B.16: Determinants of Demand for Information (Value and Blackwell)

	Ranking (Logit)			**Difference in WTP (OLS)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Difference in Value	0.800***		0.904***	0.527***		0.612***
	(0.063)		(0.080)	(0.046)		(0.055)
Blackwell		0.720***	-0.256***		0.591***	-0.147***
		(0.055)	(0.066)		(0.054)	(0.037)
Difference in Disorder	-0.317***	-0.309***	-0.312***	0.039	0.072	0.079
	(0.072)	(0.065)	(0.072)	(0.054)	(0.058)	(0.058)
Clusters	163	163	163	163	163	163

*Notes: Value denotes instrumental value as defined in equation 2.2. Blackwell denotes whether the first structure is strictly more Blackwell informative than the other in a pairwise comparison. Details can be found in Table B.11. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

Table B.17: Determinants of Demand for Information (Value and Posterior Variance)

	Ranking (Logit)			**Difference in WTP (OLS)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Difference in Value	0.800***		1.954***	0.527***		1.104***
	(0.063)		(0.193)	(0.046)		(0.133)
Difference in Variance of $P(b\|s)$		10.833***	-19.396***		7.763***	-9.827***
		(0.840)	(2.683)		(0.702)	(1.845)
Difference in Disorder	-0.317***	-0.200***	-0.683***	0.039	0.137**	-0.145**
	(0.072)	(0.068)	(0.089)	(0.054)	(0.056)	(0.064)
Clusters	163	163	163	163	163	163

*Notes: Value denotes instrumental value as defined in equation 2.2. Variance of $P(b|s)$ denotes the variance of the Bayesian posterior of the drawn ball being blue given signal s. It is a valid measure of uncertainty reduction (informativeness) according to Frankel & Kamenica (2019). Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

Table B.18: Determinants of Demand for Information (Removing Maximal Uncertainty)

	Ranking (Logit)			**Difference in WTP (OLS)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Difference in Value	0.914***		1.281***	0.550***		0.978***
	(0.078)		(0.148)	(0.048)		(0.107)
Difference in Informativeness		3.488***	-1.636***		2.076***	-1.883***
		(0.293)	(0.498)		(0.188)	(0.360)
Difference in Disorder	-0.459***	-0.242***	-0.589***	0.013	0.131**	-0.139**
	(0.077)	(0.072)	(0.087)	(0.056)	(0.058)	(0.062)
Clusters	163	163	163	163	163	163

*Notes: Value denotes instrumental value as defined in equation 2.2. Informativeness denotes entropy informativeness as defined in equation 2.1. Disorder denotes whether an information structure has a visual disorder, i.e., whether the blue and the red balls were presented out of order. Pairwise comparisons that include either of the two information structures that may generate maximally uncertain signals (i.e., information structures 12 and 15 as shown in Figure 2.2) are not included. Standard errors (clustered at the subject level) in parentheses. *** 1%, ** 5%, * 10% significance.*

Ranking Information Sources

Please drag these information sources on the screen to rank them in order from most favorite (top of the screen) to least favorite (bottom of the screen). The computer will randomly pick two information sources, and you will receive information from the one that you ranked higher before you are asked to make a guess about the color of the randomly selected ball. As in other parts, you will earn a $10 BONUS payment if your guess is correct and $0 if your guess is incorrect.

(To help with your decision, for each Information Source, we include below each Group the guess you made for that Group earlier in a corresponding Guessing Question.)

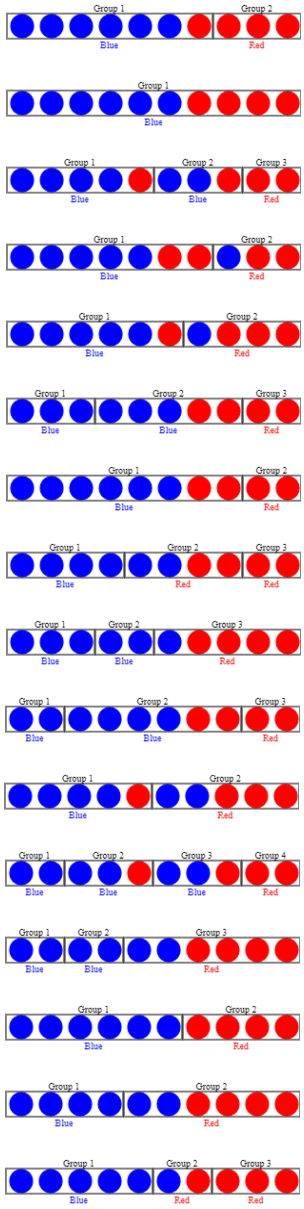


Figure B.7: Ranking with Elicited Guesses

Appendix C

Appendix for "Trustworthy by Design"

This supplement contains four parts. Appendix C.1 provides details about the behavioral model. Appendix C.2 contains proofs. Appendix C.3 reports additional data analysis. Appendix C.4 provides an English translation of the experimental instructions presented to the subjects.

C.1 Details about the behavioral model

C.1.1 Assumptions

Heterogeneity in prosociality and viewpoint of opponents' prosociality

Charness & Rabin (2002) formulate a model of social preferences that embeds difference aversion (Fehr & Schmidt 1999; Bolton & Ockenfels 2000), social-welfare, and reciprocity models. When modeling social preferences in our games, we consider the relevance of each of the above motives and trade off a parsimonious model against a complete characterization of different driving forces. First, unlike the standard trust games, in which the trustee moves after observing the trustor's offer and reciprocity is thus likely motivated, reciprocity is arguably less likely motivated in our games because of the reverse decision order. Second, player A chooses between allocation plans with an identical social surplus, and thus difference aversion should be the most prominent one if any of the above motives is aroused. Third, player B determines whether to invest in a social-welfare increasing project. Formulating a motive of increasing social welfare is useful for understanding player B's potential deviation from the standard framework with the pure selfishness assumption.[1]

Therefore, we formulate social preferences by highlighting player A's difference aver-

[1] A motive of increasing social welfare can lead player B to invest in the project even if she may sacrifice. In contrast, the directional effect of difference aversion on player B's investing decision hinges on the specific values of monetary payoff parameters.

sion and player B's increasing social welfare, respectively, and view this formulation as a simple proxy for comprehensive social preferences. Specifically, we assume that player A's social preference is captured by θ_A that relies on the payoff allocation and player B's social preference is captured by θ_B that relies on the social surplus. We also assume that social preferences are additively separable from the monetary payoff.[2]

We consider two types of each player in terms of social preferences: selfish player A ($A1$), prosocial player A ($A2$), selfish player B ($B1$) and prosocial player B ($B2$). We assume the following properties on θ_A and θ_B: (1) $\theta_{A1} \equiv 0$ and $\theta_{B1} \equiv 0$; (2) $v_0^A + \theta_{A2}(v_0^A, v_0^B) < min\{\rho_1 v + \theta_{A2}(\rho_1 v, v - \rho_1 v), \rho_2 v + \theta_{A2}(\rho_2 v, v - \rho_2 v)\}$ and $\theta_{A2}(\rho_2 v, v - \rho_2 v) - \theta_{A2}(\rho_1 v, v - \rho_1 v) < (\rho_1 - \rho_2)v$;[3] and (3) $v_0^B + \theta_{B2}(v_0^A + v_0^B) < min\{(1 - \rho_1)v + \theta_{B2}(v), (1 - \rho_2)v + \theta_{B2}(v)\}$.[4] Thus, a prosocial player A prefers $(\rho_1 v, v - \rho_1 v)$ to $(\rho_2 v, v - \rho_2 v)$ and a prosocial player B prefers investing to not investing regardless of player A's choice. Without loss of generality, we simplify the formulation by the following normalization: $\theta_{A2}(v_0^A, v_0^B) = \theta_{A2}(\rho_1 v, v - \rho_1 v) = 0$ and $\theta_{B2}(v_0^A + v_0^B) = 0$. We then need only two parameters $\theta_{A2} \equiv \theta_{A2}(\rho_2 v, v - \rho_2 v) < (\rho_1 - \rho_2)v$ and $\theta_{B2} \equiv \theta_{B2}(v) > v_0^B - (1 - \rho_2)v$.

Given the assumption about players' heterogeneity in prosociality, it is natural to expect that they may hold heterogeneous viewpoints of the opponents' prosociality. It is also clear that player B's viewpoint of the opponent's prosociality affects her decision

[2] In Charness & Rabin (2002)'s formulation, a player's preference is a weighted average of his monetary payoff and his opponent's monetary payoff, with the weight depending on whether he has higher or lower monetary payoff than his opponent has. Our formulation is essentially similar to theirs in the sense that our assumed properties on θ_A and θ_B are implications of their specific formulation.

[3] The first part of this property requires that player A's difference aversion does not overturn his preference for the project being invested. Since $(\rho_1 - \rho_2)v < 0$, the second part of this property requires that $\theta_{A2}(\rho_2 v, v - \rho_2 v) < \theta_{A2}(\rho_1 v, v - \rho_1 v)$. Given player A's difference aversion, this is a natural requirement as long as $(\rho_2 v, v - \rho_2 v)$ is a more unequal allocation than $(\rho_1 v, v - \rho_1 v)$, which holds for $\rho_1 \geq \frac{1}{2}$ as an example. This part also requires that the degree of difference aversion outweighs the difference in monetary payoffs, which is demonstrated in existing experimental studies where many subjects sacrifice to choose more equal payoffs.

[4] This property requires that prosocial player B is willing to sacrifice to increase social surplus, which is demonstrated in Charness & Rabin (2002) that: "...in our data that about half of subjects make inequality-increasing sacrifices when these sacrifices are efficient and inexpensive."

substantially. We assume that player B holds different viewpoints on the fraction of prosocial A players in the population. Specifically, player B with viewpoint π believes that the fraction of prosocial A players is π or she is playing with a prosocial player A with a probability of π. We assume that π follows a uniform distribution over $(0, \overline{\pi})$, where $\overline{\pi} \in (1-(\frac{1-\rho_1-\rho_0^B}{\rho_2-\rho_1})^2, 1)$.[5] We could also assume that A players are heterogeneous in their viewpoints of player B's prosociality, but this assumption is not useful for generating new predictions. For simplicity, we assume that all A players think that the fraction of prosocial B players is a constant $\alpha \in (0, \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A})$.[6]

Heterogeneity in strategic sophistication

In contexts involving initial responses, the level-k models pioneered by Nagel (1995) and Stahl & Wilson (1994, 1995) provide a structure for analyzing players' non-equilibrium strategic thinking. We assume that players' heterogeneous strategic thinking follows level-k reasoning. Specifically, a L_k player A believes that: his opponent is a L_{k-1} prosocial or selfish player B with viewpoint π, the chance that he encounters a prosocial player B is α, and the probability density of his encountering a player B with viewpoint $\pi \in (0, \overline{\pi})$ is $\frac{1}{\overline{\pi}}$. For a L_k player B with viewpoint π, she believes that: her opponent is a L_{k-1} prosocial or selfish player A with viewpoint α, and the chance that she encounters with a prosocial player A is π.

We assume that L_0 players make non-strategic decisions. Specifically, for L_0 players in both games, a prosocial (selfish) player A chooses $p = 1$ ($p = 0$) and a prosocial

[5]Note that selfish B players with $\pi > (<)\frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$ invest (do not invest) in Game 1. So the assumption of a range of $(0, \overline{\pi})$ instead of $(0, 1)$ is useful for generating the frequency of investing in Game 1 with greater flexibility. The assumption that $\overline{\pi} > 1-(\frac{1-\rho_1-\rho_0^B}{\rho_2-\rho_1})^2$ is necessary for some A players to optimally choose the least informative structure in Game 2.

[6]The assumption that $\alpha < \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}$ is necessary for player A to be incentivized to choose an information structure strategically. If α is very large, that is, player A believes that almost all B players are prosocial and investing, then he will have no incentive to make decisions strategically.

(selfish) player B invests (does not invest).[7] In Game 2, a L_0 player A is also assumed to choose q_1 and q_2 independently according to a uniform distribution over $[0, 1]$. As in the literature, our assumption about L_0 players does not mean that there exist L_0 players. It serves to anchor the initial belief of strategic reasoning and exists in the mind of L_1 players. Moreover, in Game 2, L_1 player A's optimal response puts no discipline on his choice of q_1 and q_2. So we additionally assume that L_1 player A also chooses q_1 and q_2 independently according to a uniform distribution over $[0, 1]$. The assumption of uniform distribution about L_0/L_1 player A's choice of q_1 and q_2 is for the sake of simplicity. In fact, the model predictions remain the same as long as L_0/L_1 player A chooses q_1 and q_2 independently according to a distribution with a support of $[0, 1]$.

Conditionally pessimistic posterior in zero-probability information sets

We assume that player B forms a posterior belief based on Bayes' rule when applicable. We assume that player B holds a conditionally pessimistic posterior belief in a zero-probability information set in the sense that she uses the "worst" posterior belief among all posteriors that are consistent with that information set.[8] We illustrate the assumption via two examples below. Suppose $(q_1 = 0.8, q_2 = 0.4, s = b)$ is a zero-probability information set for player B. Her conditionally pessimistic posterior belief about state 1 is zero because any posterior belief from $[0, 1]$ is consistent with the information set and a posterior belief of zero is the "worst". Suppose $(q_1 = 0.8, q_2 = 0, s = b)$ is a zero-probability information set for player B. Her conditionally pessimistic posterior

[7] We also investigate two other natural specifications of L_0 players: (1) L_0 selfish player B's action depends on her viewpoint π, that is, invests for $\pi \in (\frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}, \overline{\pi})$ and does not invest for $\pi \in (0, \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}]$, and other L_0 players' actions remain the same as the present specification; (2) player A chooses p, q_1 and q_2 independently according to a uniform distribution over $[0, 1]$, and player B invests or does not invest with an equal probability. The optimal strategies for higher-level players under these alternative specifications are similar to the present ones. Moreover, the optimal strategies converge to the same outcomes under the three specifications.

[8] In our equilibrium analysis, we apply PBE in subgames to specify a belief of p off the equilibrium path. The approach is not applicable for the level-k model.

belief about state 1 is one because only a posterior belief of one is consistent with the information set.

The optimal action based on such a posterior belief is similar to that prescribed by a maximin strategy, which achieves the best outcome in the worst scenario. Our assumption makes only a slight modification: the support of the worst situation is restricted to situations that are consistent with the observed information set. Given that player B has a conflicting interest with her opponent in payoff allocation, it seems plausible for her to choose such a conservative strategy when reacting to information to which she assigns probability zero.

Proposition 3 is based on the assumptions about players' heterogeneity in prosociality, viewpoints about prosociality and strategic sophistication, and the assumption about conditionally pessimistic posterior belief in zero-probability information sets. In addition, a sufficient condition that guarantees the optimality of $(p = 0, q_1 = q_2)$ for type $(S, L_2/L_3)$ is that $\overline{\pi} > \max\{1 - (\frac{1-\rho_1-\rho_0^B}{\rho_2-\rho_1})^2, (1-\alpha)\frac{\rho_2-\rho_0^A}{\rho_2-\rho_1}\frac{\rho_{0B}-(1-\rho_2)}{\rho_2-\rho_1}\}$. His optimal strategy is $p = 1, (q_1, q_2) = (1, 0)$ or $(0, 1)$ when $\overline{\pi} < (1-\alpha)\frac{\rho_2-\rho_0^A}{\rho_2-\rho_1}\frac{\rho_{0B}-(1-\rho_2)}{\rho_2-\rho_1}$.

C.1.2 Intuitions

We briefly discuss the roles of these assumptions in the proof of the proposition. In Game 1, prosocial player A chooses $p = 1$ and prosocial player B invests. In Game 2, prosocial player A chooses $p = 1$, and he also chooses (q_1, q_2) strategically to persuade selfish player B to invest. Selfish player A prefers a choice of $p = 0$ conditional on player B's investing, and it is in his best interest to mimic prosocial player A's choice of (q_1, q_2). This mimicking prevents selfish player B from separating $p = 0$ from $p = 1$, which in turn is harmful for prosocial player A. Thus, sophisticated prosocial player A chooses a more

informative (q_1, q_2) so that selfish player A finds it is in his best interest to choose $p = 1$ as well when he mimics prosocial player A's choice of (q_1, q_2). The assumption about the conditionally pessimistic posterior belief in zero-probability information sets guarantees that the most sophisticated selfish player B invests only when the underlying state is clearly state 1, and in turn the most sophisticated player A chooses full trustworthiness with the most informative structure.

Now, consider the role of the assumption about player B's heterogeneity in viewpoint. In short, it helps to predict player A's choosing zero trustworthiness with the least informative structure, i.e., $(p = 0, q_1 = q_2)$, which is opposite to the prediction of player A's choosing full trustworthiness with the most informative structure. The general intuition is that when player B is not aware of the association between trustworthiness and information structure, the probability of player B's investing is continuously responsive to player A's choice of information structure given heterogeneity in viewpoint.[9] Then, player A has an incentive to obfuscate his choice of $p = 0$ using an information structure of $q_1 = q_2$.

The intuition is elaborated below. When observing (q_1, q_2, b), a selfish player B who is not aware of the association invests if her posterior belief about the underlying state exceeds a cutoff value. Her posterior belief is increasing in her prior belief, which itself is increasing in her viewpoint π. Therefore, a player B with π exceeding a cutoff value invests. Moreover, player B's posterior belief, in this case, is also increasing in $\frac{q_1}{q_2}$. As $\frac{q_1}{q_2}$ increases, those B players with a lower π begin to invest, i.e., the cutoff value of π decreases. Given the assumption about π, this result implies that a higher $\frac{q_1}{q_2}$ increases the probability of investing. Similarly, a higher $\frac{1-q_1}{1-q_2}$ increases the probability of investing

[9] If player B's viewpoint is homogeneous, then the effect of information structure on the probability of player B's investing takes a stepwise form: all selfish B players who are not aware of the association invest when observing an information structure with $\frac{q_1}{q_2}$ ($\frac{1-q_1}{1-q_2}$) exceeding a cutoff value; otherwise, none of them invest.

for an information set (q_1, q_2, w). Based on this reasoning, a player A who chooses $p = 0$ trades off choosing a high $\frac{q_1}{q_2}$ $(\frac{1-q_1}{1-q_2})$ against choosing a high probability of generating a black (white) signal, the latter of which is equivalent to choosing a high (low) q_2. Thus, the assumption about π makes the probability of investing continuously responsive to the change in information structure and creates a trade-off problem.

We also provide an intuition for L_1 selfish player B's different behavior in Game 1 and Game 2. This difference is interesting given that L_0 player A cannot signal his trustworthiness in Game 1, and he chooses q_1 and q_2 independently according to a uniform distribution over $[0, 1]$ in Game 2. While (q_1, q_2) per se is uninformative in this case, a realized signal generated according to such an information structure does provide information about the underlying state. L_1 selfish player B observes the realized signal and updates her belief about the underlying state in Game 2. In contrast, she has no belief updating in Game 1. This difference in belief updating causes L_1 selfish player B to play differently in Game 1 and Game 2.

C.1.3 Explaining the treatment effects

Below, we highlight how the behavioral model rationalizes the treatment effects and behavioral patterns that are also predicted by the equilibrium model.

Increasing trustworthiness. Player A, depending on his type, chooses $p = 1$ or $p = 0$ in both games or chooses $p = 0$ in Game 1 and $p = 1$ in Game 2. Thus, the overall trustworthiness increases from Game 1 to Game 2 as long as type $(S, L_{k\geq 4})$ has positive mass.[10] These predictions match Pattern 3 and the treatment effect of increasing trustworthiness from Game 1 to Game 2.

[10] While earlier level-k studies find that levels higher than L_3 are rare, some recent level-k studies show that the fraction of levels $L_{k\geq 4}$ is considerable, e.g., about 20% (Crawford & Iriberri 2007*a*; Kawagoe & Takizawa 2012; Jin 2021). Also note that the strategy of type $(S, L_{k\geq 4})$ coincides with the equilibrium strategy, and the equilibrium type is specified and found in many level-k studies.

Increasing trusting acts. The probability of trusting acts remains constant from Game 1 to Game 2 for types $(P, L_{k\geq 0}, \pi)$ and (S, L_0, π). The probability increases for types $(S, L_{k\geq 1}, \pi \leq \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1})$, who do not invest in Game 1, and decreases for types $(S, L_{k\geq 1}, \pi > \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1})$, who invest in Game 1. Therefore, whether introducing information design boosts trusting acts depends on the proportion of selfish B players who invest in Game 1. Specifically, trusting acts increase under a plausible assumption that a small proportion of selfish B players invest in Game 1.

Increasing expected payoff. We work out each player type's ex ante expected payoff according to the player's expected payoff function and the optimal strategies of the player and her/his opponent. The ex ante expected payoff increases from Game 1 to Game 2 for player A with type $(P, L_{k\geq 2})$ and $(S, L_{k\geq 4})$, and player B with type $(P, L_{k\geq 5}, \pi)$ and $(S, L_{k\geq 1}, \pi)$. The ex ante expected payoff remains the same across games for the remaining player types.

C.2 Proofs

C.2.1 Proof of Proposition 1

Proof:

We first introduce a refinement of PBE and work out equilibrium strategy profiles and belief systems based on this refinement. Then, we show that the set of equilibrium actions according to this refinement equals to the set of equilibrium actions according to PBE.

In & Wright (2018) propose an equilibrium refinement in the form of reordering a class of endogenous signaling games, in which a sender's actions are partially observed. They note that when a sender makes choices of unobserved action and observed action

without gaining any new payoff-relevant information in between, the order in which the choices are made does not matter. Adopting this idea, we reorder Game 2 as follows: player A first chooses (q_1, q_2); then player A chooses p and player B decides whether to invest based on her information (q_1, q_2, s).

The reordering of Game 2 creates proper subgames $\{\Gamma_{(q_1,q_2)}\}_{(q_1,q_2)\in[0,1]\times[0,1]}$. At each proper subgame $\Gamma_{(q_1,q_2)}$, PBE specifies player A's choice of p and player B's actions at $(q_1, q_2, s = b)$ and $(q_1, q_2, s = w)$ given her posterior beliefs, which are consistent with player A's choice of p and updated using Bayes' rule when applicable. At the initial node of the entire game, player A chooses (q_1, q_2) that maximizes his expected payoffs, assuming that the PBE he prefers is played in each $\Gamma_{(q_1,q_2)}$. We dub this notion of equilibrium refinement player-A-preferred PBE.[11]

Let $p(\tilde{q_1}, \tilde{q_2})$ be player A's choice of p at his information set $(\tilde{q_1}, \tilde{q_2}) \in [0, 1] \times [0, 1]$ and let $(q_1, q_2, p(\tilde{q_1}, \tilde{q_2}))$ be player A's strategy. Player B's information set is denoted by $H = (\tilde{q_1}, \tilde{q_2}, s) \in [0, 1] \times [0, 1] \times \{b, w\}/\{(1, 1, w), (0, 0, b)\}$. Player B's strategy is denoted by σ with $\sigma(H) \in [0, 1]$ specifying her probability of investing given information set H.[12] Let σ_b and σ_w refer to $\sigma(\tilde{q_1}, \tilde{q_2}, b)$ and $\sigma(\tilde{q_1}, \tilde{q_2}, w)$, respectively, when the reference is clear.

Player B's optimal strategy

Given Game 2's payoff structure, player B's optimal choice when observing $H = (q_1, q_2, s)$ is: $\sigma(H) = 1$ if her posterior belief about state 1 is $\mu_H > \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1} \equiv \mu^*$, $\sigma(H) = 0$ if $\mu_H < \mu^*$, and $\sigma(H) \in [0, 1]$ if $\mu_H = \mu^*$[13]. Note that $\mu^* \in (0, 1)$.

Let μ_0 be player B's interim belief about state 1 in $\Gamma_{(q_1,q_2)}$. At information set $H =$

[11] In a Bayesian persuasion game with costly messages, Nguyen & Tan (2021) use a solution concept called the sender-preferred equilibrium, according to which the strategies restricted to each subgame form a PBE that the sender prefers in that subgame.

[12] Note that player A and player B cannot benefit from using mixed strategies. We allow for player B's mixed strategy because in subgames with $0 < q_1 < 1, q_2 = 0$ or $q_2 = 1$, there exists only a mixed-strategy equilibrium. We show that player B uses a pure strategy on the equilibrium path.

[13] Player B chooses to invest if $\mu_H(1 - \rho_1)v + (1 - \mu_H)(1 - \rho_2)v > v_0^B$, which is $\mu_H > \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}$.

(q_1, q_2, b), her posterior belief about state 1 (when Bayes' rule applies) is $\mu_{(q_1,q_2,b)} = \frac{\mu_0 q_1}{\mu_0 q_1 + (1-\mu_0)q_2}$. Accordingly, her optimal choice at $H = (q_1, q_2, b)$ is: $\sigma(H) \equiv \sigma_b = 1$ if $\mu_{(q_1,q_2,b)} > \mu^*$ or $\mu_0 > \frac{\rho_0^B - (1-\rho_2)}{(1-\rho_1-\rho_0^B)\frac{q_1}{q_2} + \rho_0^B - (1-\rho_2)} \equiv \mu_{0b}$, $\sigma_b = 0$ if $\mu_0 < \mu_{0b}$ and $\sigma_b \in [0,1]$ if $\mu_0 = \mu_{0b}$. Similarly, her posterior belief at $H = (q_1, q_2, w)$ is $\mu_{(q_1,q_2,w)} = \frac{\mu_0(1-q_1)}{\mu_0(1-q_1)+(1-\mu_0)(1-q_2)}$ and her optimal choice is: $\sigma(H) \equiv \sigma_w = 1$ if $\mu_0 > \frac{\rho_0^B - (1-\rho_2)}{(1-\rho_1-\rho_0^B)\frac{1-q_1}{1-q_2} + \rho_0^B - (1-\rho_2)} \equiv \mu_{0w}$, $\sigma_w = 0$ if $\mu_0 < \mu_{0w}$ and $\sigma_w \in [0,1]$ if $\mu_0 = \mu_{0w}$.

Note that Bayes' rule cannot be used to calculate the posterior belief in the following situations: (1) $\mu_0 = 0$ and $H = (0 \leq q_1 < 1, q_2 = 1, s = w)$ or $H = (0 < q_1 \leq 1, q_2 = 0, s = b)$, and (2) $\mu_0 = 1$ and $H = (q_1 = 0, 0 < q_2 \leq 1, s = b)$ or $H = (q_1 = 1, 0 \leq q_2 < 1, s = w)$. We assume that her posterior beliefs about state 1 in the two situations are respectively one and zero based on intuition, which can also be justified by allowing for an ϵ perturbation of μ_0, applying Bayes' rule, and then letting ϵ approach zero. Then, player B's optimal actions in these information sets are: $\sigma_w = 1$ at $H = (0 \leq q_1 < 1, q_2 = 1, s = w)$, $\sigma_b = 1$ at $H = (0 < q_1 \leq 1, q_2 = 0, s = b)$, $\sigma_b = 0$ at $H = (q_1 = 0, 0 < q_2 \leq 1, s = b)$ and $\sigma_w = 0$ at $H = (q_1 = 1, 0 \leq q_2 < 1, s = w)$.

Based on the above analysis, it is obvious that player B's optimal choices do not depend on her interim belief μ_0 when $(q_1, q_2) = (1,0)/(0,1)$. That is, regardless of μ_0, $\sigma_b = 1$ and $\sigma_w = 0$ in $\Gamma_{(q_1=1,q_2=0)}$, and $\sigma_w = 1$ and $\sigma_b = 0$ in $\Gamma_{(q_1=0,q_2=1)}$.

Player A's optimal strategy

Player A's decision problem is to choose (p, q_1, q_2) that maximizes his expected payoff given player B's optimal strategy. Player A's expected payoff function is: $EV_A = p * [q_1(\rho_1 v \sigma_b + v_0^A(1-\sigma_b)) + (1-q_1)(\rho_1 v \sigma_w + v_0^A(1-\sigma_w))] + (1-p) * [q_2(\rho_2 v \sigma_b + v_0^A(1-\sigma_b)) + (1-q_2)(\rho_2 v \sigma_w + v_0^A(1-\sigma_w))] = v_0^A + (\rho_1 v - v_0^A)p * [q_1\sigma_b + (1-q_1)\sigma_w] + (\rho_2 v - v_0^A)(1-p) * [q_2\sigma_b + (1-q_2)\sigma_w]$.

Clearly, he should optimally choose $p(q_1, q_2) = 1$ if $(\rho_1 v - v_0^A)[q_1\sigma_b + (1-q_1)\sigma_w] >$

$(\rho_2 v - v_0^A)[q_2\sigma_b + (1-q_2)\sigma_w]$, or equivalently $\sigma_b[(\rho_1 - \rho_0^A)q_1 - (\rho_2 - \rho_0^A)q_2] > \sigma_w[(\rho_2 - \rho_0^A)(1-q_2) - (\rho_1 - \rho_0^A)(1-q_1)]$.

Let $\alpha = (\rho_1 - \rho_0^A)q_1 - (\rho_2 - \rho_0^A)q_2$ and $\beta = (\rho_2 - \rho_0^A)(1-q_2) - (\rho_1 - \rho_0^A)(1-q_1) = \alpha + \rho_2 - \rho_1$. Note that $\beta > \alpha$. Thus, his optimal choice of p for any given (q_1, q_2) can be written as: $p(q_1, q_2) = 1$ if $\sigma_b\alpha > \sigma_w\beta$, $p(q_1, q_2) = 0$ if $\sigma_b\alpha < \sigma_w\beta$, and $p(q_1, q_2) \in [0, 1]$ if $\sigma_b\alpha = \sigma_w\beta$. Let $EV_A(q_1, q_2)$ be the corresponding optimal expected payoff in $\Gamma_{(q_1,q_2)}$.

PBE strategy profile and beliefs

We now solve PBE in $\Gamma_{(q_1,q_2)}$. Note that in any PBE in $\Gamma_{(q_1,q_2)}$, player B's posterior beliefs $\mu_{(q_1,q_2,b)}$ and $\mu_{(q_1,q_2,w)}$ must be consistent with player A's choice of p and updated using Bayes' rule when applicable. Since player A's choice of p determines the probability of the realization of state 1 and μ_0 is player B's interim belief about state 1 in $\Gamma_{(q_1,q_2)}$, this belief consistency requirement implies that $\mu_0 = p$ in any PBE.

Note that $\alpha > 0$ if and only if $\frac{q_1}{q_2} > \frac{\rho_2 - \rho_0^A}{\rho_1 - \rho_0^A}$, and $\beta > 0$ if and only if $\frac{1-q_1}{1-q_2} < \frac{\rho_2 - \rho_0^A}{\rho_1 - \rho_0^A}$. In addition, as shown above, σ_b and σ_w depend on $\frac{q_1}{q_2}$ and $\frac{1-q_1}{1-q_2}$, respectively. We categorize $\Gamma_{(q_1,q_2)}$ into 18 cases as shown in Table C.2.1. It is clear that $\alpha > 0$ in cases (1), (8), (17) and (18), and $\beta < 0$ in cases (7), (9), (13) and (14). We work out PBEa player A prefers in each $\Gamma_{(q_1,q_2)}$ as follows.

Consider a PBE with $\mu_0 = p(q_1, q_2)$. First, it is straightforward to establish that in all cases except (8), (9), (14) and (18), there is a PBE with $\mu_0 = p(q_1, q_2) = 0$, $\sigma_b = \sigma_w = 0$ and $EV_A(q_1, q_2) = v_0^A$. We now show that there is no PBE with $\mu_0 = p(q_1, q_2) = 0$ in these four cases. Specifically, if assuming $\mu_0 = 0$: in cases (8) and (18), i.e., $(0 < q_1 \leq 1, q_2 = 0)$, we know $\alpha > 0$, $\sigma_b = 1$ and $\sigma_w = 0$, and thus player A should choose $p(q_1, q_2) = 1$; in cases (9) and (14), i.e., $(0 \leq q_1 < 1, q_2 = 1)$, we know $\beta < 0$, $\sigma_b = 0$ and $\sigma_w = 1$, and thus player A should also choose $p(q_1, q_2) = 1$.

Second, we show that there is no PBE with $\mu_0 = p(q_1, q_2) = 1$ in cases (1)-(10), (11)

Table C.2.1: PBEa Player A Prefers in $\Gamma_{(q_1,q_2)}$

Case	$\Gamma_{(q_1,q_2)}$	PBEa player A prefers	$EV_A(q_1,q_2)$
(1)	$0<q_2<q_1<1, \frac{q_1}{q_2}>\frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}$	$p=\mu_{0w}, \sigma_b=1, \sigma_w=\frac{\alpha}{\beta}$	$v_0^A+(\rho_1 v-v_0^A)(q_1+(1-q_1)*\frac{\alpha}{\beta})$
(2)	$0<q_2<q_1<1, \frac{q_1}{q_2}=\frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}$	$p\in[\mu_{0b},\mu_{0w}], \sigma_b=1, \sigma_w=0$	$v_0^A+(\rho_1 v-v_0^A)q_1$
(3)	$0<q_2<q_1<1, \frac{q_1}{q_2}<\frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}$	$p\in[0,\mu_{0b}], \sigma_b=\sigma_w=0$	v_0^A
(4)	$0<q_1=q_2<1$	$p\in[0,\mu_{0b}], \sigma_b=\sigma_w=0$	v_0^A
(5)	$0<q_1<q_2<1, \frac{1-q_1}{1-q_2}<\frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}$	$p\in[0,\mu_{0w}], \sigma_b=\sigma_w=0$	v_0^A
(6)	$0<q_1<q_2<1, \frac{1-q_1}{1-q_2}=\frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}$	$p\in[\mu_{0w},\mu_{0b}], \sigma_b=0, \sigma_w=1$	$v_0^A+(\rho_1 v-v_0^A)(1-q_1)$
(7)	$0<q_1<q_2<1, \frac{1-q_1}{1-q_2}>\frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}$	$p=\mu_{0b}, \sigma_b=\frac{\beta}{\alpha}, \sigma_w=1$	$v_0^A+(\rho_1 v-v_0^A)(q_1\frac{\beta}{\alpha}+1-q_1)$
(8)	$0<q_1<1, q_2=0$	$p=\mu_{0w}, \sigma_b=1, \sigma_w=\frac{\alpha}{\beta}$	$v_0^A+(\rho_1 v-v_0^A)(q_1+(1-q_1)*\frac{\alpha}{\beta})$
(9)	$0<q_1<1, q_2=1$	$p=\mu_{0b}, \sigma_b=\frac{\beta}{\alpha}, \sigma_w=1$	$v_0^A+(\rho_1 v-v_0^A)(q_1\frac{\beta}{\alpha}+(1-q_1))$
(10)	$q_1=q_2=0$ $(q_1=q_2=1)$	$p\in[0,\mu_{0w}], \sigma_w=0$ $(p\in[0,\mu_{0b}], \sigma_b=0)$	v_0^A
(11)	$q_1=0, 0<q_2<\frac{\rho_2-\rho_1}{\rho_2-\rho_0^A}$	$p\in[0,\mu_{0w}], \sigma_b=\sigma_w=0$	v_0^A
(12)	$q_1=0, q_2=\frac{\rho_2-\rho_1}{\rho_2-\rho_0^A}$	$p\in[\mu_{0w},1], \sigma_b=0, \sigma_w=1$	$\rho_1 v$
(13)	$q_1=0, \frac{\rho_2-\rho_1}{\rho_2-\rho_0^A}<q_2<1$	$p=1, \sigma_b=0, \sigma_w=1$	$\rho_1 v$
(14)	$q_1=0, q_2=1$	$p=1, \sigma_b=0, \sigma_w=1$	$\rho_1 v$
(15)	$q_1=1, \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}<q_2<1$	$p\in[0,\mu_{0b}], \sigma_b=\sigma_w=0$	v_0^A
(16)	$q_1=1, q_2=\frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}$	$p\in[\mu_{0b},1], \sigma_b=1, \sigma_w=0$	$\rho_1 v$
(17)	$q_1=1, 0<q_2<\frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}$	$p=1, \sigma_b=1, \sigma_w=0$	$\rho_1 v$
(18)	$q_1=1, q_2=0$	$p=1, \sigma_b=1, \sigma_w=0$	$\rho_1 v$

Notes: $\mu_{0b}\equiv\frac{\rho_0^B-(1-\rho_2)}{(1-\rho_1-\rho_0^B)\frac{q_1}{q_2}+\rho_0^B-(1-\rho_2)}$, $\mu_{0w}\equiv\frac{\rho_0^B-(1-\rho_2)}{(1-\rho_1-\rho_0^B)\frac{1-q_1}{1-q_2}+\rho_0^B-(1-\rho_2)}$, $\alpha=(\rho_1-\rho_0^A)q_1-(\rho_2-\rho_0^A)q_2$, *and* $\beta=(\rho_2-\rho_0^A)(1-q_2)-(\rho_1-\rho_0^A)(1-q_1)$.

and (15). In cases (1)-(10) if $\mu_0=p(q_1,q_2)=1$, then $\sigma_b=\sigma_w=1$, which implies that $\sigma_b\alpha<\sigma_w\beta$ and in turn $p(q_1,q_2)=0$. Consider $\mu_0=1$. In cases (1)-(10), $\sigma_b=\sigma_w=1$, implying $\sigma_b\alpha<\sigma_w\beta$ and in turn $p(q_1,q_2)=0$; in case (11), $\sigma_b=0$, $\beta>0$ and $\sigma_w=1$, implying $\sigma_b\alpha<\sigma_w\beta$ and in turn $p(q_1,q_2)=0$; in case (15), $\sigma_w=0$, $\alpha<0$ and $\sigma_b=1$, implying $\sigma_b\alpha<\sigma_w\beta$ and in turn $p(q_1,q_2)=0$.

Third, we show that there is a PBE with $\mu_0=p(q_1,q_2)=1$ and $EV_A(q_1,q_2)=\rho_1 v$ in cases (12)-(14) and (16)-(18). In cases (12)-(14), $\sigma_b=0$, $\alpha<0$ and $\beta\leq 0$. $\mu_0=p(q_1,q_2)=1$ and $\sigma_w=1$ satisfy both players' optimality conditions. Similarly, in cases (16)-(18), $\sigma_w=0$, $\alpha\geq 0$ and $\beta>0$. $\mu_0=p(q_1,q_2)=1$ and $\sigma_b=1$ satisfy both players'

optimality conditions.

In the above, we have explored the PBEa with $\mu_0 = p(q_1, q_2) = 1$ or $\mu_0 = p(q_1, q_2) = 0$. We now investigate whether there is any PBE with $\mu_0 = p(q_1, q_2) \in (0, 1)$ and characterize the PBE player A prefers for each of the 18 cases. Note that given $p(q_1, q_2) \in (0, 1)$, it must be that $\sigma_b\alpha = \sigma_w\beta$.

Case (1): $(q_1, q_2) \in (0, 1) \times (0, 1)$ and $\frac{q_1}{q_2} > \frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}$. In this case, $\alpha > 0$, $\beta > 0$, and $0 < \mu_{0b} < \mu_{0w} < 1$ since $\frac{q_1}{q_2} > \frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A} > 1 > \frac{1-q_1}{1-q_2}$. $\sigma_b\alpha = \sigma_w\beta$ implies $\sigma_b = \sigma_w = 0$ and $\mu_0 \in (0, \mu_{0b}]$, or $\sigma_b = 1, \sigma_w = \frac{\alpha}{\beta}$ and $\mu_0 = \mu_{0w}$. So in PBE of this situation, $EV_A(q_1, q_2) = v_0^A$ or $= v_0^A + (\rho_1 v - v_0^A)[q_1 + (1-q_1)\frac{\alpha}{\beta}] < \rho_1 v$. Therefore, $\sigma_b = 1, \sigma_w = \frac{\alpha}{\beta}$ and $p = \mu_{0w}$ constitute the PBE player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)[q_1 + (1 - q_1)\frac{\alpha}{\beta}]$.

Case (2): $(q_1, q_2) \in (0, 1) \times (0, 1)$ and $\frac{q_1}{q_2} = \frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}$. In this case, $\alpha = 0$, $\beta > 0$, and $0 < \mu_{0b} < \mu_{0w} < 1$. $\sigma_b\alpha = \sigma_w\beta$ implies that $\sigma_b \in [0, 1]$, $\sigma_w = 0$ and $\mu_0 \in (0, \mu_{0w}]$. So in PBE of this situation, $EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)\sigma_b q_1 \leq v_0^A + (\rho_1 v - v_0^A)q_1$, where the equality holds when $\sigma_b = 1$ and $\mu_0 \in [\mu_{0b}, \mu_{0w}]$. Therefore, $\sigma_b = 1, \sigma_w = 0$ and $p \in [\mu_{0b}, \mu_{0w}]$ constitute the PBEa player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)q_1$.

Case (3): $(q_1, q_2) \in (0, 1) \times (0, 1)$ and $\frac{q_1}{q_2} \in (1, \frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A})$. In this case, $\alpha < 0$, $\beta > 0$, and $0 < \mu_{0b} < \mu_{0w} < 1$. $\sigma_b\alpha = \sigma_w\beta$ implies that $\sigma_b = 0$, $\sigma_w = 0$ and $\mu_0 \in (0, \mu_{0b}]$. So in PBE of this situation, $EV_A(q_1, q_2) = v_0^A$. Therefore, $\sigma_b = 0, \sigma_w = 0$ and $p \in [0, \mu_{0b}]$ constitute the PBEa player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A$.

Case (4): $(q_1, q_2) \in (0, 1) \times (0, 1)$ and $\frac{q_1}{q_2} = 1$. In this case, $\alpha < 0$, $\beta > 0$, and $0 < \mu_{0b} = \mu_{0w} < 1$. $\sigma_b\alpha = \sigma_w\beta$ implies that $\sigma_b = 0$, $\sigma_w = 0$ and $\mu_0 \in (0, \mu_{0b}]$. So in PBE of this situation, $EV_A(q_1, q_2) = v_0^A$. Therefore, $\sigma_b = 0, \sigma_w = 0$ and $p \in [0, \mu_{0b}]$ constitute the PBEa player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A$.

Note that case (5) is symmetric to case (3), case (6) is symmetric to case (2) and case (7) is symmetric to case (1). A similar analysis shows that in case (5), $\sigma_b = 0, \sigma_w = 0$ and

$p \in [0, \mu_{0w}]$ constitute the PBEa player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A$. In case (6), $\sigma_b = 0, \sigma_w = 1$ and $p \in [\mu_{0w}, \mu_{0b}]$ constitute the PBE player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)(1 - q_1)$. In case (7), $\sigma_w = 1, \sigma_b = \frac{\beta}{\alpha}$ and $p = \mu_{0b}$ constitute the PBE player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)(q_1 \frac{\beta}{\alpha} + 1 - q_1)$.

Case (8): $(0 < q_1 < 1, q_2 = 0)$. In this case, $\alpha > 0$, $\beta > 0$, $\sigma_b = 1$, and $0 = \mu_{0b} < \mu_{0w} < 1$. $\sigma_b \alpha = \sigma_w \beta$ implies that $\sigma_b = 1, \sigma_w = \frac{\alpha}{\beta}$ and $\mu_0 = \mu_{0w}$. So in PBE of this situation, $EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)[q_1 + (1 - q_1)\frac{\alpha}{\beta}] \in (v_0^A, \rho_1 v)$. Therefore, $\sigma_b = 1, \sigma_w = \frac{\alpha}{\beta}$ and $p = \mu_{0w}$ constitute the PBE player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)[q_1 + (1 - q_1)\frac{\alpha}{\beta}]$.

Case (9): $(0 < q_1 < 1, q_2 = 1)$. In this case, $\alpha < \beta < 0$, $\sigma_w = 1$, and $0 = \mu_{0w} < \mu_{0b} < 1$. $\sigma_b \alpha = \sigma_w \beta$ implies that $\sigma_w = 1, \sigma_b = \frac{\beta}{\alpha}$ and $\mu_0 = \mu_{0b}$. So in PBE of this situation, $EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)[q_1 \sigma_b + (1 - q_1)] \in (v_0^A, \rho_1 v)$. Therefore, $\sigma_b = \frac{\beta}{\alpha}, \sigma_w = 1$ and $p = \mu_{0b}$ constitute the PBE player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)[q_1 \frac{\beta}{\alpha} + (1 - q_1)]$.

Case (10): $(q_1 = 0, q_2 = 0)$ or $(q_1 = 1, q_2 = 1)$. Note that in the case of $(q_1 = 0, q_2 = 0)$, $\alpha = 0$ and $\beta > 0$. In the case of $(q_1 = 1, q_2 = 1)$, $\alpha < 0$ and $\beta = 0$. Additionally, her posterior is equal to her interim belief μ_0. $\sigma_b \alpha = \sigma_w \beta$ implies that $\sigma_w = 0$ ($\sigma_b = 0$) in the first (second) situation, which implies that $\mu_0 \in (0, \mu_{0w}]$ ($\mu_0 \in (0, \mu_{0b}]$). Therefore, $\sigma_w = 0$ ($\sigma_b = 0$) and $p \in [0, \mu_{0w}]$ ($p \in [0, \mu_{0b}]$) constitute the PBE player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A$.

Case (11): $(q_1 = 0, 0 < q_2 < \frac{\rho_2 - \rho_1}{\rho_2 - \rho_0^A})$. In this case, $\alpha < 0$, $\beta > 0$, $\sigma_b = 0$, and $0 < \mu_{0w} < \mu_{0b} = 1$. $\sigma_b \alpha = \sigma_w \beta$ implies that $\sigma_w = 0$ and $\mu_0 \in (0, \mu_{0w}]$. So in PBE of this situation, $EV_A(q_1, q_2) = v_0^A$. Therefore, $\sigma_b = \sigma_w = 0$ and $p \in [0, \mu_{0w}]$ constitute the PBEa player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A$.

Case (12): $(q_1 = 0, q_2 = \frac{\rho_2 - \rho_1}{\rho_2 - \rho_0^A})$. In this case, $\alpha < 0$, $\beta = 0$, $\sigma_b = 0$, and $0 < \mu_{0w} < \mu_{0b} = 1$. $\sigma_b \alpha = \sigma_w \beta$ implies that $\sigma_w \in [0, 1]$. So in PBE of this situation,

$EV_A(q_1, q_2) = v_0^A + (\rho_1 v - v_0^A)\sigma_w \leq \rho_1 v$, in which the equality holds when $\sigma_w = 1$ and $\mu_0 \in [\mu_{0w}, 1)$. Therefore, $\sigma_b = 0, \sigma_w = 1$ and $p \in [\mu_{0w}, 1]$ constitute the PBEa player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = \rho_1 v$.

Case (13): $(q_1 = 0, \frac{\rho_2-\rho_1}{\rho_2-\rho_0^A} < q_2 < 1)$. In this case, $\alpha < 0$, $\beta < 0$, $\sigma_b = 0$, and $0 < \mu_{0w} < \mu_{0b} = 1$. $\sigma_b\alpha = \sigma_w\beta$ implies that $\sigma_w = 0$ and $\mu_0 \in (0, \mu_{0w}]$. So in PBE of this situation, $EV_A(q_1, q_2) = v_0^A$. Recall that in case (13), there exists another PBE in which $\sigma_b = 0, \sigma_w = 1$, $p = 1$ and $EV_A(q_1, q_2) = \rho_1 v$. Therefore, $\sigma_b = 0, \sigma_w = 1$ and $p = 1$ constitute the PBE player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = \rho_1 v$.

Case (14): $(q_1 = 0, q_2 = 1)$. In this case, $\alpha < 0$, $\beta < 0$, $\sigma_b = 0$, $\sigma_w = 1$, and $0 = \mu_{0w} < \mu_{0b} = 1$. $\sigma_b\alpha = \sigma_w\beta$ cannot hold, which implies that there is no PBE of this situation. But recall that in case (14), there exists a PBE in which $\sigma_b = 0, \sigma_w = 1$, $p = 1$ and $EV_A(q_1, q_2) = \rho_1 v$. Hence, $\sigma_b = 0, \sigma_w = 1$ and $p = 1$ constitute the PBE player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = \rho_1 v$.

Note that case (15) is symmetric to case (11), case (16) is symmetric to case (12), case (17) is symmetric to case (13) and case (18) is symmetric to case (14). A similar analysis shows that in case (15), $\sigma_b = \sigma_w = 0$ and $p \in [0, \mu_{0b}]$ constitute the PBEa player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = v_0^A$. In case (16), $\sigma_b = 1, \sigma_w = 0$ and $p \in [\mu_{0b}, 1]$ constitute the PBEa player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = \rho_1 v$. In cases (17) and (18), $\sigma_b = 1, \sigma_w = 0$ and $p = 1$ constitute the PBE player A prefers in $\Gamma_{(q_1,q_2)}$ with $EV_A(q_1, q_2) = \rho_1 v$.

We summarize in Table C.2.1 the PBEa player A prefers in $\Gamma_{(q_1,q_2)}$.

Finally, player A's decision problem at the initial node is to find (q_1, q_2) that maximizes $EV_A(q_1, q_2)$, that is, $max_{(q_1,q_2)\in[0,1]\times[0,1]} EV_A(q_1, q_2)$. According to Table C.2.1, player A optimally chooses $(q_1 = 0, q_2 \geq \frac{\rho_2-\rho_1}{\rho_2-\rho_0^A})$ or $(q_1 = 1, q_2 \leq \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A})$ with an expected payoff of $\rho_1 v$. Therefore, in player-A-preferred PBEa, player A has two types

of equilibrium actions. (1) Full trustworthiness: $(q_1 = 1, q_2 \leq \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}, p = 1)$ or $(q_1 = 0, q_2 \geq \frac{\rho_2-\rho_1}{\rho_2-\rho_0^A}, p = 1)$. (2) Intermediate trustworthiness: $(q_1 = 1, q_2 = \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}, \underline{p} \leq p < 1)$ or $(q_1 = 0, q_2 = \frac{\rho_2-\rho_1}{\rho_2-\rho_0^A}, \underline{p} \leq p < 1)$, where $\underline{p} \equiv \frac{\rho_0^B-(1-\rho_2)}{(1-\rho_1-\rho_0^B)\frac{\rho_2-\rho_0^A}{\rho_1-\rho_0^A}+\rho_0^B-(1-\rho_2)} \in (0,1)$. And $(p(\tilde{q_1},\tilde{q_2}), \sigma(\tilde{q_1},\tilde{q_2},b), \sigma(\tilde{q_1},\tilde{q_2},w))$ off the equilibrium paths is specified in Table C.2.1. And player B's belief system is: $\mu_{(\tilde{q_1},\tilde{q_2},b)} = \frac{p(\tilde{q_1},\tilde{q_2})\tilde{q_1}}{p(\tilde{q_1},\tilde{q_2})\tilde{q_1}+(1-p(\tilde{q_1},\tilde{q_2}))\tilde{q_2}}$ and $\mu_{(\tilde{q_1},\tilde{q_2},w)} = \frac{p(\tilde{q_1},\tilde{q_2})(1-\tilde{q_1})}{p(\tilde{q_1},\tilde{q_2})(1-\tilde{q_1})+(1-p(\tilde{q_1},\tilde{q_2}))(1-\tilde{q_2})}$ when Bayes' rule applies, and $\mu_H = 0$ when $H = (\tilde{q_1} = 0, \tilde{q_2} \geq \frac{\rho_2-\rho_1}{\rho_2-\rho_0^A}, s = b)$ or $H = (\tilde{q_1} = 1, \tilde{q_2} \leq \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}, s = w)$.

Set of equilibrium actions according to PBE

It is straightforward to show that any player-A-preferred PBE must also be a PBE. Thus, the set of equilibrium actions according to player-A-preferred PBE is a subset of the set of equilibrium actions according to PBE.

Next, consider an arbitrary PBE with the following strategy profile: Player A chooses (p, q_1, q_2) and player B's strategy specifies her action (investing or not investing) at each possible information set $(\tilde{q_1}, \tilde{q_2}, s)$. Clearly, given (q_1, q_2) and player B's actions at $(q_1, q_2, s = b)$ and $(q_1, q_2, s = w)$, p is optimal for player A. Given (p, q_1, q_2), player B's actions at $(q_1, q_2, s = b)$ and $(q_1, q_2, s = w)$ are optimal for player B. We also know from the text that player A has an expected payoff of $\rho_1 v$ in any PBE. So, $EV_A(p, q_1, q_2) = \rho_1 v$. This shows that player A's choosing p and player B's actions at $(q_1, q_2, s = b)$ and $(q_1, q_2, s = w)$ constitute a PBE strategy profile in $\Gamma_{(q_1,q_2)}$ such that player A's PBE expected payoff is $\rho_1 v$. Then, we know from Table C.2.1 that player A's choosing (p, q_1, q_2) and player B's actions at $(q_1, q_2, s = b)$ and $(q_1, q_2, s = w)$ constitute equilibrium actions on a certain player-A-preferred PBE. This shows that the set of equilibrium actions according to PBE is a subset of the set of equilibrium actions according to player-A-preferred PBE, and in turn the two sets are equal.

C.2.2 Proof of Proposition 2

Proof:

We first establish two claims.

Claim 1 *Q is more informative than Q' if and only if* $[min\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}, max\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}] \subseteq [min\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\}, max\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\}]$.

Claim 2 *Q and Q' are identically informative if and only if $Q + Q' = (1, 1)$ or both are the least informative.*

By definition, Q is more informative than Q' if and only if the linear system of four equations with two unknown variables has a solution $(\alpha, \beta) \in [0, 1] \times [0, 1]$: $\alpha q_1 + \beta(1 - q_1) = q_1^{'}$; $(1 - \alpha)q_1 + (1 - \beta)(1 - q_1) = 1 - q_1^{'}$; $\alpha q_2 + \beta(1 - q_2) = q_2^{'}$; and $(1-\alpha)q_2+(1-\beta)(1-q_2) = 1-q_2^{'}$. It is clear that two of the four equations are redundant: The linear system with two equations $\alpha q_1 + \beta(1 - q_1) = q_1^{'}$ and $\alpha q_2 + \beta(1 - q_2) = q_2^{'}$ is equivalent to the original system.

First, consider the case of $q_1 > q_2$. In this case, the linear system has a unique solution: $\alpha = \frac{q_1^{'}-q_2^{'}+q_1q_2^{'}-q_2q_1^{'}}{q_1-q_2}$ and $\beta = \frac{q_1q_2^{'}-q_2q_1^{'}}{q_1-q_2}$. $\alpha \geq 0$ if and only if $\frac{q_1^{'}}{q_2^{'}} \geq \frac{1-q_1}{1-q_2}$. $\alpha \leq 1$ if and only if $(q_1 - q_2) - (q_1^{'} - q_2^{'}) \geq q_1q_2^{'} - q_2q_1^{'}$, which is equivalent to $\frac{1-q_1^{'}}{1-q_2^{'}} \geq \frac{1-q_1}{1-q_2}$. $\beta \geq 0$ if and only if $\frac{q_1}{q_2} \geq \frac{q_1^{'}}{q_2^{'}}$. $\beta \leq 1$ if and only if $\frac{q_1}{q_2} \geq \frac{1-q_1^{'}}{1-q_2^{'}}$. Second, consider the case of $q_1 < q_2$. A similar analysis shows that the linear system has a solution $(\alpha, \beta) \in [0, 1] \times [0, 1]$ if and only if $max\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\} \leq \frac{1-q_1}{1-q_2}$ and $min\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\} \geq \frac{q_1}{q_2}$. Finally, consider the case of $q_1 = q_2 = q$. In this case, the linear system becomes $\alpha q+\beta(1-q) = q_1^{'}$ and $\alpha q + \beta(1 - q) = q_2^{'}$, which has a solution $(\alpha, \beta) \in [0, 1] \times [0, 1]$ if and only if $q_1^{'} = q_2^{'}$. Then, $[min\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}, max\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}] = [min\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\}, max\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\}]$. The converse of the statement in this case is straightforward.

Claim 2 is an implication of Claim 1. Suppose that Q is more informative than

$Q'(\neq Q)$ and Q' is more informative than Q. Claim 1 implies that $max\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\} = max\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}$ and $min\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\} = min\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}$, which implies that $\frac{q_1}{q_2} = \frac{q_1^{'}}{q_2^{'}}$ and $\frac{1-q_1}{1-q_2} = \frac{1-q_1^{'}}{1-q_2^{'}}$ or that $\frac{q_1}{q_2} = \frac{1-q_1^{'}}{1-q_2^{'}}$ and $\frac{1-q_1}{1-q_2} = \frac{q_1^{'}}{q_2^{'}}$. In the first case, we can establish that $q_1 = q_2$ and $q_1^{'} = q_2^{'}$. In the second case, we can establish that $q_1 = q_2$ and $q_1^{'} = q_2^{'}$ or that $q_1^{'} + q_1 = 1$ and $q_2^{'} + q_2 = 1$. It is straightforward to establish the converse of the statement.

We then show that $u_Q = |q_1 - q_2|$. Suppose now (q_1, q_2) is more informative than (x, y), which is randomly drawn from $[0, 1] \times [0, 1]$ according to the uniform distribution. We derive the conditions that (x, y) must satisfy and compute $Pr((x, y) : (q_1, q_2)$ is more informative than $(x, y))$. It is clear from the definition that $u_Q = 0$ when $q_1 = q_2$. For notation simplicity, let $t = \frac{q_1}{q_2}$ and $t^{'} = \frac{1-q_1}{1-q_2}$.

First, consider the case of $q_1 > q_2$. In this case, $t > 1 > t^{'}$. Claim 1 implies that (x, y) satisfies the following inequalities: $\frac{x}{y} \leq t$, $\frac{1-x}{1-y} \leq t$, $\frac{x}{y} \geq t^{'}$, and $y \geq \frac{x}{t^{'}} + 1 - \frac{1}{t^{'}}$. These inequalities are equivalent to $max\{\frac{x}{t}, \frac{x}{t^{'}} + 1 - \frac{1}{t^{'}}\} \leq y \leq min\{\frac{x}{t^{'}}, \frac{x}{t} + 1 - \frac{1}{t}\}$ and it is straightforward to establish that $max\{\frac{x}{t}, \frac{x}{t^{'}} + 1 - \frac{1}{t^{'}}\} \leq min\{\frac{x}{t^{'}}, \frac{x}{t} + 1 - \frac{1}{t}\}$. Note that $\frac{x}{t} \geq \frac{x}{t^{'}} + 1 - \frac{1}{t^{'}}$ if and only if $x \leq \frac{t - tt^{'}}{t - t^{'}} \equiv x_1$. Note also that $\frac{x}{t^{'}} \geq \frac{x}{t} + 1 - \frac{1}{t}$ if and only if $x \geq \frac{tt^{'} - t^{'}}{t - t^{'}} \equiv x_2$. Clearly, $x_1 > x_2$ if and only if $t + t^{'} > 2tt^{'}$, i.e., $\frac{q_1}{q_2} + \frac{1-q_1}{1-q_2} > 2\frac{q_1}{q_2}\frac{1-q_1}{1-q_2}$.

In the sub-case of $\frac{q_1}{q_2} + \frac{1-q_1}{1-q_2} > 2\frac{q_1}{q_2}\frac{1-q_1}{1-q_2}$,

$$
\begin{aligned}
u_Q &= \int_0^{x_2} Pr(\frac{x}{t} \leq y \leq \frac{x}{t^{'}})\, dx + \int_{x_2}^{x_1} Pr(\frac{x}{t} \leq y \leq \frac{x}{t} + 1 - \frac{1}{t})\, dx \\
&\quad + \int_{x_1}^{1} Pr(\frac{x}{t^{'}} + 1 - \frac{1}{t^{'}} \leq y \leq \frac{x}{t} + 1 - \frac{1}{t})\, dx \\
&= (\frac{1}{t^{'}} - \frac{1}{t}) * \frac{1}{2}x^2|_0^{x_2} + (1 - \frac{1}{t})x|_{x_2}^{x_1} + (\frac{1}{t} - \frac{1}{t^{'}}) * \frac{1}{2}x^2|_{x_1}^{1} + (\frac{1}{t^{'}} - \frac{1}{t})x|_{x_1}^{1} \\
&= \frac{1}{2}[\frac{t - t^{'}}{tt^{'}} - \frac{t(1 - t^{'})^2}{t^{'}(t - t^{'})} - \frac{t^{'}(1 - t)^2}{t(t - t^{'})}] \\
&= \frac{(1 - t^{'})(t - 1)}{t - t^{'}} \\
&= q_1 - q_2
\end{aligned}
$$

In the sub-case of $\frac{q_1}{q_2} + \frac{1-q_1}{1-q_2} < 2\frac{q_1}{q_2}\frac{1-q_1}{1-q_2}$,

$$\begin{aligned}
u_Q &= \int_0^{x_1} Pr(\frac{x}{t} \leq y \leq \frac{x}{t'})\, dx + \int_{x_1}^{x_2} Pr(\frac{x}{t'} + 1 - \frac{1}{t'} \leq y \leq \frac{x}{t'})\, dx \\
&\quad + \int_{x_2}^{1} Pr(\frac{x}{t'} + 1 - \frac{1}{t'} \leq y \leq \frac{x}{t} + 1 - \frac{1}{t})\, dx \\
&= (\frac{1}{t'} - \frac{1}{t}) * \frac{1}{2}x^2|_0^{x_1} + (\frac{1}{t'} - 1)x|_{x_1}^{x_2} + (\frac{1}{t} - \frac{1}{t'}) * \frac{1}{2}x^2|_{x_2}^{1} + (\frac{1}{t'} - \frac{1}{t})x|_{x_2}^{1} \\
&= \frac{1}{2}(\frac{1}{t'} - \frac{1}{t})[\frac{t^2(1-t')^2}{(t-t')^2} + \frac{(t')^2(1-t)^2}{(t-t')^2} + 1] - \frac{t(1-t')^2}{t'(t-t')} - \frac{t'(1-t)^2}{t(t-t')} \\
&= q_1 - q_2
\end{aligned}$$

Now consider the sub-case of $\frac{q_1}{q_2} + \frac{1-q_1}{1-q_2} = 2\frac{q_1}{q_2}\frac{1-q_1}{1-q_2}$. The equality implies that $(q_1 - q_2)(2q_1 - 1) = 0$, which implies that $q_1 = \frac{1}{2}$. Then,

$$\begin{aligned}
u_Q &= \int_0^{x_1} Pr(\frac{x}{t} \leq y \leq \frac{x}{t'})\, dx + \int_{x_1}^{1} Pr(\frac{x}{t'} + 1 - \frac{1}{t'} \leq y \leq \frac{x}{t} + 1 - \frac{1}{t})\, dx \\
&= (\frac{1}{t'} - \frac{1}{t}) * \frac{1}{2}x^2|_0^{x_1} + (\frac{1}{t} - \frac{1}{t'}) * \frac{1}{2}x^2|_{x_1}^{1} + (\frac{1}{t'} - \frac{1}{t})x|_{x_1}^{1} \\
&= \frac{t(1-t')^2}{t'(t-t')} - \frac{1}{2}(\frac{1}{t'} + \frac{1}{t}) + 1 \\
&= \frac{1}{2} - \frac{1}{2t} \\
&= \frac{1}{2} - q_2
\end{aligned}$$

where the second to last equality comes from the fact that $t + t' = 2tt'$ in the case of $x_1 = x_2$ and in turn $\frac{t'-1}{t-t'} = -\frac{1}{2t}$.

Therefore, $u_Q = q_1 - q_2$ in the case of $q_1 > q_2$.

Second, consider the case of $q_1 < q_2$. In this case, $t < 1$ and $t' > 1$. Claim 1 implies that (x, y) satisfies the following inequalities: $\frac{x}{y} \geq t$, $\frac{1-x}{1-y} \geq t$, $\frac{x}{y} \leq t'$, and $y \leq \frac{x}{t'} + 1 - \frac{1}{t'}$. These inequalities are equivalent to $max\{ty, t'y + 1 - t'\} \leq x \leq min\{t'y, ty + 1 - t\}$ and it is straightforward to establish that $max\{ty, t'y + 1 - t'\} \leq min\{t'y, ty + 1 - t\}$. A

similar analysis shows that $u_Q = \frac{1}{2}[\frac{t^{'}-t}{tt^{'}} - \frac{t^{'}(1-t)^2}{t(t^{'}-t)} - \frac{t(1-t^{'})^2}{t^{'}(t^{'}-t)}] = q_2 - q_1$.

We now show that if Q is strictly more informative than (identically informative as) Q', then $u_Q > u_{Q'}$ ($u_Q = u_{Q'}$). Claim 2 and $u_Q = |q_1 - q_2|$ imply that if Q is identically informative as Q', then $u_Q = u_{Q'}$. Suppose now that Q is strictly more informative than Q'. It is without loss of generality to assume that $\frac{q_1}{q_2} > 1$ and $\frac{q_1^{'}}{q_2^{'}} \geq 1$. Then, Claim 1 implies that $[min\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}, max\{\frac{q_1^{'}}{q_2^{'}}, \frac{1-q_1^{'}}{1-q_2^{'}}\}] = [\frac{1-q_1^{'}}{1-q_2^{'}}, \frac{q_1^{'}}{q_2^{'}}] \subset [min\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\}, max\{\frac{q_1}{q_2}, \frac{1-q_1}{1-q_2}\}] = [\frac{1-q_1}{1-q_2}, \frac{q_1}{q_2}]$, that is, $\frac{1-q_1^{'}}{1-q_2^{'}} \geq \frac{1-q_1}{1-q_2}$ and $\frac{q_1^{'}}{q_2^{'}} < \frac{q_1}{q_2}$ or $\frac{1-q_1^{'}}{1-q_2^{'}} > \frac{1-q_1}{1-q_2}$ and $\frac{q_1^{'}}{q_2^{'}} \leq \frac{q_1}{q_2}$. It is straightforward to establish that in both cases $q_1 - q_2 > q_1^{'} - q_2^{'}$, i.e., $u_Q > u_{Q'}$.

C.2.3 Proof of Proposition 3

Proof:

We first formulate player B's prior belief and posterior belief about state 1 based on the assumptions about heterogeneity. We assume that player B forms a prior belief before she observes any information set, and forms a posterior belief after observing information set (if any) and right before making her choice. Consider a player B whose type is specified by (θ_B, L_k, π). Let E denote the event of state 1 and correspondingly E^c denote the event of state 2. Given (L_k, π), her prior belief can be characterized as:

$$Pr(E) = (1-\pi) * Pr(L_{k-1} \text{ selfish A's choice of } p) + \pi * Pr(L_{k-1} \text{ prosocial A's choice of } p)$$
$$\equiv (1-\pi) * p_{A1} + \pi * p_{A2}$$

In Game 1, player B observes no information set and in turn her posterior belief is identical to her prior belief $Pr(E)$. In Game 2, her posterior belief after observing an

information set $H = (q_1, q_2, s)$ is updated according to Bayes' rule:

$$\begin{aligned} Pr(E|H) &= Pr(E, A1|H) + Pr(E, A2|H) \\ &= \frac{Pr(A1)Pr(E|A1)Pr(H|E, A1) + Pr(A2)Pr(E|A2)Pr(H|E, A2)}{Pr(H)} \\ &= \frac{(1-\pi)p_{A1}Pr(H|E, A1) + \pi p_{A2}Pr(H|E, A2)}{Pr(H)} \end{aligned}$$

where $Pr(H)$ can be explicitly written as,

$$\begin{aligned} Pr(H) &= Pr(E, A1, H) + Pr(E, A2, H) + Pr(E^c, A1, H) + Pr(E^c, A2, H) \\ &= (1-\pi)p_{A1}Pr(H|E, A1) + \pi p_{A2}Pr(H|E, A2) \\ &\quad + (1-\pi)(1-p_{A1})Pr(H|E^c, A1) + \pi(1-p_{A2})Pr(H|E^c, A2) \\ &= [(1-\pi)p_{A1}Pr(q_1, q_2|A1) + \pi p_{A2}Pr(q_1, q_2|A2)]Pr(s|E, q_1, q_2) \\ &\quad + [(1-\pi)(1-p_{A1})Pr(q_1, q_2|A1) + \pi(1-p_{A2})Pr(q_1, q_2|A2)]Pr(s|E^c, q_1, q_2). \end{aligned}$$

When Bayes' rule is not defined, i.e., $Pr(H) = 0$, her posterior belief is specified by the assumption about conditionally pessimistic posterior belief.

The specification for L_0 players is as follows. Selfish and prosocial A players choose $p = 0$ and $p = 1$ respectively in both games. In Game 2, they also choose q_1 and q_2 independently according to a uniform distribution over $[0, 1]$. Selfish and prosocial B players choose $z = 0$ and $z = 1$ respectively in both games.

In Game 1, $p = 0$ ($p = 1$) is the dominant strategy for selfish (prosocial) player A irrespective of his level of strategic sophitication. Additionally, $L_{k\geq 1}$ prosocial player B and ($L_{k\geq 1}, \pi > \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}$) selfish player B optimally choose $z = 1$, and ($L_{k\geq 1}, \pi < \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}$) selfish player B optimally chooses $z = 0$.[14]

[14]Since π is assumed to have a continuous support, the assumption about the optimal action of selfish B players with $\pi = \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}$ and the assumption about the tie-breaking rule for player B's action generally will not affect optimal strategies of players of higher levels. For this reason, we skip the tie

We now investigate $L_{k\geq 1}$ players' optimal strategies in Game 2. For $L_{k\geq 1}$ prosocial player B, her optimal strategy is always the same: chooses $z = 1$ regardless of the information set she observes.

L_1 player's optimal strategy

L_1 selfish player A's expected payoff function is $p * [\rho_1 v\alpha + v_0^A(1-\alpha)]] + (1-p) * [\rho_2 v\alpha + v_0^A(1-\alpha)]$. So he should optimally choose $p = 0$. L_1 prosocial player A's expected payoff function is $p*[\rho_1 v*\alpha + v_0^A(1-\alpha)] + (1-p)*[(\rho_2 v + \theta_{A2})\alpha + v_0^A(1-\alpha)]$. So he should optimally choose $p = 1$. We assume that L_1 player A chooses q_1 and q_2 independently according to a uniform distribution over $[0, 1]$. In short, L_1 player A has the same optimal response as L_0 player A.

For L_1 selfish player B with viewpoint π, her prior belief about state 1 is π $(= \pi * 1 + (1-\pi) * 0)$. Her posterior beliefs about state 1 after observing (q_1, q_2, b) and (q_1, q_2, w) are $\hat{p}_b = \frac{\pi q_1}{\pi q_1 + (1-\pi) q_2}$ and $\hat{p}_w = \frac{\pi(1-q_1)}{\pi(1-q_1)+(1-\pi)(1-q_2)}$ respectively. Similar to the proof of Proposition 1, the threshold value of her posterior belief about state 1 is still $\mu^* = \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}$. Therefore, conditional on (q_1, q_2), when observing a black ball those selfish B players with viewpoint $\pi > \frac{\rho_0^B - (1-\rho_2)}{(1-\rho_1-\rho_0^B)\frac{q_1}{q_2} + \rho_0^B - (1-\rho_2)} \equiv \pi_b$ invest, and when observing a white ball those selfish B players with viewpoint $\pi > \frac{\rho_0^B - (1-\rho_2)}{(1-\rho_1-\rho_0^B)\frac{1-q_1}{1-q_2} + \rho_0^B - (1-\rho_2)} \equiv \pi_w$ invest. From the perspective of L_1 selfish player B with viewpoint π, when observing a black ball she invests if $\frac{q_1}{q_2} > \frac{\rho_0^B - (1-\rho_2)}{1-\rho_1-\rho_0^B} \frac{1-\pi}{\pi}$, and when observing a black ball she invests if $\frac{1-q_1}{1-q_2} > \frac{\rho_0^B - (1-\rho_2)}{1-\rho_1-\rho_0^B} \frac{1-\pi}{\pi}$.[15]

L_2 player's optimal strategy

L_2 selfish player A's expected payoff function from choosing $p = 1$ is $\rho_1 v\alpha + (1-\alpha)q_1[\rho_1 v Pr(\pi > \pi_b) + v_0^A Pr(\pi < \pi_b)] + (1-\alpha)(1-q_1)[\rho_1 v Pr(\pi > \pi_w) + v_0^A Pr(\pi < \pi_w)] =$

case in all the analysis below.

[15]Note that the prediction remains the same as long as L_0 player A chooses q_1 and q_2 independently according to a distribution with a support of $[0, 1]$. We can make a similar assumption about L_1 player A. The assumption of uniform distribution is for the sake of simplicity.

$v_0^A + (\rho_1 v - v_0^A)\alpha + (\rho_1 v - v_0^A)(1-\alpha)[q_1 Pr(\pi > \pi_b) + (1-q_1)Pr(\pi > \pi_w)] \equiv v_0^A + V_1$. His expected payoff from choosing $p = 0$ is $\rho_2 v\alpha + (1-\alpha)q_2[\rho_2 v Pr(\pi > \pi_b) + v_0^A Pr(\pi < \pi_b)] + (1-\alpha)(1-q_2)[\rho_2 v Pr(\pi > \pi_w) + v_0^A Pr(\pi < \pi_w)] = v_0^A + (\rho_2 v - v_0^A)\alpha + (\rho_2 v - v_0^A)(1-\alpha)[q_2 Pr(\pi > \pi_b) + (1-q_2)Pr(\pi > \pi_w)] \equiv v_0^A + V_2$.

Let $S_1 = \{(q_1, q_2) \in [0,1] \times [0,1] : V_1 > V_2\}, S_2 = \{(q_1, q_2) \in [0,1] \times [0,1] : V_1 < V_2\}$, and $S_0 = \{(q_1, q_2) \in [0,1] \times [0,1] : V_1 = V_2\}$. To solve player A's optimization problem , we first solve his optimization problem in the following four constraint sets about (q_1, q_2): (1) S_1 and $q_1 \geq q_2$, (2) S_1 and $q_1 \leq q_2$, (3) S_2 , (4) S_0.

(1) S_1 and $q_1 \geq q_2$.

In this constraint set, he should optimally choose $p = 1$. His objective then is to choose (q_1, q_2) to maximizes V_1 in the constraint set $S_1 \cap \{q_1 \geq q_2\}$. We apply the following method to solve this problem: maximize V_1 with the constraint $q_1 \geq q_2$ and then verify that the solution lies in S_1.

When $q_1 \geq q_2$, $\frac{q_1}{q_2} \geq 1 \geq \frac{1-q_1}{1-q_2}$. In this case, $\pi_b \leq \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1} \leq \pi_w$. $Pr(\pi > \pi_w) \leq Pr(\pi > \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}) \leq Pr(\pi > \pi_b)$. Note that $Pr(\pi > \pi_b) > 0$ because $\overline{\pi} \geq 1 - (\frac{1-\rho_1-\rho_0^B}{\rho_2-\rho_1})^2 > \frac{\rho_0^B - (1-\rho_2)}{\rho_2 - \rho_1}$. So $V_1 \leq (\rho_1 v - v_0^A)\alpha + (\rho_1 v - v_0^A)(1-\alpha)Pr(\pi > \pi_b)$, and the equality holds only when $q_1 = q_2$ or $q_1 = 1$. Additionally, $Pr(\pi > \pi_b)$ achieves the highest value of 1 only when $\frac{q_1}{q_2} = \infty$. Therefore, V_1 is maximized at $(q_1 = 1, q_2 = 0)$ with the value of $\rho_1 v - v_0^A$. Note that at $(q_1 = 1, q_2 = 0)$, $V_2 = (\rho_2 v - v_0^A)\alpha < V_1 = \rho_1 v - v_0^A$ since $\alpha < \frac{\rho_1 - \rho_0^A}{\rho_2 - \rho_0^A}$.

(2) S_1 and $q_1 \leq q_2$.

In this constraint set, he should optimally choose $p = 1$. His objective then is to choose (q_1, q_2) in the constraint set such that it maximizes V_1. We apply a similar method as in case (1) and establish that V_1 is maximized at $(q_1 = 0, q_2 = 1)$ with the value of $\rho_1 v - v_0^A$. Similarly, at $(q_1 = 0, q_2 = 1)$, $V_2 = (\rho_2 v - v_0^A)\alpha < V_1 = \rho_1 v - v_0^A$ since $\alpha < \frac{\rho_1 - \rho_0^A}{\rho_2 - \rho_0^A}$.

(3) S_2.

In this constraint set, he should optimally choose $p = 0$. His objective then is to choose (q_1, q_2) in the constraint set such that it maximizes V_2. Recall that π is uniformly distributed over the interval $(0, \overline{\pi})$.

We first consider the case of $max\{\pi_b, \pi_w\} \leq \overline{\pi}$. In this case, $\frac{q_1}{q_2} \geq \frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}\frac{1-\overline{\pi}}{\overline{\pi}}$ and $\frac{1-q_1}{1-q_2} \geq \frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}\frac{1-\overline{\pi}}{\overline{\pi}}$, where $\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}\frac{1-\overline{\pi}}{\overline{\pi}} < 1$ because $\overline{\pi} > \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$. Then,

$$\begin{aligned} V_2 &= (\rho_2 v - v_0^A)\alpha + (\rho_2 v - v_0^A)(1-\alpha)[q_2 * (1 - \frac{\pi_b}{\overline{\pi}}) + (1-q_2) * (1 - \frac{\pi_w}{\overline{\pi}})] \\ &= (\rho_2 v - v_0^A) - \frac{(\rho_2 v - v_0^A)(1-\alpha)}{\overline{\pi}}[\frac{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}q_2}{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} + \frac{q_1}{q_2}} + \frac{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}(1-q_2)}{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} + \frac{1-q_1}{1-q_2}}] \end{aligned}$$

Let $f(q_1, q_2) = \frac{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}q_2}{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}+\frac{q_1}{q_2}} + \frac{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}(1-q_2)}{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}+\frac{1-q_1}{1-q_2}}$. When $q_1 = q_2$, $f(q_1, q_2) = \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$.[16] Then,

$$\begin{aligned} f(q_1, q_2) - f(q_1, q_1) &= \frac{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}q_2}{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} + \frac{q_1}{q_2}} + \frac{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}(1-q_2)}{\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} + \frac{1-q_1}{1-q_2}} - \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1} \\ &= \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1} * \frac{1}{(\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} + \frac{q_1}{q_2})(\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} + \frac{1-q_1}{1-q_2})q_2(1-q_2)} * F \end{aligned}$$

where $F = \{\frac{\rho_2-\rho_1}{1-\rho_1-\rho_0^B}[\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} * q_2^2(1-q_2) + (1-q_1)q_2^2] + \frac{\rho_2-\rho_1}{1-\rho_1-\rho_0^B}[\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} * q_2(1-q_2)^2 + (1-q_2)^2 q_1] - q_2(1-q_2) * (\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} + \frac{q_1}{q_2})(\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B} + \frac{1-q_1}{1-q_2})\}$. We then show $F = (q_1 - q_2)^2$

[16] When $q_1 = q_2 \neq 0$ or 1 the result can be attained through working on the formula about $f(q_1, q_2)$; when $q_1 = q_2 = 0$ or 1 we can get the same result by tracing back to the definition about V_2.

below:

$$
\begin{aligned}
F &= \frac{\rho_2-\rho_1}{1-\rho_1-\rho_0^B}\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}q_2(1-q_2)+\frac{\rho_2-\rho_1}{1-\rho_1-\rho_0^B}(q_2^2+q_1-2q_1q_2)\\
&\quad -q_2(1-q_2)[(\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B})^2+\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}\frac{1-q_1}{1-q_2}+\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}\frac{q_1}{q_2}+\frac{q_1}{q_2}\frac{1-q_1}{1-q_2}]\\
&= \frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}(q_2-q_2^2)+\frac{\rho_2-\rho_1}{1-\rho_1-\rho_0^B}(q_2^2+q_1-2q_1q_2)\\
&\quad -\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}(q_2-q_1q_2)-\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}(q_1-q_1q_2)-(q_1-q_1^2)\\
&= q_2^2-\frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}(q_1-2q_1q_2)+\frac{\rho_2-\rho_1}{1-\rho_1-\rho_0^B}(q_1-2q_1q_2)-(q_1-q_1^2)\\
&= (q_1-q_2)^2.
\end{aligned}
$$

Therefore, $f(q_1,q_2)-f(q_1,q_1)\geq 0$, where the equality holds only when $q_1=q_2$. Since $f(q_1,q_2)$ is minimized at $q_1=q_2$, V_2 is maximized at $q_1=q_2$ with a value of $\rho_2 v - v_0^A - \frac{(\rho_2 v-v_0^A)(1-\alpha)}{\overline{\pi}} * \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$.

Next, we consider the case of $\pi_b > \overline{\pi}$. In this case, $\frac{q_1}{q_2} < \frac{\rho_0^B-(1-\rho_2)}{1-\rho_1-\rho_0^B}\frac{1-\overline{\pi}}{\overline{\pi}} < 1$, which implies that $\frac{1-q_1}{1-q_2}>1$ and in turn $\pi_w < \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1} < \overline{\pi}$. Then,

$$
\begin{aligned}
V_2 &= (\rho_2 v-v_0^A)\alpha+(\rho_2 v-v_0^A)(1-\alpha)[q_2*0+(1-q_2)*(1-\frac{\pi_w}{\overline{\pi}})]\\
&= (\rho_2 v-v_0^A)\alpha+\frac{(\rho_2 v-v_0^A)(1-\alpha)}{\overline{\pi}}\{(1-q_2)\overline{\pi}-\frac{(\rho_0^B-(1-\rho_2))(1-q_2)}{(1-\rho_1-\rho_0^B)\frac{1-q_1}{1-q_2}+\rho_0^B-(1-\rho_2)}\}\\
&\leq (\rho_2 v-v_0^A)\alpha\\
&\quad +\frac{(\rho_2 v-v_0^A)(1-\alpha)}{\overline{\pi}}*\{(1-q_2)\overline{\pi}-\frac{(\rho_0^B-(1-\rho_2))*(1-q_2)^2}{1-\rho_1-\rho_0^B+(\rho_0^B-(1-\rho_2))*(1-q_2)}\}
\end{aligned}
$$

where the equality holds if $1-q_1=1$. Let $x=1-q_2$ and we refer to the term in the curly bracket as $g(x)=\overline{\pi}x-\frac{(\rho_0^B-(1-\rho_2))x^2}{1-\rho_1-\rho_0^B+(\rho_0^B-(1-\rho_2))x}$.

Given $x \in [0,1]$, we show that $g(x)$ is maximized at $x = 1$. Note that

$$g'(x) = \overline{\pi} - [\frac{(\rho_0^B - (1-\rho_2))x}{1-\rho_1-\rho_0^B+(\rho_0^B-(1-\rho_2))x} + \frac{(1-\rho_1-\rho_0^B)(\rho_0^B-(1-\rho_2))x}{(1-\rho_1-\rho_0^B+(\rho_0^B-(1-\rho_2))x)^2}]$$

which is strictly decreasing in x. Then, $g'(x)$ is minimized at $x = 1$ with a value of $\overline{\pi} - [1 - (\frac{1-\rho_1-\rho_0^B}{1-\rho_1-\rho_0^B+\rho_0^B-(1-\rho_2)})^2] = \overline{\pi} - [1 - (\frac{1-\rho_1-\rho_0^B}{\rho_2-\rho_1})^2]$. Since $\overline{\pi} \geq 1 - (\frac{1-\rho_1-\rho_0^B}{\rho_2-\rho_1})^2$, $g'(x) \geq 0$ with the equality holding at $x = 1$. Therefore, $g(x)$ is maximized at $x = 1$ with a value of $\overline{\pi} - \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$. This shows that in the case of $\pi_b > \overline{\pi}$, $V_2 < \rho_2 v - v_0^A - \frac{(\rho_2 v - v_0^A)(1-\alpha)}{\overline{\pi}} * \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$ because $\frac{1-q_1}{1-q_2} > 1$. Similarly, we show that $V_2 < \rho_2 v - v_0^A - \frac{(\rho_2 v - v_0^A)(1-\alpha)}{\overline{\pi}} * \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$ in the case of $\pi_w > \overline{\pi}$.

This shows V_2 is maximized at $q_1 = q_2$ with a value of $\rho_2 v - v_0^A - \frac{(\rho_2 v - v_0^A)(1-\alpha)}{\overline{\pi}} * \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$. Note that at $q_1 = q_2$, $V_1 = \rho_1 v - v_0^A - (\rho_1 v - v_0^A)(1-\alpha)\frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}\frac{1}{\overline{\pi}} < \rho_2 v - v_0^A - \frac{(\rho_2 v - v_0^A)(1-\alpha)}{\overline{\pi}} * \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1} = V_2$ since $\overline{\pi} \geq 1 - (\frac{1-\rho_1-\rho_0^B}{\rho_2-\rho_1})^2 > \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1}$.

(4) S_0.

In this constraint set, we show that $V_1 = V_2$ implies that $V_1 < \rho_1 v - v_0^A$. We have shown that V_1 achieves the highest value $\rho_1 v - v_0^A$ only when $(q_1, q_2) = (1, 0)$ or when $(q_1, q_2) = (0, 1)$. In both cases, $V_2 = (\rho_2 v - v_0^A)\alpha < V_1$ since $\alpha < \frac{\rho_1-\rho_0^A}{\rho_2-\rho_0^A}$. This shows that if $V_1 = V_2$, then $V_1 \neq \rho_1 v - v_0^A$, i.e., $V_1 < \rho_1 v - v_0^A$. So player A's problem is not maximized in the set S_0.

Therefore, the solution to player A's optimization problem comes from cases (1)-(3). Specifically, he should optimally choose $(p = 0, q_1 = q_2)$ if $\rho_2 v - v_0^A - \frac{(\rho_2 v - v_0^A)(1-\alpha)}{\overline{\pi}} * \frac{\rho_0^B-(1-\rho_2)}{\rho_2-\rho_1} > \rho_1 v - v_0^A$, i.e., if $\overline{\pi} > (1-\alpha)\frac{\rho_2-\rho_0^A}{\rho_2-\rho_1}\frac{\rho_{0B}-(1-\rho_2)}{\rho_2-\rho_1}$. Note that $1 - (\frac{1-\rho_1-\rho_0^B}{\rho_2-\rho_1})^2 > (1-\alpha)\frac{\rho_2-\rho_0^A}{\rho_2-\rho_1}\frac{\rho_{0B}-(1-\rho_2)}{\rho_2-\rho_1}$ if and only if $\rho_2 - \rho_0^B + 1 - 2\rho_1 > (1-\alpha)(\rho_2 - \rho_0^A)$. Therefore, he should optimally choose $(p = 0, q_1 = q_2)$ if $\rho_2 - \rho_0^B + 1 - 2\rho_1 > (1-\alpha)(\rho_2 - \rho_0^A)$ or if $\rho_2 - \rho_0^B + 1 - 2\rho_1 \leq (1-\alpha)(\rho_2 - \rho_0^A)$ and $\overline{\pi} > (1-\alpha)\frac{\rho_2-\rho_0^A}{\rho_2-\rho_1}\frac{\rho_{0B}-(1-\rho_2)}{\rho_2-\rho_1}$. He should optimally choose $(p = 1, q_1 = 1, q_2 = 0)$ or $(p = 1, q_1 = 0, q_2 = 1)$ if $\overline{\pi} < (1-\alpha)\frac{\rho_2-\rho_0^A}{\rho_2-\rho_1}\frac{\rho_{0B}-(1-\rho_2)}{\rho_2-\rho_1}$.

L_2 prosocial player A's expected payoff from choosing $p = 1$ is the same as L_2 selfish player A's expected payoff from choosing $p = 1$. Thus, it is maximized at $(q_1, q_2) = (1, 0)$ or $(q_1, q_2) = (0, 1)$ with a value of $\rho_1 v$. His expected payoff from choosing $p = 0$ is no more than $\rho_2 v + \theta_{A2}$ $(< \rho_1 v)$. Therefore, he should optimally choose $p = 1$ and $(q_1, q_2) = (1, 0)$ or $(q_1, q_2) = (0, 1)$.

L_2 selfish player B with viewpoint π has the same optimal response as L_1 selfish player B with viewpoint π because L_1 player A has the same optimal response as L_0 player A.

L_3 player's optimal strategy

L_3 player A has the same optimal response as L_2 player A because L_2 player B has the same optimal response as L_1 player B.

Now consider L_3 selfish player B with viewpoint π. According to her belief, L_2 selfish player A optimally chooses $(q_1 = q_2, p = 0)$ and L_2 prosocial player A optimally chooses $(q_1 = 1, q_2 = 0, p = 1)$ or $(q_1 = 0, q_2 = 1, p = 1)$. So $Pr(E|A1) = Pr(E^c|A2) = 0$ and $Pr(E|A2) = Pr(E^c|A1) = 1$, i.e., $p_{A1} = 0$ and $p_{A2} = 1$, and in turn her prior belief about state 1 is $Pr(E) = \pi * 1 + (1 - \pi) * 0 = \pi$.[17] Then, consider all possible Hs she may observe.

(1) $H = (q_1 = 1, q_2 = 0, b)$ or $(q_1 = 0, q_2 = 1, w)$. Consider only the case of $H = (q_1 = 1, q_2 = 0, b)$. In this case, $Pr(q_1, q_2|A1) = Pr(s|E^c, q_1, q_2) = 0$ and $Pr(q_1, q_2|A2) = Pr(s|E, q_1, q_2) = 1$. Thus, $Pr(H) = \pi$ and $Pr(E|H) = \frac{\pi}{\pi} = 1$. So she should optimally invest. Similarly, she should optimally invest in information set $H = (q_1 = 0, q_2 = 1, w)$.

(2) $H = (q_1 = 1, q_2 = 0, w)$ or $(q_1 = 0, q_2 = 1, b)$. In this case, $Pr(H) = 0$ and Bayes' rule is not defined. Only a posterior belief of zero is consistent with the disclosed

[17] If $\overline{\pi} < (1 - \alpha)\frac{\rho_2 - \rho_0^A}{\rho_2 - \rho_1}\frac{\rho_{0B} - (1 - \rho_2)}{\rho_2 - \rho_1}$, both prosocial and selfish player A of level 2 optimally choose $(p = 1, q_1 = 1, q_2 = 0)$ or $(p = 1, q_1 = 0, q_2 = 1)$. In this case, L_3 selfish player B's prior belief about state 1 becomes one and she should observe $(q_1 = 1, q_2 = 0)$ or $(q_1 = 0, q_2 = 1)$. A similar analysis suggests that her optimal strategy remains the same: invests only when $(0 < q_1 \leq 1, q_2 = 0, b)$ or $(0 \leq q_1 < 1, q_2 = 1, w)$.

information and in turn under the assumption about conditionally pessimistic posterior belief she should optimally not invest.

(3) $H = (0 < q_1 = q_2 < 1, b/w)$, $(q_1 = q_2 = 0, w)$ or $(q_1 = q_2 = 1, b)$. In this information $Pr(E|H) = 0$ and she should optimally not invest.

(4) $H = (0 < q_2 \neq q_1 < 1, b/w)$, $(0 = q_2 < q_1 < 1, w)$, $(0 < q_1 < q_2 = 1, b)$, $(0 < q_2 < q_1 = 1, b)$ or $(q_1 = 0 < q_2 < 1, w)$. In this case, $Pr(H) = 0$ and Bayes' rule is not defined. Any posterior belief of from zero to one is consistent with the disclosed information and in turn under the assumption about conditionally pessimistic posterior belief she should optimally not invest.

(5) $H = (0 < q_2 < q_1 = 1, w)$ or $(q_1 = 0 < q_2 < 1, b)$. In this case, $Pr(H) = 0$ and Bayes' rule is not defined. Only a posterior belief of zero is consistent with the disclosed information and in turn under the assumption about conditionally pessimistic posterior belief she should optimally not invest.

(6) $H = (0 = q_2 < q_1 < 1, b)$ or $(0 < q_1 < q_2 = 1, w)$. In this case, $Pr(H) = 0$ and Bayes' rule is not defined. Only a posterior belief of one is consistent with the disclosed information and in turn under the assumption about conditionally pessimistic posterior belief she should optimally invest.

To sum up, L_3 selfish player B with viewpoint π should optimally invest only when: $(0 < q_1 \leq 1, q_2 = 0, s = b)$ or $(0 \leq q_1 < 1, q_2 = 1, s = w)$.

L_4 player's optimal strategy

L_4 selfish player A's expected payoff from choosing $p = 1$ is $(1-\alpha)q_1 * (\rho_1 v * \mathbf{1}_{0<q_1\leq 1, q_2=0} + v_0^A(1-\mathbf{1}_{0<q_1\leq 1, q_2=0})) + (1-\alpha)(1-q_1)*(\rho_1 v * \mathbf{1}_{0\leq q_1<1, q_2=1} + v_0^A(1-\mathbf{1}_{0\leq q_1<1, q_2=1})) + \rho_1 v\alpha$, which is maximized at $(q_1 = 1, q_2 = 0)$ or $(q_1 = 0, q_2 = 1)$ with a value of $\rho_1 v$. His expected payoff from choosing $p = 0$ is $v_0^A(1-\alpha) + \rho_2 v\alpha$, which is less than $\rho_1 v$ since $\alpha < \frac{\rho_1 - \rho_0^A}{\rho_2 - \rho_0^A}$. Therefore, his optimal choice is $(p = 1, q_1 = 1, q_2 = 0)$

or $(p = 1, q_1 = 0, q_2 = 1)$. For L_4 prosocial player A, his optimal choice is also $(p = 1, q_1 = 1, q_2 = 0)$ or $(p = 1, q_1 = 0, q_2 = 1)$.

Since L_3 player A has the same optimal response as L_2 player A, L_4 selfish player B has the same optimal strategy as L_3 selfish player B: invest only when $(0 < q_1 \leq 1, q_2 = 0, b)$ or $(0 \leq q_1 < 1, q_2 = 1, w)$.

It is straightforward to establish that $L_{k>4}$ player has the same optimal strategy as L_4 player.

C.3 Additional Data Analysis

C.3.1 Change in the Payoff of Player A

We do not observe significant difference in player A's average payoff between games. However, under state 1, the average payoff increases significantly from 11.75 to 13.73 (p-value<0.001, paired permutation test on the averages at the session level), while under state 2, the average payoff decreases significantly from 13.98 to 13.05 (p-value=0.023, paired permutation test on the averages at the session level). This observation indicates that introducing information design benefits only those A players who choose to be trustworthy.

We also look at trusting acts across states and games. The numbers of rounds with state 1 being realized in Game 1 and Game 2 are 191 and 351, respectively. Among these rounds, the numbers of trusting acts are 67 and 262, respectively (frequency: 35.1% and 74.6%). The increase in the frequency of trusting acts from Game 1 to Game 2 accounts for player A's payoff increase under state 1. The numbers of rounds with state 2 being realized in Game 1 and Game 2 are 609 and 449, respectively. Among these rounds, the numbers of trusting acts are 202 and 114, respectively (frequency: 33.2% and 25.4%).

The decrease in the frequency of trusting acts from Game 1 to Game 2 accounts for player A's payoff decrease under state 2. Note also that the frequency of trusting acts in Game 1 is similar between under state 1 and state 2, which is expected due to player A's no capability to signal trustworthiness in Game 1. In addition, the frequency in Game 2 is substantially larger under state 1 than that under state 2, which is expected due to player A's full capability to signal trustworthiness in Game 2.

C.3.2 Treatment Effects (Round 1 Data Only)

Table C.3.1 shows that when we consider round 1 data of each game, subsample 1 of round 1 data, in which Game 1 is played first, or subsample 2 of round 1 data, in which Game 2 is played first, the treatment effects are qualitatively consistent with the equilibrium predictions.

Table C.3.1: Treatment Effects (Round 1 Data Only)

	Game 1	Game 2	Difference
Trustworthiness (p)	0.28	0.46	0.18***
	0	$\geq \frac{1}{7}$	
	0.32	0.45	0.13**
	0.23	0.46	0.23***
Trust (z)	0.34	0.48	0.14***
	0	$\geq \frac{1}{2}$	
	0.43	0.50	0.07
	0.25	0.45	0.20***

*Notes: For each variable, rows 1-4 represent the data of round 1, equilibrium prediction, subsample 1 of round 1 data and subsample 2 of round 1 data, in which Game 1 is played first and second, respectively. The used tests are Wilcoxon signed rank test for p and paired permutation test for z. *, **and *** indicate significance at the 10%, 5% and 1% levels, respectively.*

The average aggregate payoff increases from 23.38 to 24.8 (p-value=0.013, paired permutation test on the averages at the session level). As for player A, the average payoff increases from 13.35 to 13.43 (p-value=0.619, Wilcoxon signed rank test). We

then check the change in player A's payoff conditional on the realized state. Under state 1, the average payoff increases significantly from 11.82 to 13.82 (p-value<0.001, paired permutation test on the averages at the session level), while under state 2, the average payoff decreases from 13.93 to 13.13 (p-value=0.321, paired permutation test on the averages at the session level). As for player B, the average payoff increases from 10.03 to 11.33 (p-value<0.001, Wilcoxon signed rank test).

Figure C.1 presents the data in a disaggregated form: the distributions of subjects' choices of p and z and payoff in the first round of each game.

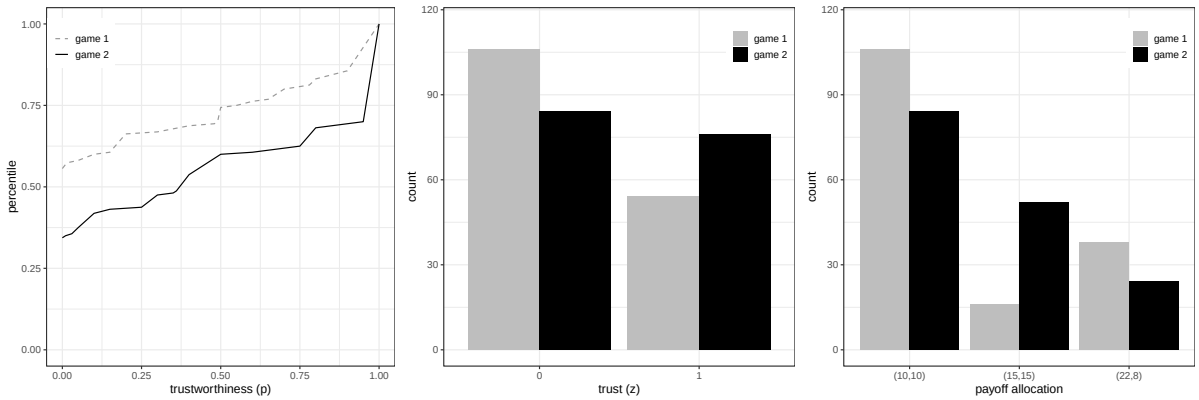


Figure C.1: Distributions of p, z and payoff in Round 1

C.3.3 Distribution of Information Sets

Since information sets that a subject observes in Game 2 are determined by her opponent, it is expected that the distribution of four categories of information sets that subjects observe in Game 2 is similar for the "trusting group" and "no trusting group", which is indeed the case according to Table C.3.2.

Table C.3.2: Distribution of Information Sets for the "trusting/no trusting group" in Game 1

		H_1	H_2	H_3	H_4
"Trusting group" in Game 1	count	123	12	51	64
	expected	113.75	15.31	49.06	71.87
"No trusting group" in Game 1	count	241	37	106	166
	expected	250.25	33.69	107.94	158.13

Notes: H_1 (H_2), unfavorable signal and information structure with a low (high) index; H_3 (H_4), favorable signal and information structure with a low (high) index. Count in each cell reports the number of the corresponding information sets that are observed. "Trusting/no trusting group": subjects who invest in no less than/less than three rounds of Game 1. Person's chi-squared test: p-value= 0.32; Fisher's exact test: p-value= 0.33.

C.3.4 Estimation Procedure

We assume that each subject of a certain type follows the predicted action subject to standard logistic errors with precision λ. The subject's choice approaches uniform randomness as $\lambda \to 0$ and approaches the predicted action as $\lambda \to \infty$. Let y index a generic value of parameters related to prosociality and strategic sophistication. Player A's type and player B's type can be represented by y and (y, π), respectively. Let Ω index the action space.[18] Let $V_{(y,g)}(c_g)$ index player A's expected payoff from choosing action c_g in Game g, and let $V_{(y,\pi,g)}(c_g|H)$ index player B's expected payoff from choosing action c_g when observing an information set H in Game g.[19] For a certain type, the choice probability or density is specified below.

$$P^A_{(y,g)}(c_g) = \frac{exp[\lambda V_{(y,g)}(c_g)]}{\int_\Omega exp[\lambda V_{(y,g)}(\tilde{c})]\, d\tilde{c}} \text{ , for player A.} \tag{C.1}$$

$$P^B_{(y,\pi,g)}(c_g|H) = \frac{exp[\lambda V_{(y,\pi,g)}(c_g|H)]}{\sum_\Omega exp[\lambda V_{(y,\pi,g)}(\tilde{c}|H)]} \text{ , for player B.} \tag{C.2}$$

Let c^i (or $c^i|H^i$) index subject i's sample, which consists of ten observations, where $c^i = \{c^i_{gr}\}_{g\in\{1,2\},r\in\{1,\cdots,5\}}$ and $H^i = \{H^i_{gr}\}_{g\in\{1,2\},r\in\{1,\cdots,5\}}$. With the assumption that a

[18] For player A, $\Omega = \{p : p \in [0,1]\}$ in Game 1 and $\Omega = \{(p, q_1, q_2) : p \in [0,1], q_1 \in [0,1], q_2 \in [0,1]\}$ in Game 2. For player B, $\Omega = \{1, 0\}$ in both games. $H \in \emptyset$ in Game 1, and $H = (q_1, q_2, s) \in [0,1] \times [0,1] \times \{b, w\}/\{(1,1,w), (0,0,b)\}$ in Game 2.

[19] The expected payoff function of each type is presented in Table C.3.3.

subject's errors are independent across rounds, the likelihood of observing a subject's sample conditional on his/her type can be specified as:

$$L_y^A(c^i) = \prod_{g=1}^{2}\prod_{r=1}^{5} P_{(y,g)}^A(c_{gr}^i) = \prod_{g=1}^{2}\prod_{r=1}^{5} \frac{exp[\lambda V_{(y,g)}(c_{gr}^i)]}{\int_\Omega exp[\lambda V_{(y,g)}(\tilde{c})]\, d\tilde{c}}, \tag{C.3}$$

$$L_{(y,\pi)}^B(c^i|H^i) = \prod_{g=1}^{2}\prod_{r=1}^{5} P_{(y,\pi,g)}^B(c_{gr}^i|H^i) = \prod_{g=1}^{2}\prod_{r=1}^{5} \frac{exp[\lambda V_{(y,\pi,g)}(c_{gr}^i|H^i)]}{\sum_\Omega exp[\lambda V_{(y,\pi,g)}(\tilde{c}|H^i)]}. \tag{C.4}$$

Let β_y be the prior probability of player A being type y, and let $\beta_{(y,\pi)}$ be the prior probability density of player B being type (y,π). According to the type classification based on the expected payoff function, player A may be classified as one of six categories and player B with viewpoint π may be classified as one of four categories. Thus, we have $\sum_{y=1}^{6}\beta_y = 1$ and $\sum_{y=1}^{4}\int_0^{\bar{\pi}}\beta_{(y,\pi)}\, d\pi = 1$. Then, the likelihood of observing a subject's sample unconditional on type can be formulated as: $L^A(c^i) = \sum_{y=1}^{6}\beta_y L_y^A(c^i)$ and $L^B(c^i|H^i) = \sum_{y=1}^{4}\int_0^{\bar{\pi}}\beta_{(y,\pi)}L_{(y,\pi)}^B(c^i|H^i)\, d\pi$.

We assume that the precision parameter λ is subject-specific and game-specific: $\{\lambda_g\}_{g=1}^2$. For each subject, we jointly estimate his/her game-specific λ_g and type probabilities. Since the likelihood function $L^A(c^i)$ is linear in β_y, the maximum likelihood estimate of player A's type probabilities sets $\beta_y = 1$ for the (generically unique) y that yields the highest $L_y^A(c^i)$, which is obtained by maximizing $L_y^A(c^i)$ over $\{\lambda_g\}_{g=1}^2$ given type y. For a similar reason, the maximum likelihood estimate of player B's type density assigns (y,π) a probability of one for the (generically unique) (y,π) that yields the highest $L_{(y,\pi)}^B(c^i|H^i)$, which is obtained by maximizing $L_{(y,\pi)}^B(c^i|H^i)$ over $\{\lambda_g\}_{g=1}^2$ given type (y,π).

How we use the estimation procedure

It is a prerequisite to specify the values of background parameters $(\theta_{A2}, \theta_{B2}, \alpha, \bar{\pi})$ before conducting MLE. Given the values of monetary payoff parameters in our experiment,

the behavioral model only requires that $\theta_{A2} < -7$, $\theta_{B2} > 2$, $\alpha < \frac{5}{12}$ and $\bar{\pi} > \frac{24}{49}$. We assume that the background parameters are role specific and apply the spirit of MLE to pick their values. Specifically, for each value of $(\theta_{A2}, \theta_{B2}, \alpha, \bar{\pi})$ taken from a range, we estimate each subject's type and precision parameters using MLE. Then, we sum up the obtained likelihoods over all A players (B players), and the value of $(\theta_{A2}, \theta_{B2}, \alpha, \bar{\pi})$ that maximizes the sum of likelihoods is chosen as player A's (player B's) background parameters.[20]

In practice, our estimation procedure proceeds as follows. We first specify a discretized range of background parameters, e.g., $(\theta_{A2}, \theta_{B2}, \alpha, \bar{\pi}) \in \{-8, -7.5\} \times \{2.5, 3\} \times \{0.1, 0.2, 0.3, 0.4\} \times \{0.5, 0.6, 0.7, 0.8, 0.9\}$. Then for each $(\theta_{A2}, \theta_{B2}, \alpha, \bar{\pi})$ taken from this range, we conduct the subject-by-subject analysis and maximize $lnL^A(c^i)$ and $lnL^B(c^i|H^i)$ over type and precision parameter. The value of $(\theta_{A2}, \theta_{B2}, \alpha, \bar{\pi})$ is picked for A players and B players respectively in a maximum likelihood fashion. Finally, each subject's type is estimated through MLE given the picked value of $(\theta_{A2}, \theta_{B2}, \alpha, \bar{\pi})$.

The procedure pins down $(\theta_{A2} = -7.5, \theta_{B2} = 2.5, \alpha = 0.3, \bar{\pi} = 0.6)$ for player A and $(\theta_{A2} = -7.5, \theta_{B2} = 2.5, \alpha = 0.3, \bar{\pi} = 0.5)$ for player B.[21] Note that θ_{B2} does not affect player A's type estimation, and θ_{A2} and α do not affect player B's type estimation because they do not enter into the corresponding player's expected payoff function.

Given the selected value of background parameters $(\theta_{A2}, \theta_{B2}, \alpha, \bar{\pi})$, we determine each subject's type in the following way. For subject i as player A, we find a $(\lambda_1^*, \lambda_2^*) \in [0, 100] \times [0, 100]$ that maximizes the likelihood $L_y^A(c^i)$ and then find a y^* that maximizes

[20] Alternatively, one may want to use a common set of values of background parameters for both players. The issue is that compared to player B, player A's action space is large and thus have extremely large absolute log-likelihood value, which in turn dominates the procedure of specifying background parameters. In other words, the value of $(\theta_{A2}, \theta_{B2}, \alpha, \bar{\pi})$ that maximizes the sum of likelihoods over all A players also maximizes the sum of likelihoods over all A players and all B players.

[21] As a robust check, we also vary the values of background parameters from the picked one. We find that the proportions of types remain similar, which indicates that a subject's estimated type roughly remains unchanged when the values of background parameters deviate from the picked ones. The details of the robust check are available upon request.

the maximum of the likelihood, that is, $y^* \in \text{argmax}_y \; max_{(\lambda_1,\lambda_2)} \, L_y^A(c^i)$. For subject j as player B, we find a $(\lambda_1^*, \lambda_2^*, \pi^*) \in [0,100]\times[0,100]\times[0.01,\bar{\pi}]$ that maximizes the likelihood $L^B_{(y,\pi)}(c^j|H^j)$ and then find a y^* that maximizes the maximum of the likelihood, that is, $y^* \in \text{argmax}_y \; max_{(\lambda_1,\lambda_2,\pi)} \, L^B_{(y,\pi)}(c^j|H^j)$. When the maximizer for a subject, y^* or (y^*,π^*), is unique, the subject is assigned a unique type; when the maximizer for a subject admits multiple values, the subject is assigned multiple types.

Table C.3.3: Expected Payoff Function of Different Types

Role	Type	Game 1	Game 2
A	(P, L_1)	$10 + [5p + (12+\theta_{A2})(1-p)] * \alpha$	$10 + [5p + (12+\theta_{A2})(1-p)] * \alpha$
	$(P, L_2/L_3)$	$10 + [5p + (12+\theta_{A2})(1-p)] * [\alpha + (1-\alpha)P_r(\pi > \frac{2}{7})]$	$10 + 5p\{\alpha + (1-\alpha)[q_1 P_r(\pi > \frac{0.4}{0.4+\frac{q_1}{q_2}}) + (1-q_1)P_r(\pi > \frac{0.4}{0.4+\frac{1-q_1}{1-q_2}})]\}$ $+(12+\theta_{A2})(1-p)\{\alpha + (1-\alpha)[q_2 P_r(\pi > \frac{0.4}{0.4+\frac{q_1}{q_2}}) + (1-q_2)P_r(\pi > \frac{0.4}{0.4+\frac{1-q_1}{1-q_2}})]\}$
	$(P, L_{k\geq 4})$	$10 + [5p + (12+\theta_{A2})(1-p)] * [\alpha + (1-\alpha)P_r(\pi > \frac{2}{7})]$	$10 + 5p[\alpha + (1-\alpha)q_1] + (12+\theta_{A2})(1-p)\alpha$ if $(q_1,q_2) \in \phi_1$; $10 + 5p[\alpha + (1-\alpha)(1-q_1)] + (12+\theta_{A2})(1-p)\alpha$ if $(q_1,q_2) \in \phi_2$; $10 + [5p + (12+\theta_{A2})(1-p)] * \alpha$ otherwise
	(S, L_1)	$10 + (12-7p) * \alpha$	$10 + (12-7p) * \alpha$
	$(S, L_2/L_3)$	$10 + (12-7p) * [\alpha + (1-\alpha)P_r(\pi > \frac{2}{7})]$	$10 + 5p\{\alpha + (1-\alpha)[q_1 P_r(\pi > \frac{0.4}{0.4+\frac{q_1}{q_2}}) + (1-q_1)P_r(\pi > \frac{0.4}{0.4+\frac{1-q_1}{1-q_2}})]\}$ $+12(1-p)\{\alpha + (1-\alpha)[q_2 P_r(\pi > \frac{0.4}{0.4+\frac{q_1}{q_2}}) + (1-q_2)P_r(\pi > \frac{0.4}{0.4+\frac{1-q_1}{1-q_2}})]\}$
	$(S, L_{k\geq 4})$	$10 + (12-7p) * [\alpha + (1-\alpha)P_r(\pi > \frac{2}{7})]$	$10 + 5p[\alpha + (1-\alpha)q_1] + 12(1-p)\alpha$ if $(q_1,q_2) \in \phi_1$; $10 + 5p[\alpha + (1-\alpha)(1-q_1)] + 12(1-p)\alpha$ if $(q_1,q_2) \in \phi_2$; $10 + (12-7p) * \alpha$ otherwise
B	$(P, L_1/L_2, \pi)$	$8 + \theta_{B2} + 7\pi$ if $z=1$; 10 if $z=0$	$8 + \theta_{B2} + 7\frac{\pi q_1}{\pi q_1 + (1-\pi)q_2}$ if $(s=b, z=1)$; $8 + \theta_{B2} + 7\frac{\pi(1-q_1)}{\pi(1-q_1)+(1-\pi)(1-q_2)}$ if $(s=w, z=1)$; 10 if $z=0$
	$(P, L_{k\geq 3}, \pi)$	$8 + \theta_{B2} + 7\pi$ if $z=1$; 10 if $z=0$	$15 + \theta_{B2}$ if $(q_1,q_2,s) \in \hat{H}$ &$z=1$; $8 + \theta_{B2}$ if $(q_1,q_2,s) \notin \hat{H}$&$z=1$; 10 if $z=0$
	$(S, L_1/L_2, \pi)$	$8 + 7\pi$ if z=1; 10 if z=0	$8 + 7\frac{\pi q_1}{\pi q_1 + (1-\pi)q_2}$ if $(s=b, z=1)$; $8 + 7\frac{\pi(1-q_1)}{\pi(1-q_1)+(1-\pi)(1-q_2)}$ if $(s=w, z=1)$; 10 if $z=0$
	$(S, L_{k\geq 3}, \pi)$	$8 + 7\pi$ if z=1; 10 if z=0	15 if $(q_1,q_2,s) \in \hat{H}$ &$z=1$; 8 if $(q_1,q_2,s) \notin \hat{H}$&$z=1$; 10 if $z=0$

Notes: $\phi_1 \equiv (0 < q_1 \leq 1, q_2 = 0)$; $\phi_2 \equiv (0 \leq q_1 < 1, q_2 = 1)$; $\hat{H} \equiv \{(\phi_1, s=b)\} \cup \{(\phi_2, s=w)\}$.

C.3.5 Causal Mediation Analysis

We follow four basic steps for the mediation analysis. First, we run a probit regression of trusting act on the treatment dummy (i.e., 1 if Game 2 and 0 if Game 1). The

regression result shows that the total treatment effect on trusting act is significantly positive. Second, we run a linear regression of trustworthiness (i.e., the mediator in our setting) on the treatment dummy and verify that the treatment effect on trustworthiness is significant. Third, we run a probit regression of trusting act on the treatment dummy, trustworthiness and their interaction term. The regression result shows that there exists both direct and indirect effects, and the indirect effect depends on the treatment status. Given the results of the latter two regressions, we use the R package *mediation* to conduct the causal mediation analysis. Table C.3.4 shows that the proportion of the indirect effect is 38.9% and correspondingly the proportion of the direct effect is 61.1%.

Table C.3.4: Causal Mediation Analysis

	Estimate	95% CI	p-value
Total Effect	0.148	[0.091, 0.190]	< 0.0001
ACME	0.057	[0.043, 0.070]	< 0.0001
ADE	0.091	[0.035, 0.130]	< 0.001
Prop. Mediated	0.389	[0.279, 0.630]	< 0.0001

Notes: ACME denotes average causal mediation effects, that is, the indirect effect of information design on trusting act that goes through trustworthiness. ADE denotes average direct effects, that is, the direct effect of information design on trusting act. Prop. Mediated describes the proportion of the indirect effect. All the estimated effects are expressed as the increase in the probability of player B's trusting act. Point estimates and 95% confidence intervals are all computed by Bootstrap.

C.4 Experimental Instructions

Welcome to today's experiment of economic decision making.[22] You will earn a considerable amount of money as long as you read experimental instructions carefully and make wise decisions.

In the paid part of the experiment, each participant is assigned a constant role, either player A or player B. Player A is paired with player B to play five rounds of Game 1 and five rounds of Game 2.

In each of ten rounds, each player is paired with an anonymously new opponent and each player has an endowment of 10 tokens. If both player A and player B agree to spend their 10 tokens on an investment project, the project will be conducted and its total payoffs will be 30 tokens. The total payoffs will be allocated between player A and player B according to either plan $(15, 15)$ or plan $(22, 8)$. The payoff table is demonstrated below.

Table C.4.1: Payoffs Table

Both invest	Allocation plan	A's payoff	B's payoff
No	endowment	10	10
Yes	$(15, 15)$	15	15
	$(22, 8)$	22	8

Game 1[23]

In Game 1, player A moves first and decides the chance of plan $(15, 15)$ that will be chosen. Specifically, player A chooses an integer number P from {0,1,2,$\cdots$,99,100 }. If $P = 0$, then the chance of plan $(15, 15)$ is 0% and correspondingly the chance of plan $(22, 8)$ is 100%. In other words, a computer will choose plan $(22, 8)$ to allocate the

[22] The experiment was conducted in Chinese. This is an English translation of the instructions.

[23] In sessions with the trustworthiness design game preceding the reverse trust game, Games 1 and 2 here are labeled as Games 2 and 1, respectively.

payoffs of the project. If $0 < P < 100$, then the chance of plan $(15, 15)$ is $P\%$ and correspondingly the chance of plan $(22, 8)$ is $(100-P)\%$. In other words, a computer will choose plan $(15, 15)$ with the chance of $P\%$ and choose plan $(22, 8)$ with the chance of $(100 - P)\%$, and then the total payoffs of the project will be allocated according to the chosen plan. If $P = 100$, then the chance of plan $(15, 15)$ is 100% and correspondingly the chance of plan $(22, 8)$ is 0%. In other words, a computer will choose plan $(15, 15)$ to allocate the payoffs of the project.

After player A moves and the allocation plan is determined, player B decides whether or not to spend his/her 10 tokens on the project without knowing the allocation plan and player A's choice of P. If player B chooses not to spend 10 tokens on the project, the project will not be conducted and both players keep their endowment of 10 tokens. If player B chooses to spend 10 tokens on the project, player A's 10 tokens will also be spent on the project and both players receive payoffs according to the determined allocation plan.

Game 2

In Game 2, player A moves first and decides the chance of plan $(15, 15)$. Specifically, player A chooses an integer number P from $\{0, 1, 2, \cdots, 99, 100\}$, which will determine the allocation plan in the same way as in Game 1. **In addition, player A decides the composition of black balls and white balls in urn 1 and urn 2**. Specifically, both urns contain 100 balls of either black or white. **Player A chooses an integer number Q_1 from $\{0, 1, 2, \cdots, 99, 100\}$ that determines the number of black balls in urn 1**. In other words, the numbers of black balls and white balls in urn 1 are chosen to be Q_1 and $100 - Q_1$ respectively. **Player A chooses an integer number Q_2 from $\{0, 1, 2, \cdots, 99, 100\}$ that determines the number of black balls in urn 2**. In other words, the numbers of black balls and white balls in urn 2 are chosen to be Q_2 and $100 - Q_2$ respectively.

After player A's move, a computer chooses the allocation plan according to P in the same way as in Game 1. When the determined plan is $(15, 15)$, urn 1 will be used. When the determined plan is $(22, 8)$, urn 2 will be used. A ball is then randomly drawn from the used urn.

Player B is neither told of the allocation plan and the used urn nor told of player A's choice of P, but observes player A's choice of Q_1 and Q_2, and also the color of the drawn ball. Based on this observation, player B decides whether or not to spend his/her 10 tokens on the investment project. If player B chooses not to spend 10 tokens on the project, the project will not be conducted and both players keep their endowment of 10 tokens. If player B chooses to spend 10 tokens on the project, player A's 10 tokens will also be spent on the project and both players receive payoffs according to the determined allocation plan.

Payoffs

Neither of both players receives feedback at the end of each round. In other words, player A is not told whether player B chooses to spend 10 tokens on the project, and player B is not told of the determined allocation plan and player A's choice of P. At the end of the experiment, one round is randomly selected out of ten rounds by the computer as a paid round. Your final payoffs are equal to your earned tokens in the paid round. Each token is equivalent to two Chinese yuan. In addition, you also receive a show-up fee of five Chinese yuan for completing the paid experiment and a six-question questionnaire after the paid experiment.

Practice

To help you become familiar with the two games in the paid experiment, we ask each participant to practice playing four rounds of games before the paid experiment starts. In the first and second rounds, you are assigned the role of player A and a computer is assigned the role of player B to play Game 1 and Game 2 respectively. In the third

and fourth rounds, you are assigned the role of player B and a computer is assigned the role of player A to play Game 1 and Game 2 respectively. Unlike a human opponent in the paid experiment, your computer opponent always makes a random decision in the practice rounds. In other words, your computer opponent as player B always randomly chooses whether or not to spend 10 tokens on the project; and your computer opponent as player A always randomly picks P in Game 1 and randomly picks P, Q_1, Q_2 in Game 2. Another difference in the practice rounds is that you are told of the determined allocation plan and your payoff at the end of each round. Your payoffs in the four practice rounds will not affect your earnings in the paid experiment.

Rules

If you have any questions during the experiment, please raise your hand. During the experiment, you must turn off your cell phone, you are neither allowed to talk with each other nor allowed to leave without permission from the investigator.

Please raise your hand now if you have any questions with the experimental instructions! We will proceed to the practice rounds after answering your questions.

Bibliography

Alaoui, L. & Penta, A. (2015), 'Endogenous Depth of Reasoning', *The Review of Economic Studies* **83**(4), 1297–1333.

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

Ambuehl, S., Ockenfels, A. & Stewart, C. (2022), 'Who Opts In? Composition Effects and Disappointment from Participation Payments', *The Review of Economics and Statistics* pp. 1–45.

Andreoni, J. (2018), 'Satisfaction guaranteed: When moral hazard meets moral preferences', *American Economic Journal: Microeconomics* **10**(4), 159–89.

Andries, M. & Haddad, V. (2020), 'Information aversion', *Journal of Political Economy* **128**(5), 1901 – 1939.

Arieli, I., Babichenko, Y. & Smorodinsky, R. (2018), 'Robust forecast aggregation', *Proceedings of the National Academy of Sciences* **115**(52), E12135–E12143.

Asriyan, V., Foarta, D. & Vanasco, V. (2023), 'The good, the bad, and the complex: Product design with imperfect information', *American Economic Journal: Microeconomics* **15**(2), 187–226.

Ba, C., Bohren, J. A. & Imas, A. (2023), 'Over- and underreaction to information', *Unpublished manuscript* .

Banovetz, J. & Oprea, R. (2023), 'Complexity and procedural choice', *American Economic Journal: Microeconomics* **15**(2), 384–413.

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.

Becker, G. M., Degroot, M. H. & Marschak, J. (1964), 'Measuring utility by a single-response sequential method', *Behavioral Science* **9**(3), 226–232.

Berg, J., Dickhaut, J. & McCabe, K. (1995), 'Trust, reciprocity, and social history', *Games and Economic Behavior* **10**(1), 122–142.

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.

Bohnet, I., Greig, F., Herrmann, B. & Zeckhauser, R. (2008), 'Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states', *American Economic Review* **98**(1), 294–310.

Boleslavsky, R. & Cotton, C. (2015), 'Grading standards and education quality', *American Economic Journal: Microeconomics* **7**(2), 248–79.

Bolton, G. E. & Ockenfels, A. (2000), 'Erc: A theory of equity, reciprocity, and competition', *American Economic Review* **90**(1), 166–193.

Börgers, T., Hernando-Veciana, A. & Krähmer, D. (2013), 'When are signals complements or substitutes?', *Journal of Economic Theory* **148**(1), 165–195.

Bracht, J. & Feltovich, N. (2009), 'Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game', *Journal of Public Economics* **93**(9-10), 1036–1044.

Brooks, B., Frankel, A. & Kamenica, E. (2023), Comparisons of signals, working paper.

Brown, M., Falk, A. & Fehr, E. (2004), 'Relational contracts and the nature of market interactions', *Econometrica* **72**(3), 747–780.

Burks, S., Carpenter, J. & Verhoogen, E. (2003), 'Playing both roles in the trust game', *Journal of Economic Behavior & Organization* **51**(2), 195–216.

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.

Caplin, A., Csaba, D., Leahy, J. & Nov, O. (2020), 'Rational Inattention, Competitive Supply, and Psychometrics', *The Quarterly Journal of Economics* **135**(3), 1681–1724.

Caplin, A. & Dean, M. (2013), Behavioral implications of rational inattention with shannon entropy, Working Paper 19318, National Bureau of Economic Research.

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

Charness, G. & Dufwenberg, M. (2006), 'Promises and partnership', *Econometrica* **74**(6), 1579–1601.

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.

Charness, G. & Rabin, M. (2002), 'Understanding social preferences with simple tests', *The Quarterly Journal of Economics* **117**(3), 817–869.

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.

Chew, S. H., Miao, B. & Zhong, S. (2017), 'Partial ambiguity', *Econometrica* **85**(4), 1239–1260.

Costa-Gomes, M. A. & Crawford, V. P. (2006), 'Cognition and behavior in two-person guessing games: An experimental study', *American Economic Review* **96**(5), 1737–1768.

Crawford, V. P. & Iriberri, N. (2007*a*), 'Fatal attraction: Salience, naïveté, and sophistication in experimental "hide-and-seek" games', *American Economic Review* **97**(5), 1731–1750.

Crawford, V. P. & Iriberri, N. (2007*b*), 'Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions?', *Econometrica* **75**(6), 1721–1770.

Danz, D., Vesterlund, L. & Wilson, A. J. (2022), 'Belief elicitation and behavioral incentive compatibility', *American Economic Review* **112**(9), 2851–83.

De Oliveira, H., Ishii, Y. & Lin, X. (2021), Robust merging of information, working paper.

Dean, M. & Neligh, N. N. (2023), 'Experimental tests of rational inattention', *Journal of Political Economy* **131**(12), 3415–3461.

Dewan, A. & Neligh, N. (2020), 'Estimating information cost functions in models of rational inattention', *Journal of Economic Theory* **187**, 105011.

Dillenberger, D. (2010), 'Preferences for one-shot resolution of uncertainty and allais-type behavior', *Econometrica* **78**(6), 1973–2004.

Eliaz, K. & Schotter, A. (2007), 'Experimental testing of intrinsic preferences for noninstrumental information', *American Economic Review* **97**(2), 166–169.

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

Ely, J., Frankel, A. & Kamenica, E. (2015*a*), 'Suspense and surprise', *Journal of Political Economy* **123**(1), 215–260.

Ely, J., Frankel, A. & Kamenica, E. (2015*b*), 'Suspense and surprise', *Journal of Political Economy* **123**(1), 215–260.

Enke, B. & Graeber, T. (2023), 'Cognitive uncertainty', *Quarterly Journal of Economics* **138**(4), 2021–2067.

Enke, B., Graeber, T. & Oprea, R. (2023), Complexity and time, 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.

Evers, E., Inbar, Y., Loewenstein, G. F. & Zeelenberg, M. (2014), Order preference, working paper.

Eyster, E. & Weizsäcker, G. (2011), Correlation neglect in financial decision-making, Discussion Papers of DIW Berlin 1104, DIW Berlin, German Institute for Economic Research.

Falk, A. & Zimmermann, F. (2022), 'Attention and dread: Experimental evidence on preferences for information', *Management Science* **Forthcoming**.

Fedyk, A. & Hodson, J. (2023), 'When can the market identify old news?', *Journal of Financial Economics* **149**(1), 92–113.

Fehr, E. (2009), 'On the economics and biology of trust', *Journal of the European Economic Association* **7**(2-3), 235–266.

Fehr, E. & Schmidt, K. M. (1999), 'A theory of fairness, competition, and cooperation', *The Quarterly Journal of Economics* **114**(3), 817–868.

Festinger, L. (1957), *A Theory of Cognitive Dissonance*, Stanford University Press.

Fischbacher, U. (2007), 'z-tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics* **10**(2), 171–178.

Frankel, A. & Kamenica, E. (2019), 'Quantifying information and uncertainty', *American Economic Review* **109**(10), 3650–80.

Frydman, C. & Jin, L. J. (2023), 'On the source and instability of probability weighting', *Unpublished Manuscript* .

Fréchette, G. R., Lizzeri, A. & Perego, J. (2022), 'Rules and Commitment in Communication: An Experimental Analysis', **90**(5), 2283–2318.

Gabaix, X. (2014), 'A sparsity-based model of bounded rationality', *The Quarterly Journal of Economics* **129**(4), 1661–1710.

Gabaix, X. & Laibson, D. (2022), Myopia and discounting, Technical report, National bureau of economic research.

Gandelman, N. & Hernández-Murillo, R. (2015), 'Risk aversion at the country level', *Federal Reserve Bank of St. Louis Review* **97**(1), 53–66.

Gennaioli, N., LaPorta, R., Lopez-de Silanes, F. & Shleifer, A. (2021), 'Trust and insurance contracts', *Review of Financial Studies* .

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. (2017*b*), 'Competition in persuasion', *The Review of Economic Studies* **84**(1), 300–322.

Goldstein, I. & Leitner, Y. (2018), 'Stress tests and information disclosure', *Journal of Economic Theory* **177**, 34–69.

Golman, R. & Loewenstein, G. (2018), 'Information gaps: A theory of preferences regarding the presence and absence of information', *Decision* **5**(3), 143–164.

Golman, R., Loewenstein, G., Molnar, A. & Saccardo, S. (2022), 'The demand for, and avoidance of, information', *Management Science* **68**(9), 6454–6476.

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

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.

Guiso, L., Sapienza, P. & Zingales, L. (2004), 'The role of social capital in financial development', *American Economic Review* **94**(3), 526–556.

Guiso, L., Sapienza, P. & Zingales, L. (2009), 'Cultural Biases in Economic Exchange?', *The Quarterly Journal of Economics* **124**(3), 1095–1131.

Halevy, Y. (2007), 'Ellsberg revisited: An experimental study', *Econometrica* **75**(2), 503–536.

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.

Huck, S., Lünser, G. K. & Tyran, J.-R. (2012), 'Competition fosters trust', *Games and Economic Behavior* **76**(1), 195 – 209.

In, Y. & Wright, J. (2018), 'Signaling private choices', *The Review of Economic Studies* **85**(1), 558–580.

Je, H. (2023), Does the size of the signal space matter, working paper.

Jin, Y. (2021), 'Does level-k behavior imply level-k thinking?', *Experimental Economics* **24**(1), 330–353.

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.

Kamenica, E. (2019), 'Bayesian persuasion and information design', *Annual Review of Economics* **11**(1), 249–272.

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

Karlan, D. S. (2005), 'Using experimental economics to measure social capital and predict financial decisions', *American Economic Review* **95**(5), 1688–1699.

Kawagoe, T. & Takizawa, H. (2012), 'Level-k analysis of experimental centipede games', *Journal of Economic Behavior & Organization* **82**(2), 548–566.

Kendall, C. & Oprea, R. (2023), On the complexity of forming mental models, working paper.

Khaw, M. W., Li, Z. & Woodford, M. (2021), 'Cognitive imprecision and small-stakes risk aversion', *Review of Economic Studies* **88**(4), 1979–2013.

Knack, S. & Keefer, P. (1997), 'Does social capital have an economic payoff? a cross-country investigation', *The Quarterly Journal of Economics* **112**(4), 1251–1288.

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

Kreps, D. M. & Porteus, E. L. (1978), 'Temporal resolution of uncertainty and dynamic choice theory', *Econometrica* **46**(1), 185–200.

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.

Masatlioglu, Y., Orhun, A. Y. & Raymond, C. (2023), Intrinsic information preferences and skewness, working paper.

Matějka, F. & McKay, A. (2015), 'Rational inattention to discrete choices: A new foundation for the multinomial logit model', *American Economic Review* **105**(1), 272–98.

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 rényi divergences and back again', *Econometrica* **89**(1), 475–506.

Nagel, R. (1995), 'Unraveling in guessing games: An experimental study', *American Economic Review* **85**(5), 1313–1326.

Nguyen, A. & Tan, T. Y. (2021), 'Bayesian persuasion with costly messages', *Journal of Economic Theory* **193**, 105–212.

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.

Oprea, R. (2020), 'What makes a rule complex?', *American Economic Review* **110**(12), 3913–51.

Oprea, R. (2023), Simplicity equivalents, working paper.

Palacios-Huerta, I. (1999), 'The aversion to the sequential resolution of uncertainty', *Journal of Risk and Uncertainty* **18**(3), 249–269.

Pearl, J. (2001), Direct and indirect effects, *in* 'Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence', UAI'01, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, p. 411–420.

Rabin, M. & Weizsäcker, G. (2009), 'Narrow bracketing and dominated choices', *American Economic Review* **99**(4), 1508–43.

Robins, J. M. & Greenland, S. (1992), 'Identifiability and exchangeability for direct and indirect effects', *Epidemiology* **3**(2), 143–155.

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.

Simon, L. K. & Stinchcombe, M. B. (1995), 'Equilibrium refinement for infinite normal-form games', *Econometrica* **63**(6), 1421–1443.

Sims, C. A. (2003), 'Implications of rational inattention', *Journal of Monetary Economics* **50**(3), 665–690.

Stahl, D. O. & Wilson, P. W. (1994), 'Experimental evidence on players' models of other players', *Journal of Economic Behavior & Organization* **25**(3), 309 – 327.

Stahl, D. O. & Wilson, P. W. (1995), 'On players' models of other players: Theory and experimental evidence', *Games and Economic Behavior* **10**(1), 218–254.

Szydlowski, M. (2021), 'Optimal financing and disclosure', *Management Science* **67**(1), 436–454.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. (2014), 'mediation: R package for causal mediation analysis', *Journal of Statistical Software* **59**(5), 1–38.

Vieider, F. (2023), 'Decisions under uncertainty as bayesian inference on choice options', *Unpublished Manuscript* .

Vieider, F. M. (2021), Noisy coding of time and reward discounting, Technical report, Ghent University, Faculty of Economics and Business Administration.

Wilson, A. & Vespa, E. (2016), Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language, working paper.

Woodford, M. (2020), 'Modeling imprecision perception, valuation and choice', *Annual Review of Economics* **12**.

Zak, P. J. & Knack, S. (2001), 'Trust and growth', *The Economic Journal* **111**(470), 295–321.