Asymmetric Failure of Bayesian Updating and Information Source Misattribution

Chun-Hou Cheng, Patrick DeJarnette, and Joseph Tao-yi Wang

October 12, 2023

Abstract

We conducted a laboratory experiment to investigate individual ability to process conflicting social information that could be potentially irrelevant, in which each subject independently draws a ball from one of two digital urns and receives information reported by another subject who may or may not have drawn from the same urn. We find that 71% of subjects who receive new information misattribute the source of the information compared to Bayesian updating. Conflicting information is overly assumed as irrelevant, and confirming information is overly assumed as relevant. This asymmetry is robust even when allowing for subjects to perceive others as reporting non-informative signals. Attributing conflicting information as irrelevant may form the foundation of stable echo chambers or equilibria where additional information has no effect on beliefs.

JEL codes: C44, D91, C91

Keywords: Bayes' rule; polarization; belief-updating; asymmetric processing; biased interpretation

Statements and Declarations: The authors declare that they have no competing interests or conflict of interest related to the research. This research was supported by a grant from Taiwan's Ministry of Science and Technology (grant number 107-2410-H-002-032-MY2).

1 Introduction

Information processing plays an important role in many life decisions, but the same information may not always be interpreted the same way. For example, a stronger belief in science has been correlated with willingness to wearing a mask during the COVID-19 pandemic, while those who identify with masculinity norms are less willing to do so (Stosic et al., 2021; Palmer and Peterson, 2020). Regarding climate change, individuals have different beliefs despite scientific consensus, and these differences persist for long periods of time (Kahan et al., 2011, 2012; Fryer Jr et al., 2019). Some even believe the earth is flat despite abundant evidence, leading to the formation of apparent echo chambers (Brazil, 2020). On an individual level, receiving a bad grade may lead some to pursue non-STEM degrees, while others may see it as a challenge to persist (Koszegi et al., forthcoming; Harris et al., 2020). In general, a failure to update beliefs in the face of conflicting information may also lead some to have overconfidence in personal ability, leading individuals to pursue self-employment (Camerer and Lovallo, 1999; Koellinger et al., 2007) or make poor financial decisions (Barber and Odean, 2001; Malmendier and Tate, 2008).

One possibility for these divergent interpretations is that personal experience and prior beliefs may play a role when processing new information. When information could be potentially untrustworthy or irrelevant, individuals may be overly inclined to discount it entirely when it conflicts with their prior beliefs. For example, if new information conflicts with prior beliefs, people may suspect that it is driven by opposing political or commercial interests and ignore it. In contrast, when new information confirms prior beliefs, it may be much harder to account for the possibility that the information is untrustworthy. As Bayes' Rules implies we should still update our posteriors even if new information is noisy, incorrect beliefs persists for much longer if people ignore information because it *may* be irrelevant.[1] This process could lead to the formation and persistence of echo chambers, where conflicting information may be essentially disregarded.

In this paper, we examine a laboratory experiment investigating people's ability to process new information from other humans, and study the difference in belief updating when new information aligns with or is against prior information. Specifically, we employ a two step procedure, in which subjects first draw objective information about one of two independent urns, and use this information to update beliefs about the state of that urn. Then, each subject learns the stated belief of another randomly chosen subject. However, the second subject's urn may or may not be the same as the first subject, allowing for the possibility for the information to be viewed as irrelevant to the assigned urn.

In the face of conflicting information, a subject should correctly infer that the other subject is more likely to have drawn from a different urn. However, this does not mean that the other subject could not have drawn from the same urn, just that it is less likely. In our neutral context of drawing balls from urns, we document that individuals asymmetrically update beliefs for conflicting and confirming information. When faced with conflicting information, subjects appear to overly attribute the source as coming from the other (irrelevant) urn. Conversely, in the face of confirming information, subjects are comparatively more likely to attribute

[1] Or worse, incorporate non-informative signals as information, as seen in Fryer Jr et al. (2019) and discussed below.

it to their own (assigned) urn.

Empirically, there is a large literature showing individuals do not perfectly Bayesian update their beliefs (Tversky and Kahneman, 1973; Grether, 1980; Holt and Smith, 2009), but more recently the literature has experimentally explored whether individuals *asymmetrically* update their beliefs.[2] One strand of this literature has focused on the asymmetric updating of personal attributes such as beauty or intelligence, as these estimates tend to be biased on average and may impact important decisions such as career choice. The results have been mixed, with Eil and Rao (2011), Ertac (2011), Grossman and Owens (2012), Möbius et al. (2014), and Coutts (2019) finding evidence that information about personal attributes leads to asymmetric updating, whereas Gotthard-Real (2017), Buser et al. (2018), and Schwardmann and Van der Weele (2019) finding no asymmetric updating about personal attributes.

Asymmetric updating of personal attributes is an important question, but many policy relevant issues are not inherently egocentric, such as climate change or health risks in a pandemic. In these settings, information does not pertain to only the receiver, but also to the state of the world. However, experimentally altering these important beliefs may be difficult, as the subject may have strong priors from a large amount of information (or disinformation).In addition, although researchers can experimentally vary information provided, it may be more difficult to manipulate subjects' perceptions of information source objectivity without deception (Ortmann and Hertwig (2002)), which is of central importance to this paper. Thus, we seek

[2] It is worth noting that psychology had noted a similar process as a subset of "confirmation bias" outside of a Bayesian framework, c.f. Lord et al. (1979).

to explore the asymmetric updating of beliefs in a more context-neutral or financial setting, to give a better picture of how initial belief evolution occurs. This also allows for a mixture of objective and social information to be analyzed as well.

Yet this paper is not the first to explore asymmetric updating in a financial or neutral context. Coutts (2019) tests for asymmetric updating across ego relevant, financial, and context-neutral settings and found consistent asymmetric updating across all three domains. Barron (2021) also focuses on the financial domain, but finds no evidence of asymmetric updating in aggregate beliefs, though note substantial heterogeneity in belief updating processes.

However, in these papers, all the information was non-social, being provided by a computer rather than another human. In comparison, this paper provides evidence that socially-processed information also results in asymmetric updating. This is not ex ante clear, as social information could be primarily ignored if one believed others were incapable of processing information. In addition, due to the two urn structure of the experimental design, we can further distinguish causes of the confirmation bias – specifically, whether subjects (i) process the new information but misattribute the source urn or (ii) completely disregard the conflicting information.

Yet ours is not the first work to study social signals and belief evolution. Oprea and Yuksel (2022) also primarily explores social evolution of beliefs in a laboratory context. The paper finds strong evidence for motivated belief updating in personal attribute settings – updating beliefs that subjects may have additional nonpecuniary incentives to believe as true. In the paper's primary experiments, each subject takes an intelligence quiz, and then is paired with a partner on the same side of the median score. Each subject reports their beliefs of their pair's intelligence in real time, first individually, and then (in the primary treatment group) with full information of their partner's real time belief. Overall, they find that the ”optimistic” partner does not move downward in response to the ”pessimistic” partner's belief, resulting in persistent overconfidence. Furthermore, their paper demonstrates that objective signals are treated in a Bayesian manner, with positive and negative news being weighted equally.

In comparison, our paper focuses on 1-way social communication, such as receiving a news report on climate change written by a journalist or a notice about mask effectiveness written by a government official.[3] One benefit of this focus is that we can isolate social belief evolution of individual beliefs rather than social signalling concerns as demonstrated in Burks et al. (2013).[4] Specifically, in 2-way communication, a "pessimistic" belief about group's intelligence status may be perceived as an insult.[5]

As for possible theoretic underpinnings for asymmetric updating, Fryer Jr et al. (2019) provides a model to depict why polarization in people's beliefs would occur in many settings where information is open to interpretation. An important theoretical prediction from this paper is that polarization increases when people interpret a (non-informative) signal as evidence for a particular state based on their current beliefs. In addition, their online Amazon Mechanical Turk experiments show that

[3] Of course, many important beliefs are more likely to involve 2-way social interaction, such as a committee collating information to make a joint decision or discussions via social media.

[4] As Oprea and Yuksel (2022) states, ”Further experiments designed prospectively to examine the relationship between signal ambiguity and social exchange of beliefs in more depth seems like an important next step in this agenda."

[5] However, in many contexts in the real world, this preference to cater to extreme views may be an important determinant for the creation of echo chambers.

when subjects observe a sequence of information, they indeed form biased interpretation of evidence in the face of ambiguous ones and results in polarization in issues like climate change and death penalty.

In comparison to Fryer Jr et al. (2019), we provide three main contributions. First, the polarizing beliefs in Fryer Jr et al. (2019) stem from non-informative signals that are incorrectly inferred to be informative. In our paper, we explore how individuals incorporate informative signals that are conflicting to their current beliefs, rather than how they misinterpret non-informative signals. Thus, even in purely informative spaces, we show improper Bayesian updating. Secondly, in our experiment, we explore a politically neutral context with objective outcomes, as opposed to the politically charged context with a subjective scale.[6]

Collectively, these results can potentially explain why, despite the general scientific consensus on climate change, individuals may form beliefs that cause them to ignore this information. In other words, given the abundance of information supporting climate change, it may be that climate deniers instead infer that this conflicting information is instead from an untrustworthy or irrelevant source, as suggested by survey evidence from Rowland et al. (2022). However, we believe extending research to matters closely connected to public policy, and further exploring the role of social signalling in echo chambers, remains an important frontier for future work.

Our paper can also be viewed as an extension of the rich literature of confirmation bias, which has been documented in economics (Babcock et al., 1995) and psychology (Lord et al., 1979). Confirmation bias describes people's tendency to interpret the

[6] As the scale employed in Fryer Jr et al. (2019) may be interpreted differently based on prior beliefs.

information in a fashion that is biased toward confirming one's prior belief. Glaeser and Sunstein (2013) introduces a model to show how balanced information can lead to polarization. They suggest that the same information have diametrically opposite effect for those who have confirming and conflicting priors. Our experiment provides additional experimental evidence and illustrates a possible mechanism for this phenomenon – information source misattribution.

2 Experimental Design

There are ten rounds in the experiment, each round consists of two phases. In each round, the subject is independently randomly assigned one of two digital urns, urn A or urn B, which have been themselves randomized (described in more detail in subsections below). In the first phase, the subject receives a piece of information about their assigned urn, and no information about the unassigned urn. With this information in hand, the subject is incentivized to truthfully report their beliefs about the rule (distribution) that their assigned urn follows, and also the rule (distribution) that the unassigned urn follows.[7]

In the second phase, each subject observes another (randomly chosen) subject's elicited beliefs about the other subject's assigned urn. This second subject's urn could be the same or differ from the original subject's urn, but this information is not revealed to the subject. With this piece of human-derived information, the subject is again incentivized to report their true beliefs about both urns. After these two phases with no feedback, the subjects begin the next round and repeat the procedure. After ten rounds, a short survey is conducted and a round is selected for payment.[8]

[7] Note, since the subject hasn't received any information about the unassigned urn, there should be no updating in the priors of the unassigned urn. This is one of several placebo tests we used to ensure subjects understand the instructions and have some basic understanding of statistics. Indeed, the vast majority of subjects report close to the prior belief of the unassigned urn at this point, as shown in the results section.

[8] Alternative experimental designs that were considered, but not implemented are listed in Appendix C.

2.1 Design Details

In the first phase, subjects are independently assigned to either urn A or urn B with equal chance. Both urns contain one hundred digital balls, labeled from 1 to 100. In each round, urn A and urn B are **independently** randomized to follow one of two "rules" with equal chance. Subjects are not told which rule the urns follows, but those assigned to the same urn experience the same rule.

While both urns have the same uniform distribution of balls, the rules of the urn influence the information that the subject actually observes. In particular, for each subject the computer draws (with replacement) two balls randomly from the assigned urn. However, the subject will only be informed about one of these two balls, depending on which rule their assigned urn follows.

If the urn is following the *Maximum Rule*, the computer will reveal the larger ball (the one with the higher value label). If the urn is following the *Minimum Rule*, the computer will reveal the smaller ball (the lower value label). As a reminder, urn A and urn B are independently randomized to either follow the *Maximum Rule* or the *Minimum Rule* with equal chance. After observing one ball, subjects are incentivized to predict the probability that the *Maximum Rule* is applied to their assigned urn. Similarly, they also are incentivized to predict the probability that *Maximum Rule* is applied to the unassigned (irrelevant) urn.[9] To be clear, the unassigned urn beliefs are also elicited with incentives described below, but less (and potentially no) information is received about this urn. And any information truly sourced from that unassigned urn is *irrelevant* to the assigned urn, as the rule

[9] Because the Maximum and Minimum rules are mutually exclusive, eliciting a single probability for each urn is sufficient. However, because the urns' rules were independently randomized, subjects must report a probability for each urn.

for each urn is independently randomized. Thus, the urn themselves are irrelevant to each other, but information coming from an unknown urn would need to be weighted properly under a Bayesian framework.

In the second phase, for each subject, the computer randomly chooses another subject, and reveals the first phase prediction of the other-subject's assigned urn. However, even though the subject observe this prediction, they do not know if this other-subject was assigned to the same (assigned) urn or the other (unassigned) urn.[10] After seeing the information from another subject, subjects again predict the probability that the *Maximum Rule* is applied to each urn. This concludes the round, followed by the next round until all 10 have been completed. Subjects are not informed of the outcomes in between rounds.[11]

2.2 Belief Elicitation

Following Holt and Smith (2016), we use an incentive compatible two-stage menu of lottery choices as the belief elicitation mechanism in the experiment. Essentially, it is the Becker-DeGroot-Marschak (BDM) pricing procedure but separated into two stages to make it easier for subjects to understand. Holt and Smith (2016) compared three mechanisms of belief elicitation and found beliefs elicited from this two-stage procedure to be more accurate and with lower average belief error in terms of Bayesian prediction.

[10] As subjects are independently randomized between Urn A and Urn B, the prior belief before seeing the prediction is that the second subject has a 50% probability to have been assigned to either urn. Subjects were informed of this statistical independence.

[11] While this reduces the scope of learning, which may be an interesting topic, the path dependence could complicate the analysis. The main benefit of completing 10 rounds was to increase variation in the signals observed, and we leave it to future work to study whether asymmetric updating could be reduced with experience.

In the first stage, subjects choose from a list of 11 lottery choices, with each row being a choice between a "random lottery" and an "event lottery". The "event lottery" is the same for all 11 rows and rewards a prize if and only if the urn in question follows the "Maximum Rule". The random lotteries vary by row and have winning probabilities ascending from 0%, 10%, ..., to 100%.

The prize for winning an "event lottery" is identical to the prize for winning a "random lottery", allowing subjects to focus on the probabilities involved. In particular, subjects compare the probability of each random lottery with their belief that the event would occur. If they have a subjective belief that there is a 55% chance the urn follows the maximum rule, then the subject would presumably[12] prefer the event lottery over a random lottery with a 20% chance of winning. Likewise, the subject would prefer the random lottery with a 80% chance of winning over the event lottery. This same logic applies for a 50% chance of winning and a 60% chance of winning, respectively.

Based on the "switching point", subjects decide a second digit of probability in the second stage. Thus, the subject might record a switching point between 50% and 60%, then report the second digit of 5, implying a subjective belief of 55% for the event (urn following maximum rule). This is conceptually identical to having 101 rows of lottery pairs (0%, 1%, 2%, etc) but saves screen space, decision fatigue, and allows for more rounds in a given time period. Because of this two-stage elicitation

[12] One might be concerned about the potential for ambiguity aversion to distort these probabilities. Though it's worth noting that the event lottery is not truly ambiguous in this experiment, though the difficulty in Bayesian calculations may make it appear so. Aside from the aforementioned research Holt and Smith (2016), we also find no such evidence of this – for example phase 1 probabilities are mostly centered around the Bayesian posterior. One can also use the 'direction' that ambiguity aversion would provide, that is, to give additional preference to the objective probabilities. Thus the switching point would tend to be shifted closer to 0% for all situations. There is no evidence that this is the case.

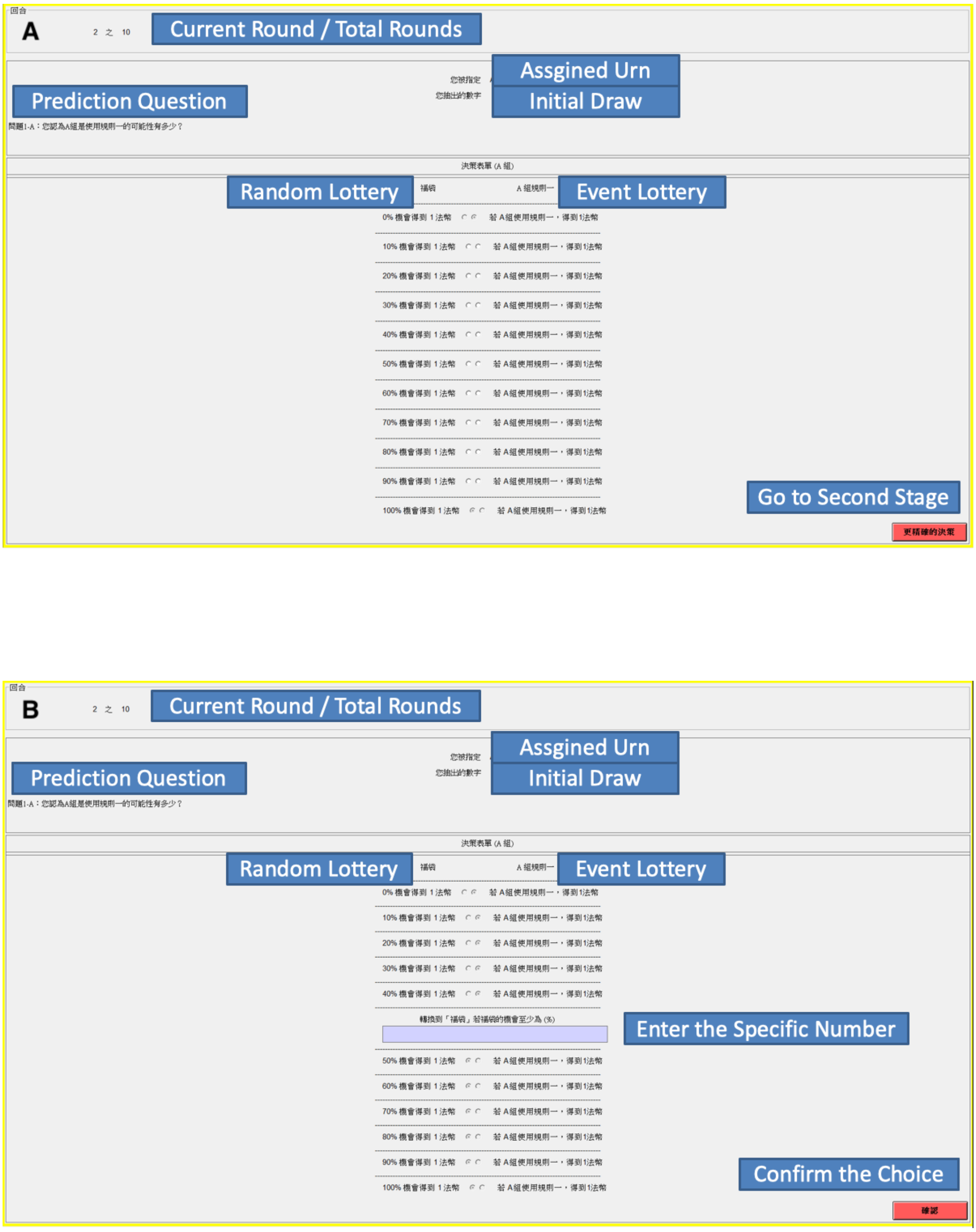
methodology however, there can only be allowed one "switching point". This removes the potential for non-monotonic behavior, though this constraint is arguably preferable for analysts and may explain why this two-stage elicitation seemed to do better at eliciting Bayesian posteriors in Holt and Smith (2016).

After all 10 rounds are finished, one belief elicitation from a single round is selected for payment. After the decision is done, the computer randomly draws one number from 0 to 100. If the number is smaller than the two-stage implied switching point, they receive the event lottery – that is, they are paid a prize only if the urn in that elicitation was indeed following the maximum rule. If the number is equal or larger than the two-stage implied switching rule, then they receive a random lottery where the probability of winning the prize is equal to the original drawn.

2.3 Experimental Procedures

All sessions were conducted at Taiwan Social Sciences Experimental Laboratory (TASSEL), National Taiwan University (NTU). Six sessions were run during October 2019 and November 2019, for a total of 123 subjects. We recruited NTU student subjects using the TASSEL website powered by ORSEE (Greiner, 2015). Each session lasted approximately 100 minutes, and average earnings were 512 NT dollars (approx. $17).[13] The experiment was programmed with z-Tree (Fischbacher, 2007) and conducted in Chinese. The experimental interfaces are shown in Figure 1A for the first stage and Figure 1B for the second stage of elicitation processes.

[13] This amount is substantial, double what students would have earned working at Taipei's minimum wage over a 100 minute period.



回合

A 2 之 10

您被指定

您抽出的數字

問題1-A：您認為A組是使用規則一的可能性有多少？

決策表單 (A 組)

福袋 A 組規則一

0% 機會得到 1 法幣 ○ ● 若 A組使用規則一，得到1法幣

10% 機會得到 1 法幣 ○ ○ 若 A組使用規則一，得到1法幣

20% 機會得到 1 法幣 ○ ○ 若 A組使用規則一，得到1法幣

30% 機會得到 1 法幣 ○ ○ 若 A組使用規則一，得到1法幣

40% 機會得到 1 法幣 ○ ○ 若 A組使用規則一，得到1法幣

50% 機會得到 1 法幣 ○ ○ 若 A組使用規則一，得到1法幣

60% 機會得到 1 法幣 ○ ○ 若 A組使用規則一，得到1法幣

70% 機會得到 1 法幣 ○ ○ 若 A組使用規則一，得到1法幣

80% 機會得到 1 法幣 ○ ○ 若 A組使用規則一，得到1法幣

90% 機會得到 1 法幣 ○ ○ 若 A組使用規則一，得到1法幣

100% 機會得到 1 法幣 ● ○ 若 A組使用規則一，得到1法幣

回合

B 2 之 10

您被指定

您抽出的數字

問題1-A：您認為A組是使用規則一的可能性有多少？

決策表單 (A 組)

福袋 A 組規則一

0% 機會得到 1 法幣 ○ ● 若 A組使用規則一，得到1法幣

10% 機會得到 1 法幣 ○ ● 若 A組使用規則一，得到1法幣

20% 機會得到 1 法幣 ○ ● 若 A組使用規則一，得到1法幣

30% 機會得到 1 法幣 ○ ● 若 A組使用規則一，得到1法幣

40% 機會得到 1 法幣 ○ ● 若 A組使用規則一，得到1法幣

轉換到「福袋」若福袋的機會至少為 (%)

50% 機會得到 1 法幣 ● ○ 若 A組使用規則一，得到1法幣

60% 機會得到 1 法幣 ● ○ 若 A組使用規則一，得到1法幣

70% 機會得到 1 法幣 ● ○ 若 A組使用規則一，得到1法幣

80% 機會得到 1 法幣 ● ○ 若 A組使用規則一，得到1法幣

90% 機會得到 1 法幣 ● ○ 若 A組使用規則一，得到1法幣

100% 機會得到 1 法幣 ● ○ 若 A組使用規則一，得到1法幣

Figure 1: Two-stage Menu of Lottery Choices: (A) 1st Stage, and (B) 2nd stage.

2.4 Bayesian Probability Predictions

For notation simplicity, we define urn A to be the assigned urn and urn B to be the unassigned (irrelevant[14]) urn. We use $\theta_{\max}$ and $\theta_{\min}$ to denote the *Maximum Rule* and *Minimum Rule* of the assigned urn; the other urn also has two mutually exclusive states, *Maximum Rule* and *Minimum Rule*, indicated by $\omega_{\max}$ and $\omega_{\min}$. The information s_1 denotes the observed ball (of the assigned urn) in the first phase, s_2 is elicited probability from another subject (of their assigned urn, which may or may not be the subject's assigned urn as well) observed in the second phase.

2.4.1 The Structure of Two States

To calculate the Bayesian probability, we consider the structure of two possible states in advance. Consider the probability $\Pr(s_1|\theta_{\max})$ of seeing s_1 under *Maximum Rule* in the assigned urn. For two randomly drawn balls S_1^1 and S_1^2, there are two mutually exclusive events: Either the first drawn ball S_1^1 is the observed ball and therefore the second drawn ball is smaller than the observed ball, or exactly the opposite, that is, the second drawn ball S_1^2 is the observed ball and the first drawn ball is equal to or smaller than the observed ball. Therefore, the probabilities of seeing a signal s_1 conditional on the assigned urn following the maximum rule is:

$$\Pr(s_1|\theta_{\max}) = \frac{2s_1 - 1}{10000} \tag{1}$$

[14] The unassigned urn itself is irrelevant to the assigned urn as the rules are independently randomized. It is not entirely irrelevant to the subject (who is incentivized to report their beliefs), though less or no information is received about the unassigned urn.

The other probability is $\Pr(s_1|\theta_{\min}) = 1 - \Pr(s_1|\theta_{max}) = (201 - 2s_1)/10000$. Therefore, the probability distribution of observing the ball s_1 is linear under both the *Maximum Rule* (increasing linearly from 0.01% when observing 1 to 1.99% when observing 100) and *Minimum Rule* (decreasing linearly from 1.99% when observing 1 to 0.01% when observing 100).

2.4.2 Phase 1

In the first phase, the processed information is the observed ball, which is only useful to infer the state of urn A. With the observed ball, the Bayesian probability prediction for urn A is as follows.

$$\begin{aligned}\Pr(\theta_{\max}|s_1) &= \frac{\Pr(s_1|\theta_{\max})\Pr(\theta_{\max})}{\Pr(s_1)} = \frac{(2s_1-1)/10000}{1/100}\cdot\frac{1}{2} \\ &= \frac{s_1}{100} - \frac{1}{200} \qquad (2)\end{aligned}$$

The Bayesian posterior for urn A shows that subjects should predict a probability slightly below their observed signal (ball) s_1 (in percentage terms). For example, observing a signal $s_1 = 100$ would imply that there is a 99.5% probability that the assigned urn is following the maximum rule.[15] However, in the experiment, fractional percentages were not allowed in the elicitation, requiring subjects to report a whole percentage term (i.e. 55% instead of 54.5%).[16] Thus reporting s_1 (in percentage

[15] This falls short of 100% because the urn draws two balls with replacement, so it's possible, though unlikely, for a minimum urn to draw the 100 ball twice and report the 'smaller' of the two balls, i.e. 100.

[16] This restriction was chosen for implementation feasibility and ease of explaining the instructions to subjects. Please see the section on Design Details for more details.

terms) would be a correct Bayesian posterior given the constraints.[17]

The intuition of this prediction is as follows. If the subject receive a signal of 75, then one of three states occurred:

- the unobserved ball was strictly less than 75 (and thus the urn must follows the *Maximum Rule*)
- the unobserved ball was strictly greater than 75 (and thus the urn must follows the *Minimum Rule*)
- the unobserved ball was exactly 75 (drawn twice due to replacement)

As the balls themselves are uniformly distributed and the rules are ex ante equally likely, the first state has a 74% chance while the second has a 25% chance. The third possibility conditionally occurs 1% of the time, but is uninformative about the urn's rule, thus it adds 0.5% to both the probability of the *Maximum Rule* and the *Minimum Rule.*

The phase 1 Bayesian posterior for unassigned urn B is straightforward since there is not yet any information about that urn. As a result, $\Pr(\omega_{\max}|s_1)$ should be 0.5.

2.4.3 Phase 2

In the second phase, we assume subjects see another ball s_2, which is either from assigned urn A or unassigned urn B with ex ante equal probability. Because the actual source is unknown, subjects are asked inferences of both urns.

[17] Likewise, reporting s_1-1 is equally correct given the constraints, though less common in the data. For example, suppose the observed ball s_1 is 30, the Bayesian probability is $\Pr(\theta_{\max}|s_1 = 30) = \frac{30}{100} - \frac{1}{200} = 29.5\%$. Thus reporting either $s_1 = 30$ or $s_1 - 1 = 29$ in percentage terms would be correct.

However, in the experiment they actually observe a signal from another human being. If the other subject is a correct Bayesian updater and the subject believes that the other subject is a correct Bayesian updater, this is theoretically equivalent.[18] For the most part, we see that the average subject is close to a Bayesian updater in phase 1. However, whether subjects believe other subjects are Bayesian updaters or not is ex ante unclear. In the next subsection and later analysis, we explore the possibility that subjects view other subjects as sending non-informative ("random") signals.

With mathematical work found in an appendix, the Bayesian probability prediction for urn A (the assigned urn) following the Maximum rule is:

$$\Pr(\theta_{\max}|s_1, s_2) = \frac{[3(2s_2 - 1) + (201 - 2s_2)]\,(2s_1 - 1)}{[3(2s_2 - 1) + (201 - 2s_2)]\,(2s_1 - 1) + [(2s_2 - 1) + 3(201 - 2s_2)]\,(201 - 2s_1)} \tag{3}$$

The Bayesian probability prediction for urn B (the irrelevant urn) following the Maximum rule is as follows.

$$\Pr(\omega_{\max}|s_1, s_2) = \frac{(2s_2 - 1)(2s_1 - 1) + 100 \cdot (201 - 2s_1)}{(2s_2 - 1)(2s_1 - 1) + 100 \cdot (201 - 2s_1) + 100 \cdot (2s_1 - 1) + (201 - 2s_2)(201 - 2s_1)} \tag{4}$$

[18] It may be worth noting that subjects have pecuniary incentives not to misreport phase 1 beliefs, and no pecuniary incentive to misreport. As interaction between subjects was limited and subjects were randomized between rounds without identification, we don't believe nonpecuniary concerns such as reputation or spite play a role here.

Although these probabilities may seem difficult at first glance, many situations outside of the laboratory are likely to have even more difficulty to properly Bayesian update. In that sense, if subjects are unable to Bayesian update in this setting where at least the priors are objective and known, they may be even less likely to Bayesian update in more ambiguous setting. Thus, we tentatively view this as a lower bound setting for Bayesian updating. In any case, the equations seem roughly equally complex for confirming and conflicting information, which is our primary research question.

2.5 Allowing for Social Information to be Non-informative

In the equations above, it is assumed that subjects believe the other subject to be a perfect Bayesian updater. However, even though a vast majority of subjects themselves report the observed signal (i.e. the correct Bayesian update), subjects have no direct knowledge whether others would do the same. This question of whether there is "common knowledge" of Bayesian updating is an important element of social information, which to our knowledge, has not been directly studied in previous experimental work, especially regarding asymmetric updating. To allow for this extension in the above framework, we take the extreme assumption that subjects perceive other subjects as sending ”non-informative” random signals.[19] To model this, we extend the model to include a third (psychic) ”urn” called the ”useless” urn, as it has no information at all about either of the other urns.

In the previous subsection, we assume the prior probability that s_2 came from

[19] It may bear repeating that truly non-informative signals may influence beliefs, as in Fryer Jr et al. (2019). This extension does not allow for this possibility, as we cannot distinguish between viewing a non-informative signal as informative vs viewing an informative signal as informative.

the assigned urn was $p_A = 0.5$ and the prior probability that s_2 came from the unassigned (irrelevant) urn was $p_I = 0.5$. Now, we allow for more flexibility with an additional p_U term, to represent the prior belief that the other subject will be so inaccurate at Bayesian updating that their signal should be discarded. However, we still employ the constraint that $p_A + p_I + p_U = 1$, and thus can substitute $p_U = 1 - p_A - p_I$ to remove the term from the Bayesian posteriors below:

$$\begin{aligned}&\Pr(\theta_{\max}|s_1, s_2)\\ &= \frac{[(2s_2-1)p_A + 100(1-p_A)](2s_1-1)}{[(2s_2-1)p_A + 100(1-p_A)](2s_1-1) + [(201-2s_2)p_A + 100(1-p_A)](201-2s_1)}\end{aligned} \tag{5}$$

$$\begin{aligned}&\Pr(\omega_{\max}|s_1, s_2)\\ &= \frac{[(2s_2-1)p_I + 100(1-p_I)](2s_1-1)}{[(2s_2-1)p_I + 100(1-p_I)](2s_1-1) + [(201-2s_2)p_I + 100(1-p_I)](201-2s_1)}\end{aligned} \tag{6}$$

In later analysis below, we estimate these parameters p_A and p_I from the data.

3 Results

3.1 Adherence to Bayesian Updating

3.1.1 Compliance After Initial Draw

Figure 2A presents elicited probabilities of the assigned urn after drawing a ball in the first phase. Each data point represents the reported belief and the initial drawn number of a subject in a particular round. In addition, the kernel density estimation shows that highest density regions are pretty close to the correct Bayesian posteriors. In fact, with nearly 90 percent of the data aligned with the theory if we allow for an errors margin of plus and minus 10 percentage points (±10%).[20] The elicited probabilities of the irrelevant urn, in which they do not have any information, are shown on Figure 2B, in which over 80% of the elicited probabilities are between 0.4 and 0.6 (50% ± 10%). Besides, the kernel density estimation extremely adheres to the correct Bayesian posteriors. Table 1 shows that a majority of choices conform with the theoretical predictions as we reduce the margin of error allowed. Even under the strictest case allowing for only 1 percentage point error (±1%), 60% and 55% of the choices are considered Bayesian in the assigned and the irrelevant urn, respectively.

Notice that there is a cluster of elicited probabilities along the 45-degree line in Figure 2B, implying that some subjects also use the initial draw to update the irrelevant urn. We find that those choices come from one-time behavior of different

[20] Alternatively, one could construct the upper and lower bounds relative to the initial draw. For example, allowing for a 10 percent error results in 50% ± 5% for the ball 50, but 10% ± 1% for the ball 10. This criteria is harsh to those who draw a very small or large ball since they have stronger information. However, under it 76% of the data are still considered to be aligned with theory.

subjects and not concentrated in particular rounds, indicating that they are not caused by particular subjects or rounds.[21] Although these choices consist of 3% to 10% of the data (depending on definition employed), they inflate the correlation between the elicited probabilities of the assigned and irrelevant urn.[22] Without these choices, the correlation is 0.003 ($p > 0.1$), indicating that the vast majority of probabilities are elicited with the knowledge that states of the two urns are independent.[23] In conclusion, most of the choices are consistent with Bayesian updating derived in section 2.4.2.[24]

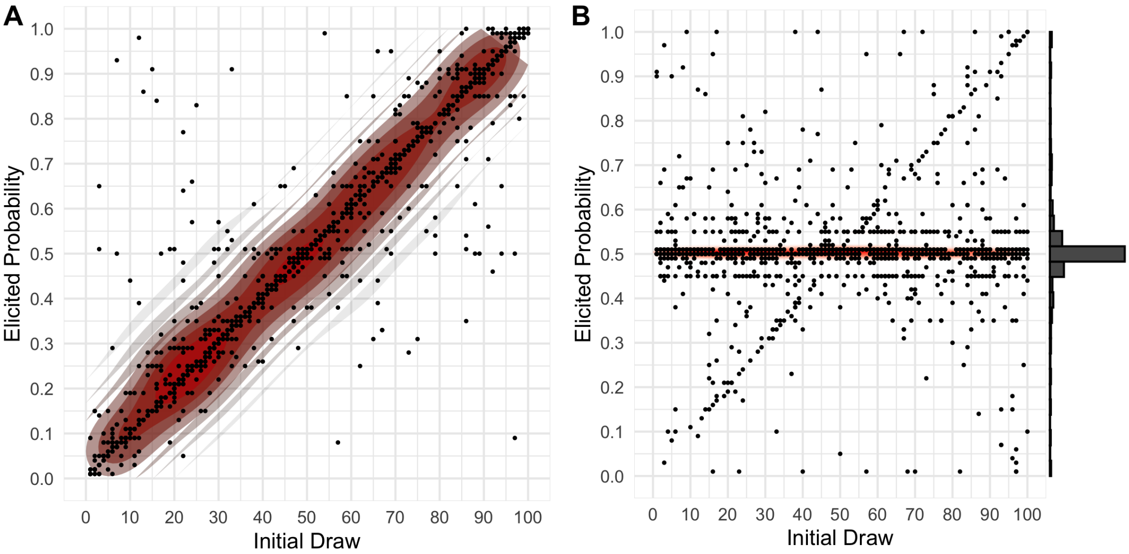


Figure 2: Elicited Beliefs in the First Phase of the (A) Assigned (B) Irrelevant Urn

[21] See Appendix A for further details. In most of these cases, both the assigned and the irrelevant urn will have elicited beliefs very close to each other, suggesting it was not just a matter of mistaking which urn was assigned to them.

[22] A total of 37 choices lie exactly on the 45-degree line excluding initial draws between 40 and 60 where we cannot easily tell if they updated beliefs of the irrelevant urn or not.

[23] Similarly, the second phase correlation between the two urns is 0.006 ($p > 0.1$). Computing with all data, the first and second phase correlations are 0.067 and 0.029, respectively.

[24] A series of additional robustness checks was conducted where these observations were dropped. Results were consistent across including or dropping the data, so we have not reported them below but are happy to provide the raw data or robustness checks upon request. Please see A for some additional detials.

Table 1: Percentage of Theory-consistent Choices Under Different Error Margins

Error Margin	Assigned Urn	Irrelevant Urn
±10 percentage points	89%	81%
±5 percentage points	81%	74%
±3 percentage points	75%	60%
±1 percentage points	66%	55%

3.1.2 Failure After Observing New Information

There exists one intuitive difference between the two possible states of the urn: When the true state is the *Maximum Rule*, the subject is more likely to observe a ball larger than 50, while under the *Minimum Rule*, the subject is more likely to observe a ball equal to or smaller than 50. This leads to a straightforward heuristic for subjects to determine whether new information in the second phase is more likely to come from an urn under the *Maximum Rule* or *Minimum Rule*. As a result, we classify the second-phase information coming from another subject, as either *confirming* or *conflicting* information. In particular, the new information is *confirming* if first and second phase information are both within 1–50 or both within 51–100, while it is *conflicting* when one is within 1–50 and the other one is within 51-100.[25]

Compared to the first phase, belief-updating in the second phase is much worse.[26] Figure 3 summarizes the distribution of Bayesian posteriors and the average deviation from them on different intervals. When the new information is *confirming*, we find that subjects deviate less in the assigned urn, but deviate more in the irrelevant urn. This suggests that it is easier to correctly process new information regarding

[25] Some information may be too close to 50 to be "confirming" or "conflicting" enough, such as initial draws or new information between 40 and 60. Excluding these cases, we expect to find stronger effects.

[26] See Figure 10 of Appendix B for the raw data plotted like Figure 2.

the assigned urn that aligns with what subjects already have. In contrast, updating behavior for the irrelevant urn is far from the Bayesian prediction as the overall deviations are larger than the assigned urn (Figure 3B).

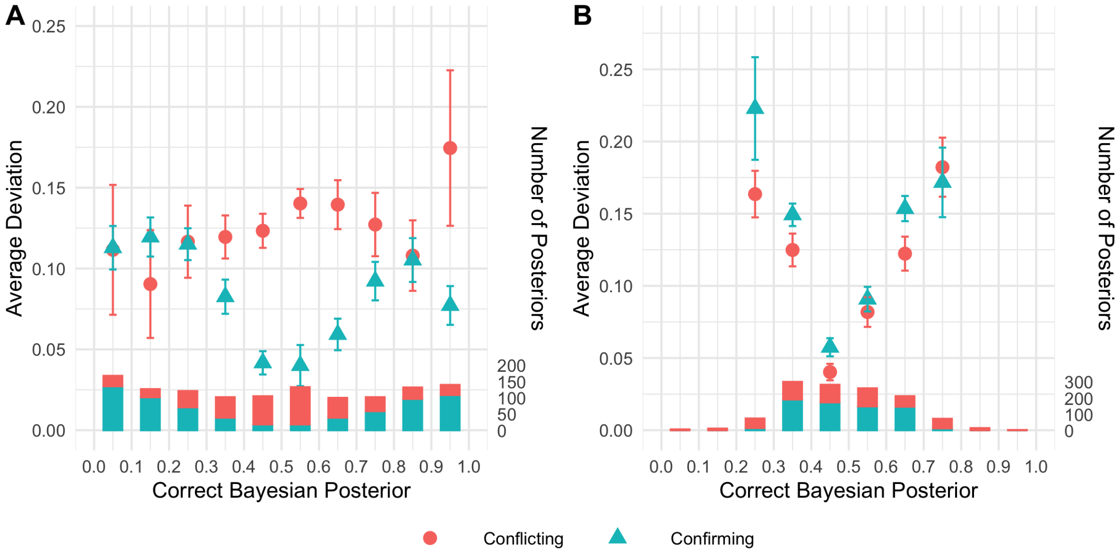


Figure 3: Elicited Beliefs Distribution in the Second Phase of (A) the Assigned, and (B) Irrelevant Urn

Furthermore, the R-squared predicting elicited probabilities using Bayesian posteriors shows that subjects perform updating well in the assigned urn when the information is *confirming* ($R^2 = 0.82$), but perform worse when it is *conflicting* ($R^2 = 0.51$). In contrast, for the irrelevant urn, subjects perform worse when the new information is confirming ($R^2 = 0.33$), but perform better when it is conflicting ($R^2 = 0.52$). The differences in R^2 are statistically significant for both urns (variance ratio test, $p < 0.001$). The results in Appendix B show that the slopes between *confirming* and *conflicting* information are not significantly different in Figure 10A ($p = 0.175$) and Figure 10B ($p = 0.434$).[27]

[27] We test the coefficient β_3 from the model: $Beliefs\beta_0 + \beta_1 Bayesian + \beta_2 Confirming + \beta_3 Interaction + \epsilon$, where the dummy variable *Confirming* indicates the new information is confirming (=1) or not (=0), *Interaction* is the interaction term of *Bayesian* and *Confirming*.

3.2 Belief Updating

In principle, subjects should update their beliefs of both urns regardless of the information received in the second phase because there is always a chance the new information could be from either urn. However, the irrelevant urn has the natural advantage that one should only update it according to the new information regarding the ball of the second phase, since the first ball only carries information about the assigned urn. Therefore, we can easily infer how subjects attribute new information to each urn in the second phase from their updating behavior.

Figure 4 plots elicited probabilities against second-phase information.[28] The red dots are elicited beliefs around 0.5, adhering to the Bayesian prediction of the first phase, indicating "fully dissociate" subjects who do not update irrelevant urn beliefs at all (and should completely attribute the new information to the assigned urn). On the other hand, the blue crosses along the 45-degree line indicate "fully attribute" types who completely ignore the fact that there is some probability that the new information is from their assigned urn.[29] These two types are strongly biased since they put extreme weight on the new information when updating the irrelevant urn. However, they account for 76.7% of the choices when we allow 5 percentage points of error. The intermediate types with more reasonable weights are shown as green triangles in Figure 4, but consist only 18.7% of the choices. This includes those who follow Bayesian updating. Lastly, the remaining 4.6% of choices in black are difficult to rationalize, and might reflect confusion or some other

[28] We drop the choices if their first phase beliefs of the irrelevant urn are out of the range, $[0.45, 0.55]$. The remaining choices plotted in the Figure 4 contain 74% of the data.

[29] The purple dot-cross symbols are overlapping area of the two types, in which we cannot distinguish their types.

information processing method. We summarize the updating behavior in the Table 2.

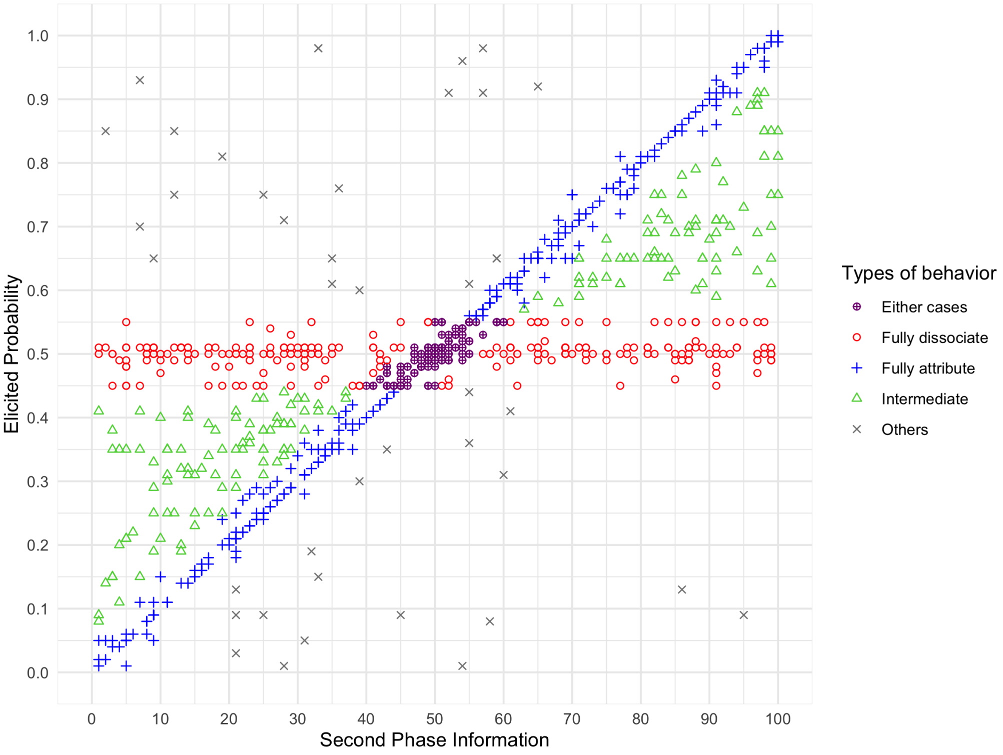


Figure 4: Types of Behavior (Irrelevant Urn)

Table 2: Types of Behavior (Irrelevant Urn)

Types of Choices	Definition	Percentage
Either	Either fully dissociate or fully attribute type.	16.3 %
Fully Dissociate	Other subject's information comes from the assigned urn.	25.4 %
Fully Attribute	Other subject's information comes from the irrelevant urn.	35 %
Intermediate	Put reasonable weights on other subject's information	18.7 %
Others	Choices cannot be classified into above four types.	4.6 %

In Figure 3, we separate second-phase information into *confirming* and *conflicting* information as defined in section 3.1.2. To compare the difference in behavior between receiving confirming and conflicting information, we use a dummy indicating confirming information to predict the occurrence of two distinct types of behavior, completely attribute the information to the assigned urn (Fully Dissociate) and the irrelevant urn (Fully Attribute). Table 3 report fixed-effect panel regression results clustered at the subject level, predicting whether the inferred prior belief fully attributes the new information to the irrelevant urn using whether information is *confirming* or not. For *confirming* information, 33.7% of the choices completely attribute the new information to the assigned urn, while 31.1% of the choices completely attribute the new information to the irrelevant urn. However, when subjects receive *conflicting* information, only 16.5% of the choices attribute new information to the assigned urn, significantly lower than that under *confirming* information. Moreover, 39% of the choices completely attribute new information to the irrelevant urn, significantly higher than that under *confirming* information. This results demonstrates a confirmation bias where subjects overweight (underweight) the possibility that new information came from the assigned urn when it confirms (refutes) their prior.

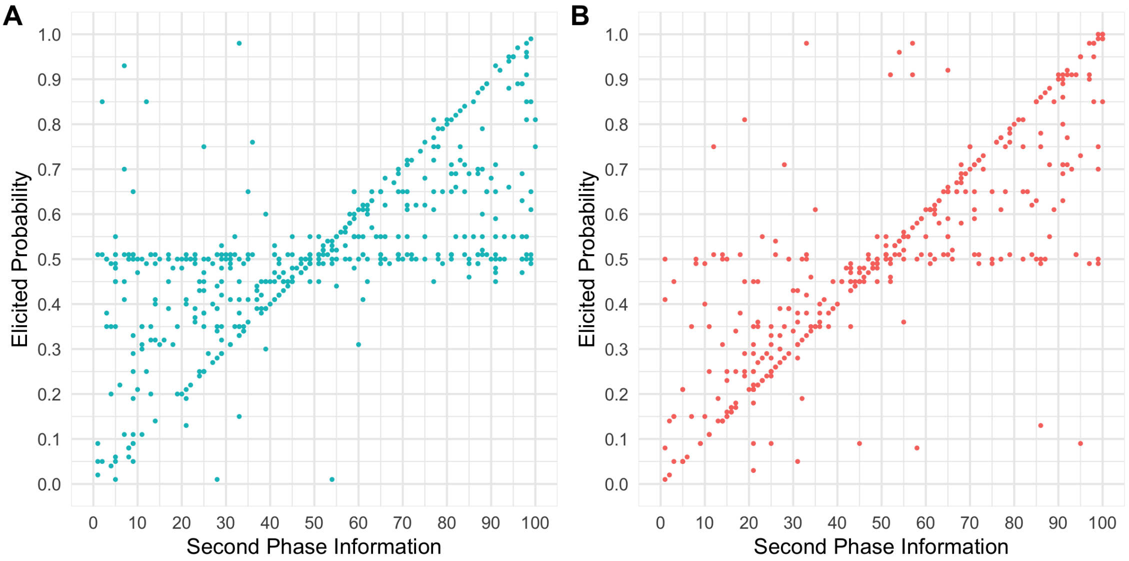


Figure 5: Elicited Beliefs of the Irrelevant Urn: (A) Confirming, and (B) Conflicting Information.

Among those who completely attribute the new information to the irrelevant urn

Table 3: Attribution of the Information

Fully Attribute to	(1) Assigned Urn	(2) Irrelevant Urn
Confirming Information	0.165***	-0.079***
	(0.022)	(0.025)
Constant	0.172***	0.390***
	(0.017)	(0.019)
N	914	914

Note: Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(Fully Attribute), their updated beliefs of the assigned urn should remain unchanged because they believe the information is coming solely from the irrelevant urn. Indeed, the posteriors of the assigned urn show that 75% do not update the assigned urn beliefs much.[30] The remaining 25% also changes their beliefs regarding the assigned urn, overreacting the new information.

In contrast, among those who completely attribute the new information to the assigned urn (Fully Dissociate), beliefs of the assigned urn should be updated as if they have two balls from that urn, resulting in a Bayesian updating process similar to equation (3) in section 2.3.2.[31] Unexpectedly, 54% of these choices stick to their first-phase posteriors of the assigned urn. This implies at least 25.4% × 54% = 13.7% of all choices completely ignore the new information and update neither urn.[32] Figure 6 plots the remaining choices after excluding those which completely ignore the new information. Figure 6A compares the elicited probabilities of fully dissociate types and the Bayesian posterior assuming that both balls came from the same urn.

[30] This number is calculated by allowing 5% error. In fact, 63% have the exact same first and second posterior beliefs.

[31] The Bayesian prediction of having two balls from the same urn is: $\Pr(\theta_{\max}|s_1, s_2) = \Pr(s_2|\theta_{\max}) \cdot \Pr(\theta_{\max}|s_1) / \left[\Pr(s_2|\theta_{\max}) \cdot \Pr(\theta_{\max}|s_1) + \Pr(s_2|\theta_{\min}) \cdot \Pr(\theta_{\min}|s_1))\right]$.

[32] 13.7% is the lower bound since 25.4% excludes choices when second phase information are close to 50 that could be either Fully Dissociate or Fully Attribute.

Even though subjects fully dissociate the information from the irrelevant urn, the updating behavior systematically under-weights the new information from the other subject, resulting in a slope of 0.67 that is significantly lower than 1 ($p < 0.001$). In fact, the elicited probabilities are closer to the Bayesian probability prediction derived in section 2.3.2 (Figure 6B), although the slope (0.78) is still lower than 1 ($p < 0.001$).

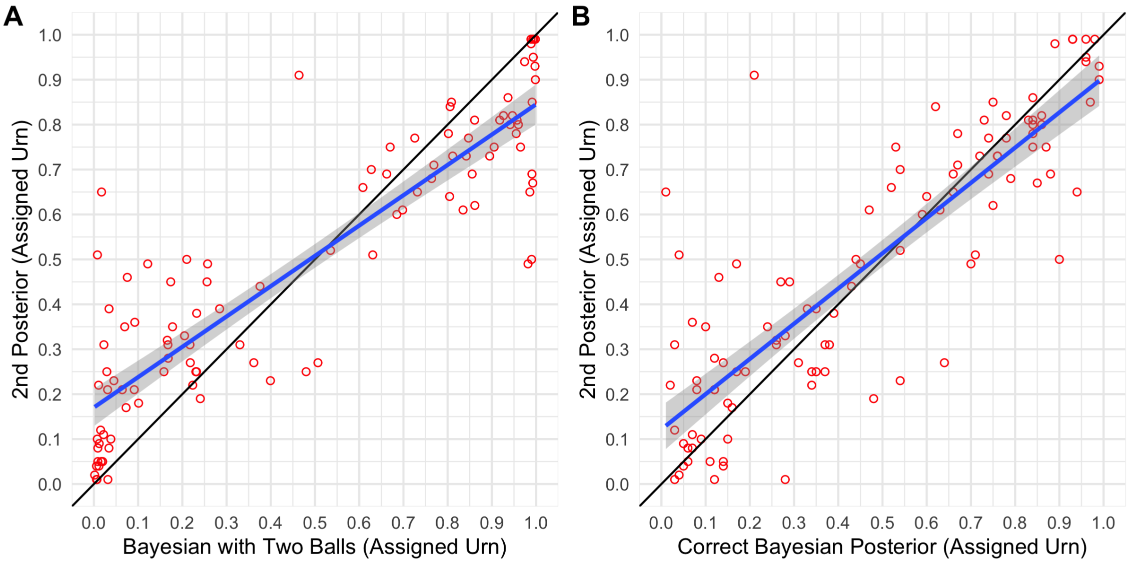


Figure 6: Fully Dissociate: (A) Two Balls from Assigned Urn. (B) Correct Bayesian.

3.3 Inferred Prior Beliefs of Other's Information

In this section, we estimate the source beliefs (p_A, p_I), probabilities subjects consider the information comes from, which reflects how subjects attribute the information to the assigned and irrelevant urn. In our experiment, it is explicitly stated that the combination of source beliefs is (0.5, 0.5). We use the four posteriors elicited (first/second phase in the assigned/irrelevant urn) to estimate subjects' (p_A, p_I) by conducting a maximum likelihood estimation.[33] We follow a structural estimation method similar to that in Costa-Gomes and Crawford (2006) but impose a logit error structure instead of spike-logit because it is hard for subjects to exactly hit the Bayesian updating prediction given the complicated Bayesian calculation.

We allow for 21 possible types, ranging from p_A = 0, 0.05..., to 1.[34] We assume that each subject's updating behavior is fixed across the 10 rounds. Formally, let $k = 0, 5, ..., 100$ (which stands for the source belief p_A from 0%, 5%, ..., to 100%) index our types, $R = 20$ denote the total number of elicited probabilities (since each round consists of two updating decisions),[35] and x_r^i denote subject i's posteriors in choice r. Given subject's type and information received, let $t_r^{i,k}$ denote the predicted posterior for a type-k subject i in round r. In order to interpret the pattern of

[33] To properly investigate individual "updating" types, we use subjects' first posteriors to calculate the target second posteriors, otherwise it could be problematic for those who deviate from the Bayesian posteriors in the first phase. For example, subject who report 60% as posteriors of the irrelevant urn and 38% as posteriors of the assigned urn in both phases is actually behaving as an "ignoring" type in the second phase. However, if we use the correct Bayesian posteriors in the first phase as benchmarks to calculate the second phase posteriors, we will mistakenly believe this subject is perfectly Bayesian.

[34] It is unnecessary to divide the types further since different p_A would map into the same combination of balls. For example, suppose one subject has the balls 30 and 70 in the first and second phase, respectively. The Bayesian posteriors are 0.38 for the assigned urn and 0.61 for the irrelevant urn if $p_A = 0.5$. If $p_A = 0.51$, the corresponding posteriors hardly change, so we cannot distinguish the subject's type.

[35] We assume that all posteriors are updated independently.

deviations from one's updating, we specify a logit error structure in which, in every particular round, a subject updates to the exact predicted posterior of one's type with highest probability, and the probability decreases as we move away from the predicted posterior. In particular, a type-k subject's assigned urn posterior in round r satisfies the logit density function $d_r^k(x_r^i, t_r^{i,k}, \lambda)$ with precision parameter λ:

$$d_r^k(x_r^i, t_r^{i,k}, \lambda) \equiv \frac{\exp\left[\lambda E(x_r^i | t_r^{i,k})\right]}{\sum_{z_r^i} \exp\left[\lambda E(z_r^i | t_r^{i,k})\right]}. \tag{7}$$

where the expected payoff $E(x|t_r^{i,k}) = x \cdot t_r^{i,k} + (1-x) \cdot (x+1)/2$, the actual payoff subjects earn in the experiment. Therefore, the density of a type-k subject with updates $\boldsymbol{x^i} \equiv (x_1^i, ..., x_R^i)$ is

$$d^k(x^i, t^{i,k}, \lambda) \equiv \prod_r d_r^k(x_r^i, t_r^{i,k}, \lambda). \tag{8}$$

Let p^k denote a subject's prior probability of being type-k, with $\sum_{k=1}^K p^k = 1$ and $\boldsymbol{p} \equiv (p^1, ..., p^K)$. By multiplying the right hand-side of (7) by p^k, summing over k and taking logarithms, the log-likelihood function for subject i becomes

$$\ln\, L(p, \varepsilon, s | x^i) = \ln \left[\sum_{k=1}^{K} p^k d^k(x^i, t^{i,k}, \lambda) \right]. \tag{9}$$

Given the estimate of λ, it is clear from (9) that the maximum likelihood estimate of p sets $p^k = 1$ for the generically unique k that yields the highest $d^k(x^i, t^{i,k}, \lambda)$. The maximum likelihood estimate of λ is the logistic scale parameter describing the spreading of subject's updating.

Figure 7A shows that on average subjects assign different weights when facing conflicting and confirming information. The weight is $p_A = 32\%$ (median = 20%) when estimated using only rounds in which the information is conflicting, but it increases to $p_A = 44\%$ (median = 45%) when using rounds in which information in confirming. The difference of subject beliefs between confirming and conflicting is significant ($44\% \gg 32\%$: t-test $p < 0.001$; Wilcoxon signed-rank test $p = 0.003$), suggesting the occurrence of an echo chamber effect.

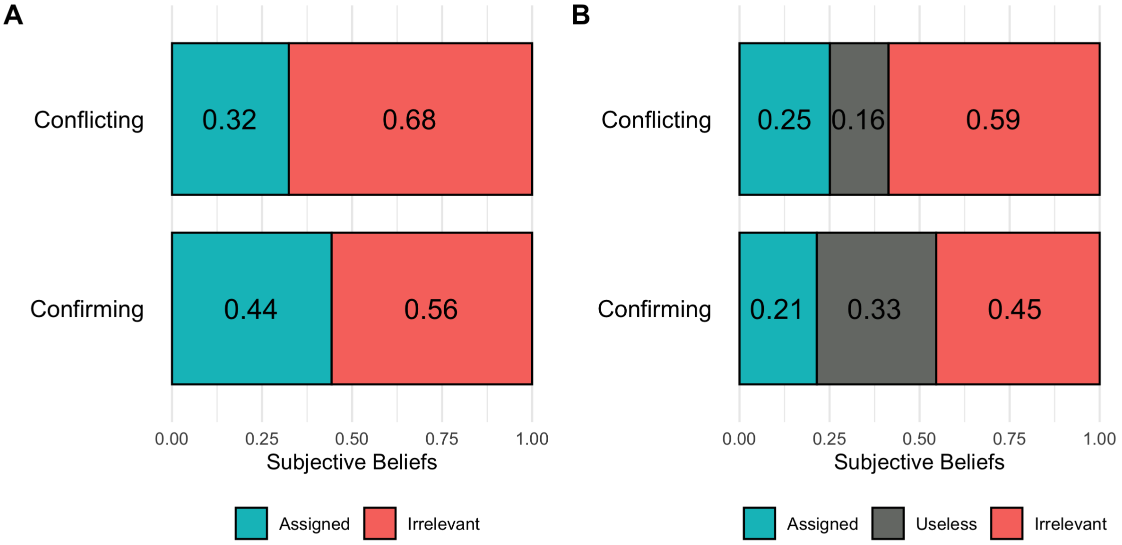


Figure 7: Models of Information Sources: (A) Two Urns (B) Three Urns.

The above model restricts the sum of p_A and p_I to necessarily equal to one, which implies the information must originate from either the assigned or irrelevant urn. This assumption adheres to our experimental design. However, people may underweight others' information. Also, notice that subjects do not always update correctly compared to Figure 2A. Therefore, subjects may believe that the information received does not coincide with a ball drawn from one of the two urns. As a result, they might decide to discount or even ignore this information completely when updating their beliefs in the second phase.

We can modify our model to accommodate the possibility of under-weighting information. Subjects may view the information as useless for making any inference, and thus ignore and attribute it to a "useless urn" added to our model to deal with such situations. If the information comes from the useless urn, each ball is drawn with equal probability. In other words, this information is completely random and not helpful to update any posteriors at all. The theoretical predictions of $\Pr(s_2|s_1, \theta_{\max})$ derived in equation (14) becomes[36]

$$\begin{aligned}
&\Pr(s_2|s_1, \theta_{\max}) \\
&\quad = \Pr(s_2|s_1, \theta_{\max}, \omega_{\max}) \cdot \Pr(\omega_{\max}|s_1, \theta_{\max}) + \Pr(s_2|s_1, \theta_{\max}, \omega_{\min}) \cdot \Pr(\omega_{\min}|s_1, \theta_{\max}) \\
&\quad = \frac{1}{2}\Big[\Pr(s_2|s_1, \theta_{\max}, \omega_{\max}, \text{Assigned } s_2) \cdot p_A + \Pr(s_2|s_1, \theta_{\max}, \omega_{\max}, \text{Irrelevant } s_2) \cdot p_I \\
&\qquad + \Pr(s_2|s_1, \theta_{\max}, \omega_{\max}, \text{Useless } s_2) \cdot p_U + \Pr(s_2|s_1, \theta_{\max}, \omega_{\min}, \text{Assigned } s_2) \cdot p_A \\
&\qquad + \Pr(s_2|s_1, \theta_{\max}, \omega_{\min}, \text{Irrelevant } s_2) \cdot p_I + \Pr(s_2|s_1, \theta_{\max}, \omega_{\min}, \text{Useless } s_2) \cdot p_U \Big].
\end{aligned} \tag{10}$$

Figure 7B shows that subjects are still significantly prone to attributing information to the irrelevant urn when it is conflicting (59% $\gg$ 45%: t-test: $p = 0.001$; Wilcoxon signed-rank test: $p = 0.002$). However, this effect disappears for the assigned urn—subject beliefs of the information source are not significantly different between conflicting and confirming information (25% $\sim$ 21%: t-test and Wilcoxon signed-rank test: $p > 0.1$). Instead, the effect is entirely on the useless urn, showing

[36] Equation 10 demonstrates how to break down the probability $\Pr(s_2|s_1, \theta_{\max})$ to three urns. We can also apply the same method to the remaining three required probabilities, $\Pr(s_2|s_1, \theta_{\min})$, $\Pr(s_2|s_1, \omega_{\max})$, and $\Pr(s_2|s_1, \omega_{\min})$.

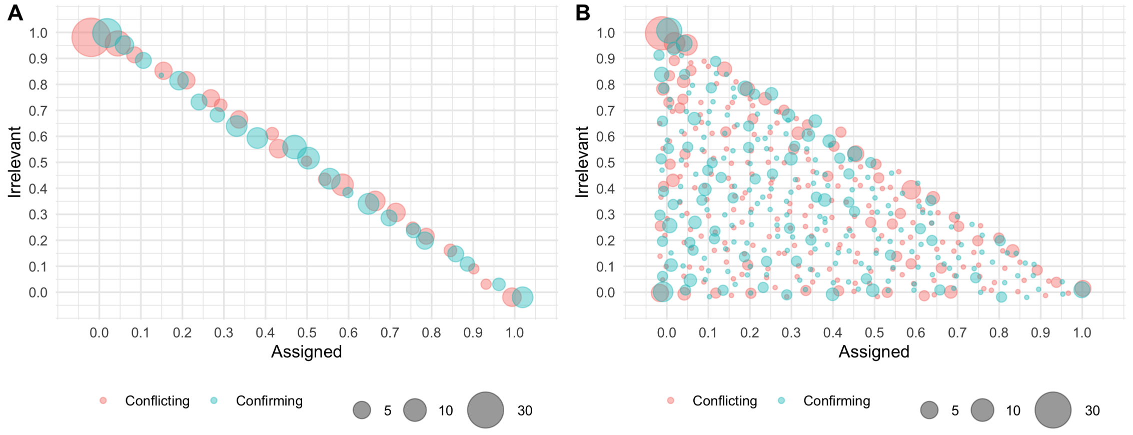


Figure 8: Information Sources Distributions: (A) Two Urns (B) Three Urns.

that subjects tend to ignore the information when it is confirming (33% $\gg$ 16%: t-test: $p < 0.001$; Wilcoxon signed-rank test: $p < 0.001$). The distributions of subjects in the two models are shown in Figure 8, and individual beliefs of the source are listed in Table 6.

To illustrate the differential processing of confirming and conflicting information, we consider three representative types: Subjects who attribute the information completely to the assigned urn ($p_A = 1$), completely to the irrelevant urn ($p_A = 0$), and those close to Bayesian ($p_A = 0.5$). Applying the same maximum likelihood estimation with these 3 types ($p_A = 0, 0.5, 1$) instead of 21 types ($p_A = 0, 0.05, \cdots, 1$), we estimate individual types and classify subjects accordingly. The results shown in Table 4 indicate that 24.4% more subjects attribute the information completely to the assigned urn when it is confirming. In contrast, 10.6% more subjects attribute the information completely to the irrelevant urn when it is conflicting. Table 4 uncovers this alternation at the individual level. Subjects along the diagonal (49.6%) are consistent under both information. Importantly, the upper triangle subjects (37.4%, underlined) put more weight on the assigned urn when moving to confirming information (from conflicting information). In other words, these subjects exhibit an "echo chamber effect," since they are more likely to believe that confirming information comes from their assigned urn and vise versa.

Table 4: Individual Type Transition: Conflicting vs. Confirming (%)

		Confirming p_A			
Conflicting	p_A	0	0.5	1	Total
	0	21.1	21.1	12.2	54.4
	0.5	6.5	25.2	4.1	35.8
	1	1.6	4.9	3.3	9.8
Total		29.3	51.2	19.5	100.0

It is apparent that subjects are not necessary consistent between belief-updating of the assigned urn and the irrelevant urn. This may be caused by the inability to properly assign probabilities between the two urns. In particular, subjects could update the two urns independently, instead of comprehensively evaluate the information and simultaneously update their beliefs about the assigned and irrelevant urn. Hence, they utilize the information and assess the probability for it to come from each urn separately. If they deem the information irrelevant, it is attributed to a useless urn, in which each ball (1 to 100) is drawn with equal chance, instead of the other urn. Therefore, subjects assign underlying beliefs (p_A, p_U) and (p_I, p_U) when assessing the assigned and irrelevant urn, respectively.

We compare underlying beliefs p_A and p_I when receiving confirming and conflicting information. Specifically, we predict underlying beliefs with a constant and the dummy for *Confirming* information to predict p_A in each round, and cluster standard errors at the subject level to control for repeated observations. We exclude choices which could only be rationalized with impossible beliefs that are not in the interval $[0, 1]$, which happens more often for the irrelevant urn. This leaves us with 846 observations for the assigned urn, in contrast to 775 observations for the irrelevant urn. Table 5 column (1) and (2) show that the directions of coefficients confirm the asymmetric updating. When the new information is aligned with their prior information, subjects put insignificantly more weight (2.4%) on the assigned urn, but significantly less (-18.7%, $p < 0.001$) weight on the irrelevant urn. However, notice that some information are more confirming or conflicting than others. For instance, when information is 51, one can hardly infers anything. Similarly, the information may not really be confirming or conflicting for subjects where the initial draws are close to 50. Thus, we regard information as strongly confirming or conflicting when neither the initial draw nor the new information are between 40 and 60. The results shown in column (3) and (4) indicated that the effects are even larger at 5.6% ($p < 0.05$) and -27.3% ($p < 0.001$) for the assigned and irrelevant urn, respectively.

Table 5: Independent Source Beliefs

Source Beliefs:	(1) Assigned Urn	(2) Irrelevant Urn	(3) Assigned Urn	(4) Irrelevant Urn
Confirming Information	0.024	-0.187***	0.056*	-0.273***
	(0.020)	(0.031)	(0.025)	(0.037)
Constant	0.155***	0.533***	0.142***	0.611***
	(0.019)	(0.028)	(0.022)	(0.033)
Stronger Confirming/Conflicting	✗	✗	✓	✓
N	846	775	555	518

Table 6: Individual Source Beliefs

	Two Urns		Three Urns					Two Urns		Three Urns			
	Conflicting	Confirming	Conflicting		Confirming			Conflicting	Confirming	Conflicting		Confirming	
ID	p_A	p_A	p_A	p_I	p_A	p_I	ID	p_A	p_A	p_A	p_I	p_A	p_I
416	0	0	0	0.25	0	1	621	0.25	0.2	0.25	0.75	0.2	0.8
111	0	0	0	1	0	1	620	0.25	0.45	0.05	0.8	0.1	0.4
115	0	0	0	1	0	1	512	0.25	0.5	0.05	0.55	0.05	0.2
508	0	0	0	1	0	1	307	0.25	0.5	0.15	0.6	0.45	0.45
519	0	0	0	1	0	1	417	0.25	0.65	0.25	0.75	0.1	0.25
604	0	0	0	1	0	1	404	0.3	0	0.3	0.6	0	1
616	0	0	0	1	0	1	109	0.3	0.45	0.3	0.7	0	0.25
212	0	0.05	0	1	0.05	0.65	503	0.35	0.35	0	0.8	0	0.25
504	0	0.05	0	1	0.05	0.95	221	0.35	0.5	0.35	0.65	0.25	0.45
511	0	0.05	0	1	0.05	0.95	407	0.35	0.55	0	0	0	0.05
217	0	0.1	0	0.95	0.05	0.85	502	0.35	0.6	0	0.4	0	0.25
301	0	0.1	0	1	0	0.85	613	0.35	0.9	0.4	0.55	0.5	0.05
313	0	0.1	0	1	0.1	0.9	316	0.4	0.4	0.4	0.6	0.4	0.6
607	0	0.2	0	0.95	0	0.55	213	0.4	0.55	0.3	0.6	0.45	0.35
210	0	0.2	0	0.95	0.1	0.8	509	0.45	0.1	0.3	0.55	0	0.65
619	0	0.2	0	1	0.2	0.75	601	0.45	0.4	0.4	0.45	0.3	0.55
218	0	0.2	0	1	0.2	0.8	317	0.45	0.45	0.45	0.55	0.2	0.45
412	0	0.35	0	0.95	0.35	0.65	617	0.45	0.5	0.45	0.55	0.4	0.35
108	0	0.35	0	1	0.05	0.15	610	0.45	0.65	0.45	0.55	0.65	0.35
614	0	0.35	0	1	0.1	0.45	517	0.45	0.75	0.45	0.55	0	0.1
611	0	0.35	0	1	0.2	0.65	117	0.5	0.55	0.5	0.5	0.3	0.25
310	0	0.4	0	0.9	0.35	0.6	214	0.55	0.5	0.5	0.45	0.1	0.2
516	0	0.4	0	1	0.15	0.45	211	0.55	0.7	0.1	0	0.05	0
320	0	0.45	0	0.45	0.1	0.4	622	0.6	0	0	0	0	1
202	0	0.45	0	1	0.3	0.5	311	0.6	0.3	0.6	0.4	0.3	0.7
314	0	0.65	0	0.85	0.35	0.2	312	0.6	0.35	0.6	0.4	0.35	0.65
103	0	0.65	0	0.95	0.6	0.25	521	0.6	0.4	0.6	0.4	0.2	0.55
414	0	0.65	0	1	0.05	0.25	319	0.6	0.45	0.6	0.4	0.45	0.55
102	0	0.7	0	1	0.65	0.2	507	0.6	0.45	0.6	0.4	0.45	0.55
624	0	0.75	0	0.5	0	0	513	0.6	0.45	0.6	0.4	0.45	0.55
306	0	0.8	0	1	0.3	0.1	625	0.6	0.55	0.05	0	0	0.2
501	0	0.85	0	1	0.5	0	603	0.6	0.55	0.35	0	0	0
208	0	1	0	1	0.4	0	118	0.65	0	0.65	0.35	0	1
203	0	1	0	1	0.5	0	114	0.65	0.35	0.55	0.25	0	0.4
216	0	1	0	1	1	0	406	0.65	0.45	0.4	0	0	0.1
205	0.05	0	0.05	0.95	0	0.8	201	0.65	0.45	0.5	0.25	0	0.35
318	0.05	0	0.05	0.95	0	1	411	0.65	0.5	0.55	0.3	0.15	0.35
615	0.05	0.25	0.05	0.95	0.05	0.65	104	0.65	0.65	0.65	0.35	0.4	0.35
321	0.05	0.3	0.05	0.95	0	0.5	403	0.65	1	0	0	0.8	0
116	0.05	0.5	0.05	0	0.3	0.5	520	0.7	0	0.2	0	0	1
606	0.05	0.5	0.05	0.95	0.5	0.5	608	0.7	0	0.7	0.3	0	1
515	0.05	0.55	0	0.95	0.1	0.35	612	0.7	0.4	0.7	0	0.4	0.6
609	0.05	0.65	0	0.95	0.2	0.15	605	0.7	0.45	0.55	0.15	0.05	0.25
206	0.05	0.7	0	0.8	0	0	408	0.7	0.85	0.7	0.25	0	0
209	0.05	0.8	0.05	0.95	0	0	113	0.7	0.95	0.6	0.1	0	0
305	0.05	0.85	0.05	0.95	0.05	0.05	207	0.75	0.65	0.6	0	0.05	0.05
409	0.05	0.9	0	0.95	0.25	0	309	0.75	0.9	0.55	0	0.75	0.05
413	0.05	1	0.05	0.95	1	0	410	0.8	0.05	0.65	0.1	0	0.95
505	0.1	0	0	0	0	0.9	303	0.8	0.25	0.8	0.2	0.25	0.75
402	0.1	0.05	0	0.75	0.05	0.95	215	0.8	0.55	0.75	0.2	0.45	0.3
623	0.1	0.25	0.05	0.8	0.25	0.75	405	0.8	0.75	0.2	0.1	0	0
415	0.1	0.85	0.05	0.85	0.4	0	602	0.85	0	0.85	0.15	0	0.85
304	0.15	0	0.05	0.75	0	0.95	302	0.85	0.3	0.85	0.15	0.3	0.7
219	0.15	0	0.15	0.85	0	0.85	220	0.9	0.2	0.9	0.1	0.2	0.8
105	0.15	0.8	0.15	0.85	0.55	0.15	107	0.95	0.25	0.95	0.05	0.05	0.55
518	0.15	0.95	0.05	0.7	0.65	0	618	1	0.35	0.7	0	0.35	0.6
315	0.15	1	0.15	0.85	0.15	0	106	1	0.5	1	0	0.4	0.5
308	0.2	0.05	0.2	0.8	0.05	0.95	112	1	0.55	1	0	0.35	0.35
110	0.2	0.4	0.2	0.65	0	0.3	401	1	0.8	1	0	0.8	0.2
204	0.2	0.4	0.2	0.8	0.1	0.5	510	1	1	0	0	1	0
101	0.2	0.7	0	0.45	0.25	0.1	506	1	1	1	0	1	0
514	0.2	0.8	0.2	0.8	0.3	0							

3.4 Log Odds Ratio Analysis

Although we believe the above frameworks are best suited to answering our research question regarding the beliefs of information sources, in this subsection we also employ an alternative structure for analysis in line with existing literature. This is to enable closer comparisons to previous findings and serves as a "robustness" check of sorts for some of our findings. This analytical model can be briefly summarized as separately identifying weights of different components contained in the log-odds ratio version of the Bayes Rule and was first described in Möbius et al. (2014).[37]

However, to our knowledge, previous papers using this model only had one source of information, i.e. one urn, and a binary signal structure (though potentially multiple binary signals). Thus, we extend the framework to allow for our two urn and 1,2,...,100 signal design.

A second (known) issue with applying this framework is that, as subjects may have persistent and incorrect biases, it's possible the error term in a new Bayesian update is correlated with the most recent posterior (the 'prior' for that round). A standard approach has been to use the past signals and randomized priors to instrument for the most recent posterior (Möbius et al. (2014); Barron (2021).)

As an additional partial solution, we do subsample analysis on observations who have correctly identified which "side" of 50 their posterior should land, given their initial signal. This corresponds to 95% of all observations. In addition, this endogeneity issue should not be a concern for the first round update of the assigned urn, or the second round update of the unassigned urn, where subjects have not received

[37] Barron (2021) also provides a concise summary of this model, parameter interpretation, and related literature.

any information.

Despite these issues above, this extension does not alter the core interpretation of the parameters meaningfully, allowing for comparison to previous literature.

3.4.1 First Stage Results

In the first round, where the subject receives a signal from their assigned urn, we can rewrite the log-odds ratio as:

$$log\left(\frac{P(\theta_{max}|s_1)}{P(\theta_{min}|s_1)}\right) = log\left(\frac{P(\theta_{max})}{P(\theta_{min})}\right) + log\left(\frac{P(s_1|\theta_{max})}{P(s_1|\theta_{min})}\right)$$

Assuming the subject understood the informed prior beliefs $P(\theta_{max}) = P(\theta_{min}) = 0.5$, the first term on the left is $log(1) = 0$. Thus, we can further simplify to

$$log\left(\frac{P(\theta_{max}|s_1)}{1 - P(\theta_{max}|s_1)}\right) = log\left(\frac{2s_1 - 1}{201 - 2s_1}\right)$$

Thus a test of Bayesian updating following the first signal would be akin to running a linear regression model:

$$log\left(\frac{P(\theta_{max}|s_1)}{1 - P(\theta_{max}|s_1)}\right) = \alpha + \beta \cdot log\left(\frac{2s_1 - 1}{201 - 2s_1}\right) + \epsilon_1$$

One benefit of the Möbius et al. (2014) framework is the ease of interpreting the coefficients – in particular, if $\hat{\beta}$ were estimated to be 2, that means that one piece of information for the experimental subject would be equivalent to 2 (independently

Table 7: First Stage Beliefs: Log-Odds Ratio Framework

Dependent Variable:	Assigned Urn		Unassigned Urn	
Log of Odds Ratio Belief	(1)	(2)	(3)	(4)
Log Odds of Signal	0.918***	0.914***	0.019	0.019
	(0.02)	(0.02)	(0.03)	(0.03)
Constant		0.112***		0.006
		(0.02)		(0.03)
Number of Observations	1216	1216	1216	1216
Number of Individuals	123	123	123	123
Adj-R^2	0.86	0.86	0.002	0.002

Notes: The dependent variable is the log odds ratio of the belief that the assigned urn (specifications 1 and 2) or unassigned urn (specifications 3 and 4) is following the maximum rule after observing the first signal. The independent variable is the log odds of the signal $log\left(\frac{2s_1-1}{201-2s_1}\right)$. All specifications report results from OLS and standard errors are given in parentheses and clustered at the subject (individual) level. Stars reference whether coefficient is significantly different from the expected coefficient of a perfect bayesian updater. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

drawn) pieces of the same information for a correct Bayesian updater.[38]

In this first stage analysis (see Table 7), we estimate $\hat{\beta} = 0.914$ or $\hat{\beta} = 0.918$ whether one constrains α to 0 or not, respectively. In either case the standard error (clustered at subject level) is 0.02, and thus $\hat{\beta}$ is significantly different from 1 ($p - value = 0.0001$). This represents slightly conservative updating from the prior, or in other words, an observed posterior belief too close to the prior of 50%.

Note that a perfect Bayesian updater would not alter their prediction of the unassigned urn, having received no information about it. Still, we could test how well this assumption fits our subjects by regressing:

$$log\left(\frac{P(\omega_{max}|s_1)}{1-P(\omega_{max}|s_1)}\right) = \zeta + \eta \cdot log\left(\frac{2s_1-1}{201-2s_1}\right) + \nu_1$$

[38] This was brought to our attention by Barron (2021), but can be demonstrated by the linear nature of the log-odds ratio.

In this case, both $\hat{\zeta}$ and $\hat{\eta}$ are statistically indistinguishable from 0 at the 5% level (clustering at the subject level). This is true regardless of whether we impose no constraints, or constrain either one of them to be 0.

In summary, for the first stage with one piece of objective information, it appears this framework shows a pattern of slight under-updating for the assigned urn and correct Bayesian inference for the unassigned urn.

3.4.2 Second Stage Results

For the second stage, where the subject learns the social signal of another subject's first stage belief elicitation, we need to use conditional probabilities, but it takes a familiar form:

$$log\left(\frac{P(\theta_{max}|s_1,s_2)}{P(\theta_{min}|s_1,s_2)}\right) = log\left(\frac{P(\theta_{max}|s_1)}{P(\theta_{min}|s_1)}\right) + log\left(\frac{P(s_2|\theta_{max},s_1)}{P(s_2|\theta_{min},s_1)}\right)$$

Note that the first term on the right is actually the log odd ratio elicited in the first stage. The second term on the right can be expanded and simplified down to:

$$log\left(\frac{P(\theta_{max}|s_1,s_2)}{1-P(\theta_{max}|s_1,s_2)}\right) = log\left(\frac{P(\theta_{max}|s_1)}{1-P(\theta_{max}|s_1)}\right) + log\left(\frac{3(2s_2-1)+(201-2s_2)}{(2s_2-1)+3(201-2s_2)}\right)$$

In Möbius et al. (2014) terminology, the left term would be $logit(\pi_2)$, the first term on the right $logit(\pi_1)$ and the right term is essentially their $log(\frac{q}{1-q})$, but in our case, q is not a constant as s_1 and s_2 are not binary signals. But otherwise they represent similar concepts (log odds ratio of posterior, prior, and signal respectively).

Thus, one equation we could estimate in this framework would then be:

$$\begin{aligned}
log\left(\frac{P(\theta_{max}|s_1,s_2)}{1-P(\theta_{max}|s_1,s_2)}\right) = \delta \cdot log\left(\frac{P(\theta_{max}|s_1)}{1-P(\theta_{max}|s_1)}\right) + \\
+\gamma_{confirm} \cdot log\left(\frac{3(2s_2-1)+(201-2s_2)}{(2s_2-1)+3(201-2s_2)}\right) \cdot 1[confirm] \\
+\gamma_{conflict} \cdot log\left(\frac{3(2s_2-1)+(201-2s_2)}{(2s_2-1)+3(201-2s_2)}\right) \cdot 1[conflict] + \epsilon_2
\end{aligned}$$

Where $1[confirm]$ refers to a binary variable for whether the signal and 100·first stage belief were on the same side of 50 (both above or both below 50), and $1[conflict]$ has the signal and 100·first stage belief on opposite sides of 50.[39]

However, as Möbius et al. (2014) and others have pointed out, there is a potential endogeneity issue of using lagged dependent variables on the right hand side. In particular, the potential concern is $E[log\left(\frac{\widehat{P(\theta_{max}|s_1)}}{1-P(\theta_{max}|s_1)}\right)\epsilon_2] \neq 0$. This may be the result of heterogeneous, but persistent, updating processes across individuals or consistent measurement error in self-reported beliefs (despite the incentivized methodology).

Therefore, in line with previous literature, we employ the signal actually observed as a instrument for the first stage belief of the assigned urn. One benefit of this methodology in our setting is that we have a very strong first stage result due to the relatively high information of our signal compared to a binary signal. This should reduce the inherent bias of employing instrumental variables. Another benefit of our setting is that the unassigned urn did not have any information in round 1, and thus the role for endogeneity is greatly diminished for that urn.

To further help correct for this endogeneity issue, we do subsample analysis on individuals who correctly (in accordance with Bayesian updating) identify which

[39] To avoid ambiguity, we drop any observations where the second signal is precisely 50.

side of the 50 their posterior should lie on in the first stage. In particular, if a subject receives a signal of 40 in round 1, according to Bayesian updating, their first round posterior should also be 40, as discussed previously. However, as long as their first round posterior lies below 50, then ”confirming” and ”conflicting” information will still have the same definitions as a Bayesian updater. Even though not everyone reports stage one beliefs identical to the first signal, the vast majority of observations do correctly infer that a signal above 50 should increase their posterior to be above 50, and a signal below 50 should decrease their posterior to be below 50. In particular 95% of first stage observations have this pattern, and we apply this subsample to further reduce the potential for endogeneity, as on this subsample, the employed definition of confirming and conflicting (which depend only on the signals) also corresponds to an alternative definition of confirming and conflicting (which depends on the prior beliefs).

Generally, as indicated in Table 8, subjects seem to have asymmetric beliefs for confirming and conflicting information in the assigned urn, and seem much more prone to attribute confirming information to their assigned urn's state than conflicting information. But in the unassigned urn, the beliefs are closer to Bayesian updating, with too little weight put on their self-reported stage 1 beliefs of urn 2. Though not significant, it appears subjects may place slightly less weight on confirming information than confirming (i.e. confirming information weight is not significantly different from 1, but conflicting information weight is significantly different from 1, $p < 0.01$).

Table 8: Second Stage Beliefs: Log-Odds Ratio Framework

Dependent Variable:	Assigned Urn				Unassigned Urn	
Log of Odds Ratio	(1)	(2)	(3)	(4)	(5)	(6)
(Belief Urn Follows Max)	OLS	OLS	IV (3SLS)	IV (3SLS)	OLS	OLS
Log Odds of	0.687***	0.604***	0.568***	0.566***	1.09	1.10
Confirming Signal	(0.09)	(0.10)	(0.07)	(0.08)	(0.13)	(0.12)
Log Odds	0.260***	0.339***	0.351***	0.364***	1.21	1.25
Conflicting Signal	(0.10)	(0.10)	(0.07)	(0.07)	(0.07	(0.08)
Log Odds of First Stage	0.796***	0.826***	0.861***	0.849***	0.494***	0.498***
Posterior Belief	(0.04)	(0.03)	(0.02)	(0.02)	(0.13)	(0.14)
Sample	Full Sample	Correct Direction	Full Sample	Correct Direction	Full Sample	Correct Direction
p-value for asymmetry	0.001	0.045	0.064	0.094	0.33	0.24
First-Stage F-Stat			6,930.6	12,665.10		
Number of Observations	1185	1126	1185	1126	1188	1129
Number of Individuals	123	123	123	123	123	123
Adj-R^2	0.73	0.86	0.73	0.75	0.44	0.46

Notes: The dependent variable is the log odds ratio of the belief that the assigned urn (specifications 1 through 4) or unassigned urn (specifications 5 and 6) is following the maximum rule after observing the first signal. The independent variable is the log odds of the signal interacted with a binary indicator for whether it was "confirming" or "conflicting" information (on the same side of 50 as the first stage posterior belief, as described in more detail on preceding pages). Due to ambiguity of definition, all second round signals of precisely 50 (neither confirming nor denying) are dropped from all specifications. The "Correct Direction" subsample refers to individuals who correctly report a first stage belief of the assigned urn in the same direction as the signal observed. All specifications report results from OLS or IV as indicated, with standard errors in paranetheses. For OLS standard errors are clustered at the subject (individual) level, for IV unfortunately the standard errors are not clustered, owing to limitations of stata command reg3. Stars reference whether coefficient is significantly different from 1.0 (i.e. perfect Bayesian updating). $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

4 Conclusion

In this experiment we set out to examine how people process potentially irrelevant social information when they already established pre-existing beliefs from objective information. We find strong evidence of “confirmation bias” in a ideologically neutral context, in which subjects asymmetrically update their beliefs when presented with information that supports those initial beliefs.

Most importantly, we try to explore the mechanism leading to this asymmetric updating. To uncover the mechanism behind confirmation bias, we ask subjects to report beliefs of an assigned urn, in which they have prior beliefs and a piece of potentially irrelevant information. Crucially, they also have to report beliefs of the irrelevant urn, by which we can visually observe the strength of weight they put on the potentially irrelevant information. We show that subjects overly tend to view this information as completely worthless in evaluating the assigned urn when it conflicts their prior beliefs, but overvalue it when it confirms their prior beliefs. This indicates that subjects may incorrectly infer that conflicting information is coming from a source that they should ignore.

When we allow subjects to consider social information as inherently inaccurate, the they still believe conflicting information is more likely to be coming from the irrelevant urn. These results are robust even if we assume subjects independently make decisions on the assigned and irrelevant urn. We find similar qualitative results when shifting our analysis to a log-odds ratio framework in line with the literature on motivated updating.

Although individuals in the real world may not be faced with digital urns when facing important policy questions, they do have to decide how much to trust different information sources when receiving news. This process of determining 'trust' level for a new information source is not independent of the information received, but closely tied to both the information and new beliefs. By explicitly modeling this 'trust' via information stemming from a potentially irrelevant urn, we highlight one possible reason people may stick to their political stance or beliefs on controversial issues, even leading to polarization. Our results suggest that dismissing new information when it conflicts with one's prior, via dismissing the information source, may cause over-persistence of beliefs.

References

L. Babcock, G. Loewenstein, S. Issacharoff, and C. Camerer. Biased judgments of fairness in bargaining. *American Economic Review*, 85(5):1337–1343, 1995.

B. M. Barber and T. Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1):261–292, 2001.

K. Barron. Belief updating: does the 'good-news, bad-news' asymmetry extend to purely financial domains? *Experimental Economics*, 24(1):31–58, 2021.

R. Brazil. Fighting flat-earth theory. *Physics World*, 33(7):35, 2020.

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

T. Buser, L. Gerhards, and J. Van Der Weele. Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, 56(2):165–192, 2018.

C. Camerer and D. Lovallo. Overconfidence and excess entry: An experimental approach. *American economic review*, 89(1):306–318, 1999.

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

A. Coutts. Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2):369–395, 2019.

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

S. Ertac. Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, 80(3):532–545, 2011.

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

R. G. Fryer Jr, P. Harms, and M. O. Jackson. t. *Journal of the European Economic Association*, 17(5):1470–1501, 2019.

E. L. Glaeser and C. R. Sunstein. Why does balanced news produce unbalanced views? Technical report, National Bureau of Economic Research, 2013.

A. Gotthard-Real. Desirability and information processing: An experimental study. *Economics Letters*, 152:96–99, 2017.

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

D. M. Grether. Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly Journal of Economics*, 95(3):537–557, 1980.

Z. Grossman and D. Owens. An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior & Organization*, 84(2):510–524, 2012.

R. Harris, M. Mack, J. Bryant, E. Theobald, and S. Freeman. Reducing achievement gaps in undergraduate general chemistry could lift underrepresented students into a “hyperpersistent zone”. *Science Advances*, 6(24):eaaz5687, 2020.

C. A. Holt and A. M. Smith. An update on bayesian updating. *Journal of Economic Behavior & Organization*, 69(2):125–134, 2009.

C. A. Holt and A. M. Smith. Belief elicitation with a synchronized lottery choice menu that is invariant to risk attitudes. *American Economic Journal: Microeconomics*, 8(1):110–39, 2016.

D. M. Kahan, H. Jenkins-Smith, and D. Braman. Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2):147–174, 2011.

D. M. Kahan, E. Peters, M. Wittlin, P. Slovic, L. L. Ouellette, D. Braman, and G. Mandel. The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, 2(10):732–735, 2012.

P. Koellinger, M. Minniti, and C. Schade. “i think i can, i think i can”: Overconfidence and entrepreneurial behavior. *Journal of economic psychology*, 28(4): 502–527, 2007.

B. Koszegi, G. Loewenstein, and T. Murooka. Fragile self-esteem. *Review of Economic Studies*, forthcoming.

C. G. Lord, L. Ross, and M. R. Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11):2098, 1979.

U. Malmendier and G. Tate. Who makes acquisitions? ceo overconfidence and the market's reaction. *Journal of financial Economics*, 89(1):20–43, 2008.

M. M. Möbius, M. Niederle, P. Niehaus, and T. S. Rosenblat. Managing self-confidence. *NBER Working paper*, 17014, 2014.

R. Oprea and S. Yuksel. Social exchange of motivated beliefs. *Journal of the European Economic Association*, 20(2):667–699, 2022.

A. Ortmann and R. Hertwig. The costs of deception: Evidence from psychology. *Experimental Economics*, 5:111–131, 2002.

C. L. Palmer and R. D. Peterson. Toxic mask-ulinity: The link between masculine toughness and affective reactions to mask wearing in the covid-19 era. *Politics & Gender*, 16(4):1044–1051, 2020.

J. Rowland, J. Estevens, A. Krzewińska, I. Warwas, and A. Delicado. Trust and mistrust in sources of scientific information on climate change and vaccines: Insights from portugal and poland. *Science & education*, 31(5):1399–1424, 2022.

P. Schwardmann and J. Van der Weele. Deception and self-deception. *Nature human behaviour*, 3(10):1055–1061, 2019.

M. D. Stosic, S. Helwig, and M. A. Ruben. Greater belief in science predicts mask-wearing behavior during covid-19. *Personality and individual differences*, 176: 110769, 2021.

A. Tversky and D. Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2):207–232, 1973.

Appendix

A First Phase Belief

The data points aligned with 45 degree line in the irrelevant urn, implying that subjects believe the initial draw can infer both urns. Figure 9A shows that a majority of these choices are made by different subjects and they only perform this behavior one time. Moreover, Figure 9B shows the occurred round of these choices. They do not concentrate on particular rounds, suggesting that such unusual behavior is randomly made throughout the experiment and is unlikely explained by learning effect.

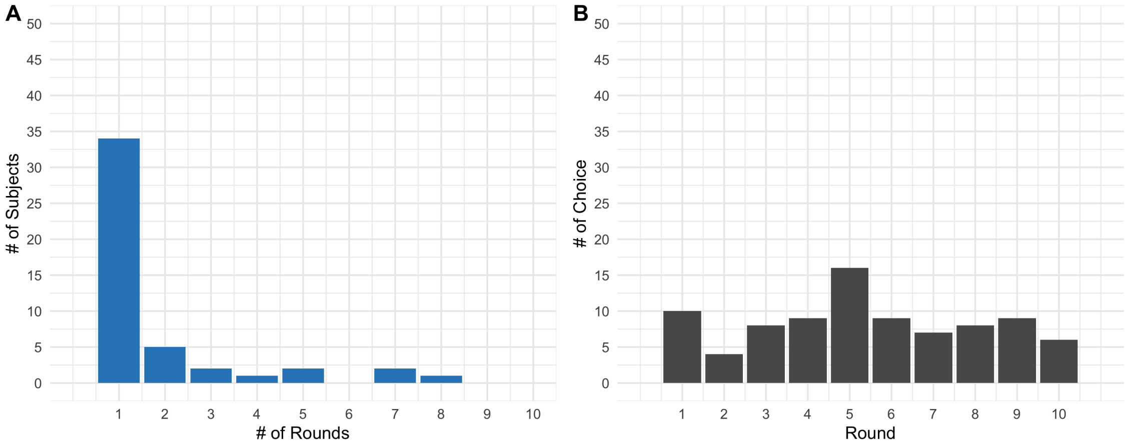


Figure 9: Beliefs Aligned with 45 Degree Line in the Irrelevant Urn. (A) the Number of Rounds (B) Occurrence Rounds.

B Second Phase Raw Data

Figure 10 shows the raw data of second phase beliefs. In particular, it is clear to see the overreaction in the irrelevant urn.

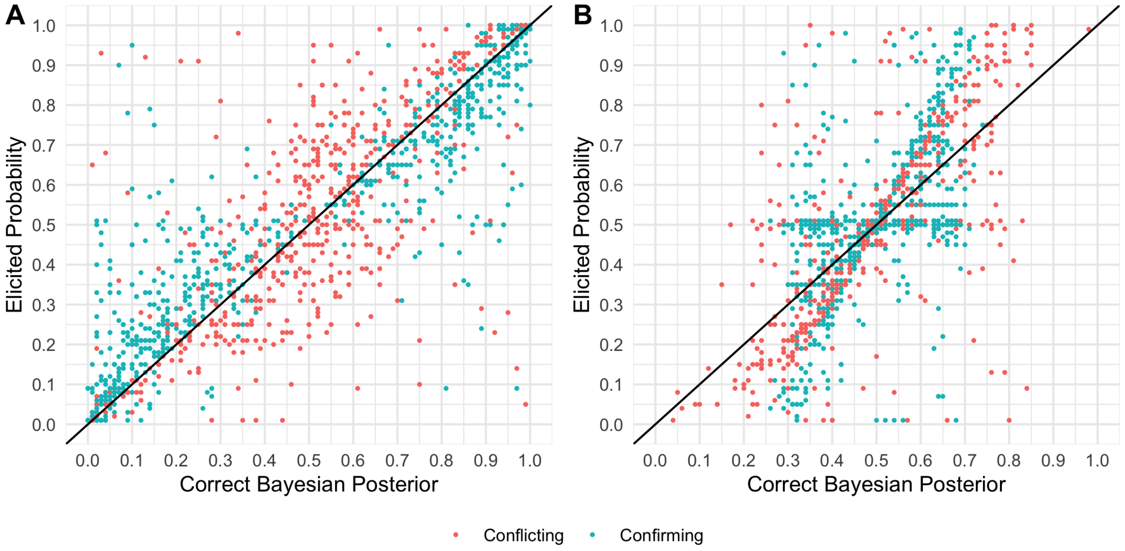
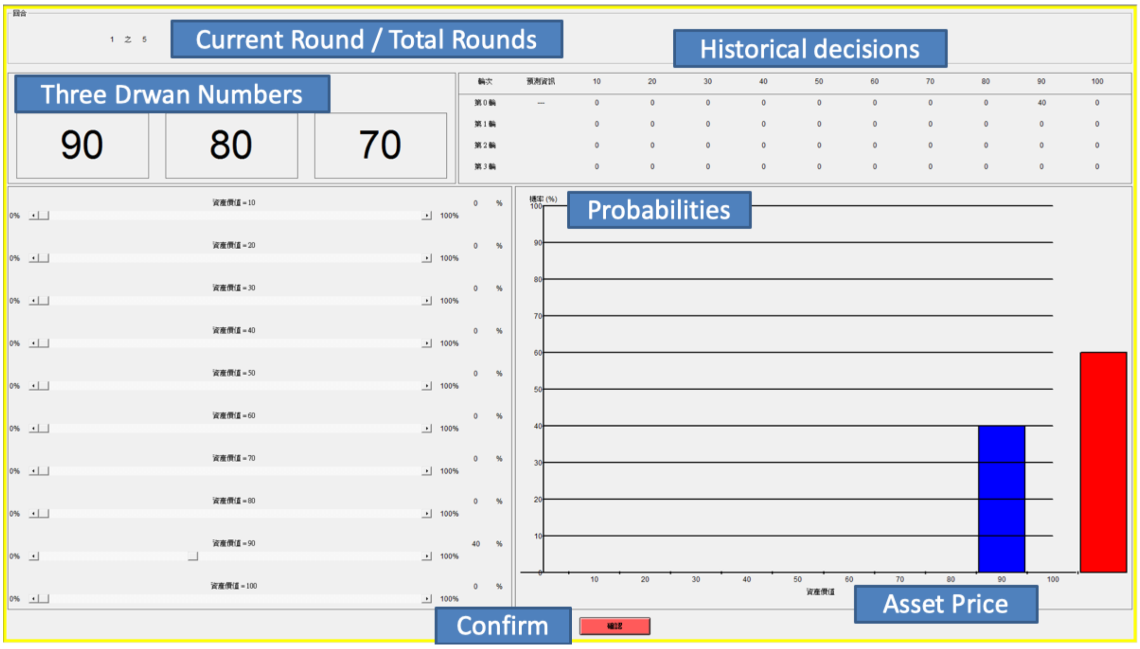


Figure 10: Elicited Beliefs in the Second Phase of the (A) Assigned (B) Irrelevant Urn

C Alternative Experimental Designs

We document alternative designs that were eventually dropped. Our first experimental design is inspired by Eil and Rao (2011). Subjects are asked to predict the real value of an asset with ten possible states. The computer randomly draws with replacement three balls from twelve, in which ten balls represent the ten possible states and the additional two balls represent the real value. Thus, the real value is drawn with probability 0.25 compared to others with 0.083. After observing their private information of three ball draws, they report their beliefs of each state that add up to 1.

Subjects then observe new information: The computer divides others into two halves, one half whose predictions are close to and the other half whose predictions are far from the subject, and randomly draws another subject from one of them to reveal his/her prediction. The procedure is repeated three times, so three other subjects' predictions will be revealed to the subject. We elicit beliefs in terms of probabilities after subjects observe each piece of information using the quadratic scoring rule. The experimental interface is shown in Figure 11.



輪次	預測資訊	10	20	30	40	50	60	70	80	90	100
第 0 輪	---	0	0	0	0	0	0	0	0	40	0
第 1 輪		0	0	0	0	0	0	0	0	0	0
第 2 輪		0	0	0	0	0	0	0	0	0	0
第 3 輪		0	0	0	0	0	0	0	0	0	0

資產價值 = 10 0% 100% 0 %

資產價值 = 20 0% 100% 0 %

資產價值 = 30 0% 100% 0 %

資產價值 = 40 0% 100% 0 %

資產價值 = 50 0% 100% 0 %

資產價值 = 60 0% 100% 0 %

資產價值 = 70 0% 100% 0 %

資產價值 = 80 0% 100% 0 %

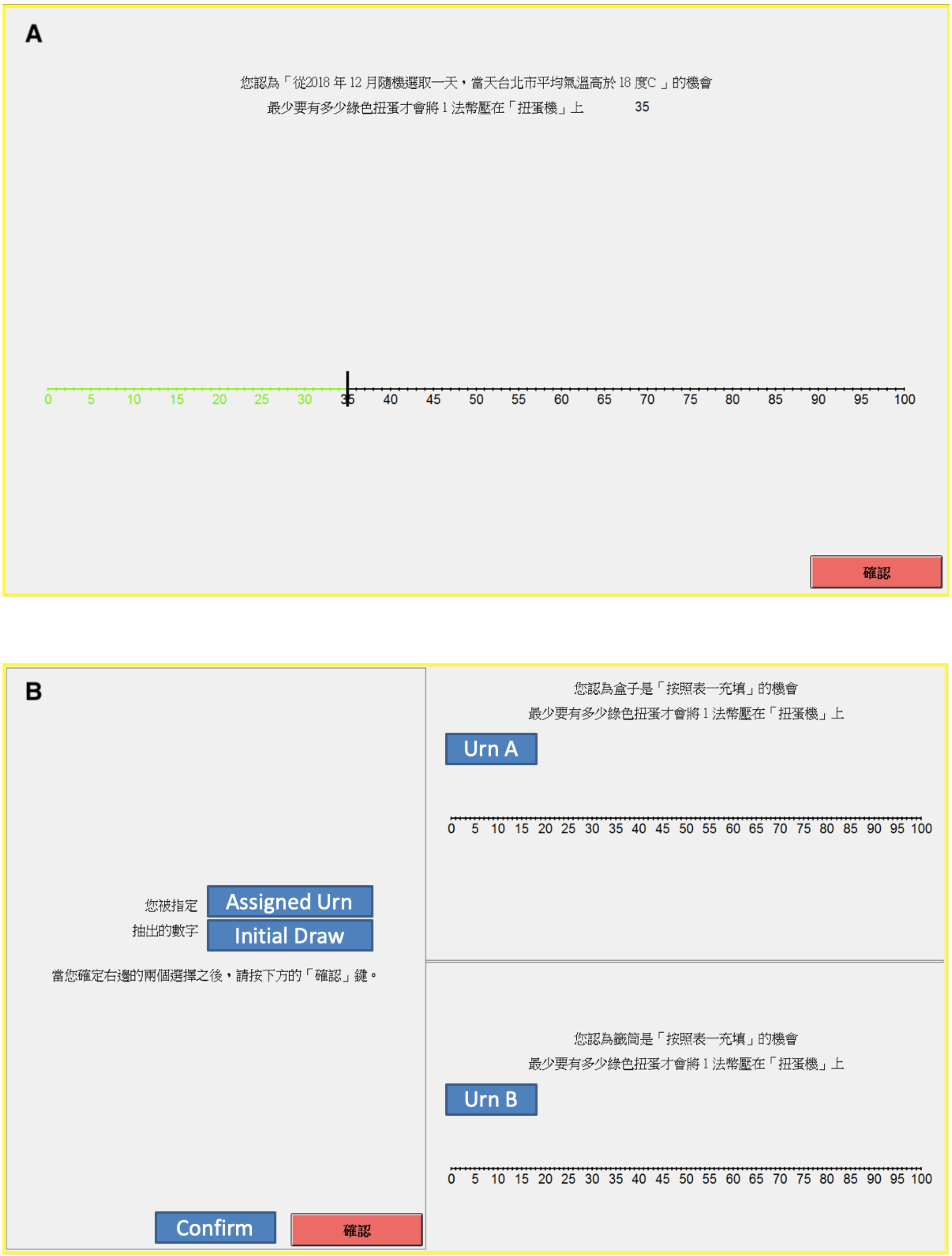
資產價值 = 90 0% 100% 40 %

資產價值 = 100 0% 100% 0 %

Figure 11: Screen Shot of the First Version Experiment.

Our second experimental design is similar to the first one, but with only two possible states. There are two urns, A and B, in the experiment. Urn A applies the *Maximum Rule* and Urn B applies the *Minimum Rule*, so each urn reports either maximum or minimum of two draws from the uniform distribution. We provide the probability table in case subjects cannot figure it out themselves. Subjects observe a ball from urn A or B with equal chance, and report the probability that the chosen urn is A. Then, subjects observe others' information and beliefs are elicited using the same design as the first version.

Our third experimental design is nearly identical to our final one implemented, but with three important differences. First of all, it is a one shot game with three stages of belief-updating, while the final experiment has ten rounds each with one stage of belief-updating. In other words, subjects observe their initial draw and then receive three other piece of information. Second, we use the BDM procedure as in Coutts (2019) to elicit beliefs, which is illustrated in Figure 12A. Finally, the probability for drawing each number under the *Maximum Rule* and *Minimum Rule* is shown in tables. The experimental interface is shown in Figure 12B.



您認為「從2018 年 12 月隨機選取一天，當天台北市平均氣溫高於 18 度C 」的機會
最少要有多少綠色扭蛋才會將 1 法幣壓在「扭蛋機」上 35

當您確定右邊的兩個選擇之後，請按下方的「確認」鍵。

您認為盒子是「按照表一充填」的機會
最少要有多少綠色扭蛋才會將 1 法幣壓在「扭蛋機」上

您認為籤筒是「按照表一充填」的機會
最少要有多少綠色扭蛋才會將 1 法幣壓在「扭蛋機」上

Figure 12: Screen Shot of Third Version Experiment.

D Bayesian Updating Equations

D.1 Round 1 Bayesian Updating

Therefore, the probability $\Pr(s_1|\theta_{\max})$ is:

$$\begin{aligned}\Pr(s_1|\theta_{\max}) &= \Pr\left(\left\{S_1^1 = s_1 \cap S_1^2 < s_1\right\} \vee \left\{S_1^1 \leq s_1 \cap S_1^2 = s_1\right\}\right)\\ &= \Pr(S_1^1 = s_1)\Pr(S_1^2 < s_1) + \Pr(S_1^1 \leq s_1)\Pr(S_1^2 = s_1)\\ &= \frac{1}{100}\cdot\frac{s_1-1}{100} + \frac{s_1}{100}\cdot\frac{1}{100} = \frac{2s_1-1}{10000} \end{aligned} \tag{11}$$

D.2 Round 2 Bayesian Updating

Their Bayesian probabilities in the second phase are:

$$\begin{aligned}\Pr(\theta_{\max}|s_1,s_2) &= \frac{\Pr(s_1\cap s_2|\theta_{\max})\cdot\Pr(\theta_{\max})}{\Pr(s_1\cap s_2)}\\ &= \frac{\Pr(s_2|s_1,\theta_{\max})\cdot\Pr(s_1|\theta_{\max})\cdot\Pr(\theta_{\max})}{\Pr(s_2|s_1,\theta_{\max})\cdot\Pr(s_1|\theta_{\max})\cdot\Pr(\theta_{\max}) + \Pr(s_2|s_1,\theta_{\min})\cdot\Pr(s_1|\theta_{\min})\cdot\Pr(\theta_{\min})}\end{aligned} \tag{12}$$

$$\begin{aligned}\Pr(\omega_{\max}|s_1,s_2) &= \frac{\Pr(s_1\cap s_2|\omega_{\max})\cdot\Pr(\omega_{\max})}{\Pr(s_1\cap s_2)}\\ &= \frac{\Pr(s_2|s_1,\omega_{\max})\cdot\Pr(s_1|\omega_{\max})\cdot\Pr(\omega_{\max})}{\Pr(s_2|s_1,\omega_{\max})\cdot\Pr(s_1|\omega_{\max})\cdot\Pr(\omega_{\max}) + \Pr(s_2|s_1,\omega_{\min})\cdot\Pr(s_1|\omega_{\min})\cdot\Pr(\omega_{\min})}\end{aligned} \tag{13}$$

where $\Pr(s_2|s_1,\theta_{\max})$

$$\begin{aligned}&= \Pr(s_2|s_1,\theta_{\max},\omega_{\max})\cdot\Pr(\omega_{\max}|s_1,\theta_{\max}) + \Pr(s_2|s_1,\theta_{\max},\omega_{\min})\cdot\Pr(\omega_{\min}|s_1,\theta_{\max})\\ &= \Pr(s_2|s_1,\theta_{\max},\omega_{\max})\cdot\frac{1}{2} + \Pr(s_2|s_1,\theta_{\max},\omega_{\min},s_2\text{ from A})\cdot p_A\cdot\frac{1}{2}\\ &\qquad + \Pr(s_2|s_1,\theta_{\max},\omega_{\min},s_2\text{ from B})\cdot p_I\cdot\frac{1}{2}\end{aligned} \tag{14}$$

Thus, we have

$$
\begin{aligned}
\Pr(s_2|s_1,\theta_{\max}) &= \frac{2s_2-1}{10000}\cdot\frac{1}{2}+\frac{2s_2-1}{10000}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{201-2s_2}{10000}\cdot\frac{1}{2}\cdot\frac{1}{2}\\
&= \frac{3}{4}\cdot\left(\frac{2s_2-1}{10000}\right)+\frac{1}{4}\cdot\left(\frac{201-2s_2}{10000}\right)\\
\Pr(s_2|s_1,\theta_{\min}) &= \frac{1}{4}\cdot\left(\frac{2s_2-1}{10000}\right)+\frac{3}{4}\cdot\left(\frac{201-2s_2}{10000}\right) \qquad (15)
\end{aligned}
$$

Equation 14 indicates the weightings that s_2 is under *Maximum Rule* or *Minimum Rule*. Since it is given the state of A is *Maximum Rule*, $\theta_{\max}$, only the state of B remains uncertain. By the settings of experimental design, there is equal chance that s_2 is either from urn A or urn B. It is the only possibility that s_2 is drawn under *Minimum Rule* when s_2 is from urn B and urn B is applied to *Minimum Rule*. Therefore, s_2 is drawn under *Maximum Rule* with 75% chance and *Minimum Rule* with 25% chance. With similar reason, we can also derive the probability in equation 15. The combination of probabilities (p_A, p_I) is the weights of the information source, indicating that the probability that new information is from the assigned urn or irrelevant urn. It is (0.5, 0.5) since the randomly drawn subject has equal chance to be assigned to urn A or B.[40]

[40] In the experiment, subjects were assigned randomly to urns independently and were informed about this. However, it may be possible for subjects to incorrectly infer that exactly half the subjects were assigned to each urn, and thus the average subject would infer ex ante s_2 is more likely to comes from the unassigned urn. Yet the sample size for each session was large, about 20 subjects, so this would result in a small modification (55% urn B and 45% urn A). Importantly, this incorrect ex ante inference would not differ by confirming and conflicting information, but to be thorough we allow for and estimate non-equal priors as discussed in the Results section.

The following equations show the results of $\Pr(s_2|s_1,\omega_{\max})$ and $\Pr(s_2|s_1,\omega_{\min})$.

$$
\begin{aligned}
&\Pr(s_2|s_1,\omega_{\max})\\
&\quad=\Pr(s_2|s_1,\omega_{\max},\theta_{\max})\cdot\Pr(\theta_{\max}|s_1,\omega_{\max})+\Pr(s_2|s_1,\omega_{\max},\theta_{\min})\cdot\Pr(\theta_{\min}|s_1,\omega_{\max})\\
&\quad=\Pr(s_2|s_1,\omega_{\max},\theta_{\max})\cdot\frac{2s_1-1}{200}+\Pr(s_2|s_1,\omega_{\max},\theta_{\min},s_2\text{ from A})\cdot p_A\cdot\frac{201-2s_1}{200}\\
&\qquad+\Pr(s_2|s_1,\omega_{\max},\theta_{\min},s_2\text{ from B})\cdot p_I\cdot\frac{201-2s_1}{200}\\
&s=\frac{2s_2-1}{10000}\cdot\frac{2s_1-1}{200}+\frac{201-2s_2}{10000}\cdot\frac{1}{2}\cdot\frac{201-2s_1}{200}+\frac{2s_2-1}{10000}\cdot\frac{1}{2}\cdot\frac{201-2s_1}{200}\\
&\quad=\frac{2s_1-1}{200}\cdot\left(\frac{2s_2-1}{10000}\right)+\frac{201-2s_1}{200}\cdot\left(\frac{1}{100}\right)
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
&\Pr(s_2|s_1,\omega_{\min})\\
&\quad=\frac{2s_1-1}{200}\cdot\left(\frac{1}{100}\right)+\frac{201-2s_1}{200}\cdot\left(\frac{201-2s_2}{10000}\right)
\end{aligned}
\tag{17}
$$

Equation 16 also shows the weightings that s_2 is under *Maximum Rule* or *Minimum Rule* but given the state of urn B, $\omega_{\max}$, instead of the state of urn A, $\theta_{\max}$. We can divide the equation into two parts, the state of urn A is either *Maximum Rule* or *Minimum Rule*. First of all, when the state of urn A is *Maximum Rule*, with the probability derived in equation 12, it is for sure that s_2 is drawn under *Maximum Rule*. Secondly, when the state of urn A is *Minimum Rule*, there is equal chance to draw s_2 under *Maximum Rule* or *Minimum Rule*. Thus, the probability of observing s_2 given states of u two urns $\omega_{\max}$ and $\theta_{\min}$ is the same as the probability of observing s_2, 1%. Equation 17 is derived by the same thoughts.