Asymmetric Failure of Bayesian Updating and the Echo Chamber Effect: An Experimental Study

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

February 15, 2022

Abstract

We conducted a laboratory experiment to investigate individual ability to process contradicting 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 randomly. 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 process-

ing; biased interpretation

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. For example, if new information conflicts with prior beliefs, people may suspect that it is driven by 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 regardless, incorrect beliefs persists for much longer if people ignore information because it *may* be irrelevant. This could lead to the formation of persistent echo chambers.

In this paper, we examine a laboratory experiment investigating people's ability to process new information, and study the difference in belief updating when information aligns with or is against the prior belief. Specifically, we consider a two step procedure, in which subjects first independently draw information from one of two 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 irrelevant.

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 (and hence, irrelevant) urn. Conversely, in the face of confirming information, subjects are comparatively more likely to attribute it to their own 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.[1] One strand of this litera-

[1]It is well worth noting that psychology had noted a similar process as a subset of "confirmation

ture 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), Schwardmann and Van der Weele (2019), and Barron (2021) 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 the COVID-19 pandemic. In these settings, information does not pertain to only the receiver. However, experimentally manipulating such belief updating may be difficult, as the subject may have strong priors from a large amount of information (or disinformation).[2] The large dimensionality of these risks also make it difficult to assess the correct measure for the belief. Lastly, it may be difficult to control subjects' prior beliefs and perceptions of information source objectivity. Thus, we seek to explore the asymmetric updating of beliefs in a more context-neutral setting, to give a better picture of how the initial belief evolution occurs. This also allows for a mixture of objective and social information to be processed as well.

Yet this paper is not the first to explore asymmetric updating in a neutral context. The most closely related is Coutts (2019), which tested for asymmetric updating across ego relevant, financial, and context-neutral settings and found consistent

bias" outside of a Bayesian framework, c.f. Lord et al. (1979).

[2]ADD FOOTNOTE ON Lord 1979 (?)

asymmetric updating across all three domains. However, in their design, all the information was non-social, being provided by a computer rather than another human. In addition, the binary information was either confirming the belief or conflicting it—there was no situation in which the information could be uninformative or irrelevant.

In comparison, this paper provides evidence that socially-processed information also results in asymmetric updating. [XXX DOUBLE CHECK XXX] 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 design, we can distinguish between whether subjects (i) process the new information but misattribute its source or (ii) completely disregard the conflicting inofrmation. In general, we find [XXX INSERT HERE XXX]

In a concurrent study, Oprea and Yuksel also explores social evolution of beliefs in a laboratory context. The paper's primary focus is on the evolution of "motivated" beliefs – beliefs that subjects may have nonpecuniary incentives to believe as true. In the paper's primary experiments, subjects take an intelligence quiz, and then paired with a partner on the same side of the median score. Each subject reports their beliefs 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 stable overconfidence. Furthermore, their paper demonstrates that public noisy signals are ineffective at reducing this bias.

In comparison, our research is more context neutral, as subjects primary motivation to be assigned the ”maximum” rule urn is pecuniary in nature. Our paper also focuses more 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]

(Burks et al., 2013)

Fryer Jr et al. (2019) introduce 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 an ambiguous signal as a certain signal for a particular state based on their current beliefs. Their online Amazon Mechanical Turk experiments show that 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 contrast, we provide three main contributions. First, the polarizing beliefs in Fryer Jr et al. (2019) stem from ambiguous information that is incorrectly inferred to be informative. In our paper, we explore how individuals incorporate information that is contradictory 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

[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 an echo chamber on social media.

[4] Specifically, in 2-way communication, a pessimistic belief about group's intelligence status may be perceived as an insult. 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, such as dinner with the in-laws.

context with objective outcomes, as opposed to the politically charged context with subjective outcomes (as the scale of interpretation in Fryer Jr et al. (2019) may be itself differently interpreted based on prior beliefs). Lastly, we provide evidence that contradictory information is not misinterpreted as consistent with beliefs, but is instead tends to be misattributed as irrelevant. 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 relative paucity of ambiguous information in climate change, it may be that individuals instead infer that unambiguous contradictory information is instead from an untrustworthy or irrelevant source.

Failure of Bayesian updating is documented in several papers, including Tversky and Kahneman (1973) and Grether (1980). They find that subjects ignore base-rate information contrary to Bayes rule, resulting in representativeness heuristic. Holt and Smith (2009) show that subjects tend to over/underweight low/high prior probabilities. Compare to cognitive incompetence to perform Bayes rule, recent studies focus on asymmetrically processing information. Eil and Rao (2011) investigate how subjects update differently between neutral and ego-relevant information like beauty and intelligence. They find that subjects respond less when the information is bad (suggesting one's beauty or intelligence is inferior) and this effect only occurs in non-neutral settings. However, Coutts (2019) do not find the "good news-bad news effect" in their experiments. How people process ego relevant information is still debating.

Besides, our paper is an extension of rich literature of confirmation bias, which was documented in economics (Babcock et al., 1995) and psychology (Lord et al., 1979). Confirmation bias describes people's tendency to interpret the 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 leads to polarization. They suggest that the same information have diametrically opposite effect for those who have confirming and conflicting priors. Our experiment provides experimental evidence and illustrates a possible mechanism of this phenomenon.

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.[5] In the second phase, each subject observes another 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 additional piece of subject-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.[6]

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.

[5]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.

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

In each round, urn A and urn B are independently randomized to follow one of two "rules" with equal chance. Subjects are not told the which rules the urn 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 urn A. Similarly, they also are incentivized to predict the probability that *Maximum Rule* is applied to urn B.[7]

In the second phase, for each subject, the computer randomly chooses another subject, and reveals the first phase prediction of the other-subject-assigned urn. However, even though the subject observe this prediction, they do not know if this other subject was assigned to urn A or urn B.[8] After seeing the information from another subject, subjects again predict the probability that the *Maximum Rule* is

[7]Because the 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.

[8]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% probablity to have been assigned to either urn. Subjects are informed of this statistical independence.

applied to each urn.

2.2 Belief Elicitation

Following Holt and Smith (2016), we use a 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.[9]

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[10] prefer the

[9] Minimizing belief error through a more

[10] One might be concerned about the potential for ambiguity aversion to distort these probabilities. Though with proper understanding of Bayesian updating, it's worth noting that the event lottery is not truly ambiguous in this experiment, though the difficulty in 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.

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 methodology however, there can only be allowed one "switching point". This removes the potential for non-monotonic behavior, though this loss is arguably a good thing and may explain why the 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.

To be clear, this method is incentive compatible. Suppose one under-reports her beliefs from her real belief, 80%, to misreported belief, 60%. The results is the same when the drawn number is less than 60 and greater than 80. However, it is disadvantage for her if the number falls into the interval between 60 and 80. Since the event lottery will be chosen if she truthfully report the belief, and in that case the probability of getting the prize is 80%. Conversely, in the case of misreporting, the random lottery will be chosen and its probability of getting the prize is between 60% and 80%. Therefore, truthfully reporting the belief is in the best interests for all subjects.

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).[11] 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.

2.4 Bayesian Probability Predictions

For notation simplicity, we let urn A be the assigned urn and urn B be the irrelevant 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 states, *Maximum Rule* and *Minimum Rule*,

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

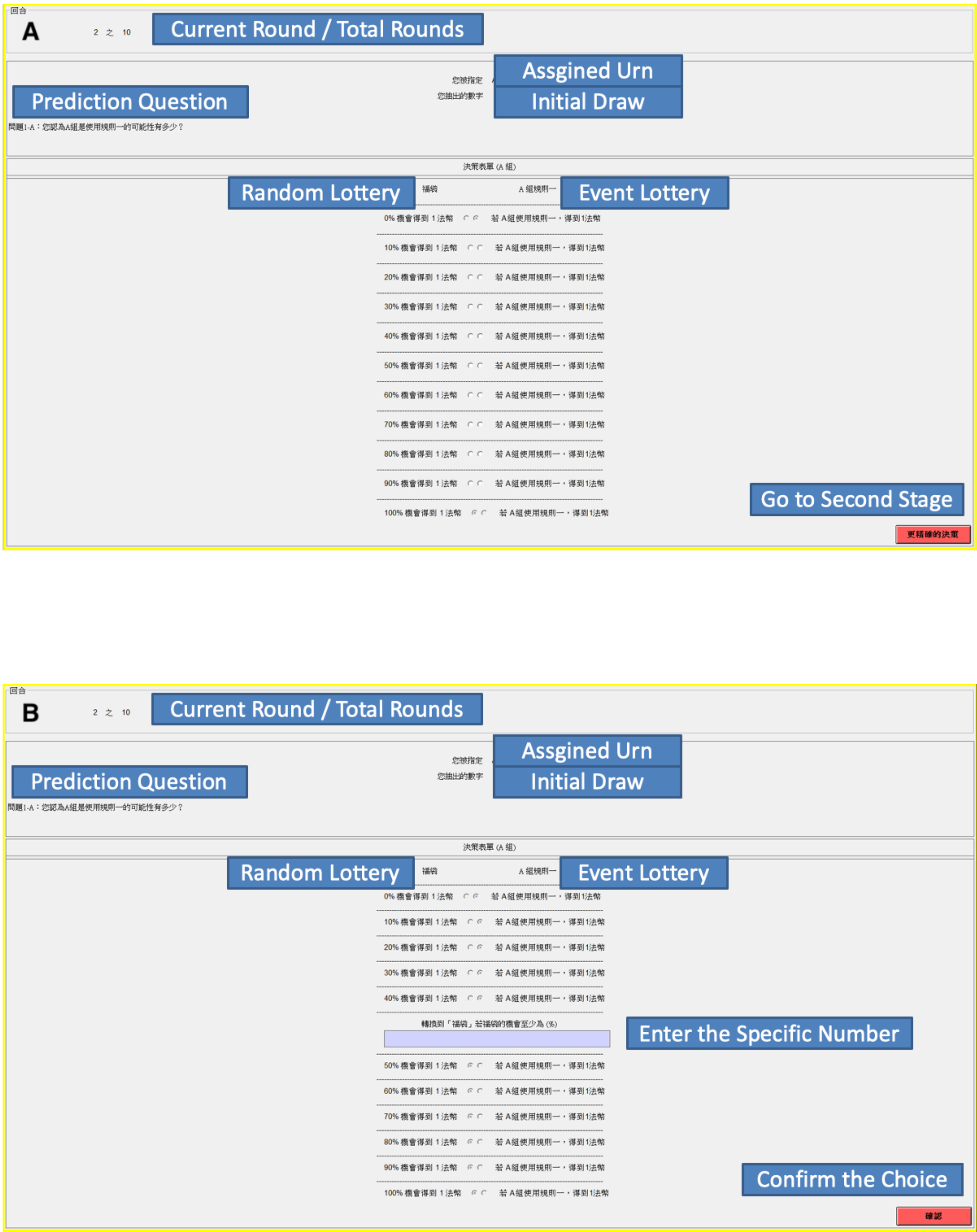

Figure 1: Two-stage Menu of Lottery Choices: (a) 1st Stage, and (b) 2nd stage.

indicated by $\omega_{\max}$ and $\omega_{\min}$. The information s_1 denotes the observed ball in the first phase, s_2 is elicited probability of the assigned urn from another subject 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 probability $\Pr(s_1|\theta_{\max})$ is:

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

Similarly, the other probability is $\Pr(s_1|\theta_{\min}) = (201 - 2s_1)/10000$. 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.

$$\Pr(\theta_{\max}|s_1) = \frac{\Pr(s_1|\theta_{\max})\Pr(\theta_{\max})}{\Pr(s_1)} = \frac{(2s_1-1)/10000}{1/100} \cdot \frac{1}{2} = \frac{s_1}{100} - \frac{1}{200} \tag{2}$$

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.[12] However, in the experiment, fractional percentages were not allowed in the elicitation, requiring subjects to round to the nearest whole percentage term (i.e. 55% instead of 54.5%).[13] Thus reporting s_1 (in percentage terms) would be a correct Bayesian posterior given the constraints.[14]

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

[12]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, that is 100.

[13]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.

[14]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.

the *Minimum Rule*)

- the unobserved ball was exactly 75 (drawn twice due to replacement)

As the balls 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.

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. To some extent, we see that the average subject is close to a Bayesian updater in phase 1. However, the belief is less clear.

Their Bayesian probabilities in the second phase are:

$$\begin{aligned}\Pr(\theta_{\max}|s_1,s_2) &= \frac{\Pr(s_1\cap s_2|\theta_{\max})\cdot\Pr(\theta_{\max})}{\Pr(s_1\cap s_2)}\\ &= \frac{\Pr(s_2|s_1,\theta_{\max})\cdot\Pr(s_1|\theta_{\max})\cdot\Pr(\theta_{\max})}{\Pr(s_2|s_1,\theta_{\max})\cdot\Pr(s_1|\theta_{\max})\cdot\Pr(\theta_{\max})+\Pr(s_2|s_1,\theta_{\min})\cdot\Pr(s_1|\theta_{\min})\cdot\Pr(\theta_{\min})}\end{aligned} \tag{3}$$

$$\begin{aligned}\Pr(\omega_{\max}|s_1,s_2) &= \frac{\Pr(s_1\cap s_2|\omega_{\max})\cdot\Pr(\omega_{\max})}{\Pr(s_1\cap s_2)}\\ &= \frac{\Pr(s_2|s_1,\omega_{\max})\cdot\Pr(s_1|\omega_{\max})\cdot\Pr(\omega_{\max})}{\Pr(s_2|s_1,\omega_{\max})\cdot\Pr(s_1|\omega_{\max})\cdot\Pr(\omega_{\max})+\Pr(s_2|s_1,\omega_{\min})\cdot\Pr(s_1|\omega_{\min})\cdot\Pr(\omega_{\min})}\end{aligned} \tag{4}$$

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

$$\begin{aligned}&= \Pr(s_2|s_1,\theta_{\max},\omega_{\max})\cdot\Pr(\omega_{\max}|s_1,\theta_{\max})+\Pr(s_2|s_1,\theta_{\max},\omega_{\min})\cdot\Pr(\omega_{\min}|s_1,\theta_{\max})\\ &= \Pr(s_2|s_1,\theta_{\max},\omega_{\max})\cdot\frac{1}{2}+\Pr(s_2|s_1,\theta_{\max},\omega_{\min},s_2\text{ from A})\cdot p_A\cdot\frac{1}{2}\\ &\qquad+\Pr(s_2|s_1,\theta_{\max},\omega_{\min},s_2\text{ from B})\cdot p_I\cdot\frac{1}{2}\end{aligned} \tag{5}$$

Thus, we have

$$\begin{aligned}\Pr(s_2|s_1,\theta_{\max}) &= \frac{2s_2-1}{10000}\cdot\frac{1}{2}+\frac{2s_2-1}{10000}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{201-2s_2}{10000}\cdot\frac{1}{2}\cdot\frac{1}{2}\\ &= \frac{3}{4}\cdot\left(\frac{2s_2-1}{10000}\right)+\frac{1}{4}\cdot\left(\frac{201-2s_2}{10000}\right)\\ \Pr(s_2|s_1,\theta_{\min}) &= \frac{1}{4}\cdot\left(\frac{2s_2-1}{10000}\right)+\frac{3}{4}\cdot\left(\frac{201-2s_2}{10000}\right)\end{aligned} \tag{6}$$

Equation 5 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 6. 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.[15]

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

$$
\begin{aligned}
&\Pr(s_2|s_1, \omega_{\max}) \\
&\quad = \Pr(s_2|s_1, \omega_{\max}, \theta_{\max}) \cdot \Pr(\theta_{\max}|s_1, \omega_{\max}) + \Pr(s_2|s_1, \omega_{\max}, \theta_{\min}) \cdot \Pr(\theta_{\min}|s_1, \omega_{\max}) \\
&\quad = \Pr(s_2|s_1, \omega_{\max}, \theta_{\max}) \cdot \frac{2s_1 - 1}{200} + \Pr(s_2|s_1, \omega_{\max}, \theta_{\min}, s_2 \text{ from A}) \cdot p_A \cdot \frac{201 - 2s_1}{200} \\
&\qquad\quad + \Pr(s_2|s_1, \omega_{\max}, \theta_{\min}, s_2 \text{ from B}) \cdot p_I \cdot \frac{201 - 2s_1}{200} \\
s &= \frac{2s_2 - 1}{10000} \cdot \frac{2s_1 - 1}{200} + \frac{201 - 2s_2}{10000} \cdot \frac{1}{2} \cdot \frac{201 - 2s_1}{200} + \frac{2s_2 - 1}{10000} \cdot \frac{1}{2} \cdot \frac{201 - 2s_1}{200} \\
&= \frac{2s_1 - 1}{200} \cdot \left(\frac{2s_2 - 1}{10000}\right) + \frac{201 - 2s_1}{200} \cdot \left(\frac{1}{100}\right)
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
&\Pr(s_2|s_1, \omega_{\min}) \\
&\quad = \frac{2s_1 - 1}{200} \cdot \left(\frac{1}{100}\right) + \frac{201 - 2s_1}{200} \cdot \left(\frac{201 - 2s_2}{10000}\right)
\end{aligned}
\tag{8}
$$

Equation 7 also shows the weightings that s_2 is under *Maximum Rule* or *Minimum*

[15] 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.

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 3, 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 8 is derived by the same thoughts.

Hence, the Bayesian probability prediction for urn A (the assigned urn) is:

$$\begin{aligned}&\Pr(\theta_{\max}|s_1, s_2)\\&= \frac{\left[3(2s_2-1)+(201-2s_2)\right](2s_1-1)}{\left[3(2s_2-1)+(201-2s_2)\right](2s_1-1)+\left[(2s_2-1)+3(201-2s_2)\right](201-2s_1)}\end{aligned} \tag{9}$$

By substituting equation 7 and 8 into 4, the Bayesian probability prediction for urn B (the irrelevant urn) is as follows.

$$\begin{aligned}&\Pr(\omega_{\max}|s_1, s_2)\\&= \frac{(2s_2-1)(2s_1-1)+100\cdot(201-2s_1)}{(2s_2-1)(2s_1-1)+100\cdot(201-2s_1)+100\cdot(2s_1-1)+(201-2s_2)(201-2s_1)}\end{aligned} \tag{10}$$

Alternatively, we can derive probabilities, $\Pr(s_2|s_1, \theta_{\max})$ and $\Pr(s_2|s_1, \omega_{\max})$ by the source of other's information. It is beneficial for analyzing how subjects consider other's information. Equation 11 and 12 show above concept. Exploiting the first ball and consequent beliefs, subjects form probabilities that second ball comes from urn A and B. Depends on two balls, subjects may distort p_A and p_I, both of which are 0.5 and p_A is equal to $(1 - p_I)$ in theory.

$$
\begin{aligned}
\Pr(s_2|s_1, \theta_{\max}) &= \Pr(s_2|s_1, \theta_{\max}, s_2 \text{ from A}) \cdot p_A + \Pr(s_2|s_1, \theta_{\max}, s_2 \text{ not from A}) \cdot (1 - p_A) \\
&= \frac{2s_2 - 1}{10000} \cdot p_A + \frac{1}{100} \cdot (1 - p_A) \qquad (11) \\
\Pr(s_2|s_1, \omega_{\max}) &= \Pr(s_2|s_1, \omega_{\max}, s_2 \text{ from I}) \cdot p_I + \Pr(s_2|s_1, \omega_{\max}, s_2 \text{ not from I}) \cdot (1 - p_I) \\
&= \frac{2s_2 - 1}{10000} \cdot p_I + \frac{1}{100} \cdot (1 - p_I) \qquad (12)
\end{aligned}
$$

In equation 5 and 7, it is assumed that subjects update posteriors of two urns together. In other words, they rationally assign probabilities p_A and p_I so that the sum of p_A and p_I is always equal to 1. Thus, if the information is considered very unlikely being drawn from urn A, subject should put higher weight on urn B. Unfortunately, subjects may not be able to allocate probabilities p_A and p_I properly. For example, even if they believe the information has 10% chance coming from their assigned urn, they might only assign 60% to the irrelevant urn. One possible and intuitive updating process is that they separately update two urns. Specifically, when they deem the information not from one urn, they do not attribute it to the other urn. In fact, it is useless to subjects when updating the belief. In this situation, it seems that the information is drawn from an urn in which each ball is drawn with equal probability. In other words, when subjects regard the information is from the "useless urn", it provide no further clue for updating. To derive the theoretical prediction, the differences are caused by $\Pr(s_2|s_1, \theta_{\max})$, $\Pr(s_2|s_1, \theta_{\min})$,

$\Pr(s_2|s_1, \omega_{\max})$, and $\Pr(s_2|s_1, \omega_{\min})$. Therefore, the theoretical results are as follows.e

$$\Pr(\theta_{\max}|s_1, s_2)$$
$$= \frac{[(2s_2-1)p_A + 100(1-p_A)](2s_1-1)}{[(2s_2-1)p_A + 100(1-p_A)](2s_1-1) + [(201-2s_2)p_A + 100(1-p_A)](201-2s_1)} \tag{13}$$

$$\Pr(\omega_{\max}|s_1, s_2)$$
$$= \frac{[(2s_2-1)p_I + 100(1-p_I)](2s_1-1)}{[(2s_2-1)p_I + 100(1-p_I)](2s_1-1) + [(201-2s_2)p_I + 100(1-p_I)](201-2s_1)} \tag{14}$$

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%).[16] 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.

The squares in Figure **??** represent the mean elicited probabilities averaged across all subjects with the same initial draw. They closely adhere to the Bayesian posteriors, especially for the assigned urn. Notice that there is a cluster of elicited

[16] 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.

probabilities along the 45-degree line in Figure ??b, 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 subjects and not concentrated in particular rounds, indicating that they are not caused by particular subjects or rounds.[17] Although these choices consists of only 3% of the data, they inflate the correlation between the elicited probabilities of the assigned and irrelevant urn.[18] 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.[19] In conclusion, most of the choices are consistent with Bayesian updating derived in section 2.4.2.

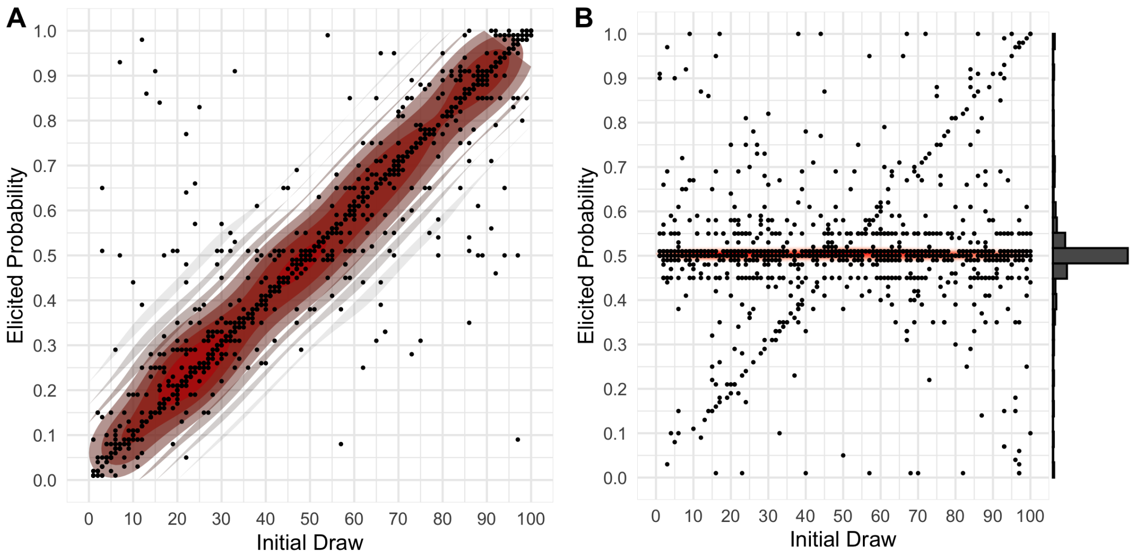

Figure 2: Elicited Beliefs in the First Phase of the (a) Assigned (b) Irrelevant Urn

[17]See Appendix A for further details.

[18]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.

[19]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.

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.[20]

Compared to the first phase, belief-updating in the second phase is much worse.[21] 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

[20]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.

[21]See Figure 10 of Appendix B for the raw data plotted like Figure ??.

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).

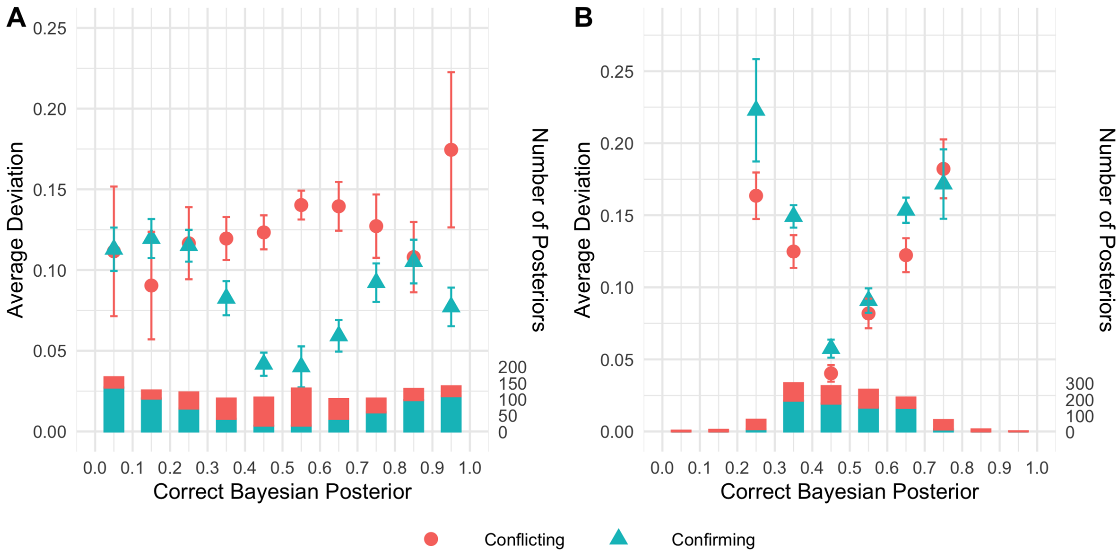

Figure 3: Elicited Beliefs Distribution in the Second Phase of (a) the Assigned, and (b) Irrelevant Urn

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 ($p = 0.434$).[22]

[22] 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 The Echo Chamber

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.[23] 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.[24] 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

[23]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.

[24]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.

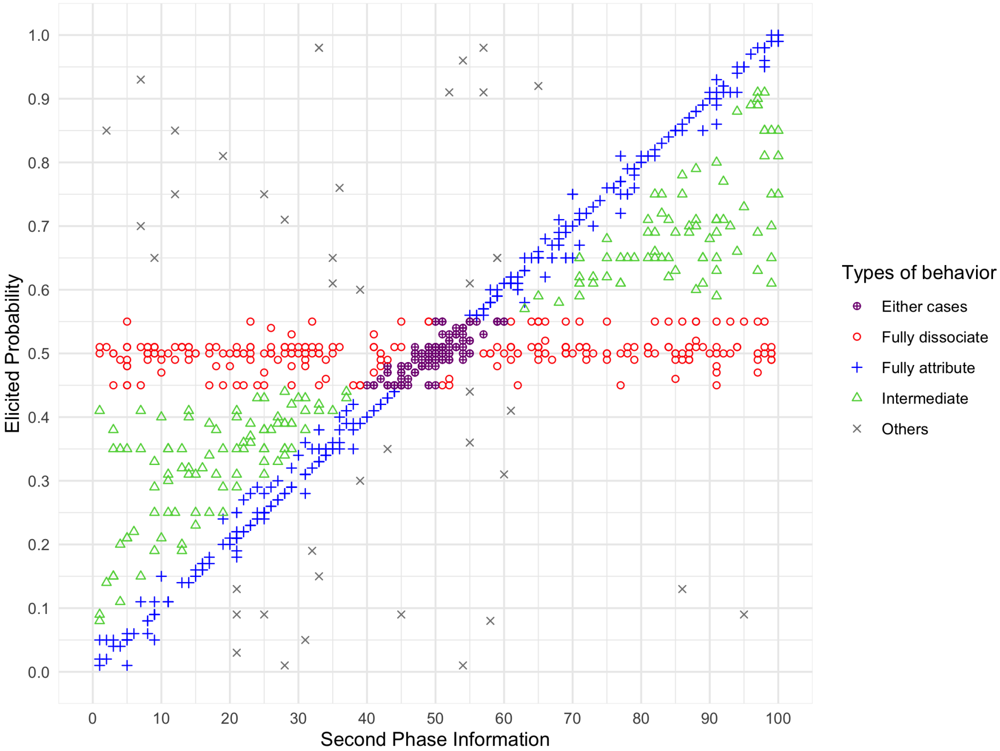

Figure 4: Types of Behavior (Irrelevant Urn)

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 5, 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 6 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.

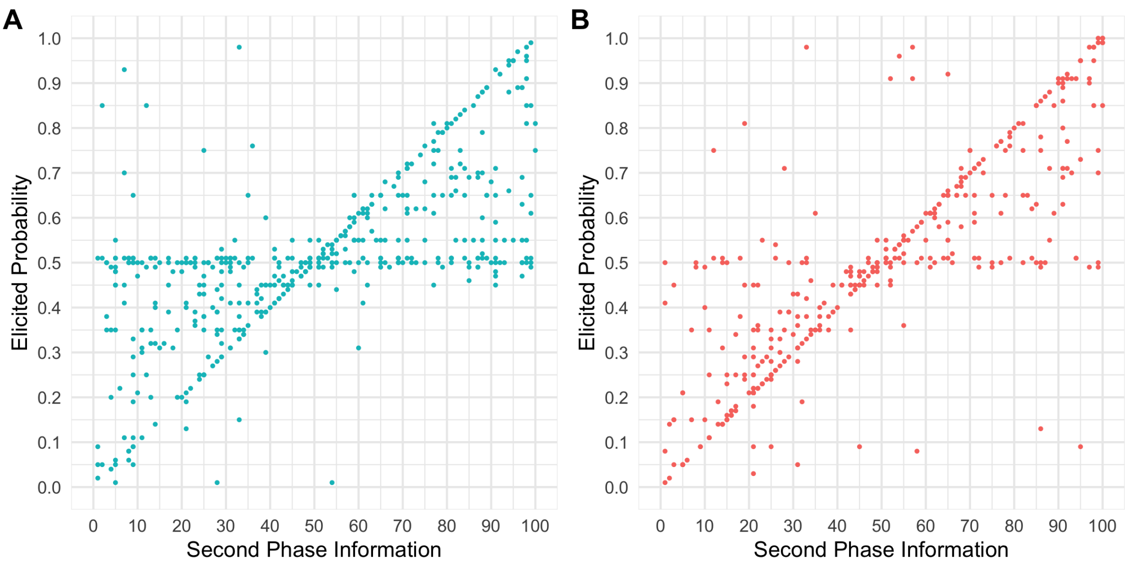

Figure 5: Elicited Beliefs of the Irrelevant Urn: (a) Confirming, and (b) Conflicting Information.

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, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Attribution of the Information (Extreme Signal)

Fully Attribute to	(1) Assigned Urn	(2) Irrelevant Urn
Confirming Information	0.264***	-0.292***
	(0.038)	(0.040)
Constant	0.135***	0.552***
	(0.027)	(0.029)
N	634	634

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.[25] 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.[26] Unexpectedly, 54% of these choices stick to their first-

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

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

Table 5: Attribution of the Information (interaction term)

Fully Attribute to	(1) Assigned Urn	(2) Irrelevant Urn
Conflict Information	-0.112*	-0.008
	(0.057)	(0.040)
Conflict*Distance	-0.001	0.006***
	(0.001)	(0.001)
Constant	0.327***	0.167***
	(0.017)	(0.024)
N	914	914

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

Table 6: Attribution of the Information (reinforce version)

Fully Attribute to	(1) Assigned Urn	(2) Irrelevant Urn
Conflict Information	-0.056	-0.138***
	(0.039)	(0.042)
Conflict*Distance	-0.004***	0.007***
	(0.007)	(0.001)
Constant	0.389***	0.283***
	(0.025)	(0.027)
N	914	914

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

phase posteriors of the assigned urn. This implies at least 25.4% $\times$ 54% = 13.7% of all choices completely ignore the new information and update neither urn.[27] 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. Even though subjects fully dissociate the information from the irrelevant urn, the updating behavior systematically under-weights the new information from the other

[27] 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.

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 ($p < 0.001$).

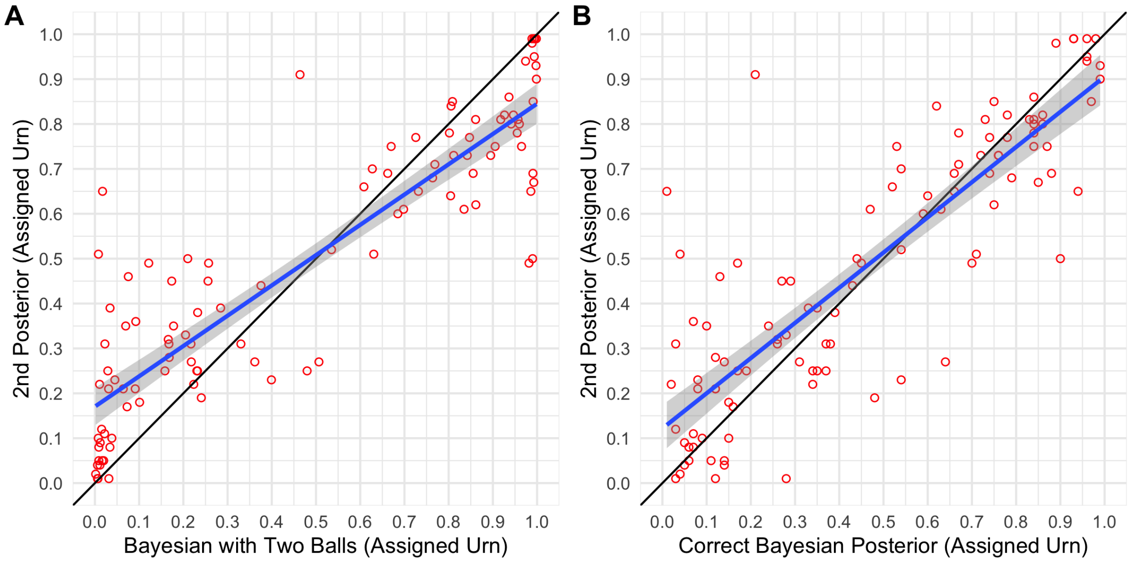

Figure 6: Fully Dissociate: (a) Two Balls from Assigned Urn. (b) Correct Bayesian.

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.[28] 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.[29] 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),[30] 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

[28]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.

[29]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.

[30]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:

$$d_r^k(x_r^i, t_r^{i,k}, \lambda) \equiv \frac{\exp\left[\lambda E(x_r^i | t_r^{i,k})\right]}{\sum_{z_r^i} \exp\left[\lambda E(z_r^i | t_r^{i,k})\right]}. \tag{15}$$

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

$$d^k(x^i, t^{i,k}, \lambda) \equiv \prod_r d_r^k(x_r^i, t_r^{i,k}, \lambda). \tag{16}$$

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 (15) by p^k, summing over k and taking logarithms, the log-likelihood function for subject i becomes

$$\ln L(p, \varepsilon, s | x^i) = \ln \left[\sum_{k=1}^K p^k d^k(x^i, t^{i,k}, \lambda) \right]. \tag{17}$$

Given the estimate of λ, it is clear from (17) 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.

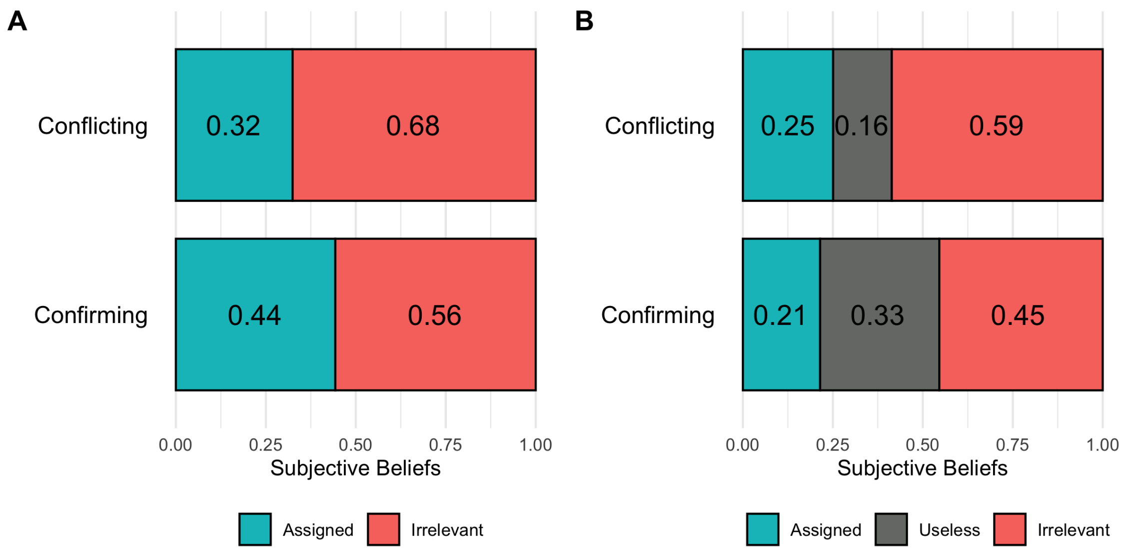

Figure 7: Models of Information Sources: (a) Two Urns (b) Three Urns.

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 **??**a. 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 (5) becomes[31]

$$
\begin{aligned}
&\Pr(s_2|s_1, \theta_{\max}) \\
&\quad = \Pr(s_2|s_1, \theta_{\max}, \omega_{\max}) \cdot \Pr(\omega_{\max}|s_1, \theta_{\max}) + \Pr(s_2|s_1, \theta_{\max}, \omega_{\min}) \cdot \Pr(\omega_{\min}|s_1, \theta_{\max}) \\
&\quad = \frac{1}{2}\Big[\Pr(s_2|s_1, \theta_{\max}, \omega_{\max}, \text{Assigned } s_2) \cdot p_A + \Pr(s_2|s_1, \theta_{\max}, \omega_{\max}, \text{Irrelevant } s_2) \cdot p_I \\
&\qquad + \Pr(s_2|s_1, \theta_{\max}, \omega_{\max}, \text{Useless } s_2) \cdot p_U + \Pr(s_2|s_1, \theta_{\max}, \omega_{\min}, \text{Assigned } s_2) \cdot p_A \\
&\qquad + \Pr(s_2|s_1, \theta_{\max}, \omega_{\min}, \text{Irrelevant } s_2) \cdot p_I + \Pr(s_2|s_1, \theta_{\max}, \omega_{\min}, \text{Useless } s_2) \cdot p_U \Big].
\end{aligned} \tag{18}
$$

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

[31]Equation 18 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 $\Pr(s_2|s_1, \omega_{\min})$.

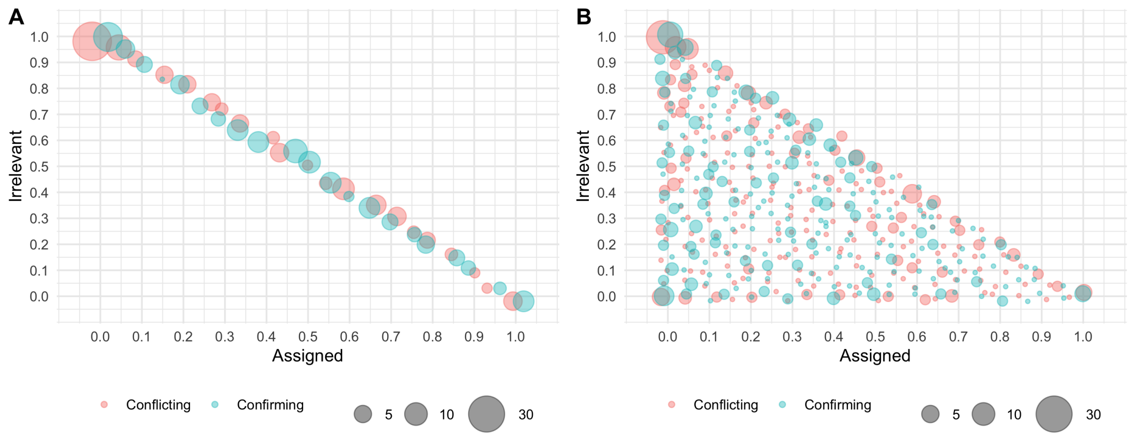

Figure 8: Information Sources Distributions: (a) Two Urns (b) Three Urns.

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 9.

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 7 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 7 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 7: Individual Type Transition: Conflicting vs. Confirming (%)

Conflicting	p_A	Confirming p_A 0	0.5	1	Total
	0	21.1	<u>21.1</u>	<u>12.2</u>	54.4
	0.5	6.5	25.2	<u>4.1</u>	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 8 column (1) and (2) show that the directions of coefficients confirm the asymmetric updating. When the information is aligned with their priors, 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 8: Independent Source Beliefs

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

Table 9: 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							

4 Conclusion

In this experiment we set out to examine how people process potentially irrelevant information when they already established certain pre-existing beliefs. To uncover the mechanism behind confirmation bias, we ask subjects to report beliefs of the 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 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. We estimate the tendency of attributing the information to the irrelevant urn. The results suggest that on average subjects believe the information is from the irrelevant urn with probabilities more regardless of the types of information. However, they increase the probabilities when the information is conflicting by 12%. When we allow subjects consider other's information might be inaccurate, the they still believe the information is more likely from the irrelevant urn when it is conflicting. These results are robust even we assume subjects independently make decisions on the assigned and irrelevant urn.

Most importantly, we try to explore the mechanism leading to the echo chamber, especially focusing on the information updating. By explicitly creating an irrelevant urn, we highlight one possible reason people usually stick to their political stance or beliefs on controversial issues, even leading to polarization. Though this may not be the only cause of the echo chamber effect, our results suggest that dismissing the information when it conflicts with one's prior is still a prominent cause.

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. Updating beliefs when evidence is open to interpretation: Implications for bias and polarization. *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. *mimeo*.

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.

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.

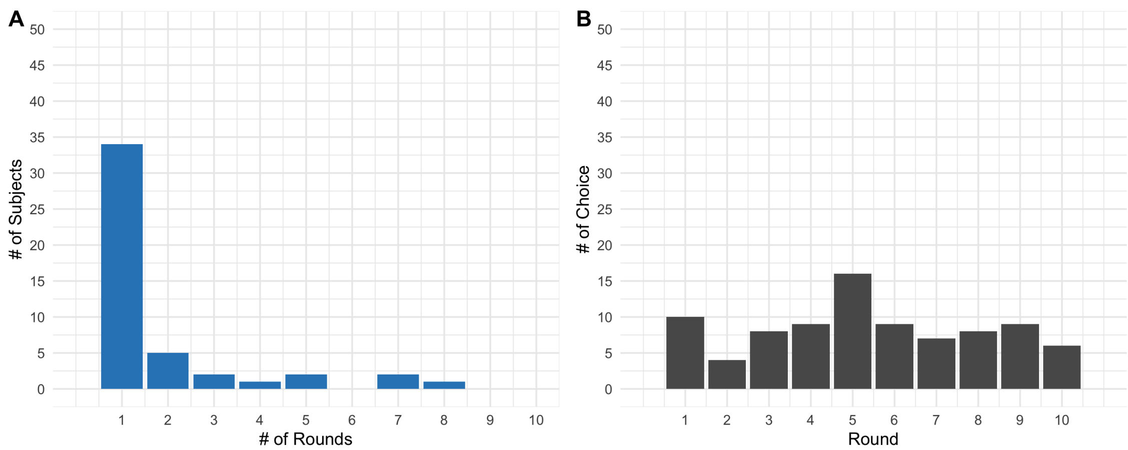

Figure 9: Beliefs Aligned with 45 Degree Line in the Irrelevant Urn. (a) the Number of Rounds (b) Occurrence Rounds.

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.

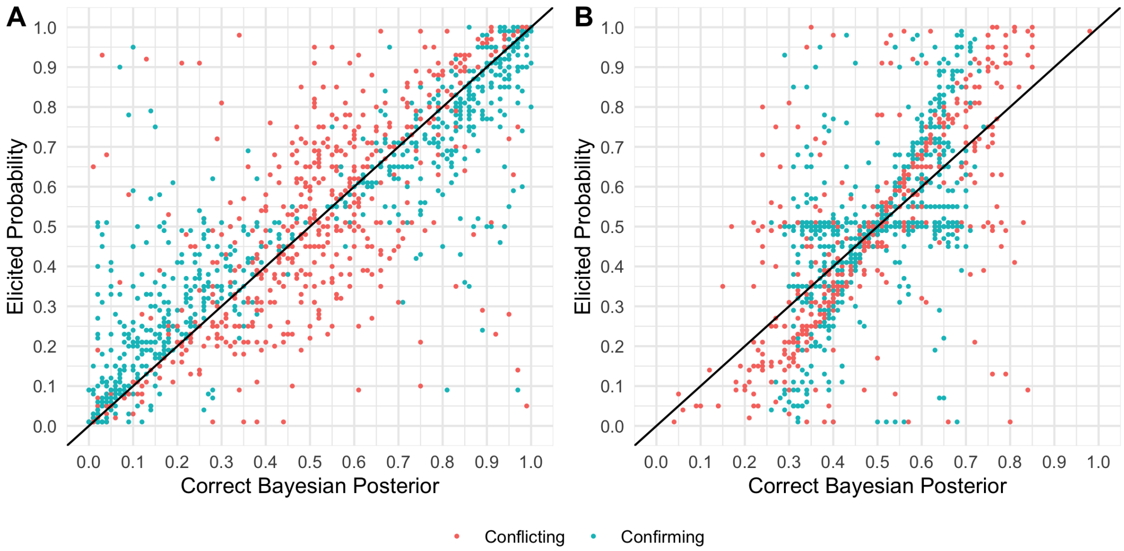

Figure 10: Elicited Beliefs in the Second Phase of the (a) Assigned (b) Irrelevant Urn

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.

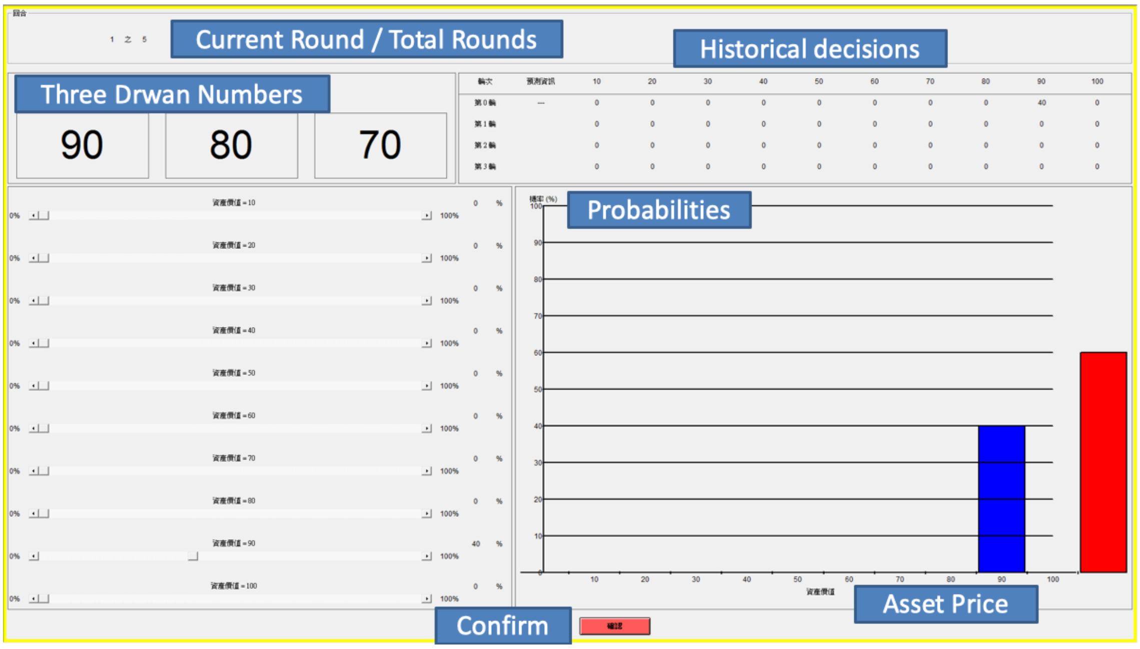

Figure 11: Screen Shot of the First Version Experiment.

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.

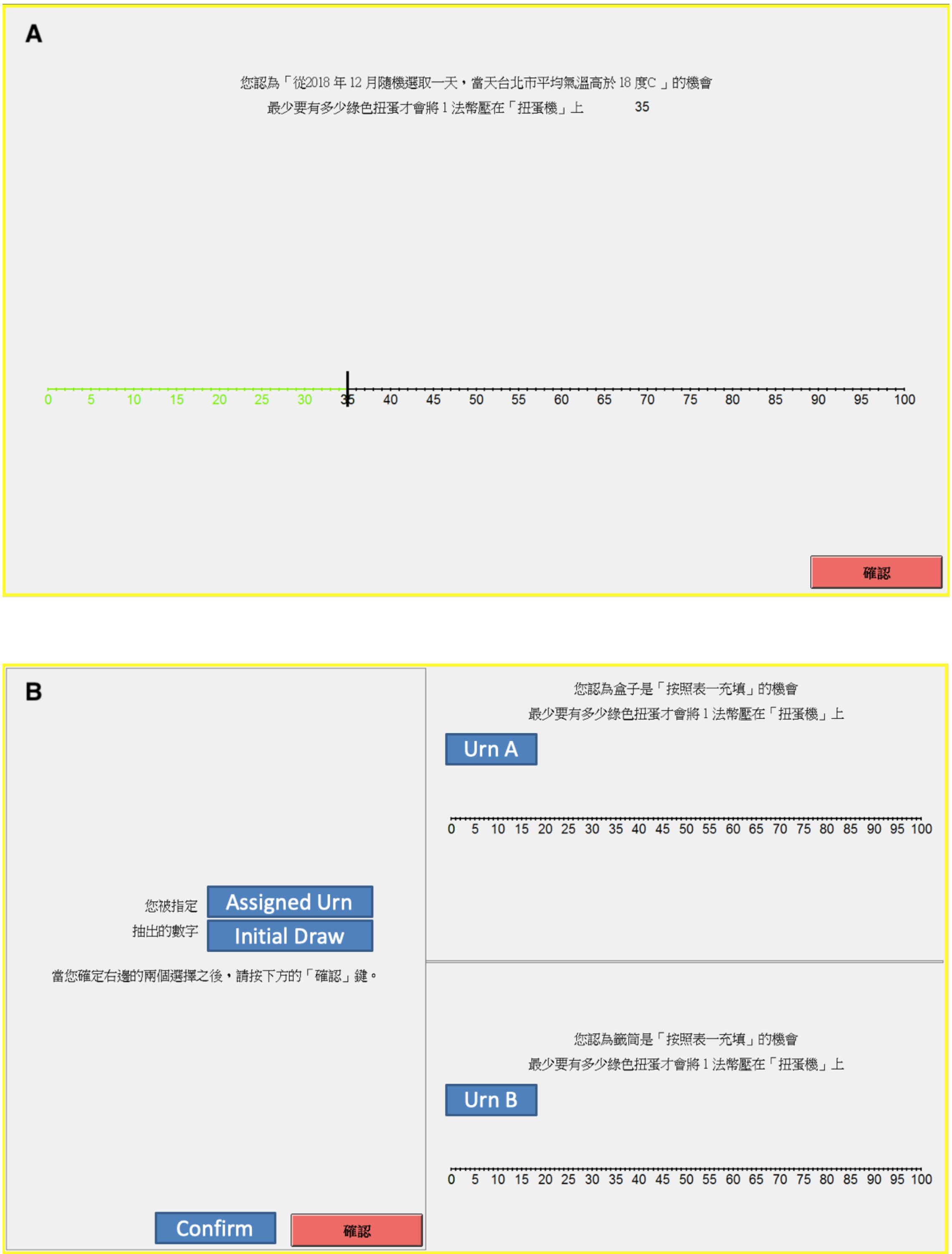

Figure 12: Screen Shot of Third Version Experiment.