Does Observation of Others' Actions Prevent Polarization? Results from Laboratory Experiments※

April 10, 2024

Kiichiro Arai, Chuo University Yasushi Asako, Waseda University*
Airo Hino, Waseda University So Morikawa, The University of Tokyo

Abstract

The political science literature has primarily neglected a fundamental and underlying mechanism of polarization, namely "belief polarization," where common interests are shared, but private information hinders a consensus. We demonstrate this mechanism through laboratory experiments by deliberately removing political contexts and investigating whether revealing others' actions can prevent it since people can infer others' private information through their actions. Our experiments have the following implications. First, when we reveal others' actions only once, polarization still occurs and increases over rounds. Second, when others' actions are revealed in all rounds of experiments, polarization does not occur. However, if subjects think others have insufficient information, polarization persists—even when others' actions are revealed in all rounds.

Keywords: belief polarization, laboratory experiments, asymmetric information, correlation neglect, social learning

JEL classification: C92, D72, D82, D83

※ The authors benefited from feedback and comments from James Andreoni, André Blais, Dominik Duell, Xin Fang, Yukihiko Funaki, Yoichi Hizen, Yoshio Kamijo, Yuko Kasuya, Toshiji Kawagoe, Daiki Kishishita, Yukio Koriyama, Ikuo Kume, Takeshi Murooka, and the seminar/session participants of Virtual Formal Model Workshop, Midwest Political Science Association, Japanese Economic Association, the Japanese Society for Quantitative Political Science, and Osaka University. The authors are grateful for the financial support received from JSPS KAKENHI (Grant Number 20H00066 and 20K01734).
* Corresponding author. yasushi.asako@waseda.jp

1. Introduction

Ideological polarization can arise even when people share common interests. This type of polarization is known as "belief polarization" and is distinct from "preference polarization," which arises from conflicting and divergent preferences. Belief polarization occurs when people hold different beliefs (defined as subjective probabilities in formal models) rather than different preferences. For example, some people may believe that COVID-19 is a severe disease, while others may believe it is just a cold; this does not necessarily mean that the two sides have a conflict of interest, but rather that they hold different beliefs about the severity of COVID-19.[1] While preference polarization arises from differences in people's backgrounds, belief polarization arises from differences in people's information, which means that some people understand the situation differently. While it might seem that belief polarization could be prevented by sharing as much information as possible, this comparison oversimplifies a complex issue; even when people have access to the same information, they may draw opposing conclusions if they possess private information.

Learning from insights based on formal models helps better understand this underlying mechanism of belief polarization. Existing formal models have shown that belief polarization can occur rationally when the dimensionality of information exceeds in a way that describes the state of the world (Dixit and Weibull, 2007; Bullock, 2009; Kondor, 2012; Acemoglu, Chernozhukov, and Yildiz, 2016).[2] To explain this, let us consider an example of dimensionality. Suppose the state of the world is one-dimensional; however, the uncertainty around it is two-dimensional, and both dimensions are essential to identify the state of the world. In this scenario, even if people commonly observe the same information as public signals about one dimension many times, polarization may

[1] Bernacer et al. (2021) show that the COVID-19 pandemic was the cause of belief polarization in Spain.

[2] Belief polarization was traditionally recognized as an "irrational" phenomenon under which people do not act in a Bayesian manner. Past studies show that such "irrational" belief polarization can occur because of confirmation bias (Rabin and Schrag, 1999; Gerber and Green, 1999; Nyhan and Reifler, 2010; Fryer, Harms, and Jackson, 2019), ambiguity aversion (Baliga, Hanany, and Klibanoff, 2013), and bounded memory, which means that individuals make decisions based on some past experiences (Wilson, 2014). However, certain studies have indicated that "irrational" belief polarization is not observed in survey experiments (e.g., Wood and Porter, 2019; Guess and Coppock, 2020). In addition, some theoretical and experimental studies have shown that (rational) inattentiveness is the primary cause of posterior polarization of beliefs, even if the players are rational and hold the same beliefs a priori (e.g., Nimark and Sundaresan, 2019; Hu et al., 2023; Novák, Matveenko, and Ravaioli, forthcoming; Bloedel and Segal, 2021). As we focus on understanding how individuals respond to others' actions when they possess information held by those individuals, we adopt a "rational" framework instead of an "irrational" one without any contextual bias and inattentiveness.

persist when they have different private information (signal) about the other dimension. In other words, polarization is caused by asymmetric information about one of two dimensions.

This paper examines whether belief polarization can (or cannot) be avoided based on the formal model. One approach to preventing belief polarization is to reveal others' actions and enable individuals to infer the private information held by others. The theoretical reasoning is as follows. If everyone shared private and public information, information asymmetry would disappear, and polarization would theoretically not occur with common interests. Moreover, since private information possessed by others can be inferred from the actions of others, polarization should disappear by revealing others' actions. Thus, this study investigates how revealing others' actions affects and prevents polarization. Through the experiments in our study, we arrive at the following three conclusions.

1. Showing others' actions only once cannot prevent polarization.
2. Showing others' actions multiple times can prevent polarization.
3. If people believe others may have insufficient information, polarization occurs—even if others' actions are revealed multiple times.

To be precise, our experiments are based on those of Andreoni and Mylovanov (2012), who simplify the belief polarization model and establish decontextualized settings of laboratory experiments (see Section 3). As the new treatment, our experiments provide information on others' actions at the aggregated level in an earlier round to see if such disclosures of the information reduce belief polarization. Nevertheless, polarization still occurs. On the one hand, this result contradicts the theoretical prediction that once people observe others' actions, polarization will cease to occur because they can infer the private information held by others. On the other hand, this result is consistent with the previous studies in behavioral economics, which show that people assign too much weight to their private information relative to the publicly observable others' actions (Nöth and Weber, 2003; Goeree et al., 2007; Weizsäcker, 2010).

We also allow the subjects to observe others' actions in all the rounds so polarization does not occur, suggesting that information on others' actions must be repeatedly provided to prevent polarization. This finding can be explained by the cognitive effect identified by behavioral economics: People's beliefs are overly influenced by "telling and retelling stories" because they tend to overlook the correlation between different sources of information. This cognitive phenomenon is

called “correlation neglect” (De Marzo, Vayanos, and Zweibei, 2003; Glaeser and Sunstein, 2009; Enke and Zimmermann, 2019). According to past studies, correlation neglect can induce people to make herding into a correct outcome (e.g., Levy and Razin, 2019) and an incorrect outcome (e.g., Eyster and Rabin, 2010). In our experiment, by “retelling” others’ actions several times, many subjects take into account their decisions even though others’ actions are correlated over rounds. Herding to a correct outcome occurred in most cases, but there were some cases with herding to an incorrect outcome.

However, polarization re-emerges with the additional treatment that some subjects cannot observe the private signal. This finding shows that if subjects think others lack sufficient information, polarization persists even though others’ actions are revealed in all rounds. Thus, people must believe that others have sufficient information to prevent polarization. It may be because informed subjects can notice that uninformed subjects are likely to form a herd. By predicting it, they do not rely on others’ actions even though the actions of others are revealed again and again.

Over the recent years, a wealth of studies has been conducted to learn why and how polarization occurs. It has been shown that higher district magnitude increases ideological polarization (Dow, 2011). A higher number of political parties increases ideological polarization, while its effect is moderated by their coalitional habits (Curini and Hino, 2012). Affective polarization is also shown to be linked to ideological polarization (e.g., Wagner, 2021). However, the political science literature has primarily overlooked an important underlying mechanism of “belief polarization.” One tends to assume that parties/candidates compete against each other over specific issues in the ideological continuum, but polarization can occur even without policy competition. To our knowledge, polarization has yet to be examined rigorously with varying treatments, as we have done in our study. Although polarization arises in different contexts, the exact mechanism can drive it, and laboratory experiments without specific contexts based on the formal model should help demonstrate them.

It is especially important to consider situations with perfectly common interests without context to investigate the effects of revealing others’ actions. First, when there are at least partially conflicting interests that induce people to have preference polarization, the source of the polarization may be rooted in the people themselves, and they do not care about others’ actions since their different actions suggest different preferences rather than different information. Thus, if people have

conflicts of interest, we cannot identify the effects of revealing others' actions on their beliefs. As a solution to this problem, our study focuses on situations wherein people have perfectly common interests by excluding conflicts of interest in our laboratory settings.

Furthermore, while many policy issues have interests shared by the entire society, they tend to generate conflicting interests among individuals. For instance, people may hold divergent beliefs about the severity of COVID-19 and have varying preferences concerning the value they place on their health or their freedom. Therefore, if we use any contextualized setting in the experiments, subjects may have (partially) conflicting interests that induce them to ignore the actions of others with different preferences. To induce subjects to have perfectly common interests, our laboratory experiments employ decontextualized settings instead of contextualized ones.

Most previous experiments of polarization adopt the contextualized setting of survey experiments such as the death penalty (Lord, Ross, and Lepper, 1979; Houston and Fazio, 1989; Schuette and Fazio, 1995), presidential debates (Katz and Feldman, 1962; Sigelman and Sigelman, 1984), and the economy (Kinder and Mebane, 1983). Through survey experiments with contextualized settings (redistribution policy), Balietti et al. (2021) show that ideological polarization can be reduced by exposing different political views of similar people. On the contrary, we investigate belief polarization in the decontextualized settings of laboratory experiments.[3]

Additionally, conspiracy theories can emerge when two people with opposing prior beliefs strengthen their beliefs after observing the same data. People who believe in conspiracy theories consume the same information as others, but their beliefs become extreme. That is, believing in conspiracy theories is a type of belief polarization. Several survey experiments have investigated how people adopt conspiracy theories by relying heavily on contextualized settings.[4] Therefore, their findings may only apply to some contexts, so we cannot investigate the effects of observing others' actions. By contrast, our study uses the decontextualized settings of laboratory experiments to understand the common features of belief polarization, including conspiracy theories.

[3] Arai et al. (2024) also investigate belief polarization with a similar experimental design. They show that belief polarization can be reduced by showing details of payoff structures in the instrument.

[4] For example, Crocker et al. (1999) show that black Americans are far more likely to endorse theories about conspiracies by the U.S. government against blacks than white Americans. Galliford and Furnham (2017) show that conservatives are more likely to believe in political and medical conspiracies than liberalists. Radnitz and Underwood (2015) show that people who have anxiety are more likely to believe the fictional conspiracy theory that the government is hiding the cause of a mysterious illness afflicting a small Midwestern town. Additional references can also be found in Douglas et al. (2019).

The rest of this paper proceeds as follows. Section 2 presents the formal model, which shows a political example of belief polarization. Section 3 outlines the experimental design. Section 4 describes the results, and Section 5 applies them to real-world political issues. Section 6 concludes.

2. Theoretical Background

2.1 Basic Model

This section formally shows the application of the belief-polarization model developed by Andreoni and Mylovanov (2012) to the political example. Our experiments are based on this model and show how belief polarization emerges and how observation of others' actions affects it.

Suppose that implementing a drastic reform policy, such as radical policy for climate change (Fryer et al., 2019, Novák et al, forthcoming), lockdown policy against infectious diseases (Bernacer et al., 2021), or joining/leaving an international union, is a controversial issue. Formally, suppose policy $x \in \{0,1\}$, where $x = 0$ is no reform and $x = 1$ is reform. Implementing such a reform policy is unnecessary to induce a desirable consequence for citizens.

There are two states of the world, $\theta \in \{g, b\}$, and each state occurs with the same probability (50%). Both citizens and the government have preferences for the policy represented by $u_{\theta x}$ and $v_{\theta x}$, respectively, when policy $x \in \{0,1\}$ is implemented in state $\theta \in \{g, b\}$. Suppose that both obtain zero from the status quo policy regardless of the state: $u_{\theta 0} = v_{\theta 0} = 0$ for any $\theta \in \{g, b\}$. Reform is desirable for citizens in state $\theta = g$ but undesirable for state $\theta = b$: i.e., $u_{g1} > 0$ and $u_{b1} < 0$. Citizens do not know the state of the world, and the prior probability of each state is 50%. If citizens can guess the state of the world, they can also know whether the reform policy is desirable for them. Thus, they try to infer the state of the world to decide whether they support the reform policy.

There are also two types of government, $\gamma \in \{c, d\}$, where $\gamma = c$ means that the government is the congruent type whose interests coincide with those of citizens. That is, $v_{g1} > 0$ and $v_{b1} < 0$. The dissonant type, $\gamma = d$, means that the government has different policy preferences from the citizen's preferences because of, for example, the strong relation with special interest groups. That is, $v_{g1} < 0$ and $v_{b1} > 0$.

Citizens do not know the government's type, and the prior probability of being each type is 50%. They obtain a *private signal* about the type of the government, $t \in \{c', d'\}$, such that $Pr(t = c'|\gamma = c) = Pr(t = d'|\gamma = d) = q \in (0.5, 1)$. That is, signal c' means that the

government is more likely to be the congruent type, and signal d' means that the government is more likely to be the dissonant type. This private signal can be interpreted as citizens' evaluations of the government based on past experiences. Each citizen should have different experiences of the government as well as check different news media. These different experiences and information provided by the media lead to their different evaluations of the government. Here, we called these past cumulative experiences and information the private signal.

While citizens cannot observe the state of the world, the government has more information about it. Suppose that the government receives an imperfect signal about the state of the world, $s \in \{g', b'\}$, such that $Pr(s = g'|\theta = g) = Pr(s = b'|\theta = b) = p \in (0.5, 1)$. That is, signal g' means that the state of the world is more likely to be g, and signal b' means that the state of the world is more likely to be b. After receiving the signal, the government sends a public message to citizens, $m \in \{0,1\}$, where $m = 0$ means that they do not support the reform and $m = 1$ means that they support it. All citizens can receive the message from the government, so it is a *public signal*. We suppose that the government truthfully supports their preferred policy in their message. When $s = g'$, the congruent type supports the reform ($m = 1$), while the dissonant type recommends keeping the status quo ($m = 0$). When $s = b'$, the congruent type sends $m = 0$, and the dissonant type sends $m = 1$.

Citizens want to infer the state of the world (a consequence of the reform) from the government's public message. However, they need to know the government's type to infer it from this message since they do not know whether the government is announcing citizens' preferred policy. Therefore, there are two dimensions of information (the government type and the signal received by government), and there is one dimension of the state of the world.

The timing over which citizens obtain each piece of information is as follows:

Period 1: Citizens obtain a private signal (about the government's type).

Period 2: Citizens obtain a public signal (a message from the government).

Suppose that the government announces that it supports reform ($m = 1$) in the second period. Then, a citizen who observed $t = c'$ in the first period believes that the state of the world is more likely to be $\theta = g$. More precisely, the probability of $\theta = g$ is $\Pr(\theta = g|t = c', m = 1) = qp + (1 - q)(1 - p) > 0.5$. On the contrary, a citizen who observed $t = d'$ believes that the state of the

world is $\theta = b$ with probability $\Pr(\theta = b|t = c', m = 1) = qp + (1-q)(1-p)$. Therefore, belief polarization emerges—even though both citizens observe the same public signal from the government—because they have different private signals.

We pointed out in the introduction that some conspiracy theories can be interpreted as a kind of belief polarization. Past studies have indicated that people with greater political distrust are more likely to believe in conspiracy theories (e.g. Swami, Chamorro-Premuzic, and Furnham, 2010; Miller, Saunders, and Farhart, 2016; Mari, et al. 2022). The model in this section shows the process of belief polarization as decision makers' level of political trust (private signals) results in different interpretations of the information produced by the government (public signals). It could be said to depict a situation where political trust makes the difference between believing in some conspiracy theories or not.

2.2 Multiple Public Signals

In the above, a citizen observes the government's message only once. Next, suppose a citizen can observe the government's messages several times. It may be strange that the government sends citizens several (possibly different) messages. In this case, there are two possible reinterpretations of the above story. First, several politicians belong to the government party who observe different signals about the state of the world. Hence, they send a message sequentially according to their observed signal, and citizens who receive $t = c'$ $(t = d')$ believe that the government party and its members are the congruent (dissonant) type with probability $q \in (0.5, 1)$. Second, several media outlets observe different signals about the state of the world. Each media outlet sends a message sequentially through reporting news according to their observed signal, and citizens who receive $t = c'$ $(t = d')$ believe that each media outlet has high (low) ability and was able to get the correct (incorrect) signal with probability $q \in (0.5, 1)$.

Under these conditions, Propositions 1 and 2 of Andreoni and Mylovanov (2012, p. 215-216) demonstrate that the probability and size of the disagreement between citizens with different private signals, $t \in \{c', d'\}$, increases as citizens receive more public signals.[5] In other words, polarization persists and may intensify with an increase in the number of public ssignals received by citizens.

[5] To be precise, the probability (expected size) of disagreement is one (unchanged) if the number of public signals is odd and increases if the number of public signals is even.

2.3 Observing Others' Beliefs

The belief polarization should disappear when the citizens observe beliefs held by others. Suppose two citizens, A and B. Citizen A receives $t = c'$, and Citizen B receives $t = d'$ in the first period. When they receive $m = 1$ in the second period, belief polarization emerges since citizen A's belief of $\theta = g$ is $qp + (1-q)(1-p) > 1/2$, and citizen B's belief is $1 - qp - (1-q)(1-p) < 1/2$. If they observe the other's beliefs after the second period (say in period 2.5), both citizens can realize that the other receives a different private signal. It means both realize there were two private signals, $t = c'$ and $t = d'$ in the first period. Then, both citizens have the same revised belief that the probability that the government is a congruent type ($\gamma = c$) is 50%, and their revised belief of $\theta = g$ is also 50%. There is no belief polarization after period 2.5.

The situation does not change even if the number of citizens increases, and they observe only the aggregate value of the beliefs of all citizens. If the aggregate value of beliefs of $\theta = g$ is above 50% after $m = 1$, most citizens should receive $t = c'$, and vice versa. If it is exactly 50%, the numbers of citizens who received $t = c'$ and $t = d'$ should be the same. Therefore, belief polarization disappears after they observe others' beliefs. Even if they continue to receive public signals after the third period, citizens continue to have the same belief about the government's type. Thus, belief polarization never emerges theoretically.

Result 1: *Suppose that people observe others' beliefs after the second period, where they already observed one private signal and one public signal. Then, belief polarization disappears.*

Because of Result 1, it does not matter whether they can observe others' beliefs after the third period. Belief polarization theoretically disappears after period 2.5, and all citizens have the same belief. Thus, the revealed others' beliefs after the third period does not contain any new information about the government's type and the state of the world.

Result 2: *If others' beliefs were revealed once after the second period, then no matter how many more times others' beliefs are revealed, it will not affect people's beliefs*

Here, we have considered the case where others' beliefs are revealed. If people's actions reflect their beliefs, then their beliefs can be inferred to some extent from their actions. Then, the above results apply to the case where others' actions are revealed.

The following section explains the details of our experiments using decontextualized settings because the above story is not directly explained to the subjects. Hence, they are not required to choose the best of the two "policies" and do not observe messages from "government." Then, we will analyze whether the above two results are valid, taking as some treatments the case where subjects observe the others' actions.

3. Experimental Design

3.1 Sessions

Six experimental sessions were conducted at Waseda University, Japan, in 2021 and 2022. 165 undergraduate and graduate students with various majors at Waseda University participated in this study. They were recruited through a sona system used exclusively by the students at Waseda University.[6] Upon arrival, the subjects were randomly allocated to a computer. Each subject had a cubicle; therefore, they could not see the computer screens of the others. They received instructions (Appendix A), which the computer read at the beginning of the experiment. The entire process is executed using a computer. oTree was used for these experiments (Chen et al., 2016).

There were 25–30 subjects in each session, and they engaged only once in the entire study. The subjects were divided into five or six teams, each comprising five members. Three sets of the same game were played successively, each containing 16 rounds. The subjects made one decision per round. The team members were randomly matched at the beginning of each set, and the team composition was changed. Subjects did not know who were playing with them on the same team.

3.2 Game Settings

Four urns were used, as shown in Figure 2. At the beginning of the set, one of the four urns was chosen, and a ball was picked up from it. Each urn was selected with a probability of 25%, and the urn chosen at the beginning of the set was used for all rounds in the set. The urns were divided into two groups. Group 1 consisted of Urns A and B, and Group 2 consisted of Urns C and D.

[Figure 2 Here]

[6] This is a system for subject management. Refer to https://www.sona-systems.com/ for more details.

The urns had two separate compartments, as Figure 2 shows. One contained red and green balls. Urns A and C had one red and three green balls. Urns B and D had one green and three red balls. A random draw from this compartment was equivalent to a signal indicating the chosen urn. When the chosen urn is A or C, the ball's color is red with a probability of 25% and green with a probability of 75%. When the chosen urn is B or D, the ball's color is red with a probability of 75% and green with a probability of 25%.

The other compartment contained white and black balls. Urns A and D had one black and three white balls. Urns B and C had one white and three black balls. A random draw from this compartment was also equivalent to a signal indicating the chosen urn. When the chosen urn is A or D, the ball's color is black with a probability of 25% and white with a probability of 75%. When the chosen urn is B or C, the ball's color is black with a probability of 75% and white with a probability of 25%.

During the experiment, the subjects had to determine whether the selected urn was from Group 1 (Urn A or B) or Group 2 (Urn C or D). They had to place bets on Groups 1, 2, or both by observing the ball's color drawn in each round. Note that after each draw, the ball was returned to the urn before making the next draw.

The first drawing was obtained from the red/green compartment. This draw was to be privately witnessed by each subject, so it is a *private signal*. A subject cannot know the color of the ball that the other members observe. The next 14 drawings were obtained from the white/black compartment. In these 14 draws, all team members saw the same color. All members publicly observed this draw; therefore, it is a *public signal*. A bet was placed after each draw in a round.

After the bets on the 15th draw, the total number of bets on each group by all the team members was revealed, and the subjects could make bets again. In the 16th round, the summations of all the points to be bet on each group from Rounds 1 to 15 are revealed in all treatments to make the same experimental designs as Andreoni and Mylovanov (2012).

3.3 Analogies with the Formal Model

These settings are analogous to the formal model in Section 2. First, a private signal from the red/green compartment is a private signal about the government's type, and green means $t = c'$ (more likely to be congruent), while red means $t = d'$ (more likely to be dissonant) with $q = 0.75$ (three balls among four balls have the "correct" color.)

Second, a public signal is a message sent by the government and white means that the government supports the reform policy ($m = 1$), while black means it does not ($m = 0$) with $p = 0.75$.

Third, each group of urns is analogous to the state of the world. Group 1 means that the reform policy is desirable for citizens ($\theta = g$), and Group 2 means that the reform policy is undesirable ($\theta = b$). Therefore, Urn A represents the case in which the government is the congruent type, meaning it is more likely to support the reform policy since $\theta = g$, and Urn B represents the case in which the government is the dissonant type, meaning it is less likely to support the reform policy. On the contrary, Urn C represents the case in which the government is the congruent type, meaning they are less likely to support the reform policy since $\theta = b$, and Urn D represents the case in which the government is the dissonant type, meaning it is more likely to support the reform policy.

3.4 Costs, Returns, and Earnings

In each round, the subjects could place between zero and nine bets in each group, meaning they could simultaneously make up to 18 bets. Only integers could be selected. Points betted for the correct group were returned: The subjects were given 10 points for every bet in the group containing the urns selected. For example, if a subject bets 7 points on the correct group, the return is 70 points. Points that were betted for the incorrect group were not returned. We called points to be returned the "return point."

Bets also incurred costs. We called them the "cost point," which marked the summation of all integers from zero to the points used in a bet for each group. The first bet made on Group 1 cost one point. The cost of each additional bet on Group 1 was one point higher than the previous bet on Group 1. For example, if a subject bet two points for Group 1, it cost an additional two points; thus, the cost was three. The third bet cost an extra three points. Thus, the total cost was $1 + 2 + 3 = 6$ points. The nth bet cost n points. Similarly, the first bet made on Group 2 in this round cost one point, and the nth cost n points. Table 1 lists the costs of each case. When subjects bet on both groups, they had to add cost points for Groups 1 and 2. For example, if a subject bet nine points in both groups, their cost points would be $45 + 45 = 90$, not 45.

[Table 1 Here]

The subjects' "earning point" was the return minus cost points in each round. Specifically, we denote b_1 and b_2 as the points to bet on Groups 1 and 2, respectively. We also denote r_1 and r_2

as the probabilities that Groups 1 and 2 contain the urns selected in this set. Then, the expected earning point is as follows:

$$10(r_1 b_1 + r_2 b_2) - \sum_{j=0}^{b_1} j - \sum_{k=0}^{b_2} k$$

Payments for subjects are determined by earning points, which can be both positive and negative because there is a possibility that the cost point is strictly higher than the return point. The bet for each round was independent of the other bets. The subjects could choose any point to bet on (between zero and nine), regardless of the points they had bet on in the previous rounds.

If individuals are risk-neutral Bayesian payoff maximizers in this setting, their bets should reveal their beliefs.[7] For instance, a subject who thinks that the likelihood of Group 1 being in the actual state is 40% should place four bets on Group 1, and a subject who believes that the likelihood of Group 2 being in the actual state is 60% should place six bets on Group 2.

All experiments lasted approximately 80–90 minutes. The subjects were informed that they would receive a participation fee of 1,000 yen in addition to any earnings they received based on the earning points (conversion rate: 1 point = 3 yen).[8] At the end of each set, two rounds were randomly selected to determine the subjects' earnings. We called this the "earning round." Payments to the subjects were decided based on both earning rounds per set. There were three sets, meaning six earning rounds in one session.[9]

To guarantee that earnings were non-negative, we gave all subjects 45 points per earning round at the end of the set because the minimum possible earning point was -45, where a subject bet nine points exclusively on the group that did not contain the selected urn. There were two earning rounds per set; subjects received 90 points per set. The subjects' final earning points were the sum of the earning points of the two earning rounds and 90 points. Their profit ranged from JPY 1,500 to JPY 2,800, including the participation fee.

---

[7] Considering the points as continuous variables, the expected earning point is $10(r_1 b_1 + r_2 b_2) - \int_{j=0}^{b_1} j dj - \int_{k=0}^{b_2} k dk = 10(r_1 b_1 + r_2 b_2) - \frac{b_1^2}{2} - \frac{b_2^2}{2}$. The value of b_t that maximizes the expected earning point is $b_t = 10r_t$, where $t = 1$ or 2.

[8] If a subject earned less than 1,500 yen, we increased the participation fee for all subjects so that each could get at least 1,500 yen.

[9] The earning-round setting prevents subjects' behavior from changing from round to round. It is known that if rewards are given in all rounds, there will be changes in behavior, such as taking more risks in the last round than in the first.

3.5 Practice Set

To ensure the subjects understood the experiment clearly, we prepared a practice set with one private signal from the red/green compartment and four public signals from the white/black compartment. The subjects decided on the points to bet for each group in each round. They tracked the draws and bets in each round using a computer interface. After all the team members completed one set, the selected urn and group, and the subject's points in each round, appeared on the screen.

3.6 Treatments

3.6.1 Providing Information on Others' Actions

Andreoni and Mylovanov (2012) provided their subjects with information about the total cumulative number of bets on each group in the final round (i.e., Round 16). We also did not inform the subjects about the total cumulative numbers of bets on each group until the final round in baseline sessions called "**Baseline**." [10]

As our primary interest was examining whether observing others' actions can prevent polarization, the following two treatments were introduced:

1. Offering information on others' actions in earlier rounds

   Like the formal model in subsection 2.3, we revealed the total cumulative number of bets on each group between the second and third rounds and called this treatment "**Round 2.5**." The timings to inform others' actions are shown in Figure 2 (a). Theoretically, belief polarization occurs after players observe both private and public signals. Therefore, the first time polarization arises right after the second round bet. After the bet in Round 2, the total points until Round 2 were announced in Round 2.5. This was a summation of all the points to be placed for a bet on each group in Rounds 1 and 2, and the subjects placed a bet after they observed it. Then, we had 13 other public signals.

   [Figure 2 Here]

2. Offering information on others' actions in all the rounds

   The total cumulative number of bets on each group was announced at the end of each round. We called this treatment "**All Rounds**." The timings to inform others' actions are shown in Figure 2

[10] Andoreoni and Mylovanov (2012) also run a different treatment in which the subjects receive both public and private signals repetitively. They confirm that polarization disappears when both signals are provided several times.

(b). The total cumulative number of bets was the summation of all the points bet on each group by all the team members in the previous rounds. Therefore, there were 15 points in time at which to reveal the other members' actions. There was no extra bet after each group's total points bet was revealed.

According to Results 1 and 2 in section 2, the theoretical prediction is that belief polarization disappears after Round 2.5 in both treatments (**Round 2.5** and **All Rounds**), and how many times others' actions are revealed will not affect polarization since all subjects can infer private signals held by others. If subjects are risk-averse or risk-loving, their bets (actions) may not reveal their beliefs directly. However, if they are at least Bayesian rational, they should prefer to bet on the group which is more likely to include the selected urn. Moreover, other team members were randomly chosen in each set, so subjects did not know the attitudes toward the risk of other members. Therefore, their actions reflect to some extent the beliefs they hold, so the theoretical prediction does not differ regardless of the risk attitude of the subjects.

3.6.2 Uninformed Subject

We employed another treatment in which one team member was not informed about the private signal. We announced to all the subjects that at least one member will not be able to observe the color of the private signal, but the number of such subjects was not announced. We called this treatment "**Uninformed**."

Theoretically, even if some subjects do not observe a private signal, they can infer the private signals observed by others through others' actions for the following reasons. If a subject observes only public signals (and no private signal and no others' actions), they must believe that the probabilities that the urns in Groups 1 and 2 are selected are the same. Thus, they should place 50/50 bets on Groups 1 and 2. Given this optimal choice, the existence of subjects who do not observe the private signal should not affect the difference in the total cumulative number of bets on each group. Therefore, if Group 1 has a higher total cumulative number of bets, more subjects observed the private signal, which induced them to believe that Group 1 was more likely to be selected.

However, the subjects may consider the other members' actions unreliable, as some may have insufficient information. If this is the case, they should rely only on their own signals, meaning that polarization cannot be prevented by announcing others' actions. This treatment was combined with **All Rounds**. Table 2 presents the treatments used in each experimental session.

[Table 2 Here]

3.4 Measurement of Polarization

3.4.1 Frequency of Disagreement

We measured the frequency and size of the polarization, which was also used by Andoreoni and Mylovanov (2012). The frequency of disagreement is defined as follows. We denote b_1^i and b_2^i as the points to bet on in Groups 1 and 2, respectively, by subject i. If $b_1^i \neq b_2^i$, we supposed that this subject strictly prefers the group on which they placed more bets. Otherwise (i.e., $b_1^i = b_2^i$), we supposed the subject was indifferent.

The team members were divided into two parties based on the color of the ball observed in the private signal: red and green parties. In a party, if the number of subjects who strictly prefer Group 1 to Group 2 (i.e., $b_1^i > b_2^i$) is higher than the number of subjects who strictly prefer Group 2 to Group 1 (i.e., $b_2^i > b_1^i$), we determined that this party strictly prefers Group 1 to Group 2. Similarly, if the number of subjects with $b_2^i > b_1^i$ is higher than those with $b_1^i > b_2^i$, we determined that this party strictly prefers Group 2 to Group 1. If an equal number of subjects strictly preferred different groups or all subjects were indifferent, we concluded that the party was indifferent to both groups. When one party strictly prefers one group, and the other prefers the other group or is indifferent to both groups, we conclude that the two parties have different preferences. The frequency of disagreement is when two parties have different preferences.

3.6.3 Value of Disagreement

The other measurement is the value of disagreement, that is, the size of the polarization. First, similar to the frequency of disagreement, team members were divided into red and green parties based on the color observed in the private signal. We then calculated the absolute value of the difference between the average number of bets made by each party on a specific group. This is the value of the disagreement. In the following section, we use the bets on Group 1 to measure the value of disagreement. Theoretically, the value of disagreement measured by the bets on Group 1 is the same as that measured by the bets on Group 2. The observed value slightly differs between them, but there is no change in the main implications, even if we use bets on Group 2 instead of those on Group 1.

3.4.3 Theoretical Predictions

Given the subjects' observations, we calculated the Bayesian beliefs about whether the urn belonged to Group 1 to derive the theoretically predicted value and frequency of disagreement. That is, the posterior probabilities after the subjects observed the colors of a few drawn balls. The Bayesian beliefs are presented in Table 3. For example, when a subject observes five white and two black balls from public signals, the white balls exceed the black balls by three. Thus, if a subject observes red, the Bayesian belief regarding whether the urn belongs to Group 1 is 15/56. If a subject observes a green, it is 41/56. When the subjects observe the same number of white and black balls or do not observe a public signal, the Bayesian belief is 1/2, regardless of whether the ball is red or green. Given these settings, the Bayesian belief is less than 3/4 and more than 1/4. The Bayesian belief converges to 3/4 or 1/4 as the subjects observe the same color from public signals more often.

[Table 3 Here]

We derived the theoretical value and frequency of disagreement using the calculated Bayesian beliefs. These represent the value and frequency at which all subjects are risk-neutral and Bayesian. The theoretical value of disagreement is the absolute value of the difference in the Bayesian beliefs among subjects who observed different colors in a private signal, multiplied by 10. The theoretical frequency of disagreement is the frequency at which the Bayesian beliefs of subjects who observe different colors from the private signals differ—that is, not 50% each—in a given round. The theoretical frequency of disagreement must be one in even rounds, as the subjects observe different numbers of black and white balls. In the odd rounds, there was no disagreement if the numbers of black and white balls observed by the subjects were the same. If they differed, there is a disagreement in the odd rounds.

Note that, in the following parts, we will show the theoretical predictions if the action of others was *not* observed. As Result 2 shows, both theoretical frequencies and values are zero after observing others' actions.

4. Experimental Results

4.1 Baseline

Figure 3 (a: left-side) presents the **Baseline**'s theoretical and observed disagreement values averaged for each set with 95% confidence intervals. In this treatment, subjects did not observe others' actions until the final round (round 16). We exclude from the data the sets in which all

members observed the same private signal in the first round, regardless of the treatment, since we cannot compare red and green parties. The observed value of disagreement under **Baseline** increased over the rounds and was approximately 3 to 3.5 in the latter half. Except for the first round, where the value is theoretically zero, the theoretical and observed values did not significantly differ, and both had similar values.

[Figure 3 Here]

Figure 3 (b: left-side) presents the **Baseline**'s theoretical and observed disagreement frequencies averaged for each set with 95% confidence intervals. Note that the variance of theoretical frequencies of disagreement is zero in the even rounds since disagreements occur for sure by observing the different numbers of white and black balls. The observed value of disagreement under **Baseline** gradually also increased over the rounds and was approximately 80-90% in the latter half. The observed frequencies do not significantly differ from the theoretical frequencies even in the even rounds.

4.2 Others' Actions

4.2.1 Offering Information in Round 2.5

The right side of Figure 3 (a) and (b) shows the value and frequency of disagreement with 95% confidence intervals under **Round 2.5**, in which the total cumulative number of bets on each group is shown in between the second and third rounds. As Figure 3 shows, just after Round 2.5, both the value and the frequency of disagreement slightly decrease, but they do not significantly differ from theoretical predictions. They increased again a few rounds later and returned to almost the same level as in **Baseline**. Therefore, against theoretical predictions (Result 1), offering information on the other members' actions early has almost no effect on polarization.

Since the total cumulative number of bets is shown only between rounds 2 and 3, the subjects may have forgotten it in the later rounds. However, the value of disagreement in **Round 2.5** significantly differs from theoretical values after Round 14, while its frequency does not differ much from theoretical ones. It means that the subjects tend to be risk-averse in the final few rounds. It suggests that the subjects might remember the information shown in the early round until Round 15.[11]

[11] Moreover, we ran another treatment in which the private signals observed by the other team members were informed between Rounds 2 and 3. With this treatment, both of the frequency and value of disagreement were much

As we discussed in the introduction, it is well known that people assign too much weight on private information than others' actions, and our experimental results are also consistent with these previous studies.

4.2.2 Offering Information in All the Rounds

Figure 4 (the left side) shows the value and frequency of disagreement with 95% confidence intervals in **All Rounds**, in which the total cumulative number of bets on each group is informed in all the rounds. The value of disagreement (Figure 4-a) is significantly lower than the theoretical predictions after round 4 except for a few rounds. That is, the polarization size in **All Rounds** is much lower than in **Baseline**.

[Figure 4 Here]

The frequency of disagreement in **All Rounds** (Figure 4-b) significantly differs from the theoretical predictions in even rounds. As we discussed, the theoretical frequency of disagreement must be one, so its variance is zero in even rounds, as the subjects observe different numbers of black and white balls. This is one possible reason why **All Round**'s observed and theoretical frequencies significantly differ. However, note that the observed and theoretical frequencies do not differ significantly, even in the even rounds of **Baseline** and **Round 2.5**. It means that the polarization frequency in **All Rounds** is much lower than in **Baseline**.

Therefore, we conclude that information on others' actions must be provided repeatedly to prevent polarization. According to the theoretical prediction (Result 2), polarization should be not occurred by showing others' actions only once, no matter how many times they are subsequently shown, because they should not bring new information. However, our experimental results here show that showing others' actions repeatedly is the way to reduce polarization. In this sense, this result contradicts the theoretical prediction. Moreover, if polarization did not disappear in **Round 2.5** because people put too much weight on their private signal, as we discussed in the previous subsection, then the **All Rounds** results above can be considered strange: Showing the behavior of others only once had no effect, but showing it repeatedly had an effect.

One possible explanation for this finding is "correlation neglect." As we discussed in the introduction, this cognitive effect implies that people neglect the level of correlation between the different sources of information. In **All Rounds**, the others' actions informed in each round are

lower than in **Baseline**. It suggests that the participants did not forget the information shown in the early round until Round 15. Appendix B shows these analyses.

naturally correlated, since they are the total number of points put up in previous rounds. Moreover, herding is likely to occur since other team members are also betting based on the actions of others. As Result 2 indicates, only the actions of others in round 2.5 reflect the private signals that they possess, and the actions of others informed in subsequent rounds do not contain usable information. In experiments, however, subjects fail to take into account this correlation between the information and the herding that occurs when they are repeatedly informed of the actions of others.[12]

4.2.3 All Rounds with Uninformed Subjects

To raise further doubts about the rationality of team members, we indicated the possibility that team members who are unaware of private signals are included. The right-hand side of Figure 4 shows the value and frequency of disagreement in **All Rounds** with **Uninformed**. The subjects who did not observe the private signal were excluded from the data. The results show that polarization reoccurs, and its value does not differ significantly from the theoretical predictions (the right side of Figure 4-a). The frequency of disagreement of **All Rounds** with **Uninformed** (the right side of Figure 4-b) significantly differs from the theoretical one in even rounds, but the differences are more minor compared to **All Rounds** without **Uninformed**.

Theoretically, the existence of uninformed subjects should not decrease the reliability of others' actions, as Section 3.3.2 discusses. In contrast, the experimental results imply that if people believe others have insufficient information, they tend to maintain opposing views.

Uninformed subjects are more likely to place bets based on others' actions, so herding is more likely to occur. All subjects understand the existence of uninformed subjects, so more subjects can understand a possibility of herding compared to the case without **Uninformed**. If so, correlation neglect is less likely occur since they perceived that the behavior of others only indicates that herding is occurring and contains no new information. As a result, polarization reoccurred.

This implication can explain an important characteristic of political polarization. Others' actions are visible in some cases, such as vaccination rates and face mask use in the recent COVID-19 pandemic. Such frequent observations can prevent polarization if people believe others have sufficient information. However, if people believe others have insufficient information, they will not rely on others' actions and maintain their opposing views.

---

[12] Enke and Zimmermann (2019) experimented with a situation in which subjects were sequential decision makers and showed that subjects did not take into account the possibility that herding might occur, especially in complex settings.

4.3 Statistical Estimations

We estimate the following using the mixed-effect model to statistically evaluate the effects of observing others' actions on the value and frequency of disagreement.

$$Observed_{ij} = \beta_0 + \beta_1 Baseline_j + \beta_2 All\ Round_j + \beta_3 Round\ 2.5_j + \beta_4 Theoretical_{ij} + \beta_5 (Baseline \times Theoretical)_{ij} + \beta_5 (All\ Round \times Theoretical)_{ij} + \beta_5 (Round\ 2.5 \times Theoretical)_{ij} + u_{0j} + e_{ij}$$

In the above model, *i* represents a subject, and j represents a team in each set (i.e., team×set). There are 97 teams×sets in which team members get the different private signals. $e_{ij} \sim N(0, \sigma_e^2)$ is the idiosyncratic error, and $u_{0j} \sim N(0, \sigma_{0u}^2)$ is the random effect of team×set on their intercept. $Baseline_j$ is the fixed effect of the treatment **Baseline**, which is one when the treatment is **Baseline**. $All\ Round_j$ and $Round\ 2.5_j$ are the fixed effects of **All Round** and **Round 2.5** treatments, respectively. $Observed_{ij}$ is the observed value or frequency of disagreement in our experiments. We also include theoretical value or frequency of disagreement in the estimation to explore the difference between theoretical and observed ones. $Theoretical_{ij}$ is the fixed effect of the disagreement's theoretical value or frequency. In the estimation, the reference group is the observed value and frequency of **All Round** with **Uninformed** treatment.

Table 4 and Figure 5 (a) show the estimation result on the value of disagreement. The observed value in **All Round** (without **Uninformed**) is significantly lower than in **Baseline**. Figure 5 (a) also demonstrates that the observed value in **All Rounds** is significantly lower than the corresponding theoretical value. On the other hand, theoretical and observed values do not differ in the other treatments, including **Baseline**, **Round 2.5**, and **All Round** with **Uninformed**.

[Table 4 and Figure 5 Here]

Table 5 and Figure 5 (b) show the estimation result on the frequency of disagreement. It confirms that the observed frequency in **All Round** significantly differs from that in **Baseline**. Observed and theoretical frequencies also differ in **All Round**, while they do not differ in other treatments. These results also show that the size and frequency of polarization become smaller only when others' actions must be provided repeatedly, and subjects believe that others have sufficient information.

[Table 5 Here]

5. Discussion

5.1 Cross-cutting Views on Social Media and Polarization

Many people rely on social media such as X and Facebook to digest news and discuss it on their networks. The rise of social media has led to concerns that people are becoming more isolated from diverse perspectives through "filter bubbles" and "echo chambers." Social media creates communities of like-minded individuals primarily exposed to only like-minded views (e.g., Sustain, 2001; Prior, 2007; Hindman, 2008). If social media induces users to communicate only with like-minded groups, this can cause greater polarization with out-group members. Indeed, empirical studies show that persistent ideological sorting exists in online communication networks, which may exacerbate political polarization (Adamic and Glance, 2005; Conover et al., 2012; Colleoni, Rozza, and Arvidsson, 2014; Lelkes et al., 2017; Boxell, Gentzkow, and Shapiro, 2017).

According to our experiments, one way to prevent such polarization is to induce individuals to observe out-group members' actions repeatedly. Some social media platforms facilitate exposure to messages from individuals who do not have like-minded views. Through such cross-cutting views on social media, individuals can infer information they would not be exposed to through offline interactions. This should prevent polarization if they check social media frequently. Indeed, empirical studies show that exposure to political diversity on social media improves political moderation (e.g., Barberá, 2015; Heatherly, Lu, and Lee, 2017; Beam, Hutchens, and Hmielowski, 2018; Nguyen and Vu, 2019; Melnikov, 2021). Our findings echo this line of recent studies.

5.2 The Cognitive Effect of Retelling Stories and Incorrect Herding

One reason polarization is diminished when others' actions are shown repeatedly is because people's beliefs are overly influenced by "telling and retelling stories." This cognitive effect is called "correlation neglect," as we discussed in the introduction and subsection 4.2.2 (De Marzo, Vayanos, and Zweibei, 2003; Glaeser and Sunstein, 2009; Enke and Zimmermann, 2019). There is some discussion on whether this correlation neglect leads to better results. This argument varies greatly depending on whether the information received by the decision maker who makes the correlation neglect is an exogenous signal or the actions of others.

Levy and Razin (2015) point out that election results tend to be more favorable to voters when voters neglect correlation and receive correlated *signals* often. In this case, voters who ignore the correlation are more likely to value the signal than rational voters who discount the correlation. As a

result, voters will make herding to the correct choice. On the other hand, Eyster and Rabin (2010) are based on an information cascade model in which people make sequential decisions. In this model, the past actions of others are correlated with each other because they are further influenced by the actions of others who acted before them. They point out that people are herding into an incorrect outcome by choosing actions that ignore that correlation.

Our experiments consider situations where the others' actions are correlated, not signals, so both correct and incorrect herding may occur, close to the situation considered in Eyster and Rabin (2010). Indeed, all the team members converged to the incorrect group in some cases in **All Rounds**. Among the 33 cases (three sets times 11 teams), all the members placed more bets on the same group (with no indifferent member) in Round 15 in **All Rounds** in six cases.[13] Of these six cases, there were two cases where all members bet on the incorrect group while correct herding emerged in other four cases. One involved a team that received more incorrect public signals. In the other case, one team member who received an incorrect private signal consistently placed 9-point bets on the wrong group from Rounds 5 to 15, leading other team members to also bet more on the wrong group by observing others' actions. Therefore, although frequently providing information on others' actions can prevent polarization, it can also induce people to make the wrong decisions.[14]

The latter case, which extremists greatly influence, suggests that people's actions can be influenced by people who say extreme things, such as conspiracy theories, which can lead to erroneous conclusions.

6. Conclusion

This study employed laboratory experiments to explore whether exposing individuals to the actions of others could prevent belief polarization. Our findings indicate that observing the actions of others only once does not significantly impact the frequency and size of polarization. However, repeated exposure to others' actions can prevent polarization. It is important to note that this effect is not observed when subjects believe others have limited information.

[13] In **Baseline**, on the contrary, this is only one case among the 33.
[14] It is not clear whether eliminating polarization improved welfare. Experimental results show that eliminating polarization does not increase earning points on average. See Appendix C for details.

These implications must be investigated in more detail in future work. First, our experiments intended to build a situation where common interests exist through decontextualized settings. However, many studies show the importance of contexts when considering polarization. Introducing such context into our experiments is an influential future research agenda. The second topic further examines the impact of reducing polarization on welfare. We showed that some teams converged to the wrong state of the world, and it is unclear whether reducing polarization by showing others' actions improves the subjects' welfare.

References

Acemoglu D., V. Chernozhukov, and M. Yildiz, 2016, "Fragility of Asymptotic Agreement under Bayesian Learning," *Theoretical Economics* 11, 187-225.

Adamic, L., and N. Glance, 2005, "The political blogosphere and the 2004 U.S. election: divided they blog," in *Proceedings of the 3rd international workshop on Link Discovery*, 36–43.

Andreoni, J., and T. Mylovanov, 2012, "Diverging Opinions," *American Economic Journal: Microeconomics* 4(1), 209–232.

Aragones E., M. Castanheira, and M. Giani, 2015, "Electoral Competition through Issue Selection," *American Journal of Political Science* 59(1), 71–90.

Arai, K., Y. Asako, A. Hino, and S. Morikawa, 2024, "Misunderstanding Payoff Structure Can Increase Polarization: Laboratory Experiments," mimeo.

Balietti, S., L. Getoor, D. G. Goldstein, and D. J. Watts, 2021, "Reducing opinion polarization: Effects of exposure to similar people with differing political views," *PINAS* 118 (52), e2112552118.

Baliga, S., E. Hanany, and P. Klibanoff, 2013, "Polarization and Ambiguity," *American Economic Review* 103(7), 3071–3083.

Barberá, P., 2015, "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S.," mimeo.

Baysan, C., 2021, "Persistent Polarizing Effects of Persuasion; Experimental Evidence from Turkey," mimeo.

Beam, M. M. Hutchens, and J. Hmielowski, 2018, "Facebook news and (de)polarization: reinforcing spirals in the 2016 U.S. election," *Information, Communication & Society* 21(7), 940–958.

Benoît, J., and J. Dubra, 2018, "When do Populations Polarize? An Explanation," mimeo.

Bernacer, J., J. Gercia-Manglano, E. Camina, and F. Güell, 2021, “Polarization of Beliefs as a Consequence of the COVID-19 pandemic: The Case of Spain,” *PLoS ONE* 16(7), e0254511.

Bloedel A. W., and I. Segal, 2021, “Persuading a Rationally Inattentive Agent,” mimeo.

Bullock, J., 2009, “Partisan Bias and the Bayesian Ideal in the Study of Public Opinion,” *The Journal of Politics* 71(3), 1109–1124.

Chen, D. L., M. Schonger, and C. Wickens, 2016, “oTree - An open-source platform for laboratory, online and field experiments,” *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.

Colleoni, E., A. Rozza and A. Arvidsson, 2014, “Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data,” *Journal of Communication* 64(2), 317–332.

Conover, M., B. Gonçalves, A. Flammini. and F. Menczer, 2012, “Partisan Asymmetries in Online Political Activity,” *EPJ Data Science* 1(1), 1–19.

Crocker, J., R. Luhtanen, S. Broadnax, and B. Blaine, 1999, "Belief in U.S. government conspiracies against blacks among black and white college students: Powerlessness or system blame?" *Personality and Social Psychology Bulletin* 25(8), 941–953.

Curini, L., and Hino, A., 2012, “Missing links in party-system polarization: How institutions and voters matter,” *The Journal of Politics* 74(02) 460–473.

De Marzo, P., D. Vayanos, and J. Zweibei, 2003, “Persuasion Bias, Social Influence, and Unidimensional Opinions,” *The Quarterly Journal of Economics* 118(3), 909-968.

Dixit A., and J. Weibull, 2007, “Political Polarization,” *Proceedings of the National Academy of Science of the United States of America* 104(18), 7351–7356.

Douglas, K., J. Uscinski, R. Sutton, A. Cichocka, T. Nefes, C. Ang, and F. Deravi, 2019, “Understanding Conspiracy Theories,” *Advanced in Political Psychology* 40(1), 3–35.

Dow, J. K., 2011, “Party-system extremism in majoritarian and proportional electoral systems,” *British Journal of Political Science* 41(02), 341–361.

Drague T., and X. Fan, 2016, “An Agenda-setting Theory of Electoral Competition,” *The Journal of Politics* 78(4), 1170–1183.

Egorov G., 2015, “Single-issue Campaign and Multidimensional Politics,” NBER working paper, No. 21265.

Enke, B., and F. Zimmermann, 2019, “Correlation Neglect in Belief Formation,” *Review of Economic Studies* 86: 313-332.

Fryer, R., P. Harms, and M. Jackson, 2019, "Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization," *Journal of European Economic Association* 17(5), 1470–1501.

Galliford, N., and A. Furnham, 2017, "Individual difference factors and beliefs in medical and political conspiracy theories," *Scandinavian Journal of Psychology* 58, 422–428.

Gerber, A., and D. Green, 1999, "Misperceptions about Perceptual Bias," *Annual Review of Political Science* 1999(2), 189–210.

Glaeser E., and C. Sunstein, 2009, "Extremism and Social Learning," *Journal of Legal Analysis* 1(1): 263-324.

Goeree, J., T. Palfrey, B. Rogers, and R. McKelvey, 2007, "Self-correcting Information Cascade," *Review of Economic Studies* 74(3), 733-762.

Guess, A., and A. Coppock, 2020, "Does Counter-attitudinal Information Cause Backlash? Result from Three Large Survey Experiments," *British Journal of Political Science* 50, 1497–1515.

Heatherly, K., Y. Lu, and J. Lee, 2017, "Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites," *New Media and Society* 19(8), 1271–1289.

Hindman, M., 2008, *The Myth of Digital Democracy*, Princeton: Princeton University Press.

Houston, D., and R. Fazio, 1989, "Biased Processing as a Function of Attitude Accessibility: Making Objective Judgments Subjectively," *Social Cognition* 7(1), 51–66.

Hu, L., A. Li, and I, Segal, 2023, "The Politics of News Personalization," *Journal of Political Economy Microeconomics* 1(3), 463-505.

Katz, E., and J. Feldman. 1962. "The Debates in Light of Research: A Survey of Surveys," in *The Great Debates: Background Perspective Effects,* Sidney Kraus eds., 173–223. Bloomington: Indiana University Press.

Kinder, D., and W. Mebane, Jr, 1983, "Politics and Economics in Everyday Life," in *The Political Process and Economic Change*, Kristen R. Monroe eds., 141–80. New York: Algora Publishing.

Kondor P., 2012, "The More We Know about the Fundamental, the Less We Agree on the Price," *Review of Economic Studies* 79, 1175–1207.

Lelkes, Y., G. Sood, S. Iyengar, 2017, "The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect," *American Journal of Political Science* 61(1), 5–20.

Levy, Gilat, and R. Razin, 2015, "Correlation Neglect, Voting Behavior, and Information Aggregation," *American Economic Review* 105(4), 1634-1645,

Mari, S., H. G. de Zúñiga, A. Suerdem, K. Hanke, G. Brown, R. Vilar, D. Boer, and M. Bilewicz, 2022, "Conspiracy Theories and Institutional Trust: Examining the Role of Uncertainty Avoidance and Active Social Media Use," *Political Psychology* 43(2): 277-296.

Melnikov, N., 2021, "Mobile Internet and Political Polarization," mimeo.

Miller, J. M., K. L. Sounders, and C. E. Farhart, 2015, "Conspiracy Endorsement as Motivated Reasoning: The Moderating Roles of Political Knowledge and Trust," *American Journal of Political Science* 60(4): 824-844.

Nguyen, A., and H. Vu, 2019, "Testing popular news disclosure on the "echo chamber" effect: Does political polarization occur among those relying on social media as their primary political polarization occur among those relying on social media as their primary politics news source?" *First Monday* 24(5).

Nimark, K. P., and S. Sundaresan, 2019, "Inattention and Belief Polarization," *Journal of Economic Theory* 180, 203-228.

Nöth, M., and M. Weber, 2003, "Information Aggregation with Random Ordering: Cascades and Overconfidence," *Economic Journal* 113(484), 166-189.

Novák, V., A. Matveenko, and S. Ravaioli, forthcoming, "The Status Quo and Belief Polarization of Inattentive Agents: Theory and Experiment," *American Economic Journal: Microeconomics*.

Nyhan, B., and J. Reifler, 2010, "When Corrections Fail: The Persistence of Political Misperceptions," *Political Behavior* 32(2), 303–330.

Prior, M., 2007, *Post-broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Election*, Cambridge: Cambridge University Press.

Rabin, M., and J. Shrag, 1999, "First Impressions Matter: A Model of Confirmatory Bias," *The Quarterly Journal of Economics* 114(1), 37–82.

Radnitz, S., and P. Underwood, 2017, "Is belief in conspiracy theories pathological? A survey experiment on the cognitive roots of extreme suspicion," *British Journal of Political Science* 47(1), 113–129.

Schuette, R., and R. Fazio, 1995, "Attitude Accessibility and Motivation as Determinants of Biased Processing: A Test of the MODE Model," *Personality and Social Psychology Bulletin* 21(7), 704–710.

Sigelman, L., and C. Sigelman, 1984, "Judgments of the Carter-Reagan Debate: The Eyes of the Beholders," *Public Opinion Quarterly* 48(3): 624–628.

Sunstein, C., 2001, *Republic.com*, Princeton: Princeton University Press.

Swami, V., T. Chamorro-Premuzic, and A. Furnham, 2010, "Unanswered Questions: A Preliminary Investigation of Personality and Individual Difference Predictors of 9/11 Conspiracist Beliefs," *Applied Cognitive Psychology* 24: 749-761.

Wagner, M., 2021, "Affective polarization in multiparty systems," *Electoral Studies* 69: 102199.

Weizsäcker, G., 2010, "Do We Follow Others when We Should?: A Simple Test of Rational Expectations," *American Economic Review* 100: 2340-2360.

Wilson, A., 2014, "Bounded Memory and Biased in Information Processing," *Econometrica* 82(6), 2257–2294.

Wood, T., and E. Porter, 2019, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence," *Political Behavior* 41, 135–163.

Table 1: Points to Bet and Cost Points

Points to Bet	Cost Points
1	1
2	3
3	6
4	10
5	15
6	21
7	28
8	36
9	45

Table 2: Summary of Experimental Sessions

Session	Day	Participants	Treatment
1	Dec. 22	25	Baseline
2	Jan. 7	25	Round 2.5
3	Jan 14	25	All Rounds
4	Jan. 17	30	All Rounds + Uninformed
5	Jun. 27	30	Baseline
6	Jul. 12	30	All Rounds

Table 3: Bayesian beliefs on whether the urn belongs to Group 1

The number of white (black) balls is more than that of the black (white) ones	Private Draw	
	Red (Green)	Green (Red)
0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{3}{8}$	$\frac{5}{8}$
2	$\frac{3}{10}$	$\frac{7}{10}$
3	$\frac{15}{56}$	$\frac{41}{56}$
4	$\frac{21}{82}$	$\frac{61}{82}$
5	$\frac{123}{488}$	$\frac{365}{488}$
6	$\frac{183}{730}$	$\frac{547}{730}$
7	$\frac{1095}{4376}$	$\frac{3281}{4376}$
8	$\frac{1641}{6562}$	$\frac{4921}{6562}$
9	$\frac{9843}{39368}$	$\frac{29525}{39368}$
10	$\frac{14763}{59050}$	$\frac{44287}{59050}$
11	$\frac{88575}{354296}$	$\frac{265721}{354296}$
12	$\frac{132861}{531442}$	$\frac{398581}{531442}$
13	$\frac{797163}{3188648}$	$\frac{2391485}{3188648}$
14	$\frac{1195743}{4782970}$	$\frac{3587227}{4782970}$

Table 4: Mixed-effects regression: Value of Disagreement

	Coefficient	Robust S.E.	95% CI	$P > z$
Baseline	0.167	0.419	[-0.654, 0.988]	0.690
All round	-0.826	0.332	[-1.477, -0.175]	0.013
2.5 Round	-0.233	0.425	[-1.066, 0.599]	0.583
Theoretical	0.130	0.329	[-0.515, 0.773]	0.693
All round*Theoretical	1.321	0.453	[0.434, 2.209]	0.004
Baseline*Theoretical	-0.121	0.444	[-0.992, 0.750]	0.785
2.5 Round*Theoretical	0.779	0.503	[-0.206, 1.764]	0.121
Constant	2.709	0.302	[2.117, 3.301]	0.000
Random-effects parameters				
Team*Session: Identity	Estimate	Robust S.E.	95% CI	
var(Constant)	0.625	0.110	[0.443, 0.883]	
var(Residual)	2.665	0.118	[2.444, 2.906]	
	ICC	S.E.	95% CI	
	0.190	0.028	[0.140, 0.252]	
Number of obs	2,625			
Number of groups	97			

Reference Group: All round + Uninformed

Table 5: Mixed-effects regression: Frequency of Disagreement

	Coefficient	Robust S.E.	95% CI	$P > z$
Baseline	0.076	0.059	[-0.040, 0.192]	0.197
All round	-0.159	0.065	[-0.287, -0.031]	0.015
2.5 Round	0.043	0.067	[-0.089, 0.175]	0.523
Theoretical	0.044	0.059	[-0.071, 0.160]	0.451
All round*Theoretical	0.193	0.079	[0.039, 0.348]	0.014
Baseline*Theoretical	-0.077	0.071	[-0.215, 0.062]	0.277
2.5 Round*Theoretical	0.027	0.080	[-0.129, 0.183]	0.737
Constant	0.726	0.050	[0.627, 0.825]	0.000
Random-effects parameters				
Team*Session: Identity	Estimate	Robust S.E.	95% CI	
var(Constant)	0.019	0.003	[0.013, 0.026]	
var(Residual)	0.165	0.006	[0.154, 0.176]	
	ICC	S.E.	95% CI	
	0.072	0.013	[0.050, 0.103]	
Number of obs	2,625			
Number of groups	97			

Reference Group: All round + Uninformed

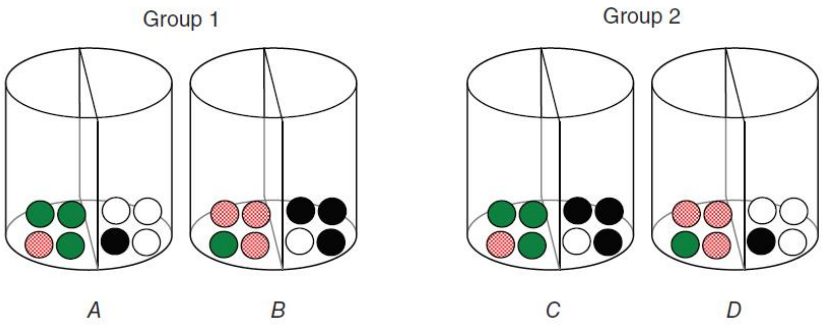
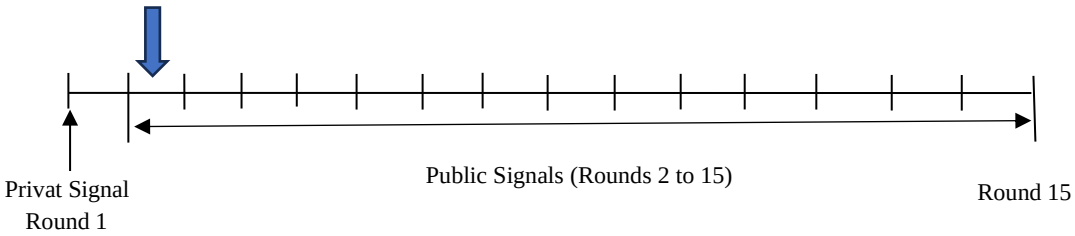
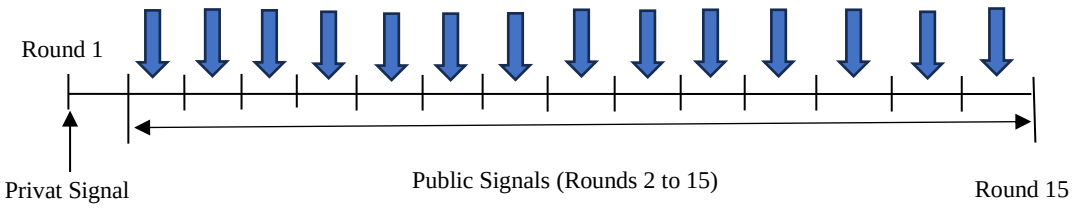


Figure 1: Four Urns

Source: Andoreoni and Mylovanov (2012)

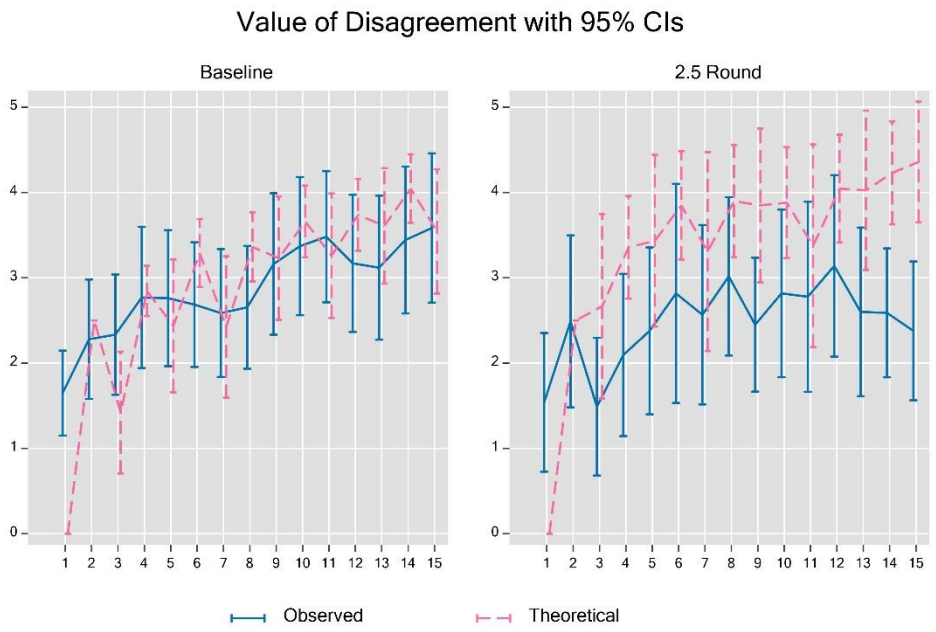


(a) Round 2.5

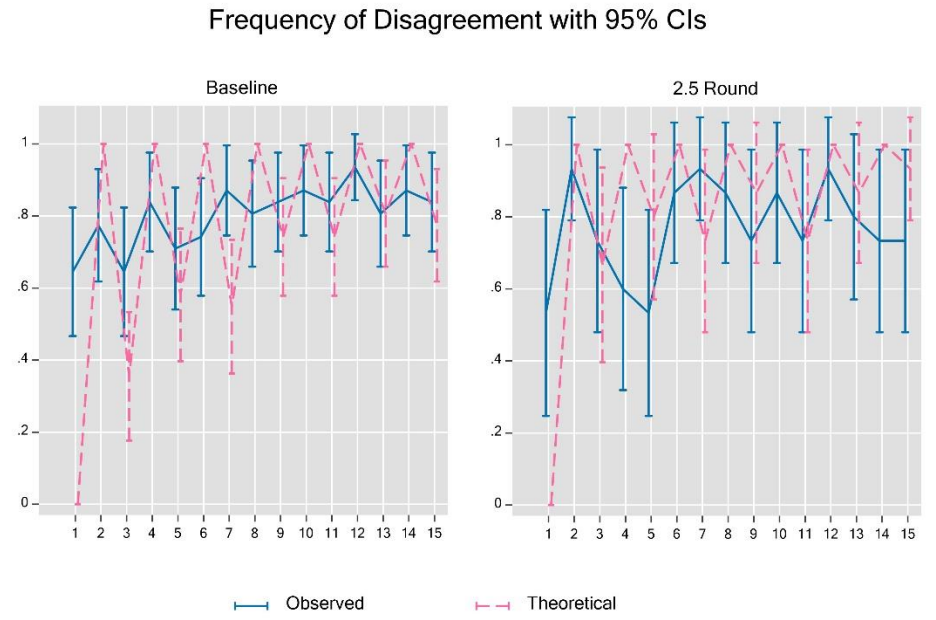


(b) All Rounds

Figure 2: The Timings to Inform Others' Actions

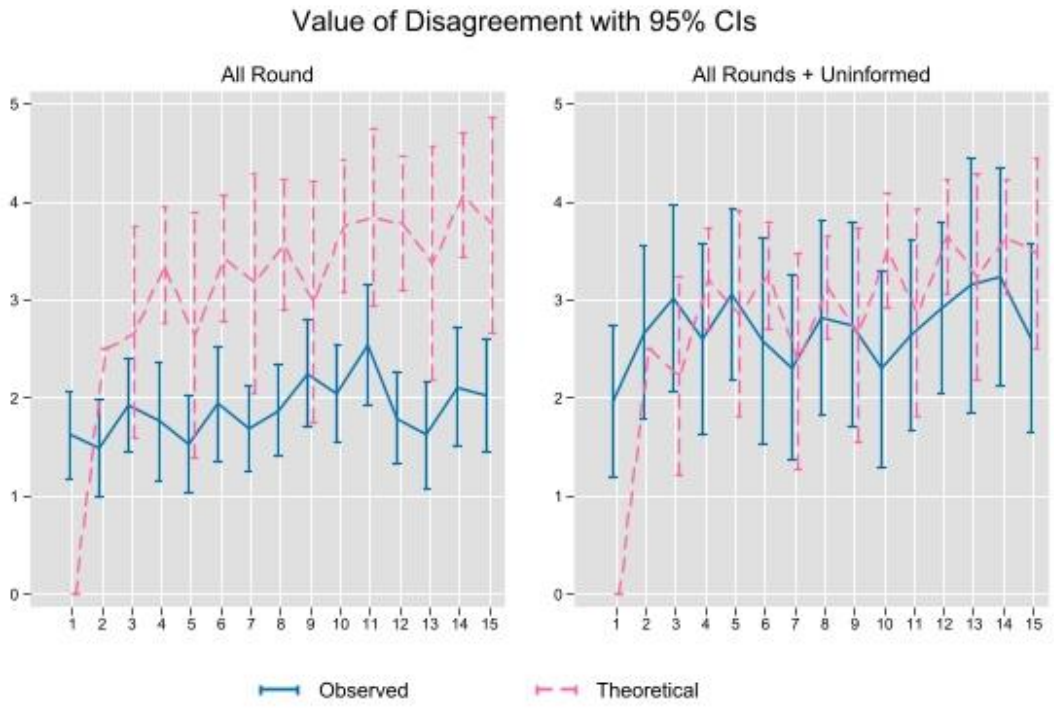


(a) Value of Disagreement

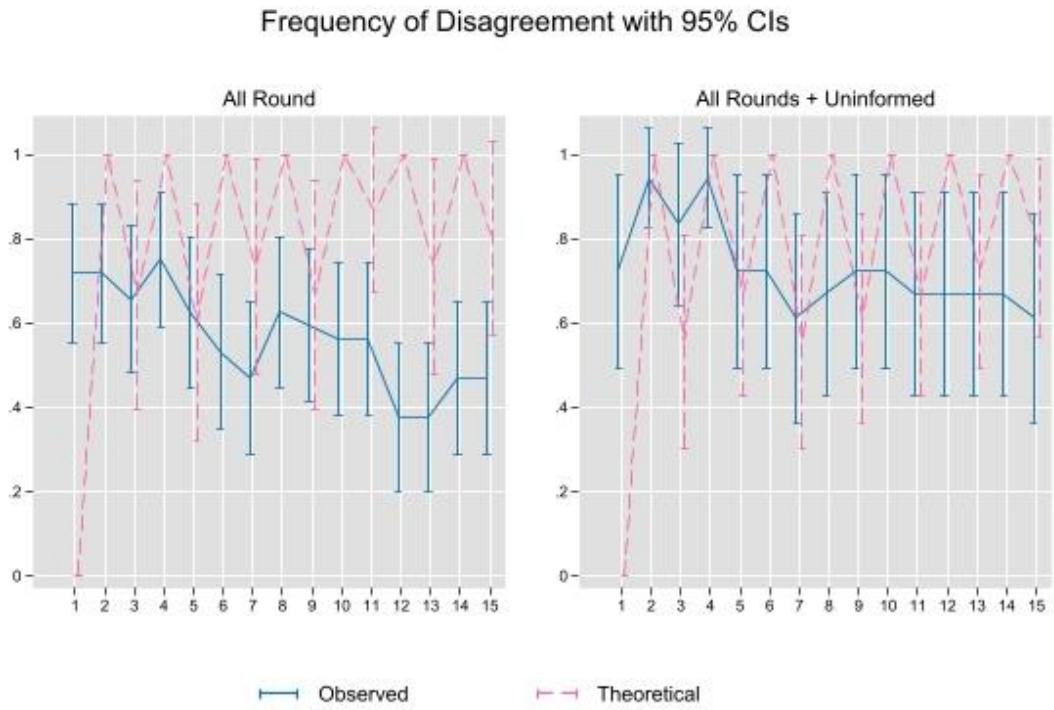


(b) Frequency of Disagreement

Figure 3: Baseline and Round 2.5

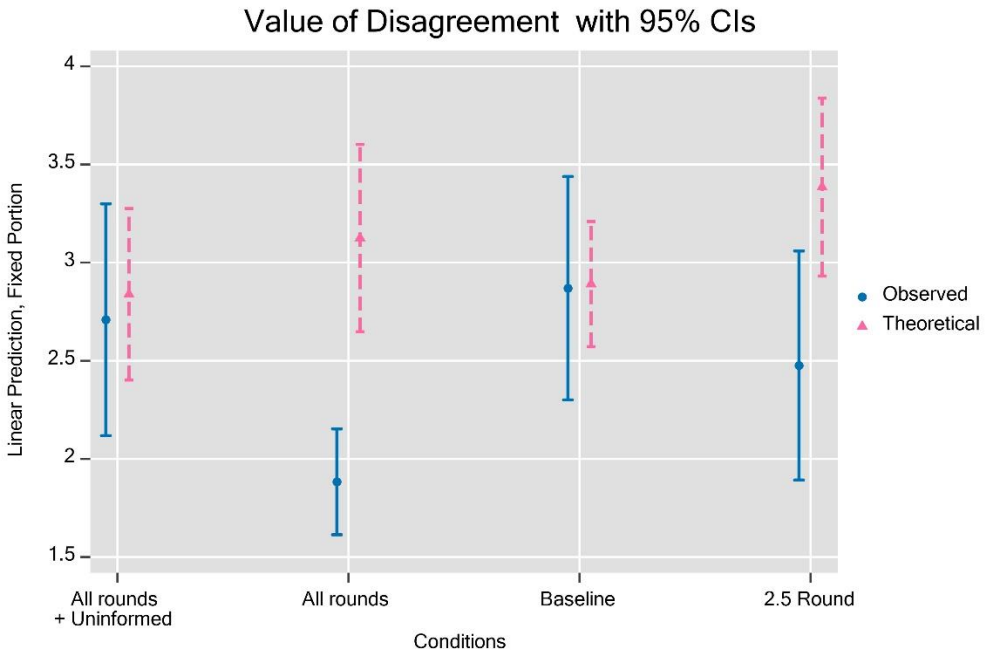


(a) Value of Disagreement

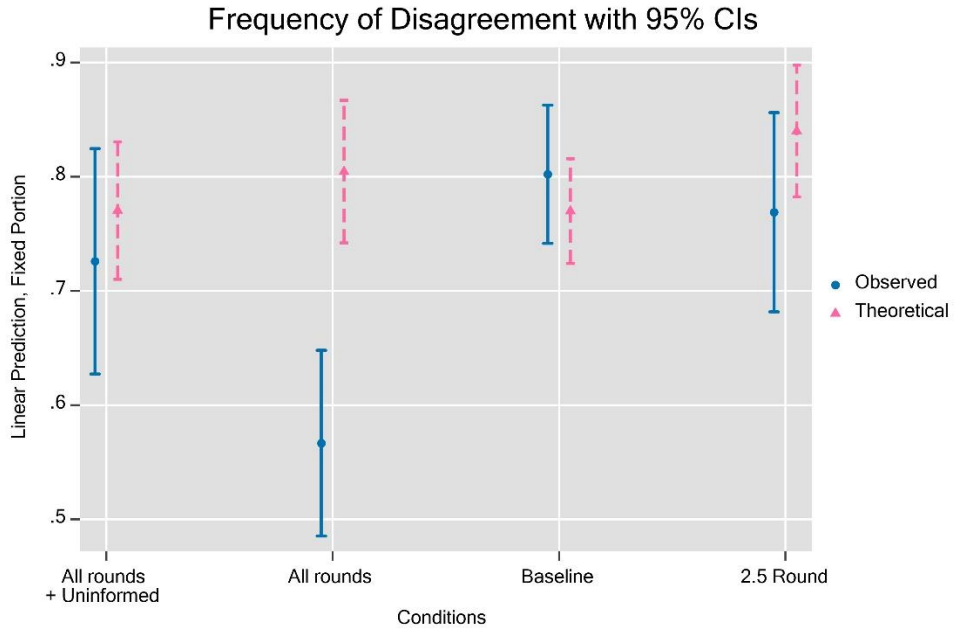


(b) Frequency of Disagreement

Figure 4: All Rounds



(a) Value of Disagreement



(b) Frequency of Disagreement

Figure 5: Linear Predictions from Random-intercept Models

Supplementary Information: Online Appendix

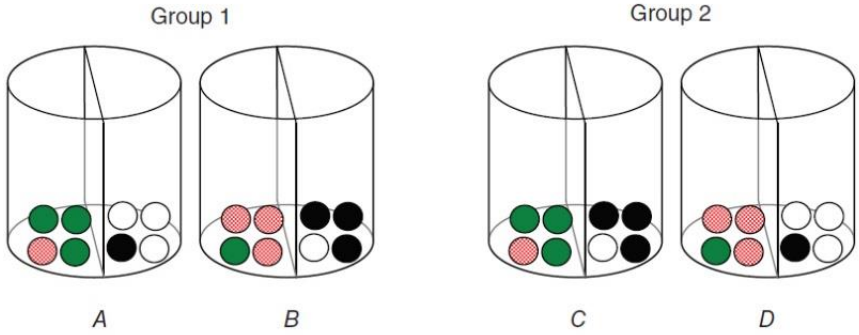
Appendix A: Instructions

Note: The paragraphs starting with parentheses [ ] are instructions for a specific session. [Round 2.5] and [All Rounds] refer to sessions where the total points of past bets are announced after the second round and in all rounds, respectively. [Uninformed] refers to a session in which there is a participant who does not observe the first draw. All other paragraphs without parentheses are common across all sessions.

Thank you for your participation in this study. You are taking part in an experiment on decision-making. You will not be allowed to talk with other participants or to take notes during the experiment. Please turn off your cellphones. At the end of the experiment, you will be given some money as compensation for your time.

There are 30 (25) participants in this experiment, divided into 6 (5) teams of 5 members each. There are three sets, and, decisions will be made among team members in each set. The details of the decision-making process are outlined below. The members of each team will be randomly matched at the beginning of each set; thus, the composition of the members will change in each set. You will not know who will be playing with you on the same team.

The details of the decision-making process in this experiment are as follows. In each round, a ball is drawn from one of four urns, as shown in the figure below. The urns are labeled A, B, C, and D and divided among the two groups: Group 1 has urns A and B, and Group 2 has urns C and D. Your task during the study will be to try to determine whether the urn we are using is in Group 1 (urn A or B) or Group 2 (urn C or D). To be precise, you are required to place bets on Group 1, Group 2, or on both groups. Each urn will be selected at a probability of 25%.



As you can see, the urns have red, green, white, and black balls. Moreover, the urns have two separate compartments: a red/green compartment and a white/black compartment. To be precise:

- Urns A and C have three green balls and one red ball.
- Urns B and D have one green ball and three red balls.
- Urns A and D have three white balls and one black ball.
- Urns B and C have one white ball and three black balls.

One of the urns will be randomly selected from which to draw balls. Based on these draws, you will have to guess whether the urn is in Group 1 or Group 2.

How the balls will be drawn:

We will make a total of 15 draws from the urn selected at the beginning of each set. Only the first drawing will be performed in front of each participant. The color of the ball drawn in front of you cannot be observed by the other members. In addition, the color of the ball drawn in front of the other members cannot be seen. Since this drawn will be privately witnessed by each participant, we will call it a "private draw." This first draw will be from the red/green compartment only.

A bet will be placed after each draw. The process will be counted until the bet consists of one round. For example, the first round will contain the first private draw and first bet.

[Uninformed] However, at least one member will not be able to observe the color of the first draw. You are not informed how many members will not be able to witness a private draw. Even if the first draw cannot be seen, a bet can be made in the first round.

In the following 14 rounds, all members of the team will see the color of the ball. In other words, all members of the team will see the same color. Since this draw will be publicly observed by all members, we will call it a "public draw." Public draws are made only from the white/black compartment.

After each draw, we will return the ball to the urn before making the next draw. Therefore, the probability of obtaining each color will be the same in all the rounds.

[except All Rounds and Round 2.5] After your bet on the 15th draw, we will reveal the total points of bets for each group in your team. This is the summation of all points to be bet on for each group from rounds 1 to 15. Then, you will have an additional opportunity to bet. Hence, there will be 16 rounds of bets for each set.

[Round 2.5] There are two timings to announce the "total point" which is the summation of all points to bet for each group by all members of your team in the previous rounds. During these two timings, you will have the chance to place a bet after the total point is announced. The first time will be round 3. After the bet of the first public draw in round 2, the total point until round 2 will be announced in the third round. This is the summation of all the points to be placed for a bet in rounds 1 and 2. The second time will be round 17. After the bet in round 16, the total point until round 16 will be announced. It is the summation of all points up until the bet from rounds 1 to 16. The following table summarizes the process.

Round 1	Private draw
Round 2	Public draw
Round 3	Announce the total point until round 2
Rounds 4 to 16	Public draws
Round 17	Announce the total point until round 17

[All Rounds] At the end of each round, the "total point" will be announced. The "total point" is the summation of all points to bet for each group by all members of your team in the previous rounds. For example, at the end of round 3, the summation of all points to place a bet for each group by team members from rounds 1 to 3 will be announced. After the total points are announced in round 15, you will have one more chance to place a bet. Thus, there will be 16 rounds of bets for each set.

<u>How to place a bet</u>

After each draw, 0 to 9 bets can be placed for each group of urns. That is, you can bet 0 to 9 points for Group 1 and 0 to 9 points for Group 2. You can place up 18 bets. Only integers can be chosen.

Points that are betted for the correct group will be returned tenfold. Points that are betted for the incorrect group will not be returned. We will call points to be returned the "return point."

However, bets also incur costs. We will call them the "cost point" which is the summation of all integers from 0 to the points used in a bet for each group. The first bet you make for Group 1 costs one point. If you bet one more point—that is, if you bet two points for Group 1—this will cost an additional two points, so the cost point is three points. The third bet will cost an additional three points, so the cost point in total is 1+2+3=6 points.

Points to bet	Cost point
1	1
2	3
3	6
4	10
5	15
6	21
7	28
8	36
9	45

The cost points for each case are shown on the table. Please note that when you bet on both groups, you must add the cost points from Group 1 and Group 2. For example, if you bet nine points in both groups, your cost point would be 45+45=90, not 45.

Your "earning points" in each round will be the return point minus the cost point. Earning points can be both positive and negative because there is a possibility that the cost point is strictly higher than the return point. The bet for each round will be independent of the other bets. You can choose any point to bet on, regardless of the points you bet on in past rounds.

We may ask you to make a decision earlier when it is too slow. Thank you for coordinating a smooth operation.

Earning money

We will select 2 of the 16 rounds (**Round 2.5**: 17 rounds) at random as "earning rounds." We will exchange 3 yen for each point of earning points that you earned in the selected earning rounds. At the same time, we will give 45 points per earning round to all participants at the end of 16 rounds (**Round 2.5**: 17 rounds). Because there are 2 earning rounds, you will obtain 90 points per set. Your final earning points are the sum of the earning points of the 2 earning rounds and 90 points. Then, we will pay you 3 yen for each point at the end of this experiment. For example, if you earned 70 earning points in 2 earning rounds, your final earning points would be 160 after adding 90 points. In this case, you would receive JPY 480.

Note: We use your points from earning rounds to determine payments. The cost and returned points in the other rounds are irrelevant to your payments. The earning round is not announced in advance.

Sets

We have described how a single set, including 16 rounds (**Round 2.5**: 17 rounds), will be run. Today, we will run three sets, where one set contains 16 rounds (**Round 2.5**: 17 rounds). In each set, we will select an urn. First, we will randomly choose one of four urns to draw from with equal probability (25%). For the next set, we will randomly choose one of four urns with equal probability (25%). The

same is true for the third set. The selection of the urn in past sets will not affect the selection of the urn in future sets. The sum of payments from each set will be paid to you at the end of the experiment. We will truncate the first place of payment.

In addition, 1,000 yen will be paid as a participation fee for all participants.

Procedure

Everything we have described will be run on a computer. The ball will not be drawn. All the experiments will be performed using a computer screen. The computer will help you keep track of the colors of balls drawn in past rounds. The computer will also tell the true urn and your earning points in all rounds at the end of the set.

Practice set

Before beginning the experiment, we will go through a practice set. The practice set will have one private draw from the red/green compartment, and four public draws from the white/black compartment. This practice set will not include any earning rounds.

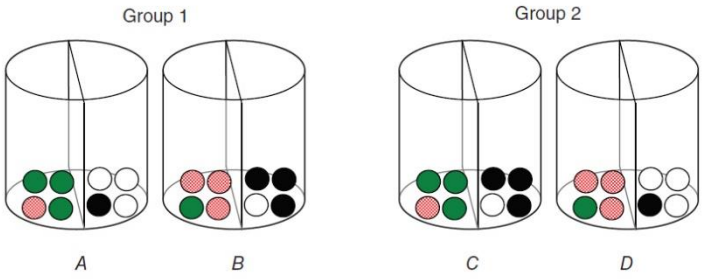
[All Rounds and Round 2.5] In the practice set, the total points will not be announced.

[Uninformed] In the practice set, all participants will be able to observe the color of the private draw.

This practice set will help you to understand this experiment better.

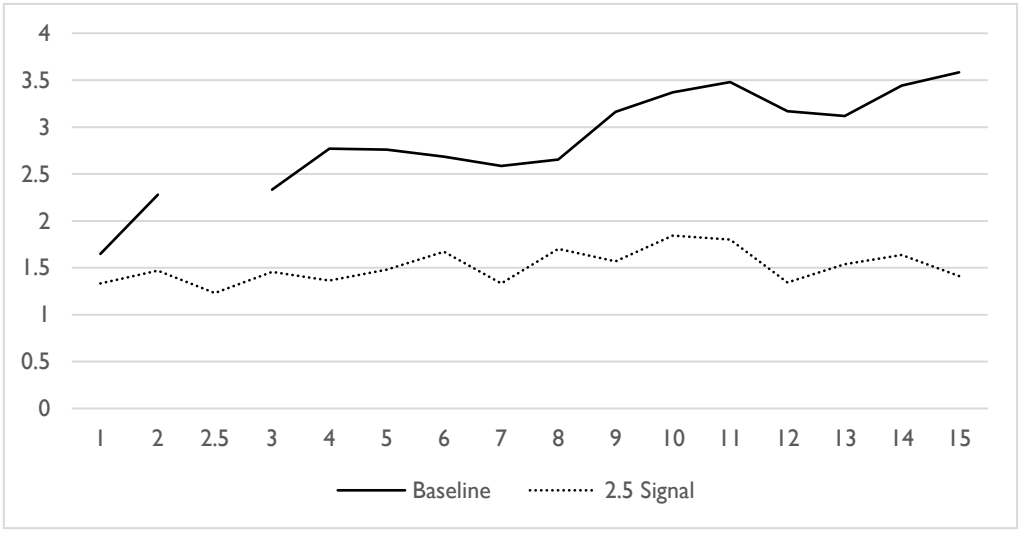
The separate sheet shows the figure for reference during the experiments. Let us begin with the practice set. If you have any questions during the experiments, please raise your hand.

Separated sheet

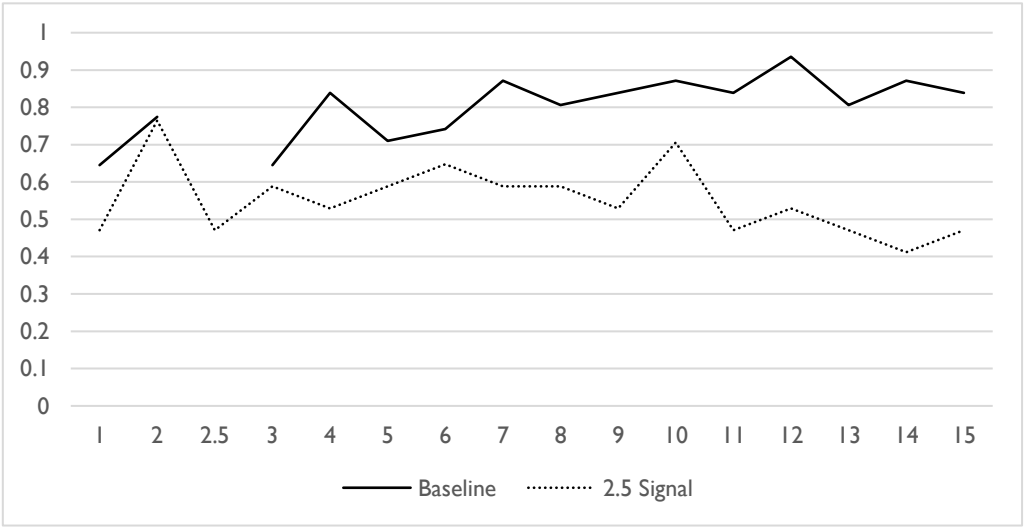


Appendix B: Observing Others' Private Signals in Round 2.5

We show the private signals observed by all the team members instead of others' actions only once between rounds 2 and 3 (called **2.5 Signal**). Figure B illustrates the value and frequency of disagreement compared to **Baseline**, and both are much lower than the ones in **Baseline**. It suggests that the participants did not forget the information shown in the early round until Round 15 and that polarization disappeared when they shared private signals.



(a) Value of Disagreement



(b) Frequency of Disagreement

Figure 4: 2.5 Round and 2.5 Signal

Appendix C: Welfare Improvement by Preventing Polarization

Figure C-1 shows the total earning points from Rounds 1 to 15, highlighting that no significant difference exists among the treatments. Therefore, we cannot conclude that reducing polarization by observing others' actions induces our subjects to obtain higher earning points.

In **All Rounds**, the subjects can observe others' actions, allowing them to infer the private signal observed by others and make more accurate predictions of the urn selected. To check whether the subjects' expectations were accurate, Table C-2 shows the rate of incorrect bets, defined as the rate at which subjects placed more bets on an incorrect group in Round 15. Betting the same points on both groups does not amount to an incorrect bet. We see no significant difference among the treatments exists. **All Rounds** do not have a lower rate of incorrect bets.

One reason the rate of incorrect bets did not decrease in **All Rounds** is that some of the subjects continued to ignore others' actions. A more important reason is that, in some cases, all the team members converged on the unselected group, as discussed in Section 5.2

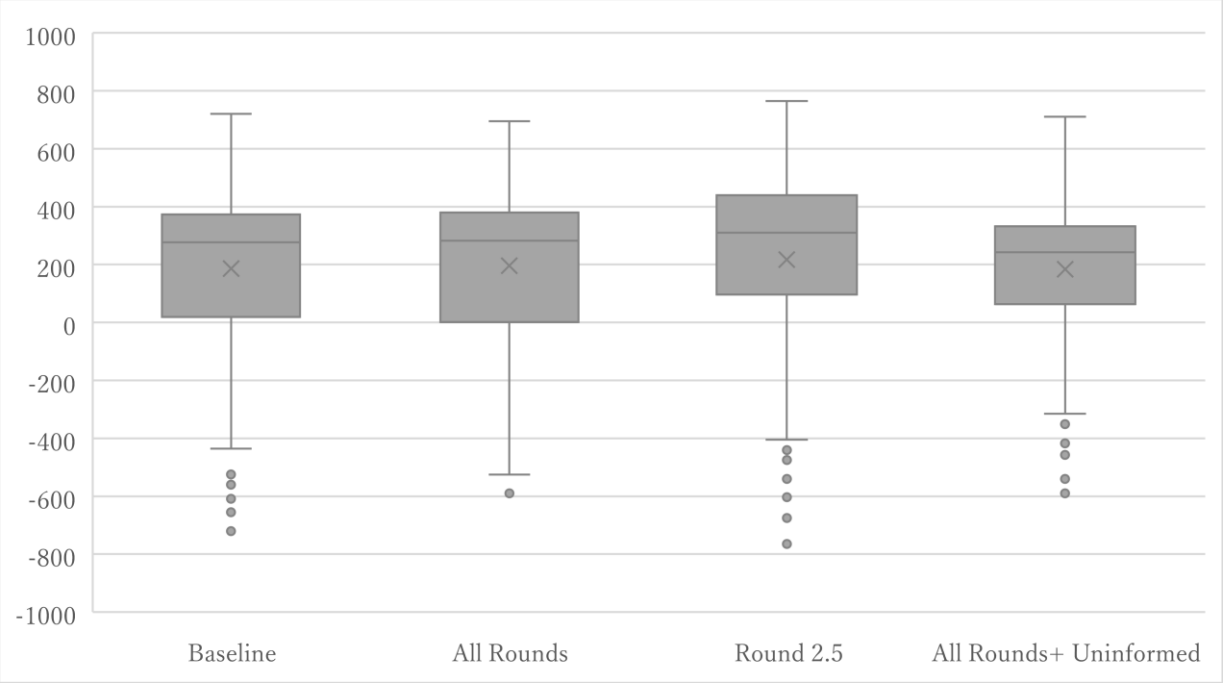


Figure C-1: Total Earning Points

Table C-2: Rate of Incorrect Bet

	All Sets	Sets 2 and 3
Baseline	0.382	0.373
All Rounds	0.364	0.436
2.5 Round	0.36	0.38
2.5 Signal	0.311	0.4