When Motivated Reasoning Grows: The Effect of Order of Information*

Hyundam Je† Sora Youn‡

May 2024

Abstract

We investigate the effect of information order on motivated reasoning. Our experimental evidence suggests that the timing of informativeness significantly affects motivated reasoning. Specifically, when participants learn about a signal's informativeness after receiving the signal, they exhibit increased motivated reasoning compared to when they learn its informativeness beforehand. Our research suggests that disclosing the informativeness earlier can serve as a simple yet effective method to mitigate motivated reasoning.

*We thank Syngjoo Choi, Catherine Eckel, Christine Exley, Charles Holt, Daeyoung Jeong, Paolina Medina, Kirby Nielsen, and Emanuel Vespa for their helpful comments. We also thank participants at the 2021 North-American ESA Meeting, the 2023 North-American ESA Meeting, the 2023 SEA Annual Meeting, the 2024 KEBEA Winter Workshop, and the NUS Applied Economics Student Workshop. This research was funded by the Department of Economics, Texas A&M University, and the College of Economics and Finance, Hanyang University. The study was pre-registered in AsPredicted (126181) and had ethics approval from Texas A&M University's Institutional Review Board (IRB2021-1435M). All errors are our own.

†Je: School of Economics, University of Sydney, hyundam.je@sydney.edu.au

‡Youn: Department of Economics, Texas A&M University, sorayoun@tamu.edu

1 Introduction

Motivated reasoning is a cognitive phenomenon where individuals are inclined to process information that aligns with their existing beliefs and desires (Kunda, 1990; Bénabou and Tirole, 2002; Epley and Gilovich, 2016; Zimmermann, 2020) In this paper, we present novel evidence that the order of information affects the level of motivated reasoning: receiving a signal before learning about its informativeness can exacerbate the extent of motivated reasoning compared to receiving it afterward.

Consider two scenarios: In the first scenario, you visit an investment website and you know its news is not credible. Then, on that website, you see a post predicting that the stock you own will soon increase in value. In the second scenario, you first come across a post stating that your stock will increase, and afterward, you find out that the source website is not credible. In both cases, you receive the same set of information: news and its credibility. The only difference is the sequence in which this information is presented.

Motivated reasoning suggests that individuals update their beliefs to align with their preferences: since you own the stock, you are more likely to believe the news is true, even if you know that the website is not trustworthy. Previous literature has provided evidence that individuals shift their beliefs to match their preferences even after realizing that the news source is not credible (Gonçalves et al., 2021; Ecker et al., 2022; Wittrock et al., 2023). In this paper, we present novel evidence that the order of information affects motivated reasoning: the level of motivated reasoning is higher in the first scenario compared to the second one.

Order effects indicate that beliefs can be influenced by the sequence in which information is received (Hogarth and Einhorn, 1992). To verify the order effect on motivated reasoning, we conducted an online experiment using a box-and-ball setting. Participants predicted the color of a Target Ball drawn from Box B, which contained two balls of either yellow or blue. To assist their decisions, a Clue Ball was also drawn either from Box A (containing one yellow and one blue ball) or from Box B. If the Clue Ball was drawn from Box A, its color was uninformative

for predicting the color of the Target Ball. However, if the Clue Ball was drawn from Box B, its color provided information about the composition of Box B and, thus, the color of the Target Ball. In the Box First treatment, participants first learned which box the Clue Ball was drawn from, and then its color. In the Color First treatment, the order was reversed: participants first saw the color of the Clue Ball and then learned which box it came from.

Results reveal that the order of information significantly impacts motivated reasoning. Participants who received the signal first exhibited a higher degree of motivated reasoning compared to those who first learned about its informativeness. This pattern was evident only when the color preference was linked to a monetary reward, suggesting that motivated reasoning is driven by monetary incentives rather than inherent color preferences.

To explain the experimental findings, we provide a framework to formalize the results. We elucidate the order effect by utilizing the concept of the primacy effect, where individuals pay more attention to the first piece of information provided (Jones et al., 1968; Ahlering and Parker, 1989; van Erkel and Thijssen, 2016; Saccardo and Serra-Garcia, 2023). According to the primacy effect, participants are more likely to focus on the informativeness of the signal when they learn about its informativeness first, thus reducing the potential for motivated reasoning.

Motivated reasoning can lead to several problems across various domains. It fosters bias and polarization, as individuals interpret information in ways that reinforce their preexisting beliefs, leading to greater ideological divides (Klaczynski, 2000; Redlawsk, 2002; Taber et al., 2009). Additionally, motivated reasoning can affect decision-making processes, resulting in skewed and partisan decisions (Bolsen and Palm, 2019). It also contributes to overconfidence, where individuals dismiss contradictory evidence, maintaining an inflated sense of certainty in their beliefs (Thaler, 2021). Overall, motivated reasoning distorts objective analysis and perpetuates cognitive biases, affecting both personal decisions and society as a whole.

Our paper presents practical implications. Previous studies have explored the adverse

effects of motivated reasoning and researched various methods to reduce it (Kunda, 1990; Beattie and Snider, 2019; Carpenter, 2019; Cotter et al., 2020). Our findings suggest that the timing of informativeness could play a key role. We propose that when individuals engage with potentially bias-inducing information, postponing the disclosure of the signal's informativeness could increase the level of motivated reasoning. Practically, it implies that simply changing the timing of informativeness could be an effective strategy to curb motivated reasoning.

The paper is organized as follows: Section 2 introduces the experimental design. Section 3 details the framework and predictions regarding the experimental results. Section 4 presents the experiment's results, and Section 5 provides the conclusion.

2 Design

There are two boxes, Box A and Box B, each containing two balls. Box A contains one yellow ball and one blue ball. Box B has two balls, each either yellow or blue with equal probability, but their composition is unknown to participants.[1] The experimental setting is summarized in Figure 1.

The computer independently draws two balls: a Target Ball and a Clue Ball. A *Target Ball* is drawn from Box B. The main experimental task for participants is to are asked to report the percentage chance (ranging between 0 and 100) that they believe the Target Ball is yellow or blue. To assist their decisions, the computer also draws a *Clue Ball* from either Box A or Box B with equal probability. If the Clue Ball is drawn from Box A, its color conveys no information about the composition of Box B or the color of the Target Ball. On the contrary, if the Clue Ball is drawn from Box B, the color of the Clue Ball hints about the color of the Target Ball: a yellow Clue Ball suggests a higher probability that the Target Ball is yellow, while a blue Clue Ball indicates a greater chance that the Target Ball is blue. In short, the

[1]That is, Box B contains either two yellow balls, two blue balls, or one yellow and one blue ball. In one version of our experiments, participants are informed about the distribution, while in the other, they are not. In the main analysis, we pool the two versions as we find no differences.

color of the Clue Ball serves as a signal about the color of the Target Ball, which *box* the Clue Ball is drawn from represents the informativeness of the signal.

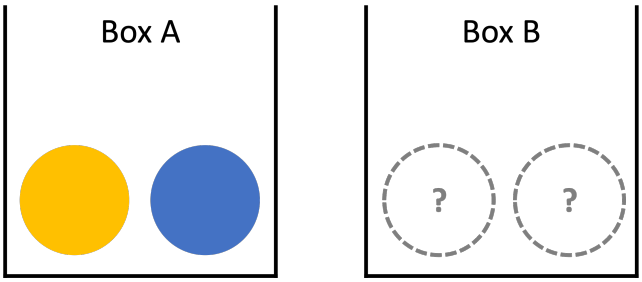


Figure 1: Experimental Setting

There are 5 rounds in total, with each round being independent of the others.[2] In each round, we elicit beliefs three times as we sequentially reveal a signal and its informativeness regarding the color of the Target Ball. The order of revelation varies across the treatments: In the *Box First* treatment, the informativeness of the signal is revealed first, whereas in the *Color First* treatment, the signal itself is revealed first. The key feature of the experimental environment is that both treatments deliver the same information — the signal and its informativeness — but in different orders.

Specifically, in the Box First treatment, participants report their *prior* beliefs about the color of the Target Ball at the beginning of every round. Then, they are informed about the informativeness of the signal: "The Clue Ball is drawn from [Box A/Box B]. You don't see the color of the Clue Ball at this point." Afterward, participants report their *interim* beliefs. They then learn about the signal itself, equipping them with complete information at the posterior stage: "The color of the Clue Ball is [Yellow/ Blue]. The Clue Ball is drawn from [Box A/ Box B]." Finally, they report their *posterior* beliefs.

The procedure of the Color First treatment is exactly the same, except for the order of

[2]To prevent potential spillover effects, we inform participants at the beginning of the experiment that it consists of two parts. During the first part, which covers the first round, participants are unaware of the task of the second part. Upon entering the second part, they are instructed to repeat the same task as in the first round four more times. We do not find evidence of a spillover effect, hence, we pool all five rounds for the data analysis.

information. After the prior stage, participants are informed about the signal, instead of its informativeness: "The color of the Clue Ball is [Yellow/Blue]. You don't know which box it was drawn from at this point." After the interim stage, they learn about its informativeness. Therefore, they have complete information at the posterior stage, the same as in the Box First treatment. Table 1 describes the procedure for each treatment.

Box First	**Color First**
Prior	Prior
↓	↓
"The Clue Ball is from **Box A/B**."	"The Clue Ball is **Yellow/Blue**."
↓	↓
Interim	Interim
↓	↓
"The Clue Ball is **Yellow/Blue**."	"The Clue Ball is from **Box A/B**."
↓	↓
Posterior	Posterior

Table 1: Treatments

We employ a binarized scoring rule to incentivize belief elicitation. At the end of the experiment, the computer randomly selects one round. Within this chosen round, it randomly selects one of three reported beliefs for payment. The accuracy of the selected belief in the chosen round determines the chance of winning a prize. This winning probability increases as participants report a higher probability of the Target Ball being yellow when it actually is yellow, and blue when it is blue. This mechanism has been widely used in studies because it ensures that any decision-maker aiming to maximize their chance of winning a prize will report their beliefs truthfully, regardless of their risk preferences (e.g., Hossain and Okui, 2013). We emphasize to the participants that the payment rule is designed to allow them to maximize their chances of winning the prize by reporting their most accurate beliefs. However, we do not provide the mathematical details of the scoring rule unless requested (Danz et al., 2022).

At the beginning of the experiment, participants are asked to choose between yellow or blue

as their preferred color. In our main experiment, the *Variable Reward* experiment, we aim to elicit motivated reasoning by varying the size of the prize. When the color of the Target Ball matches the participant's preferred color, the prize is \$1.5; otherwise, it is \$0.5. Consequently, participants who select yellow as their preferred color would hope that the Target Ball is yellow, whereas those who choose blue would prefer a blue Target Ball.

To investigate whether the motivated reasoning observed in the *Variable Reward* experiment is based solely on monetary incentives, we conducted an additional experiment. In the *Fixed Reward* experiment, all procedures remained identical except for the reward structure: participants received a fixed prize of \$1, regardless of whether the Target Ball matched their preferred color. If motivated reasoning is primarily influenced by monetary incentives rather than color preferences, then we would not expect to observe it in the *Fixed Reward* experiment.

The experiment was conducted in Prolific, an online platform for recruiting research participants. A total of 299 participants took part in the experiment, with 200 participants in the Variable Reward experiment and 99 participants in the Fixed Reward experiment. On average, participants spent 12 minutes to complete and they earned \$2.7, including the participation fee.

3 Framework

We adapt the framework of Saccardo and Serra-Garcia (2023), which explains order effect by utilizing the primacy effect. The primacy effect is a well-established cognitive bias wherein individuals tend to pay more attention to the first piece of information they encounter compared to information received later on (for a review, see Benjamin, 2019).

We assume that individuals always pay attention to and 'encode' a signal (i.e., the color of the Clue Ball). However, the informativeness of the signal (i.e., which box the Clue Ball is drawn from) may not always be encoded.[3] Instead, it depends on the timing of informativeness.

[3]This assumption reflects that individuals sometimes might fail to recall the source of a story while they

To be specific, the primacy effect suggests that individuals encode information more often when they are first presented with the informativeness of the signal, compared to when they are presented with the signal itself first. Let λ^f denote the probability that the informativeness of the signal is encoded under each treatment $f \in \{\text{Color First}, \text{Box First}\}$. Therefore, the following inequality holds: $0 < \lambda^{\text{Color First}} < \lambda^{\text{Box First}} < 1$.

3.1 Preliminaries

To simplify the analysis, we construct *ex-post* log odds of beliefs to align with the color of the (realized) Clue Ball. When the Clue Ball is yellow, the log odds capture the ratio of the probabilistic belief that the Target Ball is yellow to the probabilistic belief that the Target Ball is blue. Similarly, when the Clue Ball is blue, the log odds capture the ratio of the probabilistic belief that the Target Ball is blue to the probabilistic belief that the Target Ball is yellow. This approach ensures that the change from the log odds of prior beliefs to the log odds of posterior beliefs is symmetric, regardless of the color of the Clue Ball, when its source is fixed.

	Source of the Clue Ball	Preferred Color		Non-Preferred Color	
		Pr(Box A)	Pr(Box B)	Pr(Box A)	Pr(Box B)
Encoded	Box A	1	0	1	0
	Box B	0	1	0	1
Not Encoded	Box A	$1 - 0.5^m$	0.5^m	0.5^m	$1 - 0.5^m$
	Box B	$1 - 0.5^m$	0.5^m	0.5^m	$1 - 0.5^m$

Table 2: Subjective assessment of informativeness

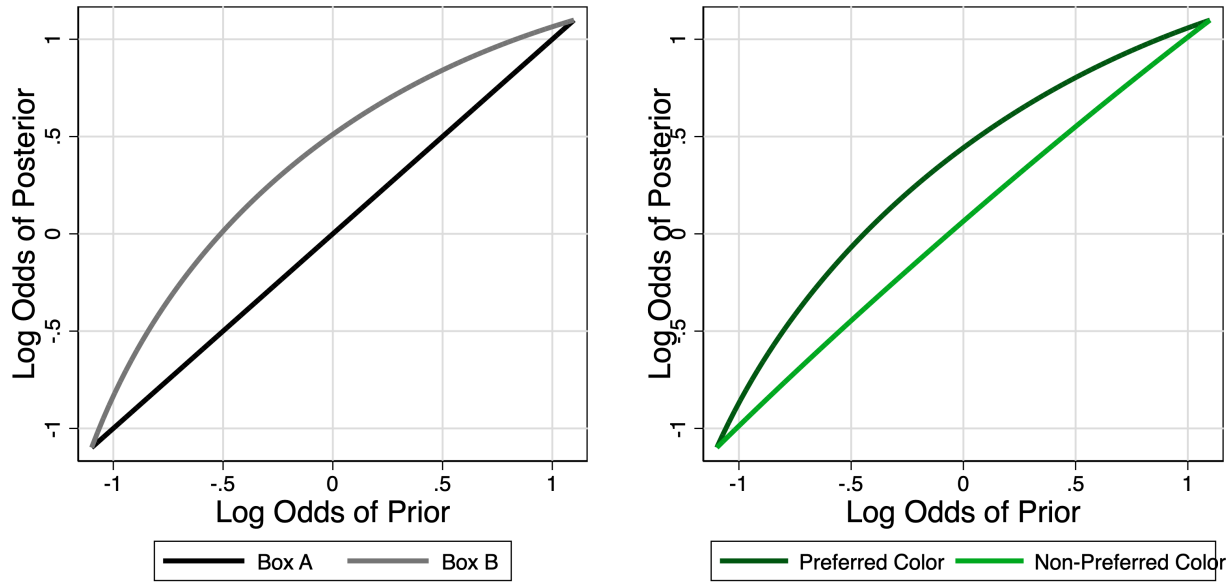
3.2 Subjective assessment of informativeness

When informativeness is encoded, the individual correctly identifies the box from which the Clue Ball is drawn. Individuals update their beliefs with certainty, and there is no room for

correctly remember the contents of the story.

distortion in assessing informativeness. The first two rows in Table 2 summarize these cases.

Figure 2a presents the benchmark cases. When the Clue Ball is drawn from Box A, its color is uninformative for guessing the color of the Target Ball. The log odds of posterior beliefs should remain unchanged from the log odds of prior beliefs, as shown by the solid 45-degree line. On the other hand, when the Clue Ball is drawn from Box B, its color provides hints about the color of the Target Ball. Thus, individual integrates this information and strengthens the log odds of posterior belief compared to the log odds of prior beliefs. This adjustment is represented as the dashed line, indicating a Bayesian benchmark. Note that whether the color is preferred or not does not affect belief updating when informativeness is encoded.



(a) Encoded

(b) Not encoded

Figure 2: Benchmarks

When informativeness is not encoded, the process of assessing information can be distorted in such a way that it reinforces individuals' preferred beliefs (e.g., Thaler, 2023). To capture this formally, we introduce a parameter that represents the levels of motivated reasoning, denoted by $m \in (0, 1)$. When $m = 1$, the individual correctly perceives that the Clue Ball comes from Box A or Box B with equal probability. However, as m approaches 0, this perception

becomes biased, favoring the individual's preferred beliefs. Specifically, when the Clue Ball's color is preferred, which induces more reward, they are more likely to believe that it comes from Box B. Conversely, when it is non-preferred, they are more likely to believe that the Clue Ball is from Box A. The last two rows in Table 2 summarize the subjective probabilities when informativeness is not encoded.

For illustration, Figure 2b presents the benchmark cases when $m = 0.2$. Unlike situations where informativeness is encoded, belief updating depends on whether the Clue Ball's color is positive or negative, rather than on which box the Clue Ball is drawn from. Controlling for the log odds of prior beliefs, the log odds of posterior beliefs respond more to a positive than to a negative color. When the Clue Ball's color is positive, individuals are more likely to update their posterior beliefs to align with their preferences.

3.3 Hypotheses

We first predict that monetary incentives are the primary source of motivated reasoning, rather than intrinsic color preferences.

Hypothesis 1. *Motivated reasoning exists in the Variable Reward experiment, but not in the Fixed Reward experiment.*

Now, we hypothesize the specific dynamics of motivated reasoning in the Variable Reward experiment. Our benchmark analysis suggests that the likelihood of the signal's informativeness being encoded is lower in the Color First treatment compared to the Box First treatment, due to the primacy effect ($0 < \lambda^{\text{Color First}} < \lambda^{\text{Box First}} < 1$). Therefore, the Color First treatment presents more space for motivated reasoning. Our first hypothesis proposes the order effect on motivated reasoning.

Hypothesis 2. *The Color First treatment exhibits a higher degree of motivated reasoning compared to the Box First treatment.*

Furthermore, as shown in Figure 2b, the degree of motivated reasoning is larger when the log odds of prior beliefs are in the middle range rather than at the two extremes. When the prior is extremely low or high, our benchmark analysis suggests less motivated reasoning. This leads to our second hypothesis.

Hypothesis 3. *The order effect on motivated reasoning is more pronounced when prior beliefs are not extreme.*

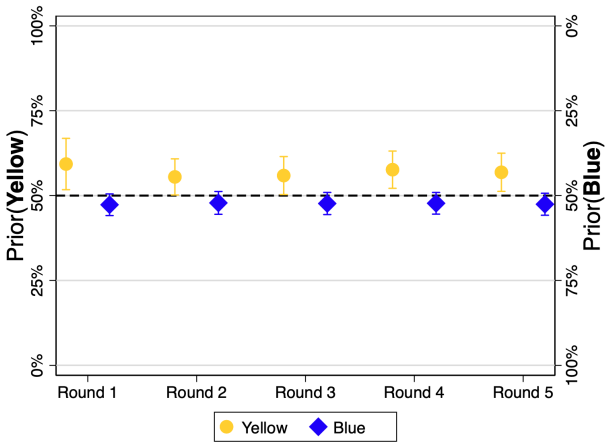
In the Variable Reward experiment, out of 200 participants, 30 (15%) chose yellow as their preferred color, while 170 (85%) chose blue.

4 Results

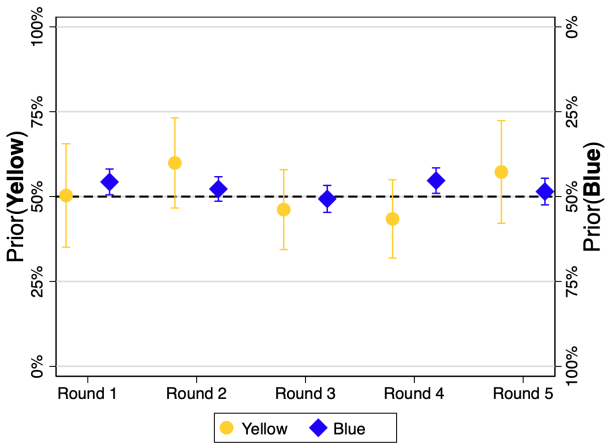
Figure 3a displays the reported prior beliefs for each round in the Variable Reward experiment. Among the 200 participants, 30 (15%) chose yellow as their preferred color, while 170 (85%) chose blue. We find evidence that the color of the preferred ball affects the priors. When the preferred ball is yellow, the belief that the Target Ball is yellow averages 57% across all five rounds. Conversely, when the preferred ball is blue, the prior belief averages 46.7% (p-value < 0.001).[4] This difference indicates that participants are more likely to believe in scenarios favorable to them.

On the contrary, in the Fixed Reward experiment, priors are not significantly different based on the preferred color (Figure 3b). Out of 99 participants, 12 (12.1%) chose yellow as their preferred color, while 87 (87.8%) chose blue. The prior beliefs of each group are 52.4% and 51.4%, respectively, which is not significantly different (p-value = 0.3465). This suggests that the difference in prior beliefs is due to the monetary incentive, not the inherent color preference.

[4]The standard errors are clustered at the individual level, and the result holds true after controlling for round fixed effects.



(a) Prior beliefs in the Variable Reward experiment

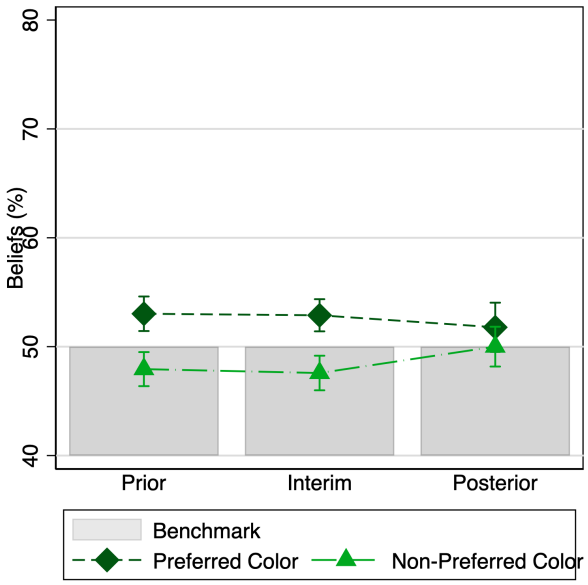


(b) Prior beliefs in the Fixed Reward experiment

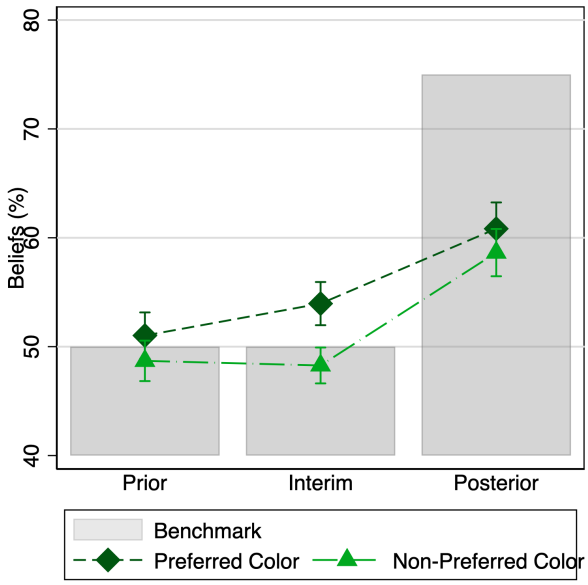
Figure 3: Prior beliefs of each round

Figure 4 illustrates the changes of participants' average prior beliefs, interim beliefs, and posterior beliefs in the Variable Reward experiment. To simplify our analysis, we align beliefs with the signals participants receive. Specifically, the belief represented in our figures corresponds to the participant's belief that the color of the Target Ball matches the color of the Clue Ball they observe. For example, if the signal indicates the Clue Ball is yellow, the belief for analysis is that the participant believes the Target Ball is also yellow.

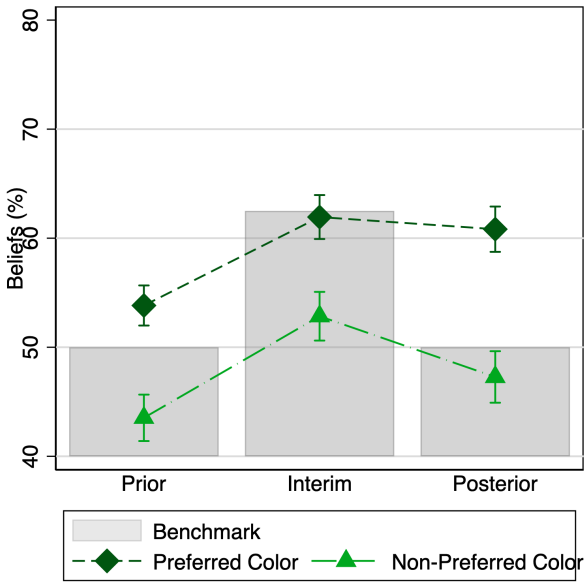
This approach has two advantages. First, regardless of whether the signal is yellow or blue, participants adjust their beliefs in a single direction. Therefore, in the analysis, we need only



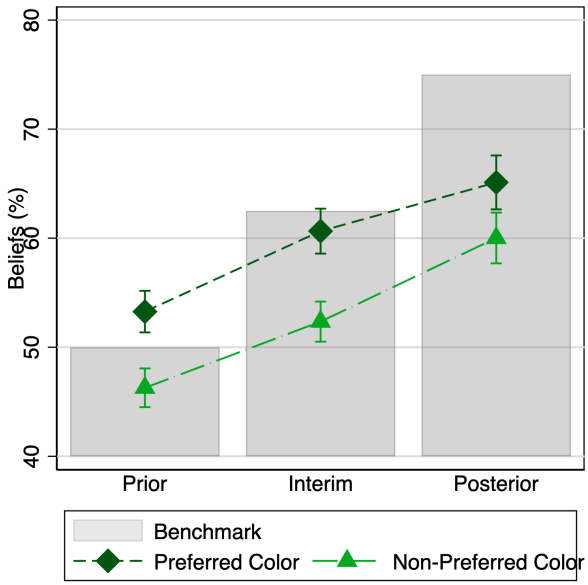
(a) Box First, the Clue Ball from Box A



(b) Box First, the Clue Ball from Box B



(c) Color First, the Clue Ball from Box A



(d) Color First, the Clue Ball from Box B

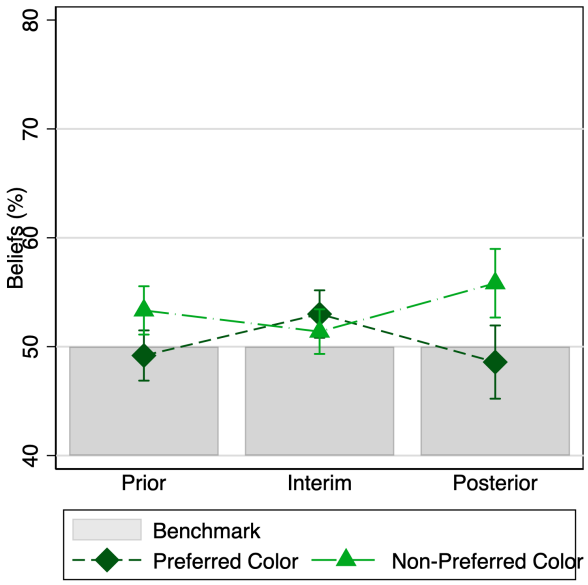
Figure 4: Patterns of belief updating in the Variable Reward experiment

consider updates in one direction. Secondly, it allows us to examine the influence of preferred color. Motivated reasoning suggests that individuals tend to adjust their beliefs towards their desired outcomes. When the signal — the color of the Clue Ball — matches the participant's

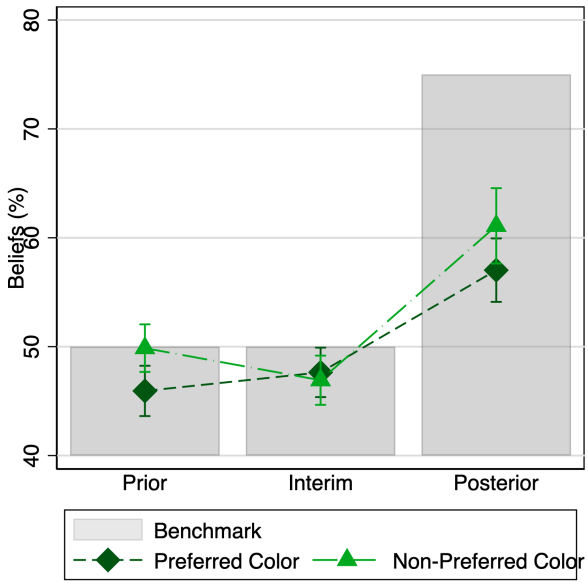
preferred color, she is more inclined to update her belief in line with the signal than when it does not.

The gray bars represent Bayesian benchmarks. At the prior stage, a Bayesian believes there is an equal chance of the Target Ball being either yellow or blue. In the Box First treatment, interim beliefs are expected to remain the same as the prior since the color of the Clue Ball has not yet been revealed. Once it is revealed, at the posterior stage, the Bayesian updates her belief to align with the color of the Clue Ball only when it is drawn from Box B, as described in Figure 4a and 4b). In the Color First treatment, at the interim stage, only the color of the Clue Ball is known. Therefore, the Bayesian updates her belief in accordance with the signal. However, considering the possibility that the Clue Ball might come from Box A, the update is not as confident as it would be with complete information. At the posterior stage, having complete information, the belief update are the same as the one in the Box First treatment (Figure 4c and 4d).

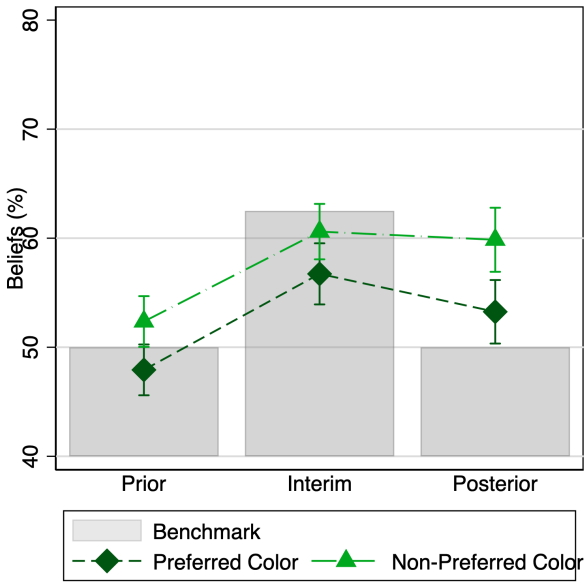
Results in Figure 4 indicate that participants generally update their beliefs in line with the Bayesian benchmark. Our primary interest, motivated reasoning, suggests that belief updating is affected by whether the signal matches the preferred color. Figure 4c provides evidence of motivated reasoning. When the Clue Ball comes from Box A, meaning the signal is uninformative, participants retract their beliefs when the signal is a non-preferred color but do not retract when it is a preferred color. Interestingly, this pattern of motivated reasoning was not observed in the Box First treatment, implying that the order of information influences motivated reasoning. Conversely, Figure 5 demonstrates that this order effect on motivated reasoning does not exist in the Fixed Reward experiment, indicating that motivated reasoning is driven by monetary incentives. Further analysis of these results will be discussed in the next section.



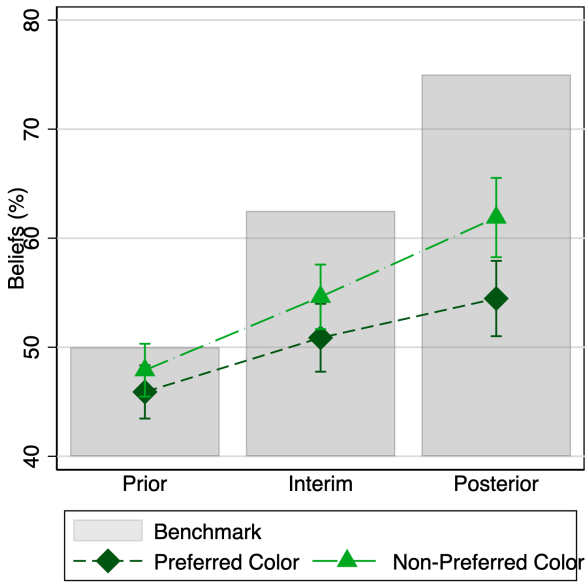
(a) Box First, the Clue Ball from Box A



(b) Box First, the Clue Ball from Box B



(c) Color First, the Clue Ball from Box A



(d) Color First, the Clue Ball from Box B

Figure 5: Patterns of belief updating in the Fixed Reward experiment

4.1 Order effect on motivated reasoning

Next, we formally investigate the order effect on the degree of motivated reasoning using the following regression model:

$$LogOddsPosterior_i = \alpha_0 + \alpha_1 LogOddsPrior_i + \alpha_2 PreferredColor_i + \alpha_3 ColorFirst_i + \alpha_4 PreferredColor_i \times ColorFirst_i + Controls_i + \epsilon_i.$$

As discussed in the previous section, we construct belief variables to align with the color of the Clue Ball. For participants observing a yellow Clue Ball, the odds of beliefs represent the ratio of the posterior probability that the Target Ball is yellow to the posterior probability that it is blue. For those observing a blue Clue Ball, the ratio reflects the posterior probability of the Target Ball being blue relative to it being yellow. This ensures that $LogOddsPosterior_i$ moves in the same direction controlling for $LogOddsPrior_i$, regardless of the color of the Clue Ball.

$PreferredColor_i$ indicates whether the color of the Clue Ball matches the participant i's preferred color, and $ColorFirst_i$ denotes whether participant i is in the Color First treatment. Therefore, the interaction term between $PreferredColor_i$ and $ColorFirst_i$ therefore captures the order effect on the degree of motivated reasoning. If motivated reasoning is stronger in the Color First treatment than in the Box First treatment, the coefficient α_4 would be positive. $Control$ include round dummies and the indicator of the chosen color.

Table 3 presents the regression estimates in the Variable Reward experiment. Columns (1) and (2) show that prior beliefs positively affect posterior beliefs, and participants update less when they receive an uninformative signal, i.e., the Clue Ball is from Box A. These regression results indicate that participants understand the experimental design well, and there is no evidence of motivated reasoning.

	Dependent Variable: Log Odds Posterior			
	(1)	(2)	(3)	(4)
Log Odds Prior	0.534***	0.531***	0.501***	0.645
	(0.076)	(0.076)	(0.128)	(0.647)
Preferred Color	0.203	-0.055	-0.070	-0.173
	(0.158)	(0.237)	(0.146)	(0.172)
Color First		0.165	0.210	0.233
		(0.223)	(0.158)	(0.172)
Preferred Color × Color First		0.483	0.417**	0.700***
		(0.327)	(0.205)	(0.241)
Uninformative Box	-0.565***	-0.566***	-0.624***	-0.575***
	(0.156)	(0.155)	(0.133)	(0.146)
Observations	1000	1000	806	611
Num. of Individuals	200	200	182	157

Table 3: The order effect and motivated reasoning (Variable Reward experiment)

According to our framework, Hypothesis 3 suggests less motivated reasoning occurs with extreme prior beliefs. Therefore, we examine the regression analysis with participants whose priors are within the 10-90 percentiles. The results, presented in Column (3), reveal a significant order effect on motivated reasoning. When participants are in the Color First treatment, they are more likely to update their beliefs in alignment with their preferred color signal. Column (4) shows the results of the same regression for participants whose priors are within the 20-80 percentiles, where the order effect on motivated reasoning becomes even more significant. Interestingly, participants with priors in the 20-80 percentiles are less likely to be influenced by their priors, while the order effect becomes more pronounced.

Table 4 presents the regression estimates for the Fixed Reward experiment. The results suggest that there is no evidence of motivated reasoning in the Fixed Reward experiment, regardless of whether the priors are extreme or not. The differences between Table 3 and Table 4 further reinforce Hypothesis 1, which posits that motivated reasoning is driven by monetary incentives.

	Dependent Variable: Log Odds Posterior			
	(1)	(2)	(3)	(4)
Log Odds Prior	0.348**	0.347**	0.914***	0.901*
	(0.136)	(0.135)	(0.295)	(0.462)
Preferred Color	-0.361	-0.209	0.024	0.076
	(0.225)	(0.307)	(0.218)	(0.245)
Color First		0.105	0.257	0.283
		(0.453)	(0.428)	(0.495)
Preferred Color × Color First		-0.315	-0.279	-0.436
		(0.434)	(0.386)	(0.454)
Uninformative Box	-0.404*	-0.399*	-0.316	-0.494*
	(0.205)	(0.216)	(0.227)	(0.251)
Observations	495	495	410	306
Num. of Individuals	99	99	96	80

Table 4: The order effect and motivated reasoning (Fixed Reward experiment)

4.2 Informativeness and the order effect

Figure 4 suggests a greater order effect when the Clue Ball originates from Box A, i.e., the uninformative source. In this section, we explore these heterogeneous effects based on the informativeness of the signal.

We divide our sample into two groups: those whose Clue Ball is drawn from Box A, and those whose Clue Ball is drawn from Box B. We then conduct identical regressions identical to those in Table 3 from the previous section. Table 5 presents the results of the Variable Reward experiment. Specifically, Table 5a shows the regression results for participants observing the Clue Ball drawn from Box A, indicating that the order effect on motivated reasoning exists, even in Column (2).

Conversely, Table 5b shows no evidence of the order effect when participants receive an informative signal. These results suggest that the order effect documented in Section 4.1 is primarily driven by cases where Box A is the source of the Clue Ball, i.e., the signal in uninformative.

Table 6 presents the results of the Fixed Reward experiment. Since there is no evidence of motivated reasoning when the color choice is not incentivized, no order effect is observed

	Dependent Variable: Log Odds Posterior			
	(1)	(2)	(3)	(4)
Log Odds Prior	0.529***	0.524***	1.463	0.745***
	(0.098)	(0.096)	(0.960)	(0.147)
Preferred Color	0.342*	-0.146	-0.347	-0.239
	(0.181)	(0.258)	(0.309)	(0.252)
Color First		-0.081	0.151	0.085
		(0.271)	(0.155)	(0.156)
Preferred Color × Color First		0.936**	0.952**	0.633*
		(0.404)	(0.399)	(0.334)
Observations	497	497	304	406
Num. of Individuals	188	188	128	167

(a) When the signal is uninformative (Box A)

	Dependent Variable: Log Odds Posterior			
	(1)	(2)	(3)	(4)
Log Odds Prior	0.537***	0.538***	-21.687**	0.273
	(0.101)	(0.101)	(10.862)	(0.225)
Preferred Color	0.055	0.049	0.055	0.127
	(0.242)	(0.369)	(0.186)	(0.189)
Color First		0.410	0.293	0.332
		(0.348)	(0.277)	(0.252)
Preferred Color × Color First		0.005	0.529	0.185
		(0.498)	(0.338)	(0.276)
Observations	503	503	257	400
Num. of Individuals	194	194	116	166

(b) When the signal is informative (Box B)

Table 5: Informativeness and the order effect (Variable Reward experiment)

regardless of whether the signal is informative or not.

	Dependent Variable: Log Odds Posterior			
	(1)	(2)	(3)	(4)
Log Odds Prior	0.258	0.260	0.776**	0.252
	(0.213)	(0.213)	(0.351)	(0.748)
Preferred Color	-0.505*	-0.379	0.368	0.618
	(0.296)	(0.460)	(0.384)	(0.472)
Color First		0.328	0.537	0.343
		(0.536)	(0.572)	(0.627)
Preferred Color × Color First		-0.224	-0.909*	-0.971
		(0.543)	(0.528)	(0.629)
Observations	242	242	205	153
Num. of Individuals	95	95	85	68

(a) When the signal is uninformative (Box A)

	Dependent Variable: Log Odds Posterior			
	(1)	(2)	(3)	(4)
Log Odds Prior	0.399***	0.397***	1.021**	1.522**
	(0.129)	(0.128)	(0.474)	(0.619)
Preferred Color	-0.159	-0.104	-0.326	-0.412
	(0.342)	(0.513)	(0.323)	(0.342)
Color First		-0.277	-0.164	0.040
		(0.535)	(0.453)	(0.510)
Preferred Color × Color First		-0.148	0.452	0.159
		(0.716)	(0.561)	(0.559)
Observations	253	253	205	153
Num. of Individuals	96	96	88	73

(b) When the signal is informative (Box B)

Table 6: Informativeness and the order effect (Fixed Reward experiment)

5 Conclusion

This paper examines how the order of information affects the degree of motivated reasoning. Our experiment, employing a variant of the classic balls-and-boxes setup, demonstrates that participants exhibit stronger motivated reasoning when they encounter information before learning about its informativeness. Specifically, receiving the signal first leads to a higher degree of motivated reasoning compared to learning about the signal's informativeness first. This finding highlights the impact of the sequence of information on motivated reasoning.

Our results suggest that exposure to information without knowing its informativeness may encourage individuals to interpret it in a way that aligns with their preexisting beliefs and preferences. This has broader implications for potential interventions to curb motivated reasoning, as simply altering the sequence of information could help mitigate its effects.

These insights pave the way for future research in several ways. Firstly, our findings can be extended to other types of motivated reasoning. For example, politicized motivated reasoning can occur even without monetary incentives (Kunda, 1990; Westen et al., 2006; Kim et al., 2010; Kahan, 2016; Thaler, 2023). In such cases, it would be valuable to investigate whether the order of information also influences the bias. Secondly, this research has practical implications in real-world settings for reducing cognitive biases. By strategically controlling the sequence in which information is presented, it may be possible to mitigate these biases, leading to better decision-making processes.

References

Ahlering, R. F. and Parker, L. (1989). Need for cognition as a moderator of the primacy effect. *Journal of Research in Personality*, 23:313–317.

Beattie, P. and Snider, D. (2019). Knowledge in international relations: Susceptibilities to motivated reasoning among experts and non-experts. *Journal of Social and Political Psychology*.

Benjamin, D. J. (2019). Chapter 2 - Errors in probabilistic reasoning and judgment biases. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics - Foundations and Applications 2*, volume 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, pages 69–186. North-Holland.

Bolsen, T. and Palm, R. (2019). Motivated reasoning and political decision making.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915.

Carpenter, C. J. (2019). Cognitive dissonance, ego-involvement, and motivated reasoning. *Annals of the International Communication Association*, 43:1 – 23.

Cotter, R., Lodge, M., and Vidigal, R. (2020). The boundary conditions of motivated reasoning.

Danz, D., Vesterlund, L., and Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, 112(9):2851–2883.

Ecker, U., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L., Brashier, N., Kendeou, P., Vraga, E., and Amazeen, M. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1:13–29.

Epley, N. and Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(2):133–140.

Gonçalves, D., Libgober, J., and Willis, J. (2021). Learning versus unlearning: An experiment on retractions. *Working Paper*.

Hogarth, R. M. and Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24(1):1–55.

Hossain, T. and Okui, R. (2013). The binarized scoring rule. *The Review of Economic Studies*, 80:984–1001.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., and Ward, L. M. (1968). Pattern of performance and ability attribution: an unexpected primacy effect. *Journal of Personality and Social Psychology*, 10(4):317–340.

Kahan, D. (2016). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. pages 1–16.

Kim, S., Taber, C., and Lodge, M. (2010). A computational model of the citizen as motivated reasoner: Modeling the dynamics of the 2000 presidential election. *Political Behavior*, 32:1–28.

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, 71(5):1347–1366.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108 3:480–98.

Redlawsk, D. P. (2002). Hot cognition or cool consideration? testing the effects of motivated reasoning on political decision making. *The Journal of Politics*, 64(4):1021–1044.

Saccardo, S. and Serra-Garcia, M. (2023). Enabling or limiting cognitive flexibility? evidence of demand for moral commitment. *American Economic Review*, 113:396–429.

Taber, C. S., Cann, D., and Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, 31(2):137–155.

Thaler, M. (2021). Gender differences in motivated reasoning. *Journal of Economic Behavior and Organization*, 191(C):501–518.

Thaler, M. (2023). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *American Economic Journal: Microeconomics, forthcoming*.

van Erkel, P. F. and Thijssen, P. (2016). The first one wins: Distilling the primacy effect. *Electoral Studies*, 44:245–254.

Westen, D., Blagov, P., Harenski, K., Kilts, C., and Hamann, S. (2006). Neural bases of motivated reasoning: An fmri study of emotional constraints on partisan political judgment in the 2004 u.s. presidential election. *Journal of Cognitive Neuroscience*, 18:1947–1958.

Wittrock, L., Strobel, M., and Tsakas, E. (2023). Belief updating with misinformation. *Working Paper*.

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, 110(2):337–361.

A Instructions and Screenshots

Figures 6 through 14 present the computer screen participants see in the experiment. Figure 6 is displayed only once at the start of the experiment. Upon entering Part 1, participants are asked to read the instructions carefully (Figures 7 - 9) and answer the comprehension questions. If participants fail to answer all the questions correctly, they will be sent back to the first page of the instructions and asked to review them. They cannot proceed until they answer all the questions correctly. After completing Part 1, participants proceed to Part 2 (Figure 14). The instructions (Figures 8 - 9) are displayed again at the beginning of Part 2.

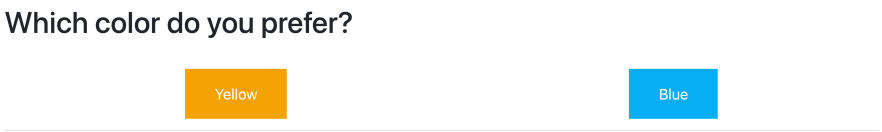


Figure 6: Screenshot before starting the experiment

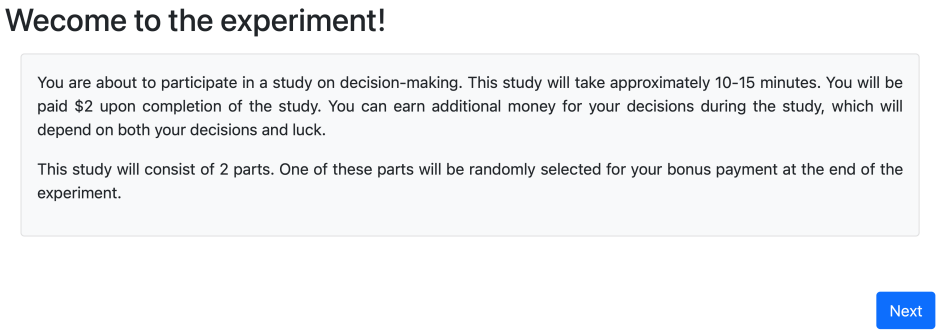
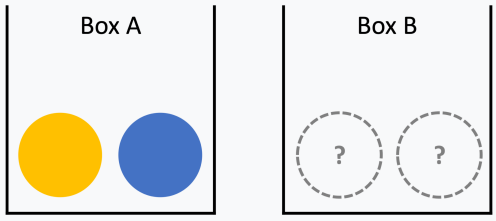


Figure 7: Instructions (1)

Part 1 Instructions (1/2)

You are presented with Box A and Box B. The computer fills the two boxes with two balls, either Yellow or Blue.



Box A always contains one Yellow ball and one Blue ball.

Box B contains two balls; however, the composition is unknown. The two balls in *Box B* could be any of the following: two Yellow balls; two Blue balls; or one Yellow ball and one Blue ball.

The computer will draw a ball from **Box B**, which we call the **Target Ball**.
Your task is to predict the color of the Target Ball. You will make a total of three predictions.

Next

Figure 8: Instructions (2)

Part 1 Instructions (2/2)

The experiment will proceed as follows:

1. **First Guess.**

   The computer randomly draws a ball from *Box B*. The ball that is drawn is referred to as the *Target Ball*. You will provide your first prediction of the color of the *Target Ball*. The results of your prediction will not yet be displayed.

   After drawing the *Target Ball*, the computer returns it to *Box B*. The contents of *Box B* therefore remains the same.

2. **Computer Generates Clue.**

   After you complete your first guess, the computer will generate a clue for you. This clue is created by the computer randomly selecting either Box A or Box B. Both boxes have equal probability in being selected. Then, the computer randomly draws a ball from the selected box. Each ball in the selected box has equal probability of being picked.

   The ball that is selected is called the *Clue Ball*. After drawing the *Clue Ball*, the computer returns it to the selected box. The contents of the selected box therefore remains the same.

3. **Second Guess.**

   You will see the color of the *Clue Ball*.

   Then, you will provide your second prediction about the color of the *Target Ball*. The results of your prediction will not yet be displayed.

4. **Third Guess.**

   Lastly, you will see which box the *Clue Ball* was drawn from.

   Then, you will provide your third prediction about the color of the *Target Ball*. The results of your prediction will not yet be displayed.

5. **Payoff.**

   Based on the accuracy of your three predictions, you will have a chance to receive a prize. The prize amount depends on the color of the *Target ball* and the color you chose in the beginning of this experiment.

   If the color of the *Target ball* is Yellow, which *matches* your color choice, the prize is $1.5.
   If the color of the *Target ball* is Blue, which *does not match* your color choice, the prize is $0.5.

   The payment rule is designed so that you can secure the largest chance of winning the prize by reporting your most accurate guess. The precise payment rule details are available by request at the end of the experiment.

   At the end of the experiment, one of the three guesses will be randomly chosen for payment. Every guess has the same chance of being selected for payment.

On the next screen, you are presented with comprehension questions. If any of your submitted answers are incorrect, you cannot proceed. Please review the instructions. Once you get familiar with the instructions, click “Confirm”.

Back Confirm

Figure 9: Instructions (3)

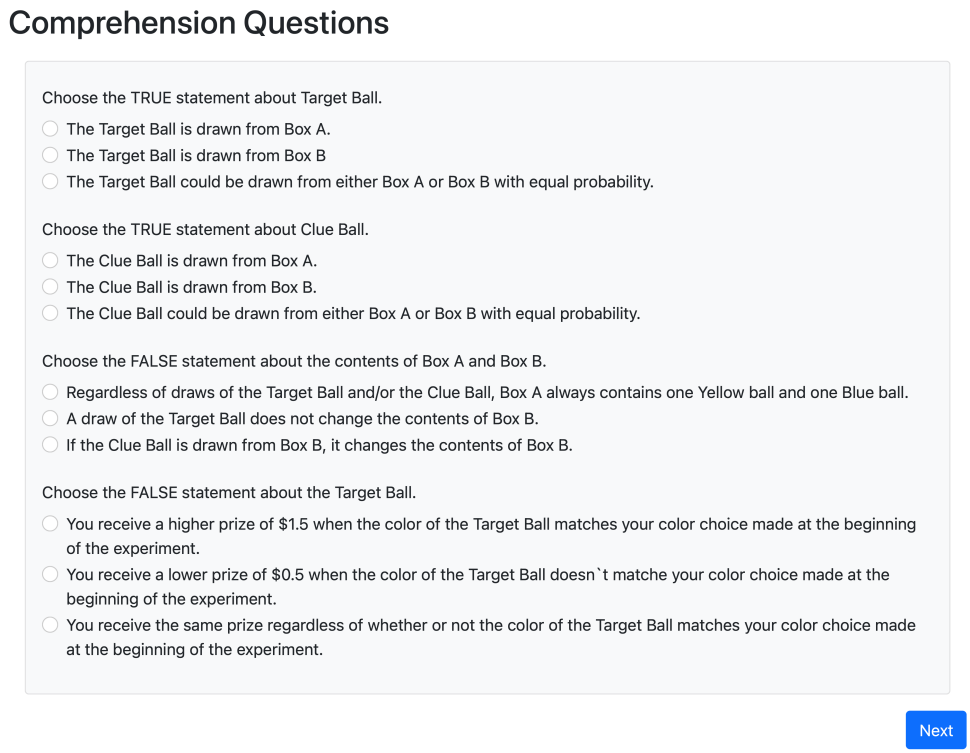


Figure 10: Comprehension questions

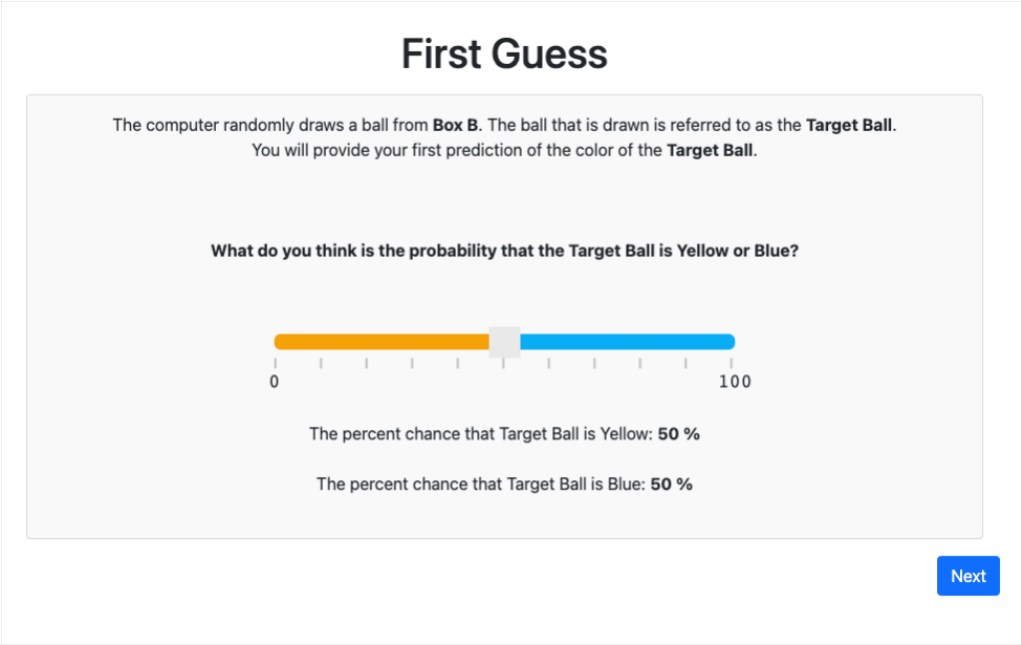


Figure 11: Screenshot for prior beliefs

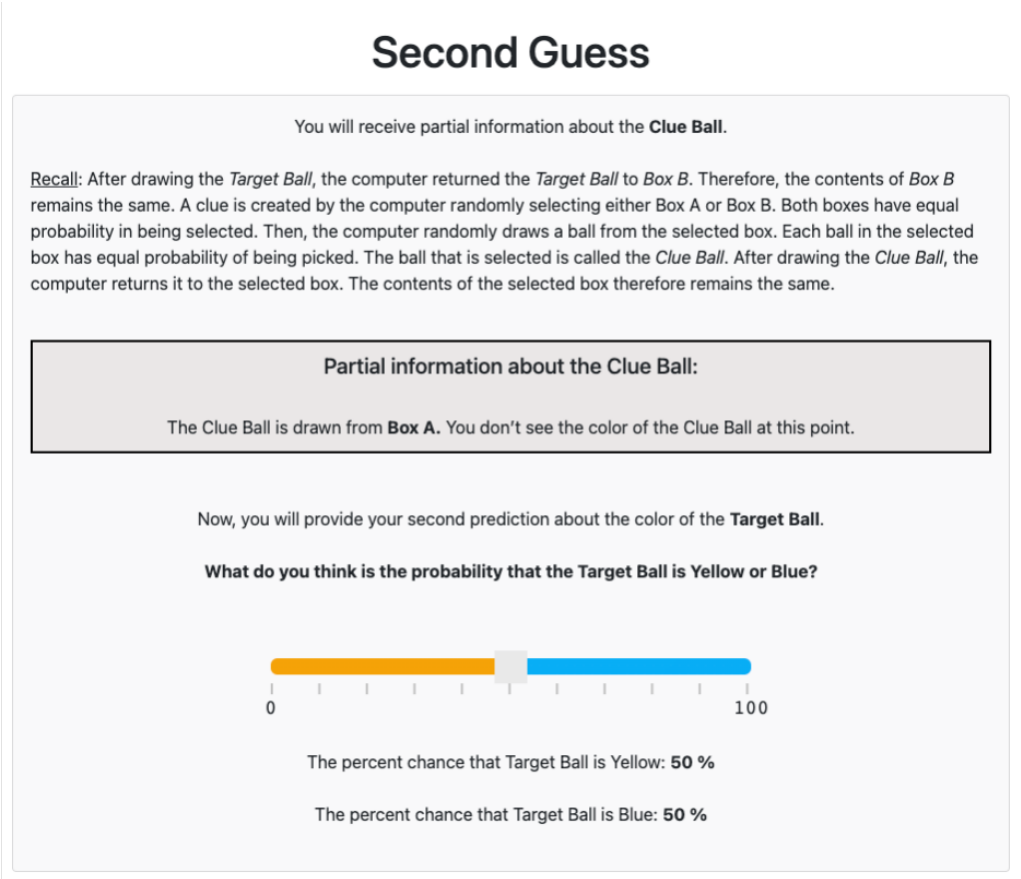


Figure 12: Screenshot for interim beliefs

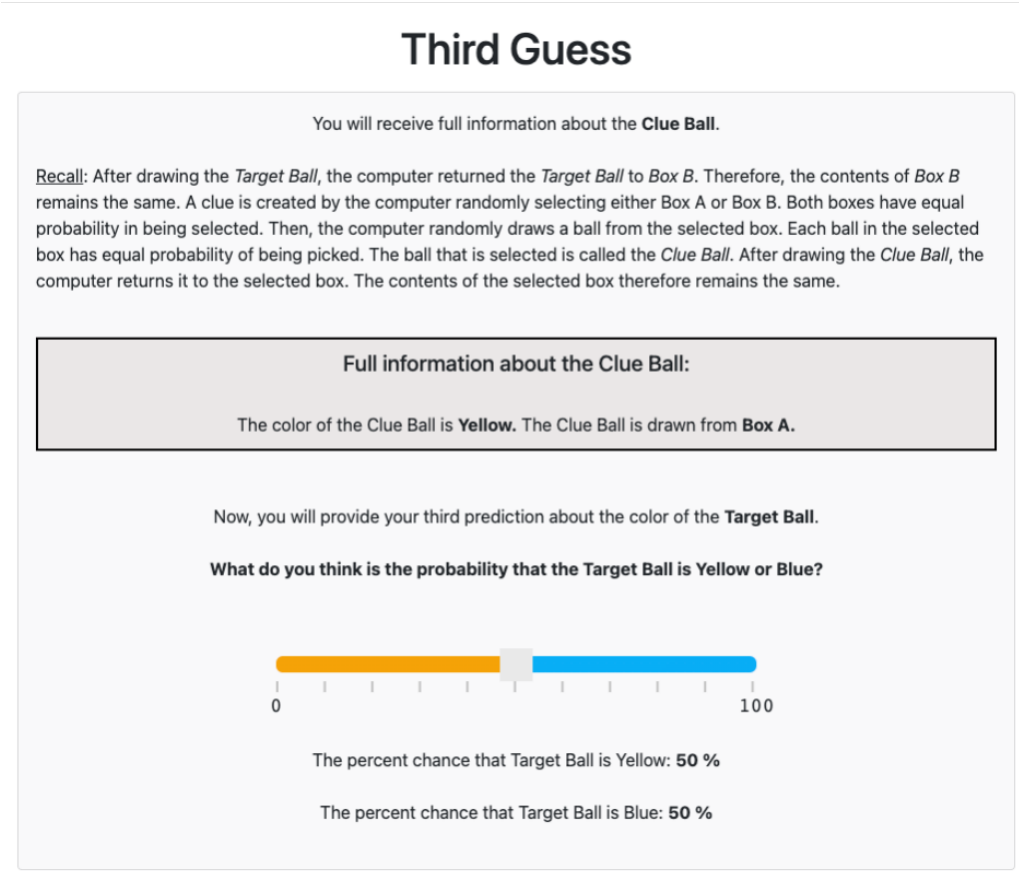


Figure 13: Screenshot for posterior beliefs

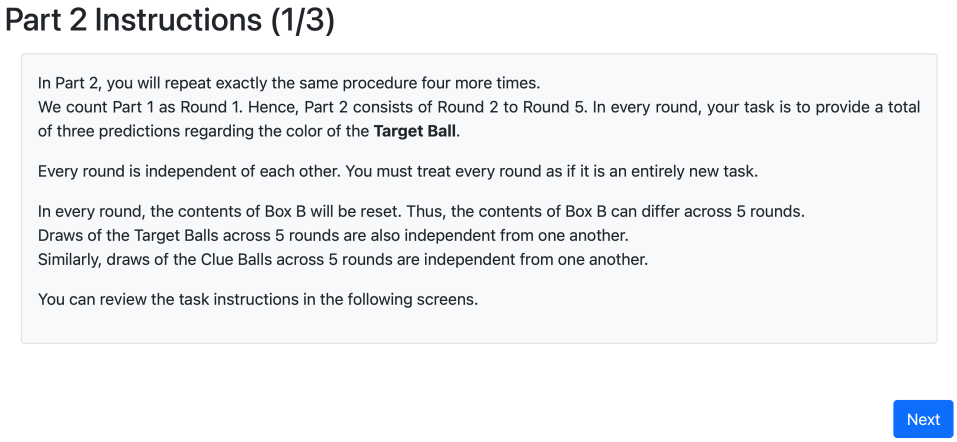


Figure 14: Instructions for Part 2