The Supply of Motivated Beliefs*

Michael Thaler†

December 14, 2021

Click here for most recent version

Abstract

When people choose what messages to send to others, they often consider how others will interpret the messages. In many environments, particularly in politics, people are motivated to hold particular beliefs and distort how they process information in directions that favor their motivated beliefs. This paper uses two experiments to study how message senders are affected by receivers' motivated beliefs. Experiment 1, conducted using an online sample of social media users, analyzes the effect of incentivizing senders to be perceived as truthful. These incentives cause senders to send less truthful messages. When incentivized, senders send more false information when it aligns with receivers' politically-motivated beliefs, controlling for receivers' current beliefs. However, receivers do not anticipate the adverse effects of senders' incentives. Experiment 2 further isolates the role that information processing plays by analyzing an environment in which receivers assess the truthfulness of messages from a computer and senders choose one of the computer's messages to determine their earnings. Senders predict that receivers distort information processing in the direction of their politics, demand information about receivers' political preferences, and condition on the receivers' politics to strategically choose less truthful computer messages.

JEL classification: C91; D83; D91

*Thank you to Roland Bénabou, Christine Exley, David Laibson, Matthew Rabin, Mattie Toma, Leeat Yariv, Sevgi Yuksel, and seminar participants at Princeton, Michigan, Nottingham, SWEET, the Belief-Based Utility Workshop, and the ESA Job Market Candidates Seminar Series for helpful comments. I am grateful for funding support from the Pershing Square Fund for the Foundations of Human Behavior. Preregistration plans for the two experiments are available on AsPredicted: (Experiment 1: https://aspredicted.org/blind.php?x=VMH_8HW; Experiment 2: https://aspredicted.org/blind.php?x=GAQ_ZJF.) Both experiments were approved by the IRB at Princeton University (13114-04 and 13114-05).

†Princeton University. Email: michael.thaler@princeton.edu.

1 Introduction

There has been a proliferation of interest in understanding how people communicate with others and why much of the news marketplace contains inaccurate information. "Fake news," defined by Allcott and Gentzkow (2017) as the intentional reporting of false information, has been shown to be a contributing factor to the undermining of trust in public health, historical misconceptions, and the state of democracy (Lazer et al. 2018; Oliver and Wood 2014; Pennycook and Rand 2021). Given these large societal costs, a better understanding of what motivates people to send false information to others can inform efforts to improve news dissemination, increase trust in credible sources, and decrease polarization.

This paper studies the determinants of false information in communication environments in which receivers of information are motivated to hold certain beliefs. While there is ample evidence that motivated beliefs affect receivers' demand for information (e.g. Oster, Shoulson, and Dorsey 2013; Peterson and Iyengar 2021), much less is known about their effects on the *supply* of information. As such, this paper analyzes the effect that receivers' motivated beliefs have on senders, highlighting the role of two factors that increase disinformation: (1) beliefs that receivers' motivated beliefs are misaligned with the truth, and (2) incentives to be perceived as truthful. Both factors are prevalent in many news transmission environments, and can play a significant role on social media, for two reasons. First, news often evokes motivated beliefs: for instance, many consumers disagree about political issues and may prefer to hold beliefs that are even farther in the direction that aligns with their political party. These desired beliefs lead them to be motivated to believe that certain pieces of news are true while others are false. Second, many social media platforms have users rate other users, incentivizing news senders to be rated well by others. For instance, Facebook implemented a policy in 2015 that enabled users to report news as false, and by 2018 began assigning users a credibility score based on the news they shared (Dwoskin 2018). This system can backfire when users send messages to appeal to receivers' motivated beliefs.

In order to have a controlled environment that is able to cleanly identify the effects of motivated beliefs and incentives, I turn to the (online) laboratory and run two large preregistered experiments. In Experiment 1, conducted among social media users in the United States, subjects send and receive messages about various factual issues. These issues are chosen to evoke politically-motivated beliefs, as described in Table 1. On each question, receivers report their prior belief about which of two answers is correct. Senders learn the true answer to the question and then choose, as a function of the receiver's prior, whether to send a message to the receiver that corresponds to the true answer or to the false answer. Finally, receivers assess the probability that the message from their sender is truthful. To

Topic	Pro-Democrat Motives	Pro-Republican Motives
Immigrants' crime rate	Lower than US citizens	Higher than US citizens
Racial discrimination	Severe in labor market	Not severe in labor market
US crime	Got worse under Trump	Got better under Trump
Media bias	Media not mostly Dems	Media mostly Dems
COVID-19 restrictions	Mostly stopped spread	Did not mostly stop spread
Gun reform	Decreased homicides	Did not decrease homicides
Unemployment	Got worse under Trump	Got better under Trump
Wages	Grew slower under Trump	Grew faster under Trump
Undocumented immigrants	Mostly overstaying visas	Mostly illegally entered US
Domestic terrorism	Mostly due to white supremacy	Mostly due to other factors
Poverty rates	Got worse under Trump	Got better under Trump
Illegal immigration	Not historically high	Historically high

Table 1: The list of political topics and hypothesized motives in the experiments.

estimate the role that incentives play, senders are randomly assigned to either be paid as a function of receivers' assessments or to have their pay not depend on receivers' assessments. To estimate the role that motivated beliefs play, senders are randomly assigned to receivers whose political party is either aligned with the truthful message, whose party is aligned with the false message, or whose party is unknown.

The main result from Experiment 1 is that senders are more likely to send false messages when incentivized and when the receiver's party is aligned with the false message. When faced with receivers with the same prior belief, incentivized senders are 7.1 percentage points (subject-level clustered s.e. 3.2 pp) more likely to send a false message when the receiver's party is aligned with the false message than when it is aligned with the true message. Senders are also more likely to send false messages when the receiver's party is aligned with the false message than when the receiver's party is unknown or when the topic is neutral instead of political. Meanwhile, unincentivized senders are not directionally affected by this treatment. Since prior beliefs are conditioned on, results seem to be due to senders appealing to receivers' party and not just receivers' priors.

Incentives have a negative effect on veracity overall. Senders are 7.3 pp (s.e. 2.5 pp) more likely to send false messages when incentivized, an increase of 34 percent. The treatment especially decreases truthfulness when the receiver's party is misaligned with the truth. There is no evidence that receivers incorporate the negative effects of incentives when rating

the veracity of senders. Neither receivers' assessments nor beliefs are statistically significantly affected by senders' incentives, and the point estimates are close to zero. Survey evidence confirms these patterns. Senders self-report choosing messages that are more aligned with the receiver's party, and less aligned with the truth, when they are incentivized, but receivers' predictions of sender behavior are not significantly affected by senders' incentives.

Experiment 2 further studies the role of motivated reasoning, in which people distort how they form posteriors in directions that favor beliefs they find more attractive, such as politics (Kunda 1990; Bénabou and Tirole 2002; Kahan 2016; Thaler 2021a). The results from Experiment 1 indicate that senders believe that receivers act in a non-Bayesian fashion to favor their motivated beliefs, but the specific structure of beliefs about motivated reasoning is hard to identify. For instance, in Experiment 1 senders may have other-regarding preferences, caring about the accuracy of receivers' assessments. This would lead senders to send news they think receivers would be better at assessing, which may be pro-party news. In addition, senders may believe that receivers expect senders to distort messages in the direction opposed to receivers' party, leading them to trust pro-party messages more (such as in Morris 2001). To identify the form of the bias that senders believe receivers have in inference, in Experiment 2 senders do not impact receivers' messages or payoffs in any way.

Instead, I identify motivated reasoning among receivers in Experiment 2 using a version of the design developed by Thaler (2021a). This design has two main steps. First, each receiver is given a variety of factual questions with numerical answers. On each question, the receiver first selects a response that they think is equally likely to be above or below the correct answer; that is, the median of their belief distribution is elicited. Second, the receiver is given two binary messages from the computer: one message is true and and the other is false. The message tells them whether the answer was above or below their median. If the message is true, it is always accurate. If the message is false, it is always inaccurate. The receiver is not told which source the message came from; instead, they are asked to make inferences about the source's veracity from the message content. Since messages relate the true answer to the receiver's median, a Bayesian would believe that it is equally likely for each source to report either message.[1] On the other hand, a receiver who engages in politically-motivated reasoning will think the news is more likely to be true if it sends a message that aligns more with the beliefs of their political party. In Thaler (2021a), I argue that this method is a well-identified and well-powered way to identify motivated reasoning.[2]

[1]That is, the receiver has stated that they believe the answer is equally likely to be greater than or less than her median; so, they believe the likelihood that a true message would report that the answer is greater than their median is 1/2, and the likelihood that a false message would report that the answer is less than their median is also equal to 1/2, leading a "greater than" or "less than" message to be completely uninformative about the veracity of the news source.

[2]While there are other experimental approaches to identifying motivated reasoning, such as those in

After receivers play the game above, each sender is given a matched receiver's median belief and are asked to select either the "greater than" or the "less than" message of the computer. The two primary treatment arms are the same as those in Experiment 1. First, senders are either incentivized to have the receiver's rating of the message be implemented for payment, or they are unincentivized. Second, senders are matched to receivers whose party is either aligned or misaligned with the truth. In particular, the median beliefs provided to the sender are selected such that at least one Democrat and at least one Republican has stated that median. This enables random assignment, for a given median belief, to a receiver who is a Democrat, is a Republican, or is equally likely to be of either party.

The main finding from Experiment 2 replicates the main finding from Experiment 1; senders who are incentivized to be perceived as truthful are more likely to choose the false message when it is aligned with the receiver's party, while unincentivized senders are not. Incentivized senders are 14.5 percentage points (subject-level clustered s.e. 4.5 pp) more likely to send a false message when the receiver's party is aligned with the false message than when it is aligned with the true message. Unincentivized senders are not directionally affected by this treatment. Incentives again lead to more false messages chosen. Senders are 8.5 pp (s.e. 2.3 pp) more likely to send false messages overall when incentivized, which is again a 34 percent increase; this effect is magnified by the condition in which the receiver's party is misaligned with the truth. This result clarifies the form of politically-motivated reasoning that determines senders' beliefs and choices of false messages.

The second finding from Experiment 2 is that senders demand information about receivers. A majority of senders are willing to pay a positive amount to learn the political party of receivers on political questions, and a majority are not willing to on neutral questions. Senders use receivers' party information in order to choose more false news, suggesting that giving news providers the option to learn about their audience may cause the audience to receive less truthful pieces of news when motivated reasoning is at play.

In each experiment, I elicit beliefs about others' assessments, finding clear evidence that people expect others to use their politically-motivated beliefs, and not just current beliefs, when inferring the truthfulness of the news they receive in the experiments. Senders in Experiment 1 expect receivers with 50-50 priors to significantly differ in their ratings of pro-party and anti-party news, and senders in Experiment 2 expect receivers in the motivated reasoning task to significantly differ in their ratings of pro-party and anti-party news. Senders' beliefs, while directionally accurate, tend to overstate the impact of party on inference.

Mobius et al. (2014) and Eil and Rao (2011), those approaches have often found it difficult to cleanly detect motivated reasoning, as identification of the bias from Bayesian updating and other inference biases is noisy (Benjamin 2019; Tappin, Pennycook, and Rand 2020a; Tappin, Pennycook, and Rand 2020b).

Theoretically, the main experimental results can be explained by a simple sender-receiver model in which the sender prefers to tell the truth and has incentives for the receiver to believe that she is telling the truth, while the receiver prefers accurately rating the sender. In such a setting, there is always a separating Bayesian Nash equilibrium (BNE) in which the sender reports the truth and the receiver believes that all sent messages are truthful. However, when the receiver engages in motivated reasoning, directionally distorting his posteriors in directions he prefers, there is no longer a separating BNE, as the sender's best response is to bias messages in the receiver-preferred direction.[3] When the sender and receiver have different beliefs about the receiver's motivated reasoning, and the receiver is unaware of this difference, incentives can lead to a gap between the strategy the sender plays and the strategy the receiver expects her to play. In particular, the sender's incentives lead her to send less truthful messages but do not lead the receiver to rate her as less truthful, predictions that are consistent with the experimental results.

This paper adds to several strands of literature. First, it contributes to the sizeable and growing literature on the causes and consequences of motivated reasoning by emphasizing information transmission. This form of motivated reasoning was first discussed in economics by Bénabou and Tirole (2002) and was further formalized in a series of subsequent papers by these authors (Bénabou and Tirole 2011; Bénabou 2013; Bénabou and Tirole 2016).[4] Bénabou (2013), Levy (2014), and McGee (2021) theoretically study beliefs about others' motivated reasoning, but to my knowledge, my paper is the first to *experimentally* study such higher-order beliefs.[5] Hagmann and Loewenstein (2018) study persuasion in a motivated setting, but focus on senders and receivers with misaligned incentives. Experiment 2 further illustrates the usefulness of the experimental design of Thaler (2021a), extending a set of papers I have written that study motivated reasoning in politics (Thaler 2021a), gender differences (Thaler 2021b), and optimism about the world (Thaler 2020).

Second, these results relate to the experimental cheap talk literature and preferences for truth-telling. As in Abeler, Nosenzo, and Raymond (2019), I find that people inherently have a preference for truth-telling, but find that this preference is malleable. Compared to most cheap-talk and information-design games in the literature (e.g. Cai and Wong 2006;

[3]More technically, I discuss a solution concept in which the receiver plays a strategy of a BNE in which he has one particular level of bias, and the sender — who is aware of the receiver's strategy and beliefs about his own bias — plays a best response to the receiver under the assumption that he has another level of bias.

[4]The "optimal beliefs" framing discussed in my model is also similar to the setups of Brunnermeier and Parker 2005 and Mobius et al. 2014).

[5]My model has overlaps with Levy (2014). He studies a setting in which a policy-maker exploits the demand for motivated beliefs by voters by choosing whether to supply information to them and finds that there is often a truthful equilibrium as well as non-truthful equilibria. In my setting, I find an even starker setting in which truthful equilibria *cannot* exist.

Frechette, Lizzeri, and Perego 2021), my game does not have misaligned incentives between senders and receivers, but still finds considerable lying and distrust.[6] A number of other papers experimentally study the determinants of lying (e.g. Erat and Gneezy 2012; Serra-Garcia, Damme, and Potters 2011; Gneezy, Kajackaite, and Sobel 2018), receivers' behavior in detecting lies in nonpolitical contexts (Serra-Garcia and Gneezy 2021); and non-motivated reasons why people may lie to be seen as truthful (Choshen-Hillel, Shaw, and Caruso 2020; Barron 2019; Shalvi et al. 2019). I add to this literature by emphasizing behavior of senders and explicitly relating the decision to lie to beliefs about others' motivated reasoning.

Third, this paper adds to the psychology literature on bias blind spots (Pronin, Lin, and Ross 2002; Pronin, Gilovich, and Ross 2004; Pronin 2007), which finds that people see themselves as less biased than others,[7] by showing that in the case of motivated reasoning these belief differences extend to higher-order reasoning. These results also relate to papers that experimentally study higher-order beliefs about motivated and unmotivated biases in social settings (e.g. Oprea and Yuksel 2020; Brownback, Burke, and Gagnon-Bartsch 2021; Gagnon-Bartsch 2021). My data provide evidence for bias blind spots in motivated reasoning, while also suggesting that people act as if *others* treat them as if they are less biased.

Finally, the findings in this paper contribute to the understanding of the spread of disinformation on social media. This paper shows that even small incentives can significantly change how truthful people are with others. It relates to a long literature that studies traditional media to show how news senders may distort messages in the direction of their audiences' current beliefs (Gentzkow and Shapiro 2006; Gentzkow and Shapiro 2010; Mullainathan and Shleifer 2005). I show that appealing to politically-motivated beliefs matters, even when current beliefs are held fixed. This finding also provides a belief-driven explanation for pandering, in which senders bias messages towards receivers' *preferences* (Maskin and Tirole 2004; Che, Dessein, and Kartik 2013; Canes-Wrone, Herron, and Shotts 2001). In my strategic setting, motivated beliefs can break informative equilibria in a way that differing priors cannot, which does not typically observably occur in individual decision making (Little 2021). The difference between motivated beliefs and current beliefs can explain why disinformation is especially prevalent in politics as opposed to unmotivated contexts.

The rest of the paper proceeds as follows: Section 2 develops the sender-receiver model. Section 3 presents the design and results of Experiment 1. Section 4 presents the design and results of Experiment 2. Section 5 concludes and proposes directions for future work. Study materials and additional results are in the appendices.

[6]This result is in many ways directionally opposite from Cai and Wong (2006), who find excess trust and truth-telling absent motivated beliefs.

[7]In economics, Fedyk (2021) finds similar results in the context of self-control problems.

2 Theory

2.1 A signaling game

This section introduces a simple sender-receiver game that forms the basis of Experiment 1. I consider a game with one sender (she) and one receiver (he). There is a state of the world which is either high (θ_H) or low (θ_L). The sender observes the true state of the world and the receiver wants to learn the true state of the world. The sender has a preference for telling the truth and may receive an additional benefit for having the receiver think she is telling the truth. This benefit is a reduced-form interpretation of reputation. For instance, the sender may prefer to be perceived as truthful so that the receiver wishes to consume more news from her or because the receiver will be more likely to "like" or "share" the message.

After observing the sender's message, the receiver is asked to assess the probability that the sender was telling the truth, receiving a payoff that depends on the accuracy of his assessment. These assessments are functionally equivalent to assessing the probability that the state is high or low. This payoff is a reduced-form interpretation of accuracy motives in which receivers choose how much to trust sources and have their choices impact utility.

More formally, the timing of the game is as follows: First, Nature chooses whether the state θ is θ_H (with probability π) or θ_L (with probability $1-\pi$). We will later interpret π as reflecting R's prior belief of the state. Next, S learns the state, and then chooses whether to send message x_H or x_L to R. After observing the message, R takes action $a \in [0,1]$.

The receiver is incentivized to report his true belief via a quadratic utility function: $u_R = 1-(1-a)^2$ if the sender's message is truthful, and $u_R = 1-a^2$ if the message is false. The sender's utility is a function of two components. First, she has an intrinsic preference to report the truth, receiving $\tau > 0$ for truth-telling. Second, she prefers that the receiver's assessment a of her truthfulness is higher. $\gamma \geq 0$ corresponds to the weight put on the receiver's assessment.

The utility matrix, as a function of a, is as follows:

	Nature chooses θ_H	Nature chooses θ_L
Sender chooses x_H	$\gamma a + \tau, 1-(1-a)^2$	$\gamma a, 1-a^2$
Sender chooses x_L	$\gamma a, 1-a^2$	$\gamma a + \tau, 1-(1-a)^2$

The first element in each cell is the sender's utility, and the second is the receiver's utility. Figure 1 describes the game in extensive form.

I consider the Bayesian Nash equilibrium (BNE) of the game. There is always a full-information separating BNE in which the sender plays a *truthful strategy* — sending x_H

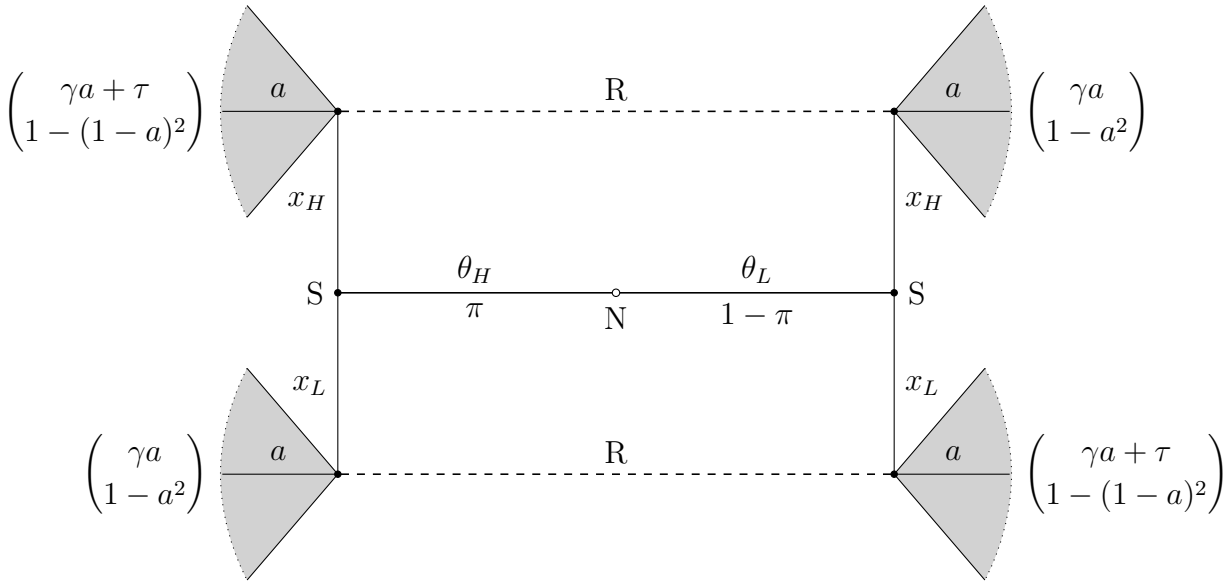


Figure 1: The extensive-form game. S's payoffs are listed on top; R's payoffs are listed on bottom. Dashed lines denote information sets.

given θ_H and sending x_L given θ_L — and the receiver chooses $a(x_H) = 1$ and $a(x_L) = 1$. The sender earns $\gamma + \tau$ and the receiver earns 1.[8] As is common in coordination games, there may be other equilibria. In this game, the existence of other equilibria depends on τ/γ being sufficiently small.[9]

2.2 Motivated reasoning

When people receive information, they often distort how they process the information in ways that benefit their motivated beliefs: *motivated reasoning*. Agents receive a binary signal that either has a "good" or "bad" realization, and motivated reasoners asymmetrically update from this information in a non-Bayesian fashion. In particular, they optimally form a posterior that trades off the benefit of believing the state is good with holding accurate beliefs, leading them to act as if they infer relatively more from signals in the good direction. This definition follows a common approach from past literature (Bénabou and Tirole 2002;

[8] I focus on BNE, instead of equilibrium refinements, for simplicity. This BNE is also a perfect Bayesian equilibrium and a sequential equilibrium, and it Pareto-dominates all other equilibria.

[9] Specifically, there is a pooling BNE in which the sender always sends x_H iff $\tau/\gamma \leq \pi$. In such a BNE, the receiver plays $a(x_H) = \pi$ and $a(x_L) = \pi'$ for $\pi' < \pi - \tau/\gamma$. Similarly, there is an x_L-pooling BNE iff $\tau/\gamma \leq 1 - \pi$. The receiver plays $a(x_H) = 1 - \pi''$ and $a(x_L) = 1 - \pi$ for $1 - \pi'' < 1 - \pi - \tau/\gamma$. There can also be mixed-strategy equilibria. If $\tau/\gamma \leq 1 - \pi$, there is a BNE in which the sender sends $x_H|\theta_H$ with probability 1, and sends $x_L|\theta_L$ with probability $\frac{1-\pi-\tau/\gamma}{(1-\pi)(1-\tau/\gamma)}$. The receiver plays $a(x_H) = 1 - \tau/\gamma$ and $a(x_L) = 1$. If $\tau/\gamma \leq \pi$ there is a BNE in which the sender sends $x_L|\theta_L$ with probability 1 and sends $x_H|\theta_H$ with probability $\frac{\pi-\tau/\gamma}{\pi(1-\tau/\gamma)}$. The receiver plays $a(x_L) = 1 - \tau/\gamma$ and $a(x_H) = 1$.

Brunnermeier and Parker 2005; Bénabou and Tirole 2011).

We consider how motivated reasoning affects behavior in the game described above. The receiver will either see an x_H signal or an x_L signal; without loss of generality, suppose that x_H is "good" and x_L is "bad." Denote $P(x_H|H) = p_H$ and $P(x_L|L) = p_L$, where $p_H \geq 1-p_L$, so that x_H is weakly more likely in state H and x_L is weakly more likely in state L.[10]

A Bayesian receiver gives the following assessments:

$$a(x_H) = P(H|x_H) = \frac{\pi p_H}{\pi p_H + (1-\pi)(1-p_L)},$$
$$a(x_L) = P(L|x_L) = \frac{(1-\pi)p_L}{\pi(1-p_H) + (1-\pi)p_L}.$$

Motivated reasoners act as if they receive additional utility for reporting a posterior that is consistent with the good state. Specifically, they receive a benefit of $\lambda \cdot a$ towards positively assessing signals that indicate that the state is high and $\lambda(1-a)$ towards negatively assessing signals that indicate the state is low:[11]

The updated utility matrix, as a function of a, is as follows:

	Nature chooses θ_H	Nature chooses θ_L
Sender chooses x_H	$\gamma a + \tau, 1-(1-a)^2 \mathbf{+\lambda a}$	$\gamma a, 1-a^2 \mathbf{+\lambda a}$
Sender chooses x_L	$\gamma a, 1-a^2 \mathbf{+\lambda(1-a)}$	$\gamma a + \tau, 1-(1-a)^2 \mathbf{+\lambda(1-a)}$

Therefore, the motivated receiver gives assessments that equal:

$$a(x_H) = P(H|x_H) = \max\left\{\lambda/2 + \frac{\pi p_H}{\pi p_H + (1-\pi)(1-p_L)}, 1\right\}$$
$$a(x_L) = P(L|x_L) = \min\left\{-\lambda/2 + \frac{(1-\pi)p_L}{\pi(1-p_H) + (1-\pi)p_L}, 0\right\}.$$

2.3 Specifying higher-order beliefs

The sender and the receiver may have have different beliefs about the receiver's type λ; they may also have different higher-order beliefs. I denote the receiver's first-order belief as a probability distribution $\Delta_R(\lambda)$. The receiver's second-order belief reflects his belief about the sender's belief $\Delta_R(\Delta_S(\lambda))$. The receiver's third-order belief reflects his belief about the sender's belief about his belief: $\Delta_R(\Delta_S(\Delta_R(\lambda)))$. And so on.

Similarly, the sender's first-order belief is a probability distribution $\Delta_S(\lambda)$. The sender's

[10] Since $\tau > 0$, this will be true in all equilibria of this game.

[11] Predictions are similar if overweighting of good news is less severe than underweighting of bad news (e.g. Bénabou and Tirole 2002; Mobius et al. 2014).

second-order belief reflects her belief about the receiver's belief: $\Delta_S(\Delta_R(\lambda))$. The sender's third-order belief reflects her beliefs about the receiver's belief about her belief: $\Delta_S(\Delta_R(\Delta_S(\lambda)))$. And so on.

Formally, I define each player's belief hierarchy in the following recursive manner:[12]

$$\Delta_{S,k+1}(\lambda) = \Delta_S(\Delta_{R,k}(\lambda)) \text{ and}$$
$$\Delta_{R,k+1}(\lambda) = \Delta_R(\Delta_{S,k}(\lambda)) \text{ for each } k = 1, 2, ...$$

For simplicity, I assume that both the sender and the receiver have point beliefs for each element in their belief hierarchy. I denote $\hat{\lambda}_{i,k}$ to be player i's point belief $\Delta_{i,k}(\lambda)$, and will omit the subscript when $k = 1$. In line with the psychological literature on bias blind spots (Pronin, Lin, and Ross 2002; Pronin, Gilovich, and Ross 2004; Pronin 2007), I assume that the receiver's belief about λ is lower than the sender's belief about λ: $\hat{\lambda}_S > \hat{\lambda}_R$.

I also assume that the receiver projects his belief onto the sender when constructing higher-order beliefs and that the sender is aware of this projection (e.g. McGee 2021):

$$\hat{\lambda}_{S,k} = \hat{\lambda}_R \text{ and}$$
$$\hat{\lambda}_{R,k} = \hat{\lambda}_R \text{ for each } k = 2, 3, ...$$

One psychology behind this formulation is that individuals believe they are "apparently unique" in their lack of bias. That is, a receiver thinks of himself as unbiased — in contrast to others — and projects his perception of his unbiasedness onto others' beliefs. Note that there is no restriction on the accuracy of *first-order* beliefs. It may be the case that λ is indeed equal to the low value of $\hat{\lambda}_R$ and that the sender overstates the receiver's bias.

2.4 Motivated equilibrium

I now consider how beliefs about λ affect equilibrium behavior. I posit that the receiver expects that his strategy is in a BNE of the game in which $\lambda = \hat{\lambda}_R$ is common knowledge, while the sender plays a best response to the receiver's strategy under her belief that $\lambda = \hat{\lambda}_S$.

The strategy profile $\Big(s_S(\theta, \{\hat{\lambda}_{S,k}\}), s_R(x, \{\hat{\lambda}_{R,k}; \lambda\})\Big)$ is a *motivated equilibrium* of the game described above if:

- $s_S(\theta, \{\hat{\lambda}_{S,k}\})$ is a best response to $s_R(x, \{\hat{\lambda}_{R,k}\})$;
- $s_R(x, \{\hat{\lambda}_{R,k}\})$ is a best response to $s_S(\theta, \{\hat{\lambda}_{R,k}\})$; and
- $s_S(\theta, \{\hat{\lambda}_{R,k}\})$ is a best response to $s_R(x, \{\hat{\lambda}_{R,k}\})$.

[12] I assume common knowledge over other game characteristics, such as over the sender's knowledge of the state. An extension could look at cases in which senders may misinfer from Nature's signals.

In a motivated equilibrium, the receiver's strategy is a best response to a hypothetical sender who holds the receiver's belief $\hat{\lambda}_R$ about λ and who plays a best response to the receiver's strategy. The actual sender, who holds belief $\hat{\lambda}_S$ about λ and is aware of the receiver's higher-order beliefs, plays a best response to the receiver's strategy.

The existence of a motivated equilibrium in which the sender chooses a truthful strategy now depends on other parameters, since she expects the receiver to give a higher rating to x_H than x_L, so she faces a tradeoff between honesty τ and incentives γ. If the receiver believes the sender will play a truthful strategy, he will play $a(x_H) = 1$ and $a(x_L) = 1 - \lambda/2$. Whether this strategy is in equilibrium relies on whether the sender believes $\lambda/2 \leq \tau/\gamma$ and on her higher-order beliefs. The receiver's belief that the sender is playing a truthful strategy similarly depends on whether he believes $\tau/\gamma \geq \lambda/2$ and on his higher-order beliefs.

When incentives are low or both players' beliefs about the receiver's bias are low — $\tau/\gamma \geq \hat{\lambda}_S/2 > \hat{\lambda}_R/2$ — then there is a motivated equilibrium in which the sender plays a truthful strategy and the receiver plays a best response to the truthful strategy (including his distortion of beliefs because of motivated reasoning).

When incentives are high or both players' beliefs about the receiver's bias are high — that is, $\tau/\gamma < \hat{\lambda}_R/2 < \hat{\lambda}_S/2$ — then in any motivated equilibrium, the sender will not play a truthful strategy and the receiver will not play a best response to the truthful strategy.[13] An x_H-pooling equilibrium exists in which the sender always plays x_H, the receiver plays $a(x_H) = \max\{\pi + \lambda/2, 1\}$, and the receiver plays $a(x_L) < \pi + \lambda/2 - \tau$ off the equilibrium path. An x_L-pooling equilibrium will also exist if $\lambda/2 + \tau < 1 - \pi$; $a(x_L) = 1 - \pi - \lambda/2$ and $a(x_H) < 1 - \pi - \lambda/2 - \tau$.

If $\hat{\lambda}_S = \hat{\lambda}_R$, then these characterize all possible cases, and receivers always play a modified best response to senders. However, if $\tau/\gamma \in (\hat{\lambda}_R/2, \hat{\lambda}_S/2)$, there is a motivated equilibrium in which the receiver plays a best response to a truthful strategy but the sender instead plays an x_H-pooling strategy. That is, the receiver plays $a(x_H) = 1$ and $a(x_L) = 1 - \hat{\lambda}_R/2$, but the sender sends x_H in both states. As above, pooling equilibria can also exist.

In the experiment, I will explore an environment and show that the latter equilibrium exists.

Hypothesis 1: When the sender is incentivized to be perceived as truthful ($\gamma > 0$), there exists τ such that there is a motivated equilibrium in which the receiver plays a best response to the separating strategy and the sender plays an x_H-pooling strategy. In this equilibrium, the receiver rates $a(x_H) > a(x_L)$ and the sender sends more false signals in the bad state: $P(x_H|\theta_L) > P(x_L|\theta_H)$.

[13] Were there such an equilibrium, the receiver would believe that the sender believes he will play $a(x_L) = 1 - \hat{\lambda}_R < 1 - \tau/\gamma$, thereby making the sender's expected utility higher for deviating to x_H given state θ_L.

Hypothesis 2: When the sender is unincentivized ($\gamma = 0$), she will not condition messages on her perception of the receiver's motivated beliefs ($\hat{\lambda}_S$).

Hypothesis 3: Increasing the sender's incentives from 0 to $\gamma \in (2\tau/\hat{\lambda}_S, 2\tau/\hat{\lambda}_R)$ leads to fewer truthful messages sent by the sender in state θ_L but leads to no change in strategy by the receiver.[14]

This analysis has thus far treated λ as exogenous from π. However, if we consider π to be the receiver's prior beliefs, there are reasons to believe that senders may condition on π. For instance, the degree to which receivers motivatedly reason may depend on π. Suppose that before entering this game, two receivers with the same prior observed the same message. Then, differences in their posteriors would reflect differences in the extent of their motivated beliefs.[15] If $\hat{\lambda}_S$ is positive and increasing in π, then the incentivized sender will be less likely to play the separating strategy as π increases. It is worth noting that other theories make similar predictions with respect to π.[16]

Lastly, note that the specification of higher-order beliefs substantially affects behavior. For instance, if

$$\hat{\lambda}_{S,k} = \hat{\lambda}_S \text{ and}$$
$$\hat{\lambda}_{R,k} = \hat{\lambda}_S \text{ for each } k = 2, 3, ...,$$

then both players expect the sender to play the strategy of the BNE in which $\lambda = \hat{\lambda}_S$, and receivers expect themselves to play as if $\lambda = \hat{\lambda}_R$. Thus, specifying higher-order beliefs is important in equilibrium predictions.

3 Experiment 1

3.1 Design

Below, I outline the timing, treatment arms, and main hypotheses of the game, which follows the setup of the model. Screenshots for the pages subjects see are in Appendix C.2.

Subjects are randomly assigned to be senders or receivers. Receivers give a prior belief about whether the answer to a factual question is greater or less than a target number. Senders learn the true answer and choose (via the strategy method) whether to send a

[14] When $\gamma = 0$, only the truthful strategy is in equilibrium.

[15] Eil and Rao (2011) initially consider this hypothesis. A more formal version of the argument is described in Thaler (2021a).

[16] For instance, under quantal response equilibrium (McKelvey and Palfrey 1995; McKelvey and Palfrey 1998), senders and receivers will play as if their utility functions have a noise term added, and receiver priors will impact sender behavior.

message that says "The answer is greater than [the target]" or "The answer is less than [the target]." For each message, receivers state (via the strategy method) how likely it is that their sender's answer is truthful.

To fix ideas, consider the following question that subjects see in the experiment:

> *The U.S. has seen a sharp rise in the share of undocumented immigrants over the past several years. Some people believe that undocumented immigrants are more likely to commit violent crimes, while others believe that undocumented immigrants are less likely to commit violent crimes.*
>
> *Texas is the only state that directly compares crime rates for US-born citizens to undocumented immigrants, and provided felony data from 2012-2018. During this time period, the felony violent crime rate was 213 per 100,000 U.S. citizens.*
>
> *This question asks about the felony violent crime rate for undocumented immigrants. Do you think it is more likely that this rate was greater or less than 213 per 100,000?*

Republicans and Democrats disagree about the answer to this question, and subjects may expect Republicans to be motivated to believe the crime rate is higher, and Democrats to be motivated to believe the crime rate is lower.[17] As such, I code the question as one on which Republicans are motivated to believe "greater than" is more likely than "less than" to be correct, and Democrats are motivated to believe "less than" is more likely than "greater than" to be correct. (For a full detailing of topics and hypothesized motives, see Table 1; for the full text of each question, see Appendix C.1.)

For each question, the following data are elicited. (Details on incentives are below.)

1. **Receivers: Prior beliefs.** Receivers are asked and incentivized to guess the percent chance that the answer to questions like the one above is "greater than" or "less than" a particular value, or to guess the percent chance that a given quote is "accurate" or "inaccurate," using a scale from 0 to 100. Their reports are restricted to be in multiples of ten: 0 percent, 10 percent, 20 percent ..., or 100 percent. They are incentivized to state their true belief.

2. **Senders: Message choice.** Senders are provided with the correct answer to the question and are asked to choose one of two messages to send to a receiver. The message they can send either says "the answer is **greater than** [the particular value]" or "the answer is **less than** [the particular value]." Senders make choices via the strategy

[17] Indeed, the two parties have differing prior beliefs. Republican receivers' average prior is 56 percent (subject-level clustered s.e. 3.5 pp) and Democratic receivers' average prior is 35 percent (s.e. 3.1 pp).

method, choosing one message *for each possible prior belief that the receiver can report.* That is, senders choose 11 messages for each question. They choose one message if the receiver has a prior of 0 percent, one message if the receiver has a prior of 10 percent, and so on.

On the questions about quote accuracy, senders choose a message that says "the statement is **accurate**" or "the statement is **inaccurate**," and cannot condition on the receiver's prior. Senders choose one message for each of these questions.

3. **Receivers: Message assessment.** Receivers are asked to assess the percent chance that each message was truthful using the strategy method and the 0-100 scale. That is, they assess two messages for each question. They are incentivized to state their true belief.

For senders, the main outcome of interest is whether they choose to send the true message or the false message. For receivers, the main outcome of interest is how they assess the truthfulness of the senders' messages.

Treatments

Senders and receivers are each randomized into several treatments. The main treatment arms are:

1. **Topic arm:** For both senders and receivers, the topics are varied within subject. They are either political or neutral.
2. **Information arm:** For senders, information about the receiver's party is varied within subject. Either senders know the receiver's party or they do not. In addition, senders are randomly matched with Democratic and Republican receivers, so the matched receiver's political motives are randomly either aligned with the true message or aligned with the false message. Receivers are honestly told whether their own party is or is not revealed to the sender (and this is randomized between subjects).
3. **Incentives arm:** The senders' incentives are randomly varied between subjects. In the main treatment arms, the sender is incentivized to be assessed as truthful by the receiver or the sender is unincentivized.[18] Receivers are honestly told what the sender's incentives are.

For senders, this is a 2x3x2 design; for receivers, this is a 2x2x2 design.[19] Subjects play the

[18]There is also a competition incentives arm in which receivers compare the truthfulness of two senders and each sender is incentivized to be rated as more truthful than their competitor. As preregistered, only the main treatment arm is in the primary analyses.

[19]I additionally randomize whether the receiver knows what party the sender is; this treatment arm is not a focus of the current paper.

Round	Senders' Block	Receivers' Block
Pre-randomization	Practice, CRT, demographics 1	Practice, CRT, demographics 1
1-7	Prior beliefs block 1	Send messages block 1
1-7	News assessment block 1	—
8	Attention check	Attention check
9-12	Prior beliefs block 2	Send messages block 2
9-12	News assessment block 2	—
End of experiment	Belief elicitations	Belief elicitations
End of experiment	Demographics 2, solutions	Demographics 2, solutions

Table 2: The timing of treatments in Experiment 1.

rounds in the order as described in Table 2. Within a block, the questions are presented in a random order.[20]

Before their role is revealed, senders and receivers give their demographic information, including political party preference.[21] Subject select a party or party-lean. Their party is defined as Republican (including leaner) if they state that they are Republican or an independent who leans towards the Republican party, and their party is defined as Democratic (including leaner) otherwise. Next, all subjects play a practice round, playing in both the sender and receiver roles for the given incentives arm they are in. After the practice round, subjects' roles are revealed. The timing of the practice round was chosen to ensure that subjects would be familiar with both roles in the experiment instead of only focusing on their specific role.

Incentives

I first describe the mapping of "points" into payoffs and then describe how subjects earn points. All subjects are incentivized using a version of the binarized scoring rule (Hossain and Okui 2013).[22] Throughout the experiment, subjects earn between 0 and 100 points for each incentivized response (e.g. an answer to a question, assessment of a message, or choice of a message) they give. At the end of the experiment, ten subjects are randomly selected to

[20] I do not see evidence that the order of questions impacts treatment effects.

[21] Other demographics are age, gender, politics, education, and race. They are also asked to do a three-item cognitive reflection task (modified from Frederick 2005).

[22] This earnings system is a version of the most broadly incentive-compatible one discussed in Azrieli, Chambers, and Healy (2018).

receive a bonus payment. If a subject is selected, they either receive \$10 or \$100 depending on their responses. One response is randomly selected to determine payment; the percent chance that they win the bonus is equal to their points earned by this response. When a response is chosen for payment, a sender and receiver are randomly matched for the relevant question, and only the relevant choices in the strategy method are used.

Receivers' prior beliefs and assessments are incentivized by the quadratic scoring rule. For each question, they report a prior $\pi \in [0, 1]$ about the answer. If the answer is "greater than," they earn $100(1-(1-\pi)^2)$ points; if the answer is "less than," they earn $100(1-\pi^2)$ points. A similar scoring rule is used to incentivize their assessments about their matched senders' messages; receivers who state that a message is truthful with probability a earn $100(1-(1-a)^2)$ points if the message is truthful and $100(1-a^2)$ points if the message is false.[23] Receivers maximize expected points by stating the closest multiple of 0.1 to their true belief. They are given a table with the points earned as a function of each assessment and news type and are told that providing honest assessments is the best way to maximize expected earnings.

Senders in the incentivized condition are incentivized based on the receivers' assessments; the points they earn for a given choice equals their matched receiver's assessment of the percent chance their message was truthful. Senders in the unincentivized condition do not have this round chosen for payment.[24] Subjects in all roles and treatments are incentivized to give accurate answers to the beliefs questions (whose answers are between 0 and 100). They are incentivized using a quadratic scoring rule. If they guess g and the correct answer is c, they earn $\max(0, 100-(c-g)^2)$ points.

This design makes substantial use of the strategy method. The strategy method has two clear advantages and two clear disadvantages in this design.[25] The first advantage is that it removes much of the role that other subjects' perceptions play for reasons outside the model. For instance, a sender may otherwise send messages that they would like the receiver to see because they want the receiver to find her entertaining or likeable (e.g. Serra-Garcia and Gneezy 2021).[26] The second advantage is statistical power in detecting effects; senders choose eleven messages on each question (instead of one). The data can also estimate, within-subject and within-question, the effect of priors on messages sent. The first disadvantage is

[23]Receivers in the competition treatment who state that Sender X is more truthful than Sender Y earn $100(1-(1-a)^2)$ points if Sender X has chosen more truthful messages over the course of the experiment and earn $100(1-a^2)$ if Sender Y has.

[24]Instead, unincentivized senders are solely paid based on their responses to the beliefs questions.

[25]Brandts and Charness (2011) is a classic paper that discusses differences in behavior between direct elicitation and the strategy method in other games, generally finding that such differences are modest.

[26]Relatedly, with this version of the strategy method, subjects will not use others' behavior in previous rounds to predict messages in the current round, limiting learning effects.

that the strategy method does not reflect how people send messages in practice, but it is not clear why this would affect the role of hypothesized mechanisms. The second disadvantage is choice overload, which may lead subjects to attend less to any particular choice of message. Choice overload would likely push the rate of false news closer to 1/2, and would dampen treatment effects, if senders behave more randomly.[27]

3.2 Data

750 subjects were recruited from Prolific (prolific.co) in September 2021 and passed a simple attention check.[28] Prolific is an online platform that was designed by social scientists in order to attain more representative subject samples; it has been shown to perform well relative to other subject pools (Gupta, Rigotti, and Wilson 2021). The subject pool was restricted to regular social media users. Specifically, Prolific asks platform users which websites they use "on a regular basis (at least once a month)," and the study was only available to Prolific users who say they regularly use Facebook, Twitter, Instagram, Reddit, or LinkedIn. Subjects were recruited so that 375 were Democrats or Independents who lean towards the Democratic Party and the other 375 were Republicans or Independents who lean towards the Republican Party. They were additionally required to have had prior experience on the platform.[29]

The subjects were split evenly into 375 senders and 375 receivers. Within each role, subjects were randomly chosen to be in each of the three incentives treatments. Overall, there were 254 subjects (34 percent) in the unincentivized treatment, 248 (33 percent) in the main incentives treatment, and 248 (33 percent) in the competition treatment.

Over the course of the experiment, senders made 30,340 choices of messages.[30] Most analyses restrict to subjects in the unincentivized treatment (10,107 choices) and the incentivized treatment (10,036 choices). Note that for analyses in this paper, I include all messages *chosen* by senders in the data. That is, I include each strategy-method choice in analyses: {Send x if R has prior of 0/10, send x if R has prior of 1/10, send x if R has prior of 2/10, ...}. An alternative approach would be to restrict analyses to messages that correspond to

[27] Effects do not noticeably change over the course of the experiment, suggesting that choice overload is not a primary confound.

[28] 13 subjects (2 percent) were excluded for incorrectly answering the attention check. Results are robust to the inclusion of these subjects.

[29] All subjects needed to have completed 100 prior studies and have at least a 90-percent approval rating. Most subjects were required to have registered for Prolific prior to July 2021. These specifications were preregistered; however, due to an unexpectedly slow sign-up rate, 19 percent of subjects were eventually recruited from a larger sample that included registrations after July 2021.

[30] This is 99.9 percent of the targeted number of 30,375 choices. The remaining 0.1 percent are instances where the subject did not select an answer.

the receiver's actual prior. While such an approach may be natural to interpret, it would limit the number of observations dramatically (typically by a factor of 11), and all estimates would become substantially noisier.

Receivers gave 4,125 prior beliefs and 8,237 news assessments.[31] Most analyses focus on the unincentivized treatment (2,832 assessments) and the incentivized treatment (2,726 assessments). As with senders, I include all messages assessed by receivers in the data. That is, I include both {Assess truthfulness of message if it says x_H} and {Assess truthfulness of message if it says x_L}.

I analyze treatment balance for the main incentives treatment for senders (Appendix Table 9) and receivers (Appendix Table 10). I do not find significant differences by treatment.

3.3 Main results

The effects of incentives and receivers' party are evident from the raw data. The top panel of Figure 2 shows that incentives lead to more false messages chosen by senders. The bottom panel of Figure 2 shows that both the receiver's party and prior affect senders' truthfulness.

To causally identify the effect of the receiver's party and prior on the senders' behavior, I run within-subject regressions. The main specification regresses an indicator for sending the false message on an indicator that equals one if the false message is aligned with the receiver's party (Party-False Aligned) for each subject i, question topic q, and round r. I control for the receiver's prior belief and include fixed effects for i, q, and r.

$$SendFalse_{iqr} = \alpha + \beta_1 \cdot 1(PartyFalse)_{iqr} + \beta_2 \cdot PriorFalse_{iqr} + \nu FE_i + \delta FE_q + \zeta FE_r + \epsilon_{iqr}$$

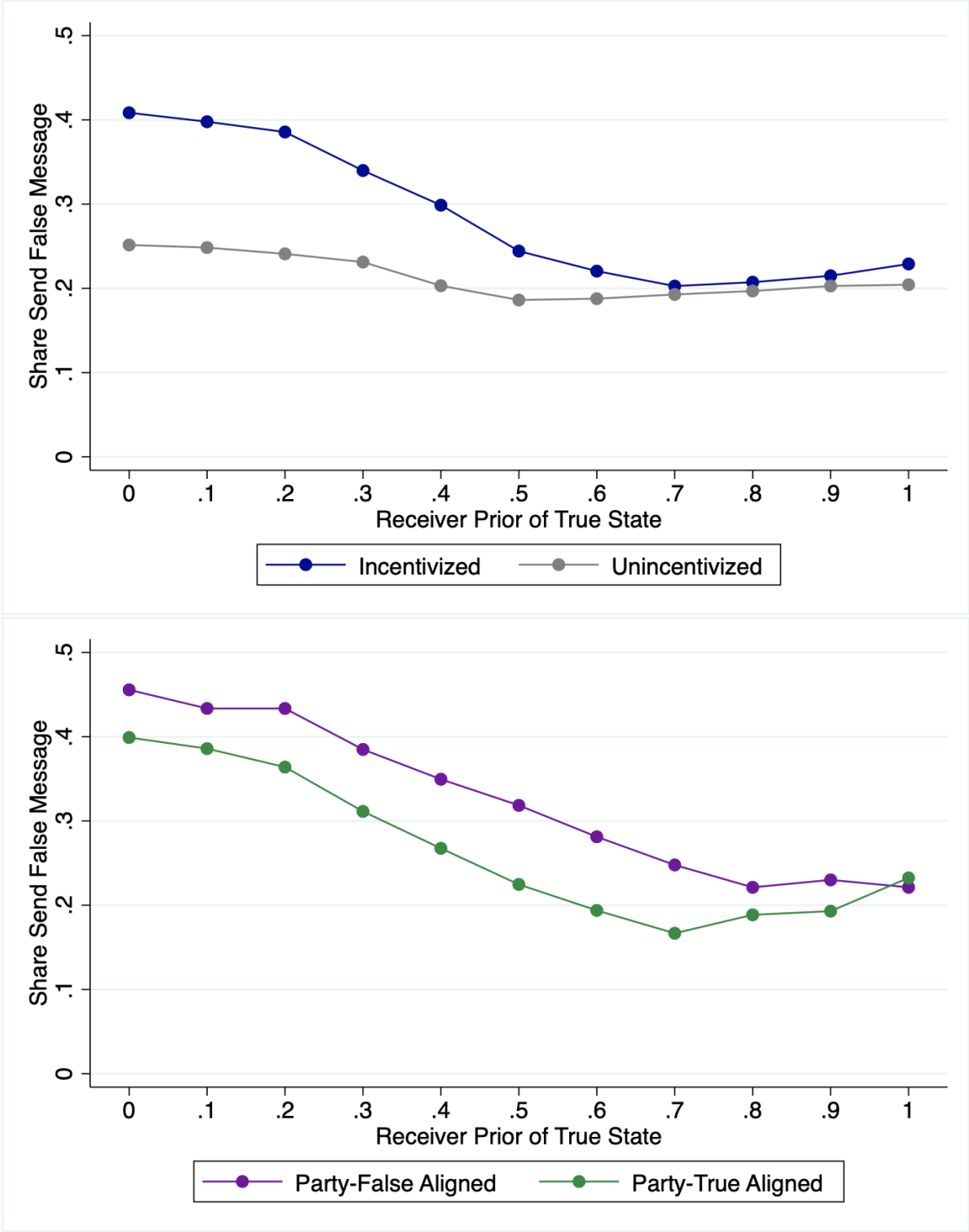
Table 3 shows, consistent with Figure 2, that incentivized senders are more likely to send the false message when good news for the receiver's party is false than in other treatments, controlling for the receiver's prior over the true state.

The specifications differ in the group that Party-False Aligned is compared to. The first column shows that subjects send more false messages when Party-False is aligned than when Party-True is aligned on political topics. The second column shows that subjects send more false messages when Party-False is aligned than when the receiver's party is unknown. The third column suggests that subjects send more false messages when Party-False is aligned than they do on neutral topics. Finally, the last column shows that subjects send more false messages when Party-False is aligned than the three categories aggregated.[32]

[31] Priors constituted 100 percent of the targeted number of 4,125. News assessments constituted 99.8 percent of the targeted number of 8,250.

[32] Note that since I am comparing political to neutral topics, the last two specifications do not include question-level fixed effects.

Figure 2: The Effect of Senders' Incentives, Receivers' Prior, and Receivers' Party on Sending False News



Notes: Receiver Prior of True State: the receiver's belief that the true state is correct. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message. Both panels restrict to political questions. The bottom panel restricts to senders who are incentivized and learn the party of the receiver.

Table 3: Factors that lead incentivized subjects to send false messages

	Vs. Party-True Aligned	Vs. No Info	Vs. Neutral Topics	Vs. All Others
Party-False Aligned	0.071**	0.072***	0.049*	0.053**
	(0.032)	(0.023)	(0.029)	(0.021)
Prior-False Aligned	0.248***	0.229***	0.226***	0.216***
	(0.053)	(0.059)	(0.060)	(0.049)
Question FE	✓	✓		
Subject FE	✓	✓	✓	✓
Round FE	✓	✓	✓	✓
Vs. Party-True Aligned	✓			✓
Vs. No Info		✓		✓
Vs. Neutral Topics			✓	✓
Observations	4990	4705	4055	8782
Subjects	124	123	124	124
Mean	0.296	0.297	0.322	0.292

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Vs. lines indicate the comparison group. Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message. Prior-False Aligned: the receiver's prior belief that the incorrect answer is true. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. No Info: indicator for the receiver's party not being revealed. By chance, one subject happened to always learn the receiver's party. Only includes observations where the sender conditions on the receiver's prior.

Table 3 also shows that incentivized senders are more likely to send the false message when receivers' priors are false. The first column shows that the treatment effect of having the receiver's party aligned with the false state rather than the true state is equivalent to the treatment effect of the receiver's prior changing by 0.071 / 0.248 = 29 pp.

Appendix Table 11 repeats the analysis in Table 3, but does not control for whether the receiver's prior is aligned with the false message. Senders were not given the receiver's prior on all questions, so this version modestly increases the sample size. Results are qualitatively similar, and the treatment effect estimates are slightly larger. The increased estimates may be because senders inferred receivers' prior beliefs from their political party on questions where they could not condition on priors.

The effect that senders expect receivers' party to have on inference is similar when beliefs are explicitly elicited. Senders are asked to predict the average assessment of receivers whose prior is 1/2 when they receive good news for their party or bad news for their party. Incentivized senders estimate that receivers will have an average gap of 30 pp (s.e. 3 pp).

Next, we turn to the effect that incentives have on the truthfulness of messages. Table 4 compares behavior of the incentivized senders to behavior of the unincentivized senders using a between-subject specification. This specification regresses $SendFalse_{iqr}$ on an indicator for the incentives treatment and a set of demographic and treatment controls (X_i) as well as fixed effects for q and r.

$$SendFalse_{iqr} = \alpha + \beta \cdot 1(Incentivized)_i + \eta X_i + \delta FE_q + \zeta FE_r + \epsilon_{iqr}$$

Table 4 shows that there is a negative effects of incentives on message truthfulness.

Table 4: The effect of incentives on sending false messages

	(1)	(2)	(3)	(4)	(5)	(6)
Incentivized	0.072***	0.073***	0.109***	0.107***	0.050*	0.051*
	(0.026)	(0.026)	(0.037)	(0.037)	(0.029)	(0.028)
Party-False Aligned					0.061**	0.062**
x Incentivized					(0.026)	(0.026)
Party-False Aligned					-0.005	-0.003
x Unincentivized					(0.027)	(0.027)
Question FE	✓	✓	✓	✓	✓	✓
Round FE	✓	✓	✓	✓	✓	✓
Subject controls		✓		✓		✓
All Questions	✓	✓			✓	✓
Only Party-False Aligned			✓	✓		
Observations	14248	14248	4702	4702	14248	14248
Subjects	249	249	220	220	249	249
Mean	0.250	0.250	0.272	0.272	0.250	0.250

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Subject controls: Gender, race, age, own party, education, CRT score, and whether the receiver knows the sender's party. Only includes observations where the sender conditions on the receiver's prior.

In particular, columns (1) and (2) of Table 4 show that senders are more likely to send the false message when they are incentivized to be perceived as truthful, both with and without demographic controls. As shown by columns (3) and (4), this treatment effect is more pronounced when constrained to the questions on which receivers' party is misaligned with the truth, both with and without demographic controls. Columns (5) and (6) show, both with and without demographic controls, that incentivized senders send more false messages

when receivers' party is misaligned with the truth (as in Table 3), but that unincentivized senders are not directionally affected by this condition. The inclusion of controls does not affect the estimates.

3.4 Unincentivized senders and own-party effects

These effects are driven by explicit incentives rather than by an innate preference to be rated as truthful. In particular, the main treatments do not significantly affect unincentivized subjects, as shown in Appendix Table 12, which replicates Table 3 among the unincentivized senders. There is no statistically significant effect of receivers' party in any specification, and the estimates are all close to zero. This result further demonstrates the important role that incentives play; unincentivized senders do not inherently value aligning their messages with the receiver's motivated beliefs, but they are directionally affected by the incentives to be perceived as truthful.

However, there is still a non-negligible share of unincentivized senders who choose false messages, at 21 percent. Part of this 21 percent could be due to randomness: for instance, sometimes senders click randomly, do not read the answer correctly, or misclick. In addition, even pure coordination games may lead to communication problems if players disagree about the meaning of messages (e.g. Farrell and Rabin 1996). However, some of these senders may be sending false information because of expressive preferences: they prefer to send news that aligns with their own party, even when it is false. Table 5 provides supporting evidence for this expressive-preferences mechanism, showing that unincentivized senders (column 2) send false messages significantly more often when it aligns with their own party.[33]

In other exploratory analyses, I also find evidence that false news effects are stronger when senders and receivers are of the same party. On political questions in the unincentivized and incentivized groups, senders send false news to their own party 23.0 percent of the time (s.e. 1.7 pp) and false news to the opposing party 28.0 percent of the time (s.e. 1.7 pp); that is, they send 5.0 pp more true messages to their own party (s.e. 1.9 pp, $p = 0.009$). Similarly, receivers rate political messages that come from their own party as being 6.0 pp more likely to be truthful (s.e. 0.8 pp, $p < 0.001$).

[33]Meanwhile, incentivized senders (column 1) are not statistically significantly affected by their own party, and are (as described above) instead affected by the other subject's party. However, these estimates are not precisely measured and should not be overinterpreted.

Table 5: The effect of senders' own party on messages

	(1)	(2)	(3)
Own Party-False Aligned	0.010	0.053**	0.054**
	(0.025)	(0.021)	(0.021)
Other's Party-False Aligned	0.092***	0.005	0.008
	(0.028)	(0.028)	(0.027)
Own Party-False Aligned x Incentivized			-0.039
			(0.033)
Other's Party-False Aligned x Incentivized			0.088**
			(0.040)
Question FE	✓	✓	✓
Subject FE	✓	✓	✓
Round FE	✓	✓	✓
Incentivized subjects	✓		✓
Unincentivized subjects		✓	✓
Observations	5486	5136	10622
Subjects	124	125	249
Mean	0.296	0.212	0.255

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Own Party-False Aligned: indicator for the sender's party being aligned with the false message. Other's Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message.

3.5 Receivers' behavior and higher-order beliefs

Next, we turn to receivers' actual assessments of these messages and discuss both players' higher-order beliefs. Receivers in the sender-incentivized condition assess pro-party news to be true with probability 58.4 percent (subject-level clustered s.e. 1.2 pp) and anti-party news to be true with probability 51.3 percent (s.e. 1.3 pp). This gap is 7.1 pp (s.e. 1.8 pp) and statistically significant ($p < 0.001$).

The effect may be due to updating in directions that are consistent with prior beliefs. Receivers believe that the pro-party state is true with probability 55.8 percent (s.e. 1.1 pp), which is statistically significantly larger than 50 percent ($p < 0.001$). As shown by Appendix Figure 5, receivers' priors are consistently, but modestly, in the pro-party direction on essentially every topic. Once prior beliefs are controlled for, the gap in assessments is reduced

to 1.0 pp (s.e. 1.3 pp) and is statistically insignificant.[34] The positive-but-null effect is consistent with other findings in the literature. As summarized by Benjamin (2019), designs in which people receive informative signals are mixed in their ability to detect motivated reasoning.[35] In addition, the sender-receiver setting may confound identification of motivated reasoning from strategic considerations. Because of these limitations, Section 4 uses a different experimental strategy that is able to more cleanly identify receivers' motivated reasoning from differences in prior beliefs, showing that motivated reasoning does indeed play an important role.

Receivers' strategies are largely unaffected by senders' incentives, even though receivers are informed about the incentives in each round. Receivers who are matched with incentivized senders give similar levels of assessments when compared to receivers who are matched with unincentivized senders. Receivers in the sender-unincentivized condition assess pro-party news to be true with probability 57.9 percent (s.e. 1.1 pp) and anti-party news to be true with probability 50.8 percent (s.e. 1.2 pp). This gap is 7.1 pp (s.e. 1.6 pp; $p < 0.001$). Each of these estimates is nearly identical to those in the incentivized condition.

Figure 3 shows the treatment effect on senders and receivers graphically. In particular, it shows that senders are significantly more likely to send false messages when incentivized, with a particularly negative effect when the receiver's party is aligned with the false message.[36] However, receivers who are matched with incentivized senders give similar assessments to receivers who are matched with unincentivized senders in each of these conditions, suggesting naivete to the effect that senders' incentives have.

In exploratory analyses, I use survey questions to study what senders and receivers believe determines the behavior of each player in the game. Appendix Figure 6 shows that incentivized senders report that they rely less on the truth and more on the party of the receiver, as compared to unincentivized senders. However, receivers do not state significantly different reports when they have been faced with incentivized or unincentivized senders. Neither senders nor receivers respond significantly differently about receivers' behavior when they are incentivized versus unincentivized, though the effects are noisy, weakly suggesting that both players believe receivers will not respond much to senders' incentives.

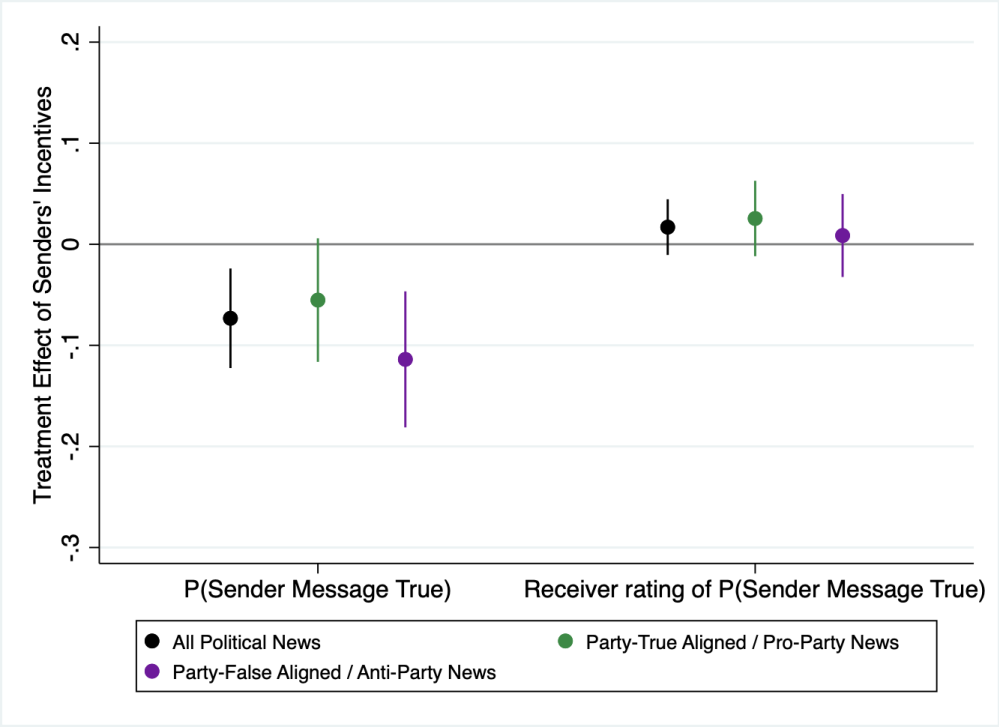
Next, I find systematic differences between receivers' behavior and senders' beliefs about

[34]This estimate comes from regressing assessments on pro-party vs. anti-party within-subject, controlling flexibly for prior beliefs.

[35]Mobius et al. (2014); Eil and Rao (2011); and Charness and Dave (2017) find statistically-significant evidence that people update more from good news than bad news in their contexts, while Ertac (2011); Kuhnen (2014); Buser, Gerhards, and Weele (2018); Coutts (2018); and Barron (2020) do not find such evidence.

[36]In addition, the interaction between the incentives and the party-alignment treatments is in the expected direction, though the estimate is only suggestively significant ($p = 0.088$).

Figure 3: Senders' Incentives Affect Messages Sent but Not How They Are Assessed



Notes: OLS regression coefficients, errors clustered at subject level. Coefficients are from a regression of message truthfulness (for senders) or assessments about message truthfulness (for receivers) on being in the S-incentivized treatment. Controls for age, race, gender, education, CRT score, own party and fixed effects for question and round number are included. This figure shows that senders choose more false messages when incentivized, while receivers do not anticipate more false messages when the sender is incentivized. Only receivers whose party is revealed to senders are included. Error bars correspond to 95 percent confidence intervals.

receivers' behavior, as senders overstate the relative role of politics versus priors in receivers' inference. On average, senders are asked to state their beliefs about Republican and Democratic receivers' assessments of pro-Republican and pro-Democratic messages when the receivers have a prior of 1/2. Senders believe that the gap between the pro-party and the anti-party assessments will be 30 pp. (These beliefs are predictive of the treatment effects of incentives, as shown in Appendix Table 13.)

Senders' beliefs about the party gap are substantially larger than the true gap of 2 pp. As mentioned above, senders treat the effect of the receiver's party as being similar to an effect of a change of 29 pp in the receiver's prior. While it is not possible to determine what the optimal *level* of truth-telling is for a sender, these results suggest that, conditional on the receivers' behavior, senders would be better off by being *relatively* more sensitive to priors and less sensitive to politics.

In the context of the model, these findings about higher-order beliefs are consistent with

the model in which $\hat{\lambda}_S > \lambda \approx \hat{\lambda}_R$. That is, receivers may engage in motivated reasoning to a small extent, and they are aware of this, but senders overstate the bias.[37] When senders' incentives go from $\gamma = 0$ to an intermediate level $\gamma > 0$, receivers do not realize that senders overstate the bias, leading to the form of motivated equilibrium described in Section 2 in which this increase in incentives affect senders but not receivers.

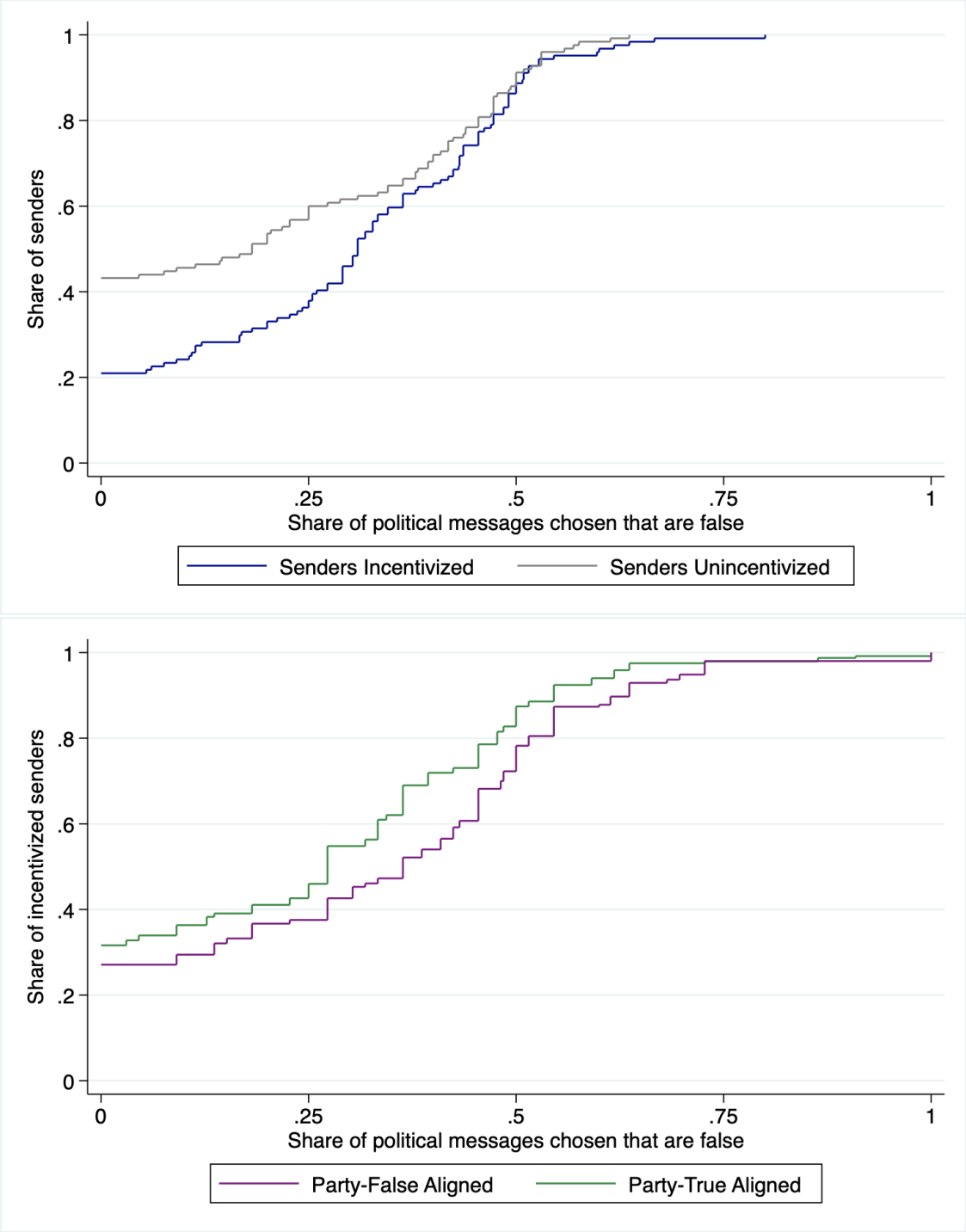
3.6 Discussion and Robustness

The results above are presented at the aggregate level; I next explore differences at the individual level. Figure 4 plots these data with CDFs by incentivized condition and by the alignment of the receiver's party. Figure 4 shows that incentive effects are primarily driven by the more-truthful part of the sender distribution. 43 percent of unincentivized senders never send a false message, while only 21 percent of incentivized senders are always truthful. However, similar shares of senders send false messages over half the time, and these shares are small (9 percent for unincentivized; 11 percent for incentivized). Meanwhile, the effect that the receiver's party has on senders is more evenly dispersed. In particular, it is not clearly clumped at the low end of the distribution. This distributional difference may be because senders condition both on priors and party, and therefore are willing to send false messages that are misaligned with the receiver's party when they are aligned with the receiver's prior. Despite these suggestive differences, these distributions are not sufficiently precisely estimated to indicate more.

There are no distinguishable effects of the receiver's party on unincentivized senders at the individual level. Appendix Figure 7 plots the same CDF as the one in the second panel of Figure 4 but restricts to observations from unincentivized senders. In Appendix Figure 7, the CDFs lie on top of each other, indicating that there are little distributional differences for unincentivized senders by the receiver's party-truth alignment. In addition, the median share of false messages in each condition is zero, indicating that the majority of unincentivized senders never or rarely send false messages.

[37] In Experiment 2, I argue that while λ may be of modest size, it is greater than zero.

Figure 4: CDF of Individual-Level False News by Senders' Incentives and Receivers' Party



Notes: CDF plots of the average share of messages chosen by senders. For instance, the top panel shows that half of incentivized senders send false messages at least 31 percent of the time. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message. Both panels restrict to political questions for which senders condition on receivers' priors. The bottom panel restricts to senders who are incentivized and learn the party of the receiver.

Next, I consider what types of incentives lead to more false messages chosen. I have so far focused on a very particular type of incentivization scheme: senders have direct incentives to have their messages be perceived as truthful. However, there are more indirect forms of incentives that news suppliers often face. One such incentive is due to competition with other senders. To test the effects of competition, I run an additional treatment in which receivers assess which of two senders they believe was more truthful over the course of the experiment as a function of their message in one round, and each sender is incentivized to be rated as more truthful than her competitor.[38] I find that competition incentives do not have a significant impact on the overall truthfulness of messages (Appendix Table 14), and the receiver's prior plays more of a role than party for senders (Appendix Table 15). These results suggest that competition incentives do not substantially affect behavior, but since the null effects are noisy, it is not possible to rule out modest increases in false messages.

Senders may believe that receivers systematically misreport their priors or report their priors with noise. If senders believe that receivers' priors are accentuated more in ways that favor their party, the effects of prior may be overestimated and the effects of party may be understated for senders. On the other hand, senders may expect a Democratic and Republican receiver who each give the same prior to have different beliefs because the prior is rounded to the nearest 0.1. This would affect estimates in the direction that overestimates the party effect. However, the magnitude of the party effect is large enough that, to fully account for these effects, priors would need to be biased by 15 pp.[39] Similarly, if senders believe that receivers' priors are stated with noise, then they may infer something about the receiver from his party. In either case, to fully explain these results, senders would need to expect, on average, that a Democratic and Republican receiver who each state the same $\hat{\pi}$ actually hold priors π_D and π_R that differ by 29 pp. Senders may also believe that receivers' assessments do not reflect their true beliefs, but rather a form of expressive preferences. In this case, the results can be interpreted as saying that senders cater to their beliefs about receivers' expressive beliefs.

Lastly, while the main results provide evidence that senders use the direction of receivers' political beliefs in determining what messages to send them, it is difficult to isolate what form of motivated reasoning is the root cause. For instance, senders may believe that receivers are more *accurate* at assessing senders' messages when they are aligned with the receivers'

[38] In particular, receivers predict the likelihood that each sender is more truthful if the two senders choose the same message and if the two senders choose opposite messages. Mirroring the main incentives treatment, senders earn points equal to the receiver's assessment of the percent chance they are more truthful.

[39] In particular, I run the main specification from Table 4, but replace the prior π with $\min\{\pi + x, 1\}$ for party-true aligned and with $\max\{\pi - x, 0\}$ for party-false aligned beliefs. The coefficient of the party effect necessarily decreases in x, and crosses zero at $x = 0.15$.

party. If senders have other-regarding preferences, caring about the accuracy of receivers' assessments, senders would distort in the direction of the receivers' party. (However, this cannot fully explain why incentives have an effect on senders.) In addition, there may be higher-order belief explanations for these results; for instance, senders may believe that receivers believe that senders distort how they send messages towards being misaligned with the receivers' party. Such a mechanism is similar to the one discussed in Morris (2001).

Therefore, in order to isolate senders' beliefs about the particular bias that receivers have in processing information, Experiment 2 complements the above analyses by (1) identifying receivers' motivated reasoning using assessments of messages sent by a computer and (2) studying the behavior of senders choosing which computer messages to be paid for.

4 Experiment 2

4.1 Design

Experiment 2 is designed to isolate the role that beliefs about others' motivated reasoning play in choosing messages. Receivers are asked to assess the truthfulness of messages from a computer in a task that is designed to elicit their motivated reasoning. Senders decide which of the two computers' messages they want to choose; the messages affect their payoff through the receivers' assessments, but do not affect what the receivers see or how many points they score. That is, senders are essentially betting on messages. Breaking the interaction between senders and receivers allows for identification of motivated reasoning among receivers, which in turn enables us to disentangle the strategic effects present in Section 3 from the motivated-reasoning effects of senders' choice of news. The game is described below. Screenshots for the pages subjects see are in Appendix D.

Receivers play a simplified version of the game in Thaler (2021a). They play the game in two steps:

1. **Median belief:** Receivers are given questions like the one in Section 3.1. The only difference is that, instead of being asked for the probability that the answer is greater or less than a particular number, they are asked to guess what the number actually is. I elicit their median beliefs; receivers report a guess μ such that they believe that the answer is equally likely to be greater than or less than μ. (Details on incentives are below.)
2. **News assessment:** After stating their median beliefs, receivers are given two messages from a computer. One message says "the answer is **greater than** [median]" and the other message says "the answer is **less than** [median]." One of these messages is true

and one is false, and receivers do not know which one is true. They are asked whether they believe that the first message is more likely to be true, the second message is more likely to be true, or believe that both messages are equally likely to be true.[40]

Senders are asked to choose one of the two computer messages on each question. Specifically, on each question senders are matched with two receivers, one Democrat and one Republican, who have the same median belief. Senders are told the true answer and then asked to choose either: "the answer is **greater than** [receiver's median]" or "the answer is **less than** [receiver's median]."

Treatments

The main treatment arms for senders are similar to those in Experiment 1.

1. **Topic arm:** The topics are varied within subject. They are either political or neutral.
2. **Information arm:** Information about the receiver's party is varied within subject. Either senders know the receiver's party or they do not. In addition, senders are randomly matched with Democratic and Republican receivers (with the same median belief), so the matched receiver's political motives are randomly either aligned with the true message or aligned with the false message.
3. **Incentives arm:** The senders' incentives are randomly varied between subjects. The sender is either incentivized to choose a computer message that is assessed as truthful by the receiver or the sender is unincentivized.

This is a 2x3x2 design. Senders play the rounds in the order as described in Table 6, and see questions in a random order.

After senders play six rounds in which they choose messages, they play four rounds in which they choose whether to "purchase information" about the receiver by conditioning their message on the receiver's political party. In these rounds, senders see the receiver's question and are asked to choose one of the following two options: (1) Be able to condition their message choice on the receivers' party with probability 1/2 and receive a slightly-higher payoff, or (2) Be able to condition their message choice on the receivers' party with probability 1 and receive a slightly-lower payoff. They are asked the information-purchasing questions after the main treatment block in order to enable them to have a chance to determine how much they value party information.

[40]This setup has two differences from the setup in Thaler (2021a), both of which serve to simplify the environment. First, in this experiment I elicit beliefs about the relative likelihood of the two messages instead of beliefs about only one of the two messages. Second, in this experiment I only ask receivers to choose which message is more truthful instead of the probability that a given message is truthful.

Round	Senders' Block
1	Sample question in role of Receiver
2-7	Choose messages
8-11	Demand for information
12	Attention check
End of experiment	Belief elicitation
End of experiment	Demographics and solutions

Table 6: The timing of treatments for senders in Experiment 2.

In two of the four rounds, senders can condition on the receiver's party regardless of their choice; in the other two rounds, they can condition on the receiver's party if and only if they have purchased the information.

Incentives

Subjects earn points in the experiment; points translate into earnings using the same binarized scoring rule as in Section 3. The only difference is that this experiment has five winners instead of ten.

Receivers' median beliefs are incentivized by a linear scoring rule. For each question, they give a guess μ about the answer. They earn $\max\{100 - |\mu - \text{answer}|, 0\}$ for their guess. Receivers' news assessments are incentivized by a simple concave scoring rule. Receivers who guess that Message X is more likely to be true than Message Y earn 100 points if they are correct and 0 points if they are incorrect; receivers who guess that both messages are equally likely to be true earn 55 points regardless of the true answer. They maximize points by guessing that X is true iff they believe P(X true) ≥ 0.55, by guessing that Y is true iff they believe P(X true) ≤ 0.45, and by guessing that they are equally likely iff P(X true) $\in [0.45, 0.55]$. A Bayesian, whose belief remains at 1/2, would guess that they are equally likely. Systematic differences in news source ratings are attributed here to motivated reasoning.

Senders in the incentivized condition are incentivized to choose the one of the two messages that the receiver was more likely to think is true. They earn 100 points if the receiver guesses their message is true, 50 points if the receiver guesses both messages are equally likely, and 0 points if the receiver guesses the other message is true. All senders are given these incentives in the demand-for-information rounds. Subjects in both treatments are in-

centivized to give accurate answers to the beliefs questions (whose answers are between 0 and 100). They are incentivized using a quadratic scoring rule. If they guess g and the correct answer is c, they earn $\max(0, 100 - (c - g)^2)$ points.

Comparing the two experimental designs

The main difference in the experiments is that, while in Experiment 1 the sender and receiver both affect each others' payoffs, in Experiment 2 the sender does not impact the receiver. Experiment 1 is able to identify the role of higher-order beliefs and receivers' beliefs about senders more broadly. There are a few additional differences. While Experiment 1 uses fixed target values, Experiment 2 uses median beliefs. As such, while Experiment 1 elicits senders' beliefs about receivers' belief updating, Experiment 2 elicits senders' beliefs about receivers' motivated reasoning. Only Experiment 1 studies receivers' beliefs about senders. More subtly, Experiment 1 varies, within topic, the effect of answers being too Dem or too Rep, while Experiment 2 only does this between topics. Lastly, only Experiment 2 looks at senders' demand for information about receivers.

Given the relative contributions of the two studies, the emphasis in the analysis of Experiment 2 is on senders' behavior. For a deeper discussion on receivers in a similar context, and what the requirements are for this design to be able to identify motivated reasoning, see Thaler (2021a), which uses an expanded version of this design with a sample of approximately 1,000 receivers. What is important to know for this paper is that, assuming that receivers report their true median beliefs, a Bayesian would always say that the two messages were equally likely to be true.

4.2 Data

550 subjects were recruited from Prolific (prolific.co) in May-June 2021 and passed a simple attention check. The subject pool included the general United States population, and the pool was restricted to subjects who had prior experience on the platform.[41] To emphasize that the focus is on sender behavior, the sample consists of 500 senders and 50 receivers. Receivers participated first. After receivers took the experiment, on each question I chose a median belief that both a Democratic receiver and a Republican receiver stated.[42] Half of median beliefs were too far in the Democratic direction, and the other half were too far in the Republican direction. These were the median beliefs that were presented

[41] All subjects needed to have completed 100 prior studies and have at least a 90-percent approval rating.

[42] On each question, such a median belief existed. I would not have included a question if there was no median that both a Democratic and a Republican receiver had stated.

to senders. See Appendix D for the list of topics, median beliefs, and truthful computer messages.

48 receivers (96 percent) and 492 senders (98 percent) stated a preference or lean for one party versus the other.[43] Of the receivers, 26 (52 percent) were Republicans and 22 (44 percent) were Democrats. Of the senders, 244 (49 percent) were Republicans and 248 (50 percent) were Democrats. Analyses are restricted to receivers with a party preference; these receivers made 383 news assessments on political topics. In the total sample of senders, there were 2,999 messages chosen.[44] Senders are split into the incentives treatment and the unincentivized treatment, with 245 (49 percent) being incentivized. In Appendix Table 17, I show the balance table for the incentives treatment among senders. I find modest political differences, but no other substantial differences, by treatment.

4.3 Main results

Senders in Experiment 2 choose computer messages in a similar manner to how senders in Experiment 1 choose messages to send to receivers. The nearly-exact replication of these results suggests that beliefs about motivated reasoning are an important determinant in understanding the results in the full sender-receiver game.

Table 7 tests the impact of the receiver's party on the truthfulness of senders' choices. The specification is identical to the specification in Appendix Table 11. It is similar to Table 3, but there are no receiver priors to control for since senders see the same median belief regardless of the receiver's party.

Table 7 shows that the effects of party-false alignment are qualitatively identical to the effects in Experiment 1; party-false alignment leads to more false computer messages chosen in each comparison group.

[43] Unlike in Experiment 1, the party question included an option to select "Independent (do not lean towards either party)."

[44] The targeted sample was 3,000; one subject did not choose a message on one question.

Table 7: Factors that lead incentivized senders to choose false computer messages

	Vs. Party-True Aligned	Vs. No Info	Vs. Neutral Topics	Vs. All Others
Party-False Aligned	0.145***	0.098**	0.090**	0.100***
	(0.045)	(0.041)	(0.041)	(0.033)
Question FE	✓	✓		
Subject FE	✓	✓	✓	✓
Round FE	✓	✓	✓	✓
Vs. Party-True Aligned	✓			✓
Vs. No Info		✓		✓
Vs. Neutral Topics			✓	✓
Observations	779	789	638	1470
Subjects	229	223	206	245
Mean	0.336	0.335	0.352	0.307

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Vs. lines indicate the comparison group. Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message. Prior-False Aligned: the receiver's prior belief that the incorrect answer is true. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. No Info: indicator for the receiver's party not being revealed.

Appendix Table 18 shows that, as in Appendix Table 12, unincentivized senders are not affected by these treatments. Subjects are not statistically-significantly more likely to send false computer messages when the messages are aligned with receiver's party in any comparison. As in Experiment 1, there is still a non-negligible share of unincentivized senders who choose false computer messages, at 24 percent. Appendix Table 19 shows that part of this effect is driven by a form of expressive preferences in which senders prefer to select the option of their own party. As in Table 5, this effect plays a significant role in the unincentivized condition and does not play a significant role in the incentivized condition.

The overall share of false messages chosen is modestly larger in both conditions in Experiment 2 compared to Experiment 1, suggesting that the strategic element present in Experiment 1 tampers false news. However, the false message rate in both experiments is substantially below 1/2, indicating that senders still have a preference for choosing truthful messages, even though in Experiment 2 the messages are not sent to receivers.

Table 8 shows that, as in Table 4, the incentives treatment causes senders to choose more false computer messages. This effect is again largely driven by the condition in which party and false messages are aligned.

Table 8: The effect of incentives on choosing false computer messages

	(1)	(2)	(3)	(4)	(5)	(6)
Incentivized	0.081***	0.078***	0.172***	0.167***	0.037	0.035
	(0.023)	(0.022)	(0.036)	(0.036)	(0.025)	(0.025)
Party-False Aligned					0.136***	0.135***
x Incentivized					(0.032)	(0.032)
Party-False Aligned					0.006	0.006
x Unincentivized					(0.025)	(0.025)
Question FE	✓	✓	✓	✓	✓	✓
Round FE	✓	✓	✓	✓	✓	✓
Subject controls		✓		✓		✓
All Questions	✓	✓			✓	✓
Only Party-False Aligned			✓	✓		
Observations	2421	2421	822	822	2421	2421
Subjects	500	500	429	429	500	500
Mean	0.274	0.274	0.318	0.318	0.274	0.274

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Subject controls: Gender, race, age, own party, education, and CRT score. Own party takes 0.5 if subject does not lean towards either party.

4.4 Receivers' behavior and senders' beliefs

Receivers assess pro-party news to be more likely than anti-party news to be true 38.4 percent of the time (subject-level clustered s.e. 2.5 pp), less likely to be true 29.2 percent of the time (s.e. 2.3 pp), and equally likely to be true 32.4 percent of the time (s.e. 2.4 pp). The difference between pro-party and anti-party news is statistically significant, and the point estimate is similar to that of the substantially-larger sample in Thaler (2021a) (9.1 pp; s.e. 4.2 pp; $p = 0.036$).

Senders are asked to predict each of these three percentages. They predict that receivers will assess pro-party news to be more likely to be true true 42.3 percent of the time (s.e. 0.8 pp), less likely to be true 29.8 percent of the time (s.e. 0.7 pp), and equally likely 27.9 percent of the time (s.e. 0.7 pp). The difference between pro-party and anti-party news is statistically significant (12.5 pp; s.e. 1.4 pp; $p < 0.001$), and the point estimate is suggestively larger than that of receivers' behavior. The results are qualitatively similar, but not as stark, as those in Experiment 1.[45]

[45]The correlation with the incentives treatment effect is positive but statistically insignificant ($p = 0.347$).

4.5 Demand for information

The majority of senders choose to pay to condition on the party of the receivers on each of the political questions. There are no sizeable differences between topics; on every political topic, between 55.6 percent and 63.1 percent of senders choose to pay. Since there is no effect on other parts of the experiment, this result indicates that senders value this information for instrumental reasons. Meanwhile, less than half of senders choose to pay on either of the neutral topics, indicating that senders particularly value this information on political topics. The gap in the share demanding information between political and neutral topics is 17.7 pp (s.e. 2.7 pp; $p < 0.001$).

Senders' information choices are predictive of their behavior. Recall that in two of the four rounds, senders are allowed to condition on the receiver's party regardless of their choice to the demand-for-information question. Comparing behavior from high-demand and low-demand senders, there is a correlation between demand for information and willingness to choose false computer messages. Senders who demand the information choose false messages 40.6 percent of the time (s.e. 1.2 pp), and senders who do not demand the information choose false messages 30.1 percent of the time (s.e. 1.7 pp). The difference is large and statistically significant (10.5 pp; s.e. 2.0 pp; $p < 0.001$). Suggestively, senders who do not demand information on political topics choose false messages a similar amount to subjects who send messages on neutral topics (30.8 percent).

In addition to these correlations, there is causal evidence from the main treatment block that speaks to this relationship. Incentivized senders who randomly receive the receiver's party information are 6.5 pp more likely to choose false messages (s.e. 2.9 pp; $p = 0.026$). These results suggest that senders causally condition on the party of their receiver to strategically choose more of the false computer messages.

5 Conclusion

Understanding the root causes of disinformation is critical in determining how best to combat it. This paper demonstrates that in settings that evoke motivated beliefs, and on political issues in particular, incentivizing senders to be perceived as truthful can lead to greater disinformation. Incentivized senders strategically distort messages both in order to appeal to receivers' current beliefs and to appeal to receivers' politically-motivated beliefs.

There are a number of potential avenues for future work. The experimental designs provided in this paper can be portable across domains, allowing applied researchers to test

Differences may be due to senders' different perceptions about the two tasks, different behavior from receivers, or elements of both.

whether senders believe that receivers motivatedly reason on any issue, and then to test whether senders asymmetrically send receivers false information. They can also be used to study how these effects play out in field settings like on social media.

Lastly, these results suggest two possible levers for reducing disinformation in political settings: either change the structure of incentives for news suppliers or change the perceived impact that motivated reasoning plays. While incentives are fixed in many news environments, this paper suggests that the latter may be malleable. Treatments that debias receivers in a way that changes senders' higher-order beliefs may reduce both trust in, and therefore the supply of, false information.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2019). "Preferences for Truth-telling." In: *Econometrica*.

Allcott, Hunt and Matthew Gentzkow (2017). "Social Media and Fake News in the 2016 Election." In: *Journal of Economic Perspectives*.

Azrieli, Yaron, Christopher Chambers, and Paul Healy (2018). "Incentives in Experiments: A Theoretical Analysis." In: *Journal of Political Economy*.

Barron, Kai (2019). "Lying to appear honest." In: *Working Paper*.

Barron, Kai (2020). "Belief updating: does the 'good-news, bad-news' asymmetry extend to purely financial domains?" In: *Experimental Economics*.

Bénabou, Roland (2013). "Groupthink: Collective Delusions in Organizations and Markets." In: *The Review of Economic Studies*.

Bénabou, Roland and Jean Tirole (2002). "Self-Confidence and Personal Motivation." In: *Quarterly Journal of Economics*.

Bénabou, Roland and Jean Tirole (2011). "Identity, Morals, and Taboos: Beliefs as Assets." In: *Quarterly Journal of Economics*.

Bénabou, Roland and Jean Tirole (2016). "Mindful Economics: The Production, Consumption, and Value of Beliefs." In: *Journal of Economic Perspectives*.

Benjamin, Daniel (2019). "Errors in Probabilistic Reasoning and Judgment Biases." In: *Chapter for the Handbook of Behavioral Economics*.

Brandts, Jordi and Gary Charness (2011). "The strategy versus the direct-response method: a first survey of experimental comparisons." In: *Experimental Economics*.

Brownback, Andy, Nathaniel Burke, and Tristan Gagnon-Bartsch (2021). "Inference from Biased Polls." In: *Working Paper*.

Brunnermeier, Markus and Jonathan Parker (2005). "Optimal expectations." In: *The American Economic Review*.

Buser, Thomas, Leonie Gerhards, and Joël van der Weele (2018). "Responsiveness to feedback as a personal trait." In: *Journal of Risk and Uncertainty*.

Cai, Hongbin and Joseph Tao-Yi Wong (2006). "Overcommunication in strategic information transmission games." In: *Games and Economic Behavior*.

Canes-Wrone, Brandice, Michael Herron, and Kenneth Shotts (2001). "Leadership and Pandering: A Theory of Executive Policymaking." In: *American Journal of Political Science*.

Charness, Gary and Chetan Dave (2017). "Confirmation bias with motivated beliefs." In: *Games and Economic Behavior*.

Che, Yeon-Koo, Wouter Dessein, and Navin Kartik (2013). "Pandering to Persuade." In: *American Economic Review*.

Choshen-Hillel, Shoham, Alex Shaw, and Eugene Caruso (2020). "Lying to appear honest." In: *Journal of Experimental Psychology*.

Coutts, Alexander (2018). "Good news and bad news are still news: Experimental evidence on belief updating." In: *Experimental Economics*.

Dwoskin, Elizabeth (2018). "Facebook is rating the trustworthiness of its users on a scale from zero to 1." In: *The Washington Post*. URL: https://www.washingtonpost.com/technology/2018/08/21/facebook-is-rating-trustworthiness-its-users-scale-zero-one/.

Eil, David and Justin Rao (2011). "The good news-bad news effect: asymmetric processing of objective information about yourself." In: *American Economic Journal: Microeconomics*.

Erat, Sanjiv and Uri Gneezy (2012). "White Lies." In: *Management Science*.

Ertac, Seda (2011). "Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback." In: *Journal of Economic Behavior and Organization*.

Farrell, Joseph and Matthew Rabin (1996). "Cheap Talk." In: *Journal of Economic Perspectives*.

Fedyk, Anastassia (2021). "Asymmetric Naivete: Beliefs About Self-Control." In: *Working Paper*.

Frechette, Guillaume, Alessandro Lizzeri, and Jacopo Perego (2021). "Rules and commitment in communication: An experimental analysis." In: *Working Paper*.

Frederick, Shane (2005). "Cognitive Reflection and Decision Making." In: *Journal of Economic Perspectives*.

Gagnon-Bartsch, Tristan (2021). "Taste Projection in Models of Social Learning." In: *Working Paper*.

Gentzkow, Matthew and Jesse Shapiro (2006). "Media Bias and Reputation." In: *Journal of Political Economy*.

Gentzkow, Matthew and Jesse Shapiro (2010). "What Drives Media Slant? Evidence from US Daily Newspapers." In: *Econometrica*.

Gneezy, Uri, Agne Kajackaite, and Joel Sobel (2018). "Lying Aversion and the Size of the Lie." In: *The American Economic Review*.

Gupta, Neeraja, Luca Rigotti, and Alistair Wilson (2021). "The Experimenters' Dilemma: Inferential Preferences over Populations." In: *Working Paper*.

Hagmann, David and George Loewenstein (2018). "Persuasion with motivated beliefs." In: *Working Paper*.

Hossain, Tanjim and Ryo Okui (2013). "The binarized scoring rule." In: *The Review of Economic Studies*.

Kahan, Dan (2016). "The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It." In: *Emerging Trends in Social and Behavioral Sciences*.

Kuhnen, Camelia (2014). "Asymmetric Learning from Financial Information." In: *The Journal of Finance*.

Kunda, Ziva (1990). "The case for motivated reasoning." In: *Psychological Bulletin*.

Lazer, David, Matthew Baum, Yochai Benkler, Adam Berinsky, Kelly Greenhill, Filippo Menczer, Miriam Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven Sloman, Cass Sunstein Emily Thorson, Duncan Watts, and Jonathan Zittrain (2018). "The Science of Fake News." In: *Science*.

Levy, Raphael (2014). "Soothing Politics." In: *Journal of Public Economics*.

Little, Andrew (2021). "Directional motives and different priors are observationally equivalent." In: *Working Paper*.

Maskin, Eric and Jean Tirole (2004). "The Politician and the Judge: Accountability in Government." In: *American Economic Review*.

McGee, Dan (2021). "Emergence of Stereotypes Under Group Competition." In: *Working Paper*.

McKelvey, Richard and Thomas Palfrey (1995). "Quantal response equilibria for normal form games." In: *Games and Economic Behavior*.

McKelvey, Richard and Thomas Palfrey (1998). "Quantal response equilibria for extensive form games." In: *Experimental Economics*.

Mobius, Markus, Muriel Niederle, Paul Niehaus, and Tanya Rosenblat (2014). "Managing self-confidence: Theory and experimental evidence." In: *Working Paper*.

Morris, Stephen (2001). "Political Correctness." In: *Journal of political Economy*.

Mullainathan, Sendhil and Andrei Shleifer (2005). "The Market for News." In: *American Economic Review*.

Oliver, J. Eric and Thomas Wood (2014). "Medical Conspiracy Theories and Health Behaviors in the United States." In: *JAMA Internal Medicine*.

Oprea, Ryan and Sevgi Yuksel (2020). "Social exchange of motivated beliefs." In: *Journal of the European Economic Association*.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey (2013). "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." In: *American Economic Review*.

Pennycook, Gordon and David Rand (2021). "Examining false beliefs about voter fraud in the wake of the 2020 Presidential Election." In: *The Harvard Kennedy School Misinformation Review*.

Peterson, Erik and Shanto Iyengar (2021). "Partisan Gaps in Political Information andInformation-Seeking Behavior: Mo tivated Reasoningor Cheerleading?" In: *American Journal of Political Science*.

Pronin, Emily (2007). "Perception and misperception of bias in human judgment." In: *Trends in Cognitive Sciences*.

Pronin, Emily, Thomas Gilovich, and Lee Ross (2004). "Objectivity in the eye of the beholder: divergent perceptions of bias in self versus others." In: *Psychological Review*.

Pronin, Emily, Daniel Lin, and Lee Ross (2002). "The Bias Blind Spot: Perceptions of Bias in Self Versus Others." In: *Personality and Social Psychology Bulletin*.

Serra-Garcia, Marta, Eric van Damme, and Jan Potters (2011). "Hiding an inconvenient truth: Lies and vagueness." In: *Games and Economic Behavior*.

Serra-Garcia, Marta and Uri Gneezy (2021). "Mistakes, Overconfidence, and the Effect of Sharing on Detecting Lies." In: *Working Paper*.

Shalvi, Shaul, Ivan Soraperra, Joel van der Weele, and Marie Claire Villeval (2019). "Shooting the Messenger? Supply and Demand in Markets for Willful Ignorance." In: *Working Paper*.

Tappin, Ben, Gordon Pennycook, and David Rand (2020a). "Bayesian or biased? Analytic thinking and political belief updating." In: *Cognition*.

Tappin, Ben, Gordon Pennycook, and David Rand (2020b). "Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often undermine causal inference." In: *Current Opinion in Behavioral Science*.

Thaler, Michael (2020). "Do People Engage in Motivated Reasoning to Think the World Is a Good Place for Others?" In: *Working Paper*.

Thaler, Michael (2021a). "The Fake News Effect: Experimentally Identifying Motivated Reasoning Using Trust in News." In: *Working Paper*.

Thaler, Michael (2021b). "Gender Differences in Motivated Reasoning." In: *Journal of Economic Behavior and Organization*.

A Additional Tables and Figures for Experiment 1

Table 9: Balance Table for Senders

	Incentivized	Unincentivized	Inc. vs. Uninc.	p-value
Age	35.431	33.227	2.165	0.148
	(1.074)	(1.038)	(1.490)	
White	0.766	0.745	0.022	0.694
	(0.038)	(0.039)	(0.055)	
Female	0.516	0.624	-0.108	0.087
	(0.045)	(0.044)	(0.063)	
Education	15.064	14.849	0.215	0.402
	(0.177)	(0.186)	(0.257)	
CRT score	1.387	1.554	-0.167	0.262
	(0.106)	(0.104)	(0.148)	
Party	0.508	0.512	-0.004	0.944
	(0.045)	(0.045)	(0.063)	
Other's party revealed	0.693	0.657	0.036	0.108
	(0.016)	(0.016)	(0.022)	
Own party revealed	0.654	0.672	-0.019	0.754
	(0.043)	(0.042)	(0.060)	
Others' have	0.500	0.516	-0.156	0.584
party-truth aligned	(0.022)	(0.018)	(0.028)	
Self has	0.512	0.526	-0.015	0.344
party-truth aligned	(0.011)	(0.011)	(0.016)	
N	10,036	10,107	20,143	

Notes: Standard errors in parentheses. Education is in years. CRT score is number of correct answers on the cognitive reflection task. Party is 1 if subject is Republican or Republican-leaning. Other party known only pertains to Rounds 1-7 (in later rounds the party is always revealed). Party-truth alignment is defined in the main text. Party-truth alignment restricted to observations where party is known.

Table 10: Balance Table for Receivers

	Incentivized	Unincentivized	Inc. vs. Uninc.	p-value
Age	34.096	33.635	0.461	0.772
	(1.172)	(1.075)	(1.587)	
White	0.718	0.729	-0.012	0.837
	(0.041)	(0.039)	(0.056)	
Female	0.572	0.558	0.014	0.823
	(0.045)	(0.044)	(0.062)	
Education	15.072	14.828	0.244	0.327
	(0.173)	(0.179)	(0.248)	
CRT score	1.354	1.296	0.058	0.679
	(0.098)	(0.102)	(0.141)	
Party	0.500	0.512	-0.012	0.849
	(0.045)	(0.044)	(0.063)	
Other's party revealed	0.681	0.650	0.019	0.206
	(0.018)	(0.016)	(0.024)	
Own party revealed	0.670	0.682	-0.012	0.834
	(0.042)	(0.041)	(0.059)	
N	2,726	2,832	5,558	

Notes: Standard errors in parentheses. Education is in years. CRT score is number of correct answers on the cognitive reflection task. Party is 1 if subject is Republican or Republican-leaning. Other party known only pertains to Rounds 1-7 (in later rounds the party is always revealed).

Table 11: Factors that lead incentivized senders to send false messages: No controls for prior

	Vs. Party-True Aligned	Vs. No Info	Vs. Neutral Topics	Vs. All Others
Party-False Aligned	0.092***	0.075***	0.051*	0.067***
	(0.028)	(0.023)	(0.028)	(0.020)
Question FE	✓	✓		
Subject FE	✓	✓	✓	✓
Round FE	✓	✓	✓	✓
Vs. Party-True Aligned	✓			✓
Vs. No Info		✓		✓
Vs. Neutral Topics			✓	✓
Observations	5486	4963	4313	9278
Subjects	124	124	124	124
Mean	0.296	0.303	0.327	0.292

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Vs. lines indicate the comparison group. Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. No Info: indicator for the receiver's party not being revealed.

Table 12: Factors that lead unincentivized senders to send false messages

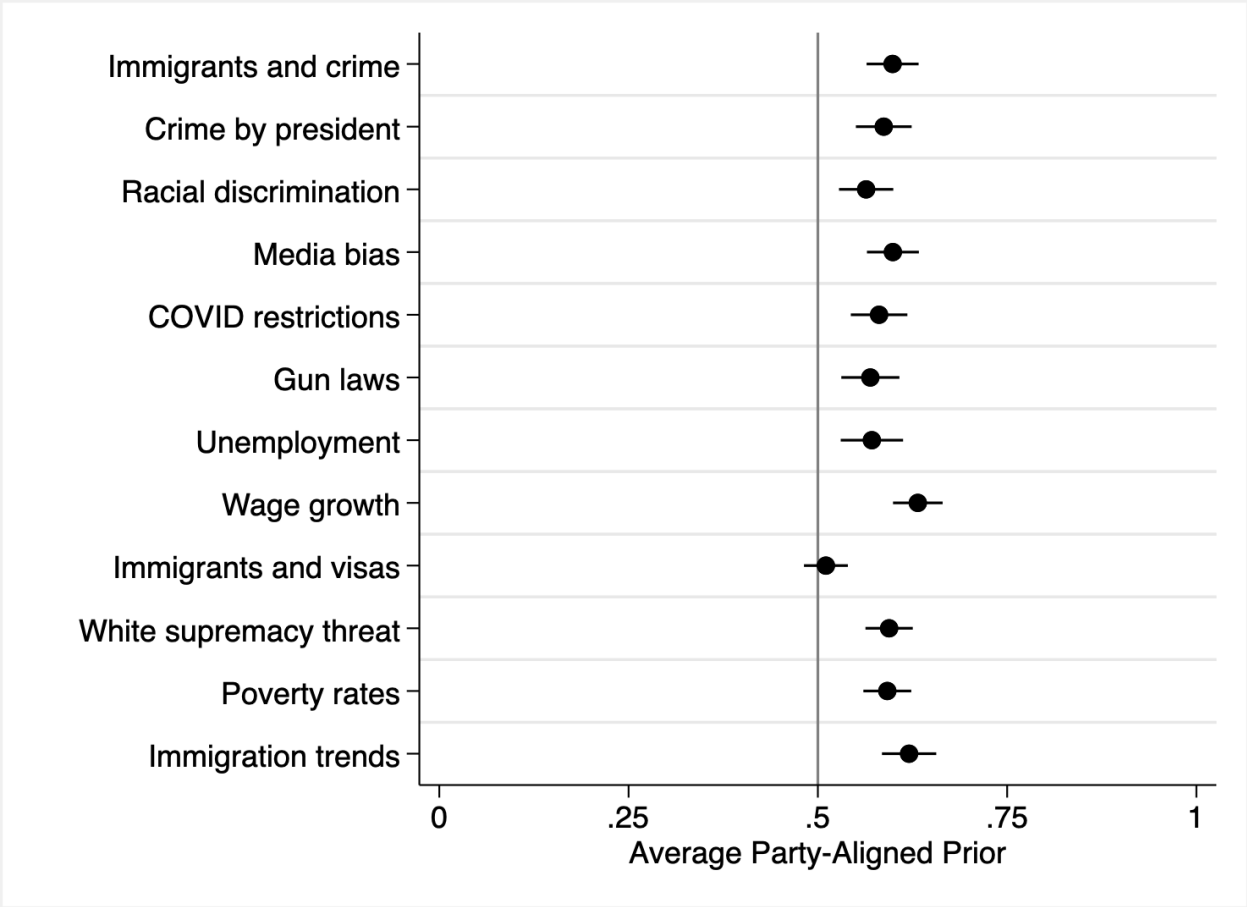
	Vs. Party-True Aligned	Vs. No Info	Vs. Neutral Topics	Vs. All Others
Party-False Aligned	0.005	-0.038	-0.011	-0.006
	(0.032)	(0.030)	(0.024)	(0.023)
Prior-False Aligned	0.063	0.074*	0.071	0.054
	(0.039)	(0.040)	(0.045)	(0.034)
Question FE	✓	✓		
Subject FE	✓	✓	✓	✓
Round FE	✓	✓	✓	✓
Vs. Party-True Aligned	✓			✓
Vs. No Info		✓		✓
Vs. Neutral Topics			✓	✓
Observations	4636	4619	3920	8739
Subjects	125	125	125	125
Mean	0.211	0.215	0.203	0.209

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

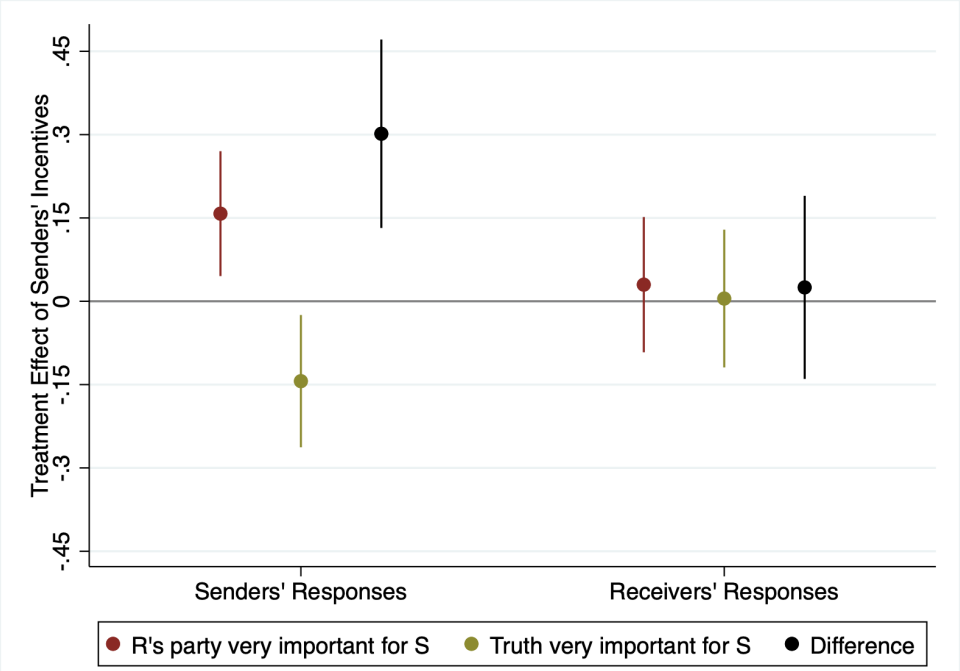
Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Vs. lines indicate the comparison group. Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message. Prior-False Aligned: the receiver's prior belief that the incorrect answer is true. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. No Info: indicator for the receiver's party not being revealed.

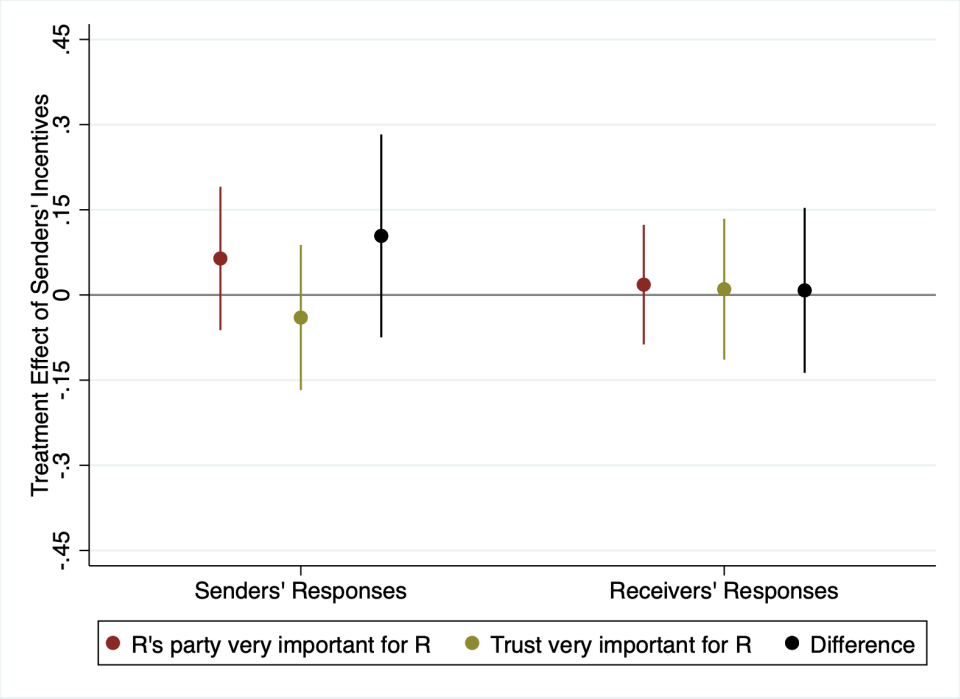
Figure 5: Receivers' Prior Beliefs by Topic



Notes: Standard errors clustered at subject level. Party-Aligned Prior denotes the prior belief that the receiver has that the pro-party state is true, as described in Table 1. Error bars correspond to 95 percent confidence intervals.

Figure 6: The Effect of Incentives on Survey Beliefs about Senders and Receivers





Notes: OLS regression coefficients, robust standard errors. DV takes 1 if subject answers "very important" or "extremely important" and 0 otherwise. Exact questions are provided in the experimental materials. Controls for age, race, gender, education, CRT score, and own party are included. The top panel shows that incentivized Ss believe that R's party matters more, and the truth matters less, in their decisions, while Rs are unaffected by S's incentives. The bottom panel shows that incentives do not significantly affect senders' or receivers' beliefs about the impact of R's party and R's trust on R behavior. Error bars correspond to 95 percent confidence intervals.

Table 13: The interaction between beliefs and incentives

	(1)	(2)	(3)	(4)
Incentivized	0.021	-0.008	0.018	0.000
	(0.036)	(0.042)	(0.054)	(0.074)
S's Belief of R's Updating	-0.121**	-0.132**	-0.198**	-0.206*
	(0.059)	(0.063)	(0.088)	(0.107)
Incentivized x S's Belief	0.195**	0.239**	0.323**	0.447**
	(0.082)	(0.097)	(0.124)	(0.172)
Question FE	✓	✓	✓	✓
Round FE	✓	✓	✓	✓
Subject controls	✓	✓	✓	✓
All Questions	✓	✓		
Only Party-False Aligned			✓	✓
Only 50-50 Priors		✓		✓
Observations	14248	1294	4702	427
Subjects	249	249	220	220
Mean	0.250	0.216	0.250	0.216

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. S's belief: sender's belief about party gap in receiver assessments. Subject controls: Gender, race, age, own party, education, and CRT score. Only 50-50 Priors: only observations where the receiver's prior is 1/2. Only includes observations where the sender conditions on the receiver's prior.

Table 14: The effect of competition incentives on choosing false messages

	(1)	(2)	(3)	(4)	(5)	(6)
Competition	0.020	0.027	0.022	-0.005	0.018	0.030
Incentives	(0.028)	(0.037)	(0.036)	(0.051)	(0.032)	(0.039)
Party-False Aligned					0.003	-0.002
x Incentivized					(0.027)	(0.027)
Party-False Aligned					-0.003	0.007
x Unincentivized					(0.027)	(0.026)
Question FE	✓	✓	✓	✓	✓	✓
Round FE	✓	✓	✓	✓	✓	✓
Subject controls		✓		✓		✓
All Questions	✓	✓			✓	✓
Only Party-False Aligned			✓	✓		
Observations	14214	14214	4682	4682	14214	14214
Subjects	251	251	225	225	251	251
Mean	0.224	0.224	0.223	0.223	0.224	0.224

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Subject controls: Gender, race, age, own party, education, and CRT score. Only includes observations where the sender conditions on the receiver's prior.

Table 15: Factors that lead senders with competition incentives to send false messages

	Vs. Party-True Aligned	Vs. No Info	Vs. Neutral Topics	Vs. All Others
Party-False Aligned	0.012	0.019	0.040	0.011
	(0.028)	(0.028)	(0.026)	(0.024)
Prior-False Aligned	0.167***	0.202***	0.166***	0.172***
	(0.043)	(0.045)	(0.046)	(0.035)
Question FE	✓	✓		
Subject FE	✓	✓	✓	✓
Round FE	✓	✓	✓	✓
Vs. Party-True Aligned	✓			✓
Vs. No Info		✓		✓
Vs. Neutral Topics			✓	✓
Observations	4750	4891	4123	8836
Subjects	126	125	124	126
Mean	0.230	0.238	0.222	0.230

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Vs. lines indicate the comparison group. Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message. Prior-False Aligned: the receiver's prior belief that the incorrect answer is true. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. No Info: indicator for the receiver's party not being revealed.

Table 16: Factors that lead senders with competition incentives to send false messages: No controls for prior

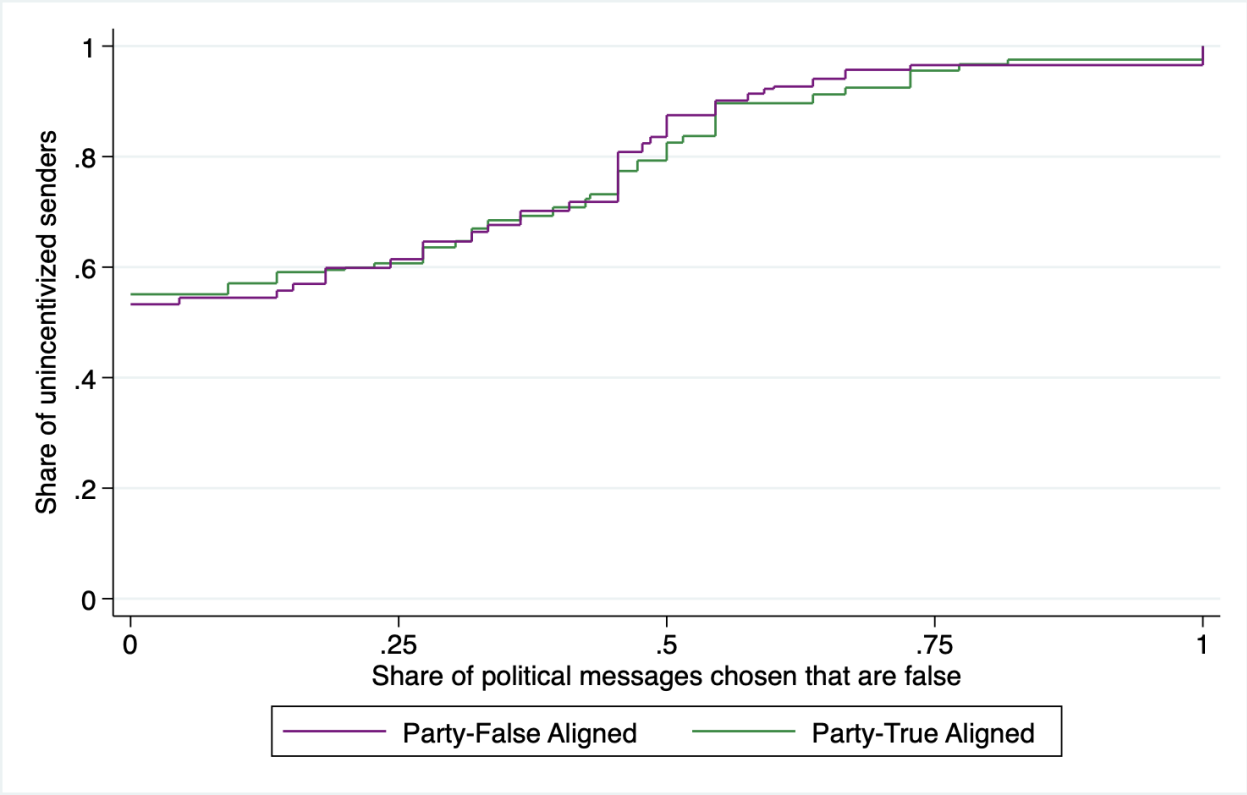
	Vs. Party-True Aligned	Vs. No Info	Vs. Neutral Topics	Vs. All Others
Party-False Aligned	0.024	0.017	0.041	0.017
	(0.026)	(0.028)	(0.026)	(0.023)
Question FE	✓	✓		
Subject FE	✓	✓	✓	✓
Round FE	✓	✓	✓	✓
Vs. Party-True Aligned	✓			✓
Vs. No Info		✓		✓
Vs. Neutral Topics			✓	✓
Observations	5254	5137	4368	9340
Subjects	126	126	125	126
Mean	0.228	0.239	0.224	0.228

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Vs. lines indicate the comparison group. Prior-False Aligned: the receiver's prior belief that the incorrect answer is true. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. No Info: indicator for the receiver's party not being revealed.

Figure 7: CDF of Individual-Level False News when Senders are Unincentivized



Notes: CDF plot of the average share of messages chosen by senders. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message. Data are restricted to political questions for which senders condition on receivers' priors. Senders are included if they are unincentivized and learn the party of the receiver.

B Additional Tables and Figures for Experiment 2

Table 17: Balance Table for Senders

	Incentivized	Unincentivized	Inc. vs. Uninc.	p-value
Age	38.706	38.344	0.362	0.766
	(0.878)	(0.839)	(1.213)	
White	0.767	0.788	-0.021	0.578
	(0.027)	(0.026)	(0.037)	
Female	0.469	0.545	-0.076	0.089
	(0.032)	(0.031)	(0.045)	
Education	15.261	15.243	0.019	0.920
	(0.132)	(0.130)	(0.185)	
CRT score	1.514	1.467	0.047	0.651
	(0.075)	(0.073)	(0.105)	
Party	0.551	0.443	0.108	0.015
	(0.032)	(0.031)	(0.044)	
Other's party revealed	0.659	0.664	-0.005	0.791
	(0.013)	(0.012)	(0.018)	
Others' have	0.492	0.488	0.005	0.856
party-truth aligned	(0.019)	(0.017)	(0.025)	
Self has	0.499	0.490	0.008	0.547
party-truth aligned	(0.009)	(0.010)	(0.014)	
N	1,470	1,529	2,999	

Notes: Standard errors in parentheses. Education is in years. CRT score is number of correct answers on the cognitive reflection task. Party is 1 if subject is Republican or Republican-leaning and 1/2 if subject is Independent (no lean). Party-truth alignment is defined in the main text. Party-truth alignment is restricted to observations where party is revealed.

Table 18: Factors that lead unincentivized senders to choose false computer messages

	Vs. Party-True Aligned	Vs. No Info	Vs. Neutral Topics	Vs. All Others
Party-False Aligned	-0.022	0.025	-0.065*	-0.009
	(0.028)	(0.034)	(0.033)	(0.024)
Question FE	✓	✓		
Subject FE	✓	✓	✓	✓
Round FE	✓	✓	✓	✓
Vs. Party-True Aligned	✓			✓
Vs. No Info		✓		✓
Vs. Neutral Topics			✓	✓
Observations	803	812	695	1529
Subjects	235	243	224	255
Mean	0.241	0.230	0.253	0.243

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Vs. lines indicate the comparison group. Party-False Aligned: indicator for the receiver's party being revealed and aligned with the false message. Prior-False Aligned: the receiver's prior belief that the incorrect answer is true. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message. No Info: indicator for the receiver's party not being revealed.

Table 19: The effect of own party on messages

	(1)	(2)	(3)
Own Party-False Aligned	0.070*	0.190***	0.189***
	(0.039)	(0.029)	(0.029)
Other's Party-False Aligned	0.155***	-0.023	-0.017
	(0.045)	(0.027)	(0.028)
Own Party-False Aligned x Incentivized			-0.108**
			(0.049)
Other's Party-False Aligned x Incentivized			0.165***
			(0.053)
Question FE	✓	✓	✓
Subject FE	✓	✓	✓
Round FE	✓	✓	✓
Incentivized subjects	✓		✓
Unincentivized subjects		✓	✓
Observations	763	790	1553
Subjects	225	231	456
Mean	0.339	0.234	0.286

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: OLS, errors clustered at subject level. Dependent variable: indicator for sender choosing the false message. Prior-False Aligned: the receiver's prior belief that the incorrect answer is true. Party-True Aligned: indicator for the receiver's party being revealed and aligned with the true message.

C Study Materials for Experiment 1

C.1 Question Wordings

Crime Under Trump

The Trump administration campaigned on tough-on-crime policies. Some people believe that the Trump administration's policies were effective at reducing violent crime, while others believe that his rhetoric provoked more violence.

This question asks how violent crime rates changed during the Trump administration. In 2016 (before Trump became president), the violent crime rate was 386.6 per 100,000 Americans.

In 2020 (at the end of Trump's presidency), do you think it is more likely that the violent crime rate was greater or less than [300 or 500] per 100,000 Americans?

Correct answer: 366.7 per 100,000

Source linked on results page: http://bit.ly/us-crime-rate

Undocumented Immigrants

The U.S. has seen a sharp rise in the share of undocumented immigrants over the past several years. Some people believe that undocumented immigrants are more likely to commit violent crime, while others believe that undocumented immigrants are less likely to commit violent crimes.

Texas is the only state that directly compares crime rates for US-born citizens to undocumented immigrants, and provided felony data from 2012-2018. During this time period, the felony violent crime rate was 213 per 100,000 U.S. citizens.

This question asks about the felony violent crime rate for undocumented immigrants. Do you think it is more likely that this rate was greater or less than [90 or 213] per 100,000?

Correct answer: 96.2 per 100,000

Source linked on results page: http://bit.ly/crime-by-immigrant-status

Racial Discrimination

In the United States, white Americans have higher salaries than black Americans on average. Some people attribute these differences in income to differences in education, training, and culture, while others attribute them more to racial discrimination.

In a study, researchers sent fictitious resumes to respond to thousands of help-wanted ads in newspapers. The resumes sent had identical skills and education, but the researchers gave half of the (fake) applicants stereotypically White names such as Emily Walsh and Greg Baker, and gave the other half of the applicants stereotypically Black names such as Lakisha Washington and Jamal Jones.

This question asks how the callback rates differed between White- and Black-sounding names. 9.65 percent of the applicants with White-sounding names received a call back. Do you think it is

more likely that the percent of the applicants with Black-sounding names who received a call back was greater or less than [5.0 or 8.5] percent?

Correct answer: 6.45 percent

Source linked on results page: http://bit.ly/labor-market-discrimination

Media Bias

Some people believe that the media is filled with Democrats and unfairly biased towards the Democratic Party, while some believe the media is more balanced, and others believe it is biased towards Republicans.

This question asks whether journalists are significantly more likely to be Democrats than Republicans.

A representative sample of journalists were asked about their party affiliation. Compared to the number of Republicans, do you think it is more likely that the number of journalists who said they were Democrats was greater or less than [2 or 5] times as much?

Correct answer: 4 times as much

Source linked on results page: http://bit.ly/journalist-political-affiliation

COVID-19 Restrictions

In the face of the coronavirus pandemic, some places mandated strict lockdowns, while other places allowed for more activity and opened up sooner. This question asks how effective lockdowns were at preventing the spread of the coronavirus.

A recent study estimated how cases would have changed during the early stages of the pandemic if all areas implemented stay-at-home orders on March 17, 2020.

This question asks about the percent reduction in cases by April 30, 2020 if all areas implemented stay-at-home orders on March 17, 2020. Do you think it is more likely that this reduction was greater or less than [10 or 50] percent?

Correct answer: 19.5 percent

Source linked on results page: http://bit.ly/covid-restrictions-effect

Gun Laws

The United States has a homicide rate that is much higher than other wealthy countries. Some people attribute this to the prevalence of guns and favor stricter gun laws, while others believe that stricter gun laws will limit Americans' Second Amendment rights without reducing homicides very much.

After a mass shooting in 1996, Australia passed a massive gun control law called the National Firearms Agreement (NFA). The law illegalized, bought back, and destroyed almost one million

firearms by 1997, mandated that all non-destroyed firearms be registered, and required a lengthy waiting period for firearm sales.

Democrats and Republicans have each pointed to the NFA as evidence for/against stricter gun laws. In the five years before the NFA (1991-1996), there were 320 homicides per year in Australia. In the five years after the NFA (1998-2003), do you think it is more likely that the average number of homicides in Australia was greater or less than [220 or 320] per year?

Correct answer: 318.6 per year

Source linked on results page: http://bit.ly/australia-homicide-rate and http://bit.ly/impact-australia-gun-laws.

Unemployment Rate

Some people believe that Donald Trump's policies improved the jobs situation in the United States, while others believe that his policies hindered employment.

This question asks whether the unemployment rate increased or decreased during the Trump administration as compared to the end of the Obama administration.

In the last two years of the Obama administration (Jan 2015-Jan 2017), the average unemployment rate was 5.1 percent. Do you think it is more likely that the average unemployment rate during the Trump administration was greater or less than [3.2 or 5.1] percent?

Correct answer: 5.04 percent

Source linked on results page: http://bit.ly/unemployment-rate-data

Wage Growth

Some people believe that the Trump administration did a better job at increasing wages for most Americans, and some people believe that the Obama administration did a better job of wage growth.

In the last two years of the Obama administration (Jan 2015-Jan 2017), the median growth in Americans' wages was 3.28 percent on average.

Do you think it is more likely that the average median growth in Americans' wages during the Trump administration was greater or less than [3.28 or 4] percent?

Correct answer: 3.49 percent

Source linked on results page: http://bit.ly/median-wage-growth

Center of the US

The U.S. National Geodetic Survey approximated the geographic center of the continental United States. (This excludes Alaska and Hawaii, and U.S. territories.)

This question asks how far North the U.S. is located. For reference, the continental U.S. lies in the Northern Hemisphere, the Equator is 0 degrees North, and the North Pole is 90 degrees North.

Do you think it is more likely that this geographic center is greater or less than [30 or 45] degrees North?

Correct answer: 39.833 degrees North

Source linked on results page: http://bit.ly/center-of-the-us

Random Number

A computer randomly generated a number between 0 and 100, decimals included. What number do you think the computer chose?

As a reminder, it is in your interest to guess an answer that is close to the computer's choice, even if you don't perfectly guess it.

Do you think it is more likely that this number is greater or less than [40 or 60]?

Correct answer: 33.54026

Performance on a CRT Task

Previously in this study, you were asked three quiz questions that some people use as a measure of cognitive ability.

At the end of the study, your score on this test will be compared to the scores among all participants. This question asks you to predict how your score compared to others.

Do you think it is more likely that your score was greater or less than the average score?

The average score was between 1 and 2, so subjects who scored 2 or 3 scored greater than the average, and subjects who scored 0 or 1 scored less than the average.

Quote from Biden: Visas and Immigrants

In 2021, Joe Biden said that there are "over 11 million undocumented folks – the vast majority are here overstaying visas."

Do you think this statement is accurate or inaccurate?

Correct answer: Inaccurate

Source linked on results page: https://bit.ly/undocumented-mostly-visas

Quote from Biden: White Supremacists

In 2020, Joe Biden said that "[Donald Trump's] FBI chief has said the greatest domestic threat to terrorism are white supremacists."

Do you think this statement is accurate or inaccurate?

Correct answer: Accurate

Source linked on results page: https://bit.ly/white-supremacists-threat

Quote from Trump: Poverty Rates

In 2018, Donald Trump said that "The poverty rates for African Americans and Hispanic Americans ... it's been incredible, they've all reached their lowest levels."

Do you think this statement is accurate or inaccurate?

Correct answer: Accurate

Source linked on results page: https://bit.ly/poverty-rates-black-hispanic

Quote from Trump: Illegal Immigration

In 2021, Donald Trump said that there has been "a massive flood of illegal immigration into our country, the likes of which we have never seen before."

Do you think this statement is accurate or inaccurate?

Correct answer: Inaccurate

Source linked on results page: https://bit.ly/record-illegal-immigration

Attention Check: Current Year

In 1776 our fathers brought forth, upon this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal. What is the year right now?

This is not a trick question and the first sentence is irrelevant; this is a check to make sure you are paying attention. If you get this question incorrect, you will not be eligible to receive a bonus payment.

Correct answer: 2021.

Source linked on results page: http://bit.ly/what-year-is-it

C.2 Screenshots

Figure 8: Overview

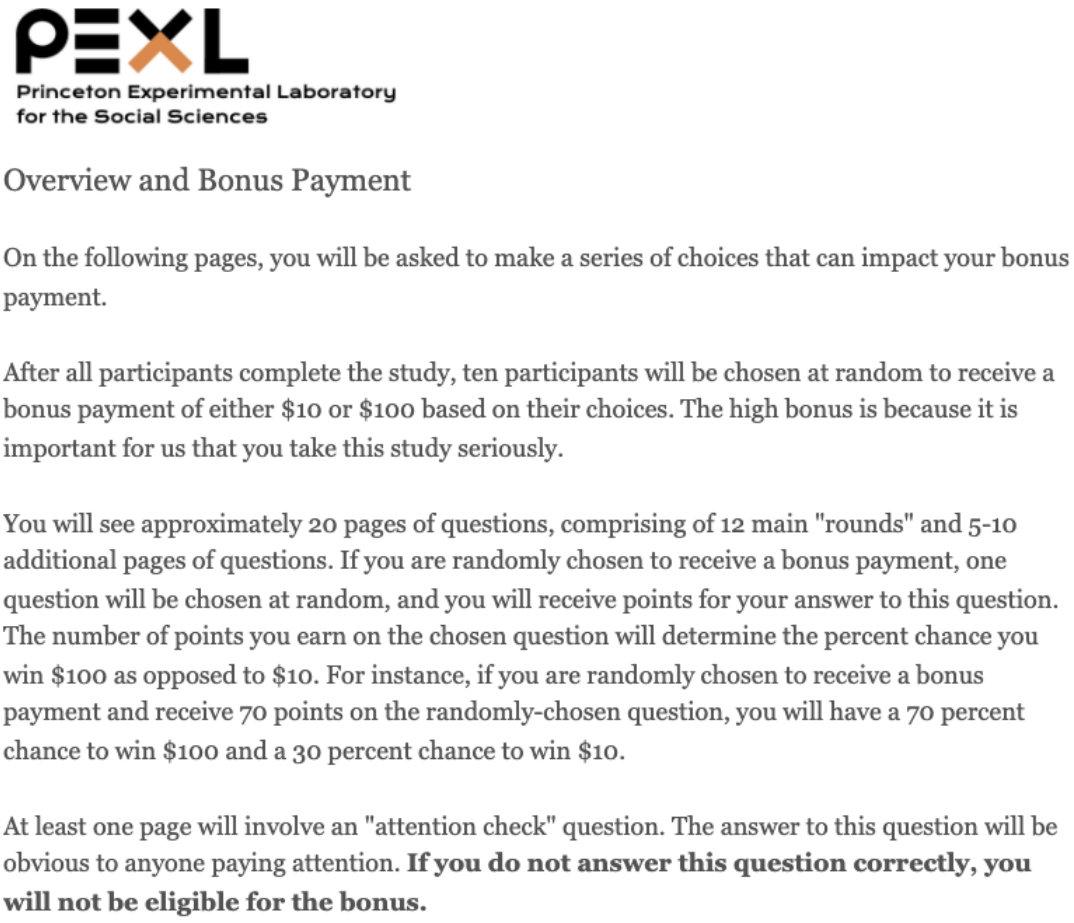


Figure 9: Demographics

Demographics

This section will ask you a few questions about yourself. Your earnings in the study are not affected by your answers to these questions.

What is your age?

What is your gender?

- Male
- Female
- Other / Nonbinary
- Prefer not to answer

In politics today, do you consider yourself a Republican, a Democrat, or an Independent?

- Strongly Democratic
- Weakly Democratic
- Independent (lean Democratic)
- Independent (lean Republican)
- Weakly Republican
- Strongly Republican

What is your highest level of education?

- ◯ Did not graduate high school
- ◯ High school graduate, diploma, or equivalent (such as GED)
- ◯ Began college, no degree
- ◯ Associate's degree
- ◯ Bachelor's degree
- ◯ Postgraduate or professional degree

What race/ethnicity best describes you?

- ◯ Black or African American
- ◯ American Indian
- ◯ Asian
- ◯ White
- ◯ Hispanic or Latino
- ◯ Two or more of these
- ◯ Other / Prefer not to answer

Figure 10: Cognitive reflection task

Quiz Question 1

A computer and a keyboard cost $350 in total. The computer costs $300 more than the keyboard.

How much does the computer cost?

- ◯ $325
- ◯ $25
- ◯ $50
- ◯ $300

Quiz Question 2

If it takes 7 machines 7 minutes to make 7 widgets, how long would it take 70 machines to make 70 widgets?

- ◯ 70 minutes
- ◯ 490 minutes
- ◯ 7 minutes
- ◯ 700 minutes

Quiz Question 3

In a community, there is a rapidly-spreading virus. Every day, the virus infects twice as many people. If it takes 42 days for the virus to infect the entire community, how long would it take the virus to infect half the community?

- ◯ 42 days
- ◯ 41 days
- ◯ 21 days
- ◯ 2 days

Figure 11: Overview for practice questions

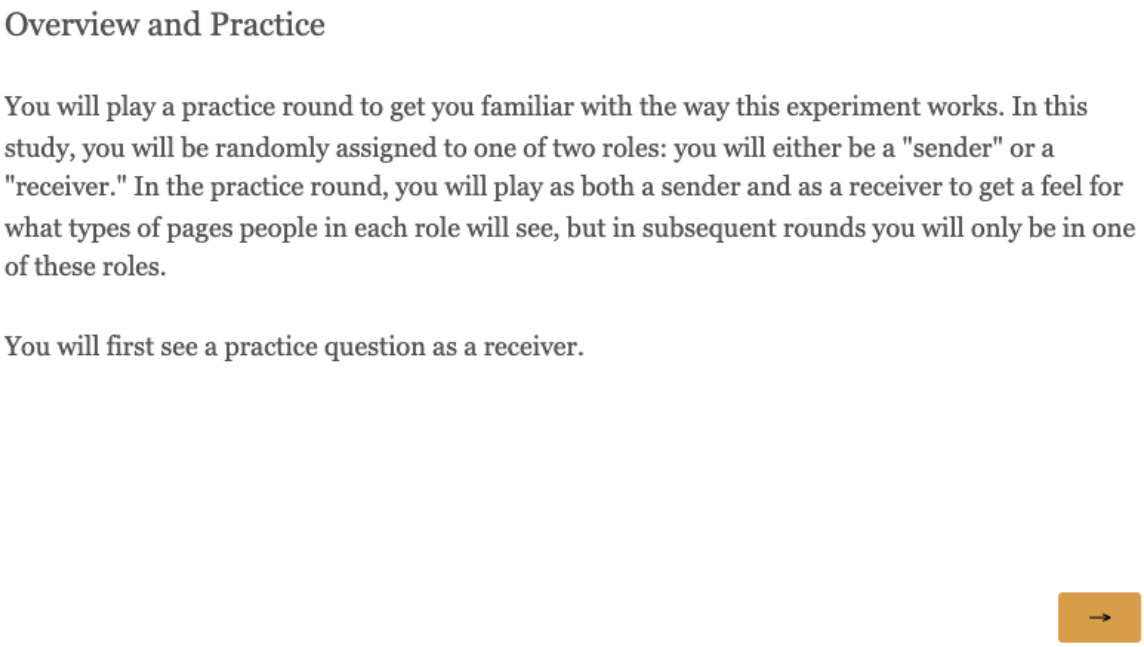


Figure 12: Instructions for receiver questions

Receiver: Instructions for Questions

Throughout this study, receivers will see several types of pages, including Question pages.

On each Question page, you will be given a factual question and be asked to guess whether the answer is more likely to be *greater than* or *less than* a given number; each question has a correct answer. If you are randomly selected to win a bonus and a Question page is chosen, **your answer to that question will determine the chance you win the additional prize.**

The details of the point system used to determine your chance of winning the high bonus are a bit complicated, but are explained below if you are interested. **What is important to know is that the way your earnings are determined ensures that your chances of winning the high bonus are maximized by carefully and honestly answering these questions.**

Regardless of your guesses, a list of correct answers will be provided at the end of the study.

You will see a practice question on the next page.

Points for your guess:

Your guess about the chance that **the answer is greater**	Points if answer is greater	Points if answer is less
100 percent	100	0
90 percent	99	19
80 percent	96	36
70 percent	91	51
60 percent	84	64
50 percent	75	75
40 percent	64	84
30 percent	51	91
20 percent	36	96
10 percent	19	99
0 percent	0	100

The practice question is omitted since it looks the same as questions in the main rounds.

Figure 13: Practice receiver question

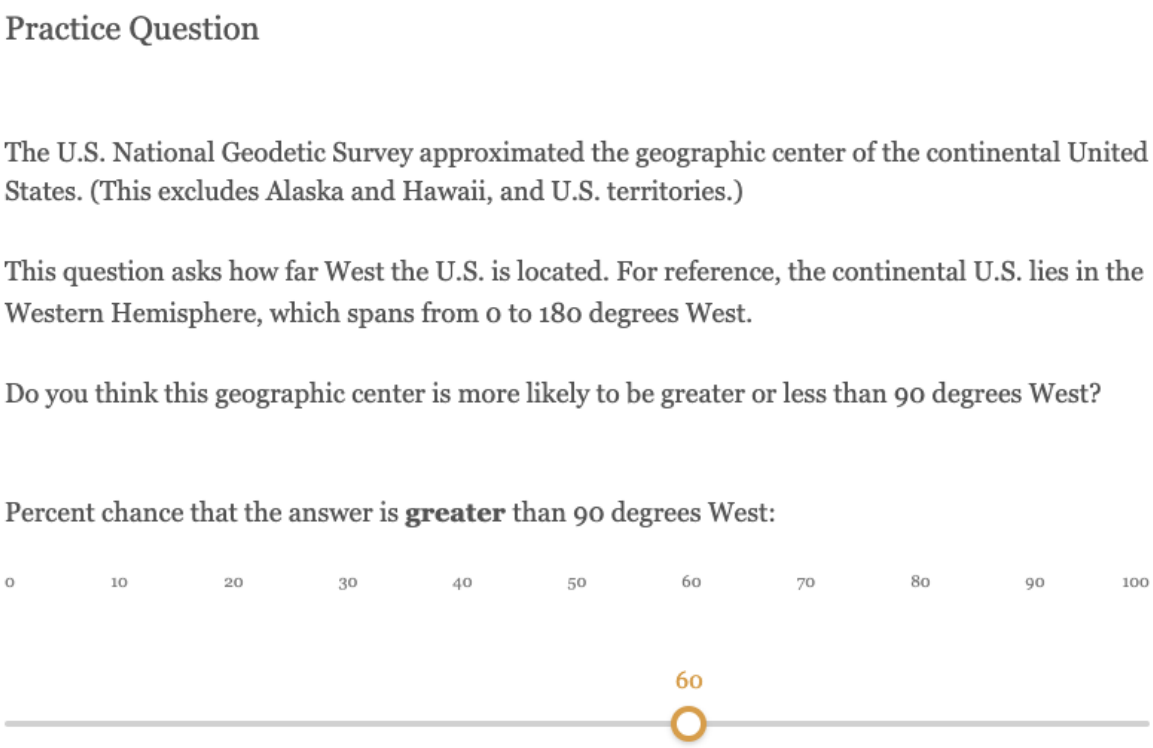


Figure 14: Instructions for sending messages: Incentivized treatment

Sender Instructions: Sending Messages

Now, you will play a practice round as a sender. In Sending Messages rounds, you will be asked to send one of two messages to another participant (the "receiver").

If this question is selected for payment, **the chance that you win the high bonus will depend on whether the other participant said that your chosen message was truthful or false.**

- If the receiver said your message was 100% likely to be truthful, you would earn 100 points.
- If the receiver said your message was 90% likely to be truthful, you would earn 90 points.
- If the receiver said your message was 80% likely to be truthful, you would earn 80 points.
- ...
- If the receiver said your message was 10% likely to be truthful, you would earn 10 points.
- If the receiver said your message was 0% likely to be truthful, you would earn 0 points.

You will maximize the chance you win the high bonus if you choose a message that the receiver thinks is truthful.

The receiver will rate both messages, regardless of your choice. A list of correct answers will be provided to you and all receivers at the end of the study.

You will be matched with different receivers in different rounds. You will be equally likely to match with a Democrat or a Republican (including leaners). You will sometimes be told whether the receiver is a Democrat or a Republican.

Subjects in the unincentivized treatment are instead told: *"**These questions will not affect your bonus payment.**"* and are not given the point system.

Subjects in the competition treatment are instead told that *"If this question is selected for payment, **the chance that you win the high bonus will depend on whether the receiver said that you were more likely to have sent more truthful messages over the course of the study.**"*

Figure 15: Practice sending messages: Incentivized treatment

Practice Sending Messages

The other participant ("the receiver") was asked the following question:

The U.S. National Geodetic Survey approximated the geographic center of the continental United States. (This excludes Alaska and Hawaii, and U.S. territories.)

This question asks how far West the U.S. is located. For reference, the continental U.S. lies in the Western Hemisphere, which spans from 0 to 180 degrees West.

Is this geographic center greater or less than 90 degrees West?

The correct answer is greater than 90 degrees West.

For each possible answer the receiver can give, you will be asked to choose one of the two messages below. ***You will maximize the chance you win the bonus if you choose a message that the receiver thinks is truthful.***

Message G: *The answer is* ***greater than*** 90 degrees West.

OR

Message L: *The answer is* ***less than*** 90 degrees West.

You will be asked below to send Message G or Message L for each possible answer the receiver can give.

Please choose one answer in each row. For instance, if you wanted to send Message G for all receivers, you would do the following:

If you wanted to send Message G for all receivers who believe the answer is at least 80% likely to be greater than 90 degrees West, and send Message L to receivers who do not, you would do the following:

You may choose any combination you like, as long as you choose **one message in every row.** Please choose your messages below:

	Send Message G ("greater than 90 degrees West")	Send Message L ("less than 90 degrees West")
If receiver says: 0% chance the answer is greater	○	○
If receiver says: 10% chance the answer is greater	○	○
If receiver says: 20% chance the answer is greater	○	○
If receiver says: 30% chance the answer is greater	○	○
If receiver says: 40% chance the answer is greater	○	○
If receiver says: 50% chance the answer is greater	○	○
If receiver says: 60% chance the answer is greater	○	○
If receiver says: 70% chance the answer is greater	○	○
If receiver says: 80% chance the answer is greater	○	○
If receiver says: 90% chance the answer is greater	○	○
If receiver says: 100% chance the answer is greater	○	○

Subjects in the unincentivized treatment are instead told: *"**Your chance of winning the bonus is not affected by how you answer this question.**"*

Subjects in the competition treatment are instead told that *"**The receiver will predict, based on your message and another sender's message on this question, which sender sent more truthful messages over the course of the experiment. You will maximize the chance you win the bonus if the receiver believes that you sent more truthful messages.**"*

Figure 16: Instructions for assessing messages

Receiver: Instructions for News Assessments

Now, you will see one more practice screen in the receiver role. After each Question page, receivers will see a News Assessment page.

There has been a growing debate about the accuracy of news sources. This part of the study is testing whether people can recognize when others send them truthful and false information.

On a News Assessment page, you will see the previous Question page and be given a message from another participant (the "sender") about the correct answer. The sender will answer the question like the one you saw on the previous screen. They will:

1. Learn what the true answer is.
2. Be told that they have a higher chance of winning a bonus if *you* say they're telling the truth.
3. Choose which message to send.

The sender can either choose to send you a truthful message or a false message, and your goal is to predict which one they chose.

You will be matched with different senders in different rounds. You will be equally likely to match with a Democrat or a Republican (including leaners), and will sometimes be told whether the receiver is a Democrat or a Republican.

If a news assessment round is chosen for payment, **your assessment will determine the chance you win the bonus.**

The details of the point system used to determine your chance of winning the bonus are a bit complicated, but are explained below if you are interested. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by carefully and honestly answering these questions.**

Regardless of your guesses, a list of correct answers will be provided at the end of the study.

You will see a practice news assessment on the next page.

The practice question is omitted since it looks the same as questions in the main rounds.

Figure 17: Treatment revelation page

Your Role in this Study

You have been randomly assigned to be a **Receiver** in this study. You will see your first Question page on the next screen.

Based on your answer to the party question earlier, you have been coded as a **Democrat** (or Democrat-leaner).

When you are matched with senders, they **will not be given this information about you**.

Figure 18: Receiver: Prior beliefs

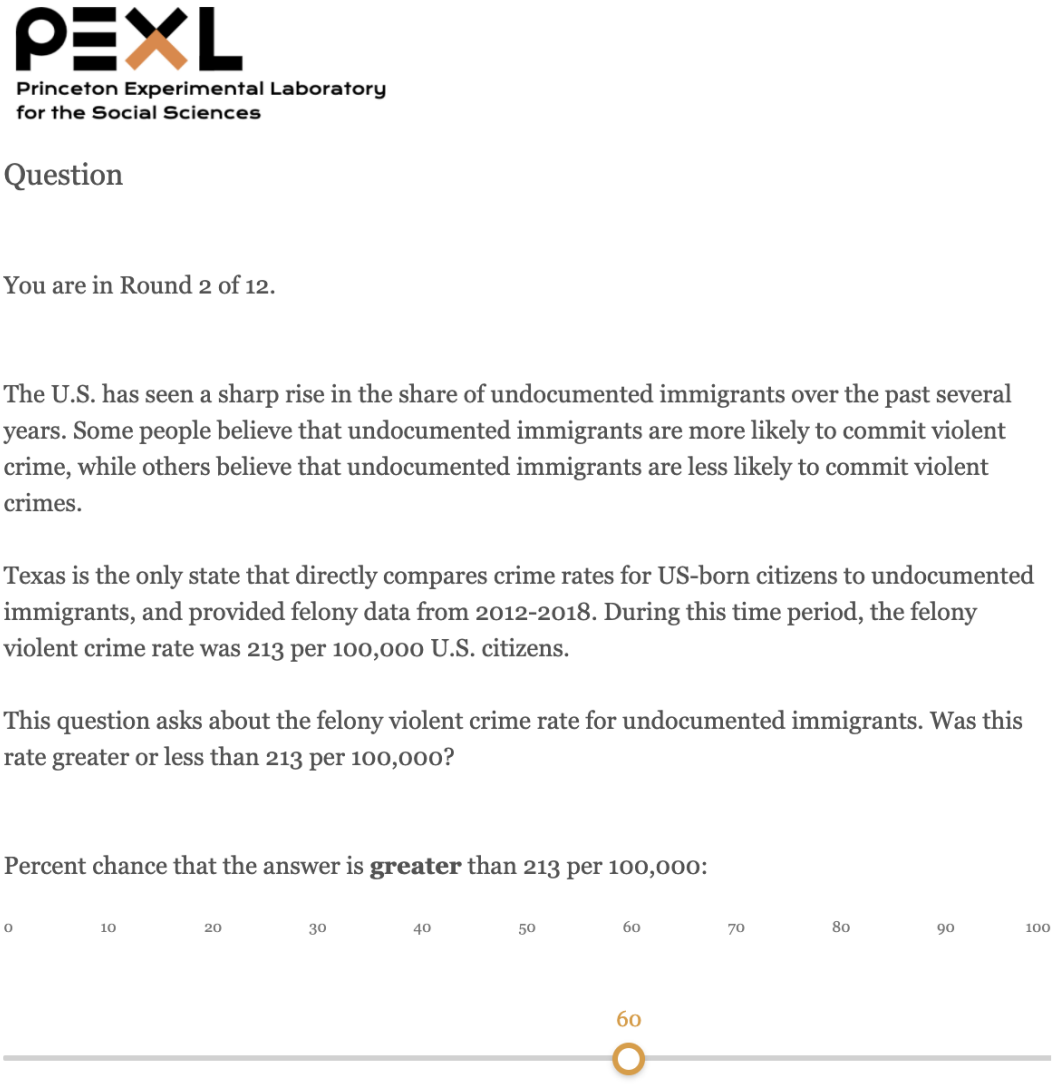
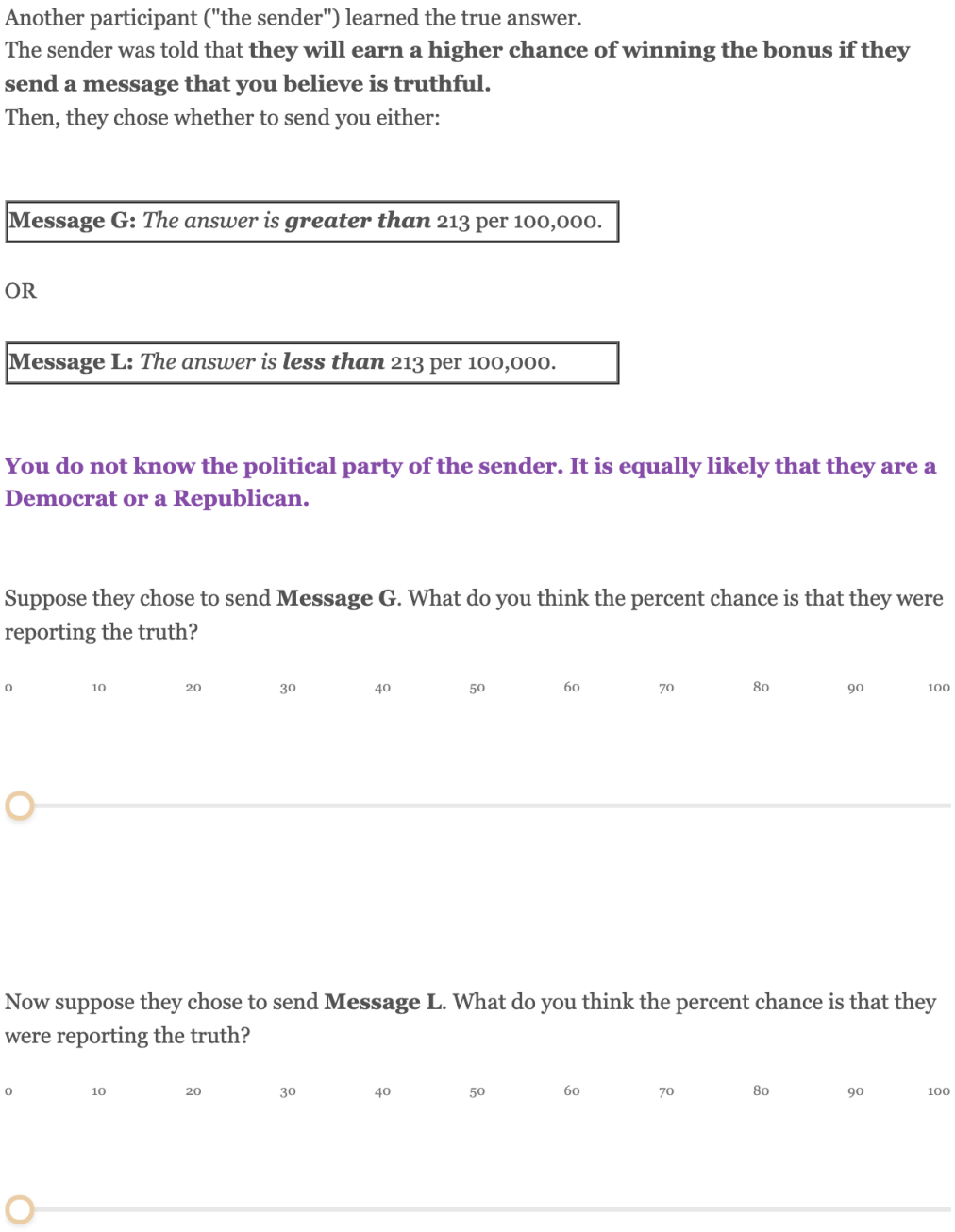


Figure 19: Receiver: News assessments



Receivers in the unincentivized treatment do not see the line about the sender's bonus.

Receivers in the competition treatment are instead asked: *Now suppose that **Sender X** chose to send **Message G** and that **Sender Y** chose to send **Message L**. What do you think is the percent chance is that **Sender X** sent more truthful signals over the course of the experiment?*

Figure 20: Sender: Sending messages

Sending Messages

You are in Round 1 of 12.

The other participant ("the receiver") was asked the following question:

The U.S. has seen a sharp rise in the share of undocumented immigrants over the past several years. Some people believe that undocumented immigrants are more likely to commit violent crime, while others believe that undocumented immigrants are less likely to commit violent crimes.

Texas is the only state that directly compares crime rates for US-born citizens to undocumented immigrants, and provided felony data from 2012-2018. During this time period, the felony violent crime rate was 213 per 100,000 U.S. citizens.

This question asks about the felony violent crime rate for undocumented immigrants. Was this rate greater or less than 213 per 100,000?

The receiver is a Democrat.

The correct answer is less than 213 per 100,000.

For each possible answer the receiver can give, you will be asked to choose one of the two messages below. ***You will maximize the chance you win the bonus if you choose a message that the receiver thinks is truthful.***

Message G: *The answer is* ***greater than*** 213 per 100,000.

OR

Message L: *The answer is* ***less than*** 213 per 100,000.

For each of the receiver's answers, would you like to send Message G or Message L?

	Send Message G	Send Message L
If they say there is a 0 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 10 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 20 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 30 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 40 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 50 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 60 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 70 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 80 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 90 percent chance the answer is greater than 213 per 100,000	○	○
If they say there is a 100 percent chance the answer is greater than 213 per 100,000	○	○

Senders in the unincentivized treatment are instead told: *"**Your chance of winning the bonus is not affected by how you answer this question.**"*

Senders in the competition treatment are instead told that *"**The receiver will predict, based on your message and another sender's message on this question, which sender sent more truthful messages over the course of the experiment. You will maximize the chance you win the bonus if the receiver believes that you sent more truthful messages.**"*

Figure 21: Attention check

Question

You are in Round 8 of 12.

In 1776 our fathers brought forth, upon this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal. What is the year right now?

This is not a trick question and the first sentence is irrelevant; this is a check to make sure you are paying attention. If you get this question incorrect, you will not be eligible to receive a bonus payment.

- ◯ 1776
- ◯ 1984
- ◯ 2021
- ◯ 1746

Figure 22: Receiver: Quote page instructions

Additional Question and News Assessment Pages

You will continue to see Question pages as a receiver. In the next four rounds, you will see quotes from either President Joe Biden or former President Donald Trump. The content of each quote has been verified to be accurate or inaccurate, and you will rate the likelihood that each one is accurate.

You will then see messages from other participants (senders) as before. Senders will learn the correct answer and be asked to choose a message to send. The messages will either say

Message A: *The statement is* ***accurate****.*

OR

Message I: *The statement is* ***inaccurate****.*

Unlike in previous rounds, each sender will only make one message choice instead of one message choice for each possible initial belief you may have. Each sender will be told whether your party (or lean) was towards the Republican Party or towards the Democratic Party.

The scoring rules to determine the chance you win the high bonus payment are the same as the ones in previous rounds.

Figure 23: Sender: Quote page instructions

Additional Question and News Assessment Pages

You will continue to see Sending Messages rounds as a sender. In the next four rounds, you will see quotes from either President Joe Biden or former President Donald Trump. The content of each quote has been verified to be accurate or inaccurate.

You will be told whether each statement is accurate and be asked to send one of two messages to another participant (the "receiver"):

Message A: *The statement is* ***accurate****.*

OR

Message I: *The statement is* ***inaccurate****.*

Unlike in previous rounds, you will only make one choice to send a message to each receiver, instead of one choice for each possible belief they may have.

Each receiver will be told whether your party (or lean) was towards the Republican Party or towards the Democratic Party.

Figure 24: Receiver: Quote question

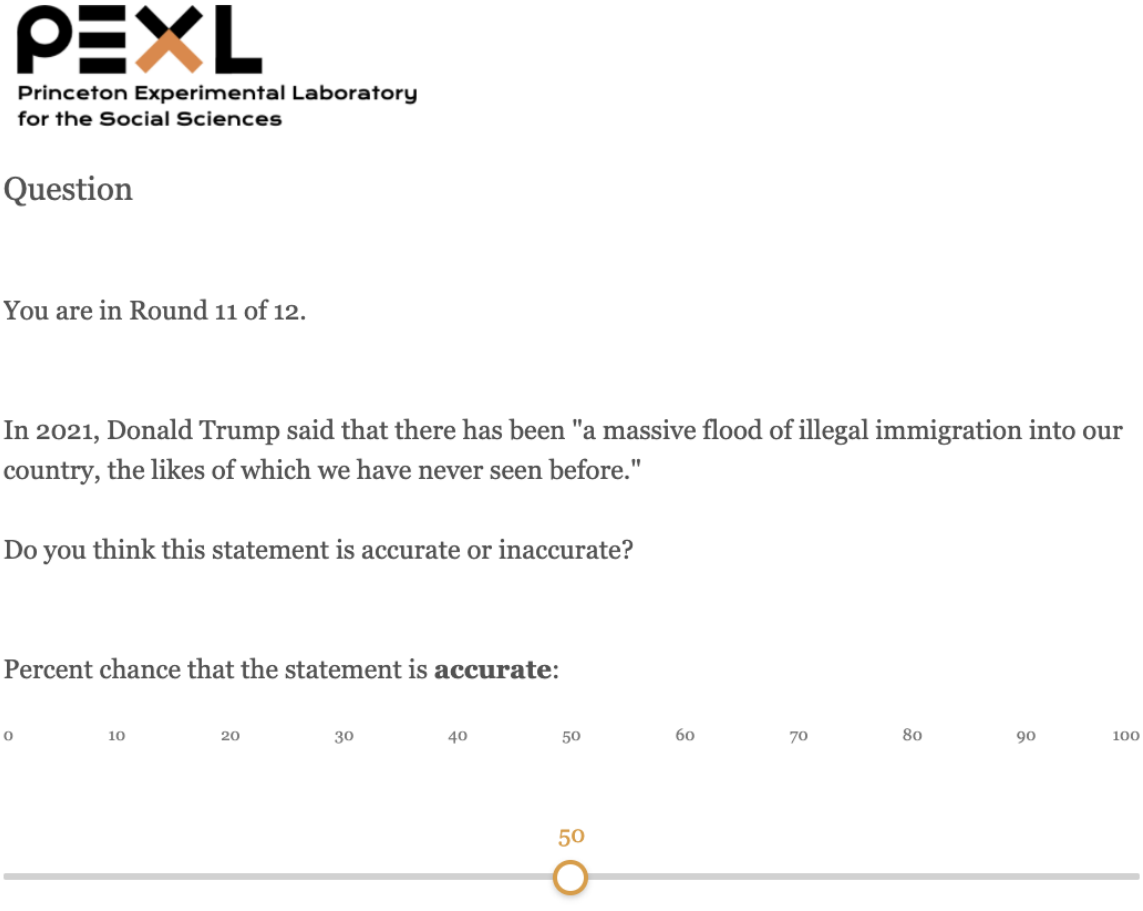
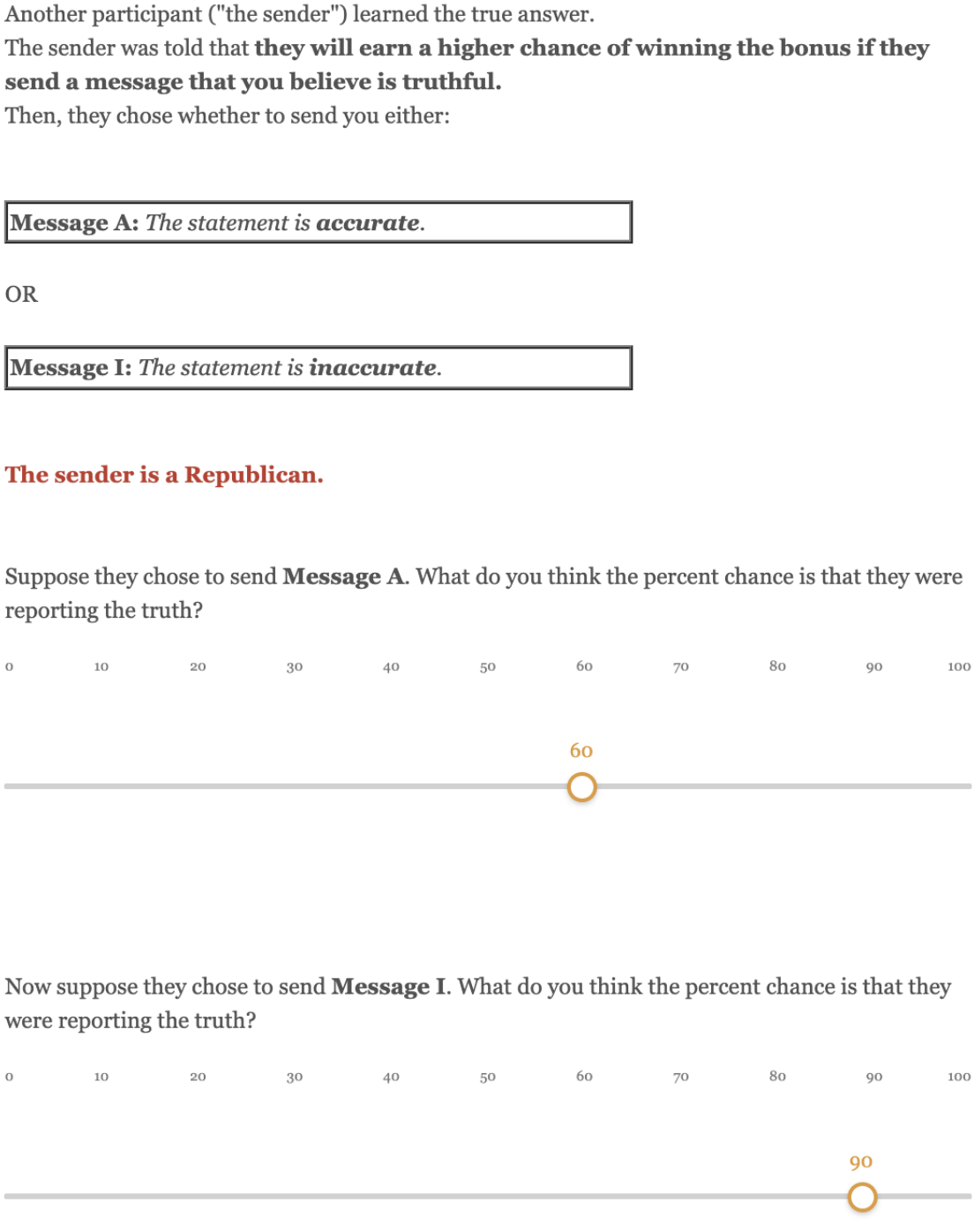


Figure 25: Receiver: Quote news assessment



Receivers in the unincentivized treatment do not see the line about the sender's bonus.

Receivers in the competition treatment are instead asked: *Now suppose that **Sender X** chose to send **Message G** and that **Sender Y** chose to send **Message L**. What do you think is the percent chance is that **Sender X** sent more truthful signals over the course of the experiment?*

Figure 26: Sender: Quote message choice

Sending Messages

You are in Round 10 of 12.

The other participant ("the receiver") was asked the following question:

In 2021, Donald Trump said that there has been "a massive flood of illegal immigration into our country, the likes of which we have never seen before."

Do you think this statement is accurate or inaccurate?

The receiver is a Republican.

The statement is inaccurate.

You will be asked to choose one of the two messages below. ***You will maximize the chance you win the bonus if you choose a message that the receiver thinks is truthful.***

Message A: *The statement is* ***accurate****.*

OR

Message I: *The statement is* ***inaccurate****.*

Would you like to send Message A or Message I?

◯ Send Message A

◯ Send Message I

Senders in the unincentivized treatment are instead told: *"**Your chance of winning the bonus is not affected by how you answer this question.**"*

Senders in the competition treatment are instead told that *"**The receiver will predict, based on your message and another sender's message on this question, which sender sent more truthful messages over the course of the experiment. You will maximize the chance you win the bonus if the receiver believes that you sent more truthful messages.**"*

Figure 27: Receiver: Predictions

Predictions About Senders

As you may have noticed, several of these questions were on topics that tend to be politically charged: racial discrimination, media bias, COVID stay-at-home orders, crime, undocumented immigrants, gun laws, unemployment, and wages. On each of these questions, Democrats and Republicans tend to have systematically different beliefs as to what the true answer is.

This question asks you how often senders in this study sent truthful messages, and whether their messages differed when they were matched with Democrats versus Republicans.

If you are randomly selected to win a bonus payment and this question is randomly chosen, one of your answers on this question will be randomly selected to determine your payment. The details of the point system used to determine your chance of winning the bonus are a bit complicated, but are explained below if you are interested. It is in your best interest to guess an answer that is close to the correct answer. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by carefully and honestly answering these questions.**

When the two messages say:

Message D: *The answer is* ***[more in the Democratic direction]****.*

Message R: *The answer is* ***[more in the Republican direction]****.*

(Participants did not see the bracketed statements, but instead saw the "greater than" or "less than" statements as before.)

When senders in this study:

- Were matched with a **Democrat** who thought that the answer was equally likely to be greater or less than the stated number,
- And were told that the truthful message was **Message D**.

What percent of the time did senders send **Message D (the truthful message)**?

When senders in this study:

- Were matched with a **Democrat** who thought that the answer was equally likely to be greater or less than the stated number,
- And were told that the truthful message was **Message R**.

What percent of the time did senders send **Message R (the truthful message)**?

When senders in this study:

- Were matched with a **Republican** who thought that the answer was equally likely to be greater or less than the stated number,
- And were told that the truthful message was **Message D**.

What percent of the time did senders send **Message D (the truthful message)**?

When senders in this study:

- Were matched with a **Republican** who thought that the answer was equally likely to be greater or less than the stated number,
- And were told that the truthful message was **Message R**.

What percent of the time did senders send **Message R (the truthful message)**?

0 10 20 30 40 50 60 70 80 90 100

Point system for your guess:

If a question is chosen for payment, you will receive between 0 and 100 points for your guess. The closer your guess is to the correct answer, the more likely it is that you'll win the prize.

If you guess the answer correctly, you will receive 100 points (the maximum).
If your guess is more than 10 away from the answer, you will receive 0 points.
If your guess is less than 10 away from the answer, you will receive points equal to 100 minus the squared distance to the correct answer.
For instance, if your guess is off by 7, you will receive 100 minus 7 times 7, or 51 points.

Figure 28: Sender: Predictions

Predictions About Receivers

As you may have noticed, several of these questions were on topics that tend to be politically charged: racial discrimination, media bias, COVID stay-at-home orders, crime, undocumented immigrants, gun laws, unemployment, and wages. On each of these questions, Democrats and Republicans tend to have systematically different beliefs as to what the true answer is.

This question asks you whether receivers in this study thought that messages were more likely to be truthful if the messages said that the answer was more in the direction that Democrats believed or more in the direction that Republicans believed.

If you are randomly selected to win a bonus payment and this question is randomly chosen, one of your answers on this question will be randomly selected to determine your payment. The details of the point system used to determine your chance of winning the bonus are a bit complicated, but are explained below if you are interested. It is in your best interest to guess an answer that is close to the correct answer. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by carefully and honestly answering these questions.**

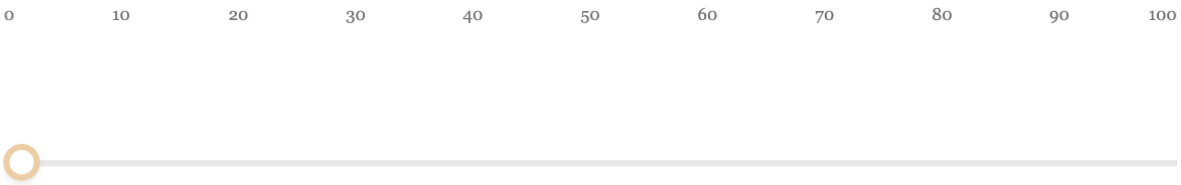
When the two messages say:

Message D: *The answer is* ***[more in the Democratic direction]****.*

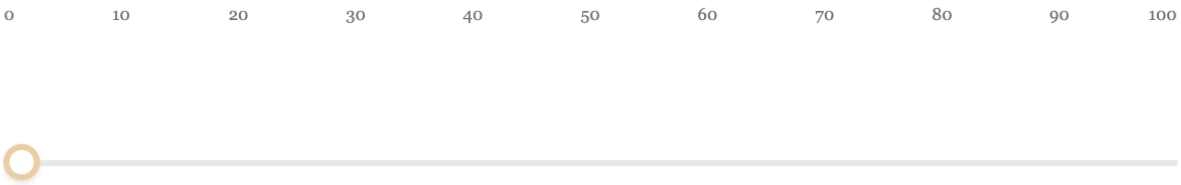
Message R: *The answer is* ***[more in the Republican direction]****.*

(Participants did not see the bracketed statements, but instead saw the "greater than" or "less than" statements as before.)

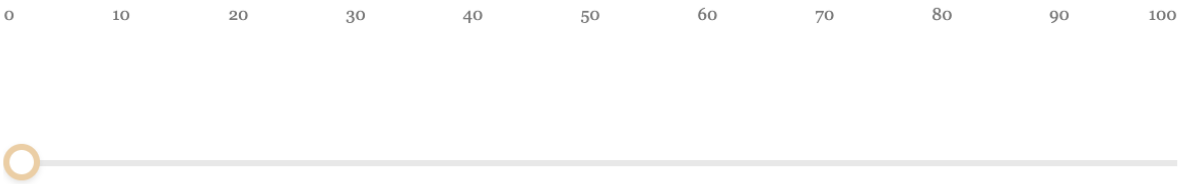
When a **Democratic** receiver in this study believed that the answer was equally likely to be greater than or less than the stated number, what was the percent chance that the receiver believed that a sender who sent **Message D** was being truthful?



When a **Democratic** receiver in this study believed that the answer was equally likely to be greater than or less than the stated number, what was the percent chance that the receiver believed that a sender who sent **Message R** was being truthful?



When a **Republican** receiver in this study believed that the answer was equally likely to be greater than or less than the stated number, what was the percent chance that the receiver believed that a sender who sent **Message D** was being truthful?



When a **Republican** receiver in this study believed that the answer was equally likely to be greater than or less than the stated number, what was the percent chance that the receiver believed that a sender who sent **Message R** was being truthful?

0 10 20 30 40 50 60 70 80 90 100

Point system for your guess:

If a question is chosen for payment, you will receive between 0 and 100 points for your guess. The closer your guess is to the correct answer, the more likely it is that you'll win the prize.

If you guess the answer correctly, you will receive 100 points (the maximum).
If your guess is more than 10 away from the answer, you will receive 0 points.
If your guess is less than 10 away from the answer, you will receive points equal to 100 minus the squared distance to the correct answer.
For instance, if your guess is off by 7, you will receive 100 minus 7 times 7, or 51 points.

Figure 29: Receiver: Survey beliefs

How important were each of the following factors in determining which messages **you believed were truthful** in this study?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
My general trust in others to tell me the truth	○	○	○	○	○
Whether the message was aligned with my own political preferences	○	○	○	○	○
Whether the message was aligned with what I already believe	○	○	○	○	○
Whether the message was aligned with the sender's political preferences	○	○	○	○	○

How important were each of the following factors in determining which messages **senders chose to send** in this study?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Whether the message was aligned with their own political preferences	○	○	○	○	○
Whether the message was aligned with the receiver's political preferences	○	○	○	○	○
Whether the message was aligned with what the receiver already believed	○	○	○	○	○
Their general inclination to tell others the truth	○	○	○	○	○

Figure 30: Sender: Survey beliefs

How important were each of the following factors in determining which messages **you chose to send** in this study?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Whether the message was aligned with the receiver's political preferences	○	○	○	○	○
My general inclination to tell others the truth	○	○	○	○	○
Whether the message was aligned with my own political preferences	○	○	○	○	○
Whether the message was aligned with what the receiver already believed	○	○	○	○	○

How important were each of the following factors in determining which messages **receivers believed were truthful** in this study?

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Whether the message was aligned with their own political preferences	○	○	○	○	○
Whether the message was aligned with the sender's political preferences	○	○	○	○	○
Their general trust in others to tell me the truth	○	○	○	○	○
Whether the message was aligned with what they already believe	○	○	○	○	○

D Study Materials for Experiment 2

D.1 Questions

Receivers saw a subset of the questions from Experiment 1. Instead of being given a target number, they were asked to input their guess. For instance, the end of the question about violent crime among undocumented immigrants said "What was the felony violent crime rate per 100,000 undocumented immigrants?"

Senders saw each question with the following median beliefs of receivers:

Topic	**Median belief**	**Truthful computer message**
US crime	500.0	Less than
Immigrants' crime	213.0	Less than
Racial discrimination	8.50	Less than
Media bias	65	Greater than
COVID-19 restrictions	50.0	Less than
Gun reform	220.0	Greater than
Unemployment	3.20	Greater than
Wages	4.00	Less than
Latitude of US	45.0	Less than
Random number	50.0	Greater than

D.2 Screenshots

Figure 31: Overview page

Overview and Bonus Payment

On the following pages, you will be asked to make a series of choices that can impact your bonus payment.

After all participants complete the study, 5 participants will be chosen at random to receive a bonus payment of either $10 or $100 based on their choices. The high bonus is because it is important for us that you take this study seriously.

You will see 15-20 pages of questions, comprising of approximately 12 "rounds" and 5 additional pages of questions. Many of these questions will earn you points based on the answers you give. If you are randomly chosen to receive a bonus payment, one question will be chosen at random and your payment will be determined by the number of points you earn. The number of points will determine the percent chance you win $100 as opposed to $10. For instance, if you are randomly chosen to receive a bonus payment and receive 70 points on the randomly-chosen question, you will have a 70 percent chance to win $100 and a 30 percent chance to win $10.

In addition, one page will involve an "attention check" question in the study. The answer to this question will be obvious to anyone paying attention. Participants who do not answer the attention check question correctly will still receive their show-up payment, but will not be eligible for any bonus payment.

Instructions for the first type of question are on the following page.

Figure 32: Instructions for median beliefs

Instructions for Questions

Throughout this study, you will see several types of pages, including Question pages.

On a Question page, you will be asked to guess the answer to a factual question; each question has a correct numerical answer. In addition to your guaranteed payment, you will have a chance to win an additional bonus of $100 based on your answers to these questions.

At least one question is an "attention check" for which the correct answer will be obvious. **If you do not answer this question correctly, you will not be eligible for the bonus.**

The details of the point system used to determine your chance of winning the bonus are a bit complicated, but are explained below if you are interested. It is in your best interest to guess an answer that is the "median" (in the middle) of what you believe is likely. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by carefully and honestly answering these questions.**

Regardless of your guesses, a list of correct answers will be provided at the end of the study.

Point system for your guess:

If a question is chosen for payment, you will receive between 0 and 100 points for your guess. The closer your guess is to the correct answer, the more likely it is that you'll win the prize.

If you guess the answer correctly, you will receive 100 points (the maximum).
If your guess is more than 100 away from the answer, you will receive 0 points.
If your guess is less than 100 away from the answer, you will receive points equal to 100 minus the distance to the correct answer.

It is in your best interest to guess an answer that is the "median" (in the middle) of what you believe is likely. For example, if you think the answer is equally likely to be 10, 40, and 60, you should guess 40.

Figure 33: Instructions for news assessments

Instructions for News Assessments

You have just seen a Question page. After a Question page, you will see a News Assessment page.

There has been a growing debate about the accuracy of news sources. This part of the study is testing whether people can recognize Fake News and True News.

On a News Assessment page, you will see the previous Question page and be given a message related to your previous guess. The message will either be from a True News source or from a Fake News source. If this question is selected for payment, the chance you win the prize depends your answer.

One source will say: "The answer is **greater than** your previous guess"
and the other source will say: "The answer is **less than** your previous guess."

One of these messages will be truthful, and the other will not. **The True News source always tells you the truth, while the Fake News source will never tell the truth.**

After each question, you will assess whether you think each message is more likely to come from True News, come from Fake News, or about equally likely of each. If this round is chosen for payment, **your assessment will determine the chance you win the prize.**

- If you guess True News and the message comes from True News, you will receive 100 points.
- If you guess True News and the message comes from Fake News, you will receive 0 points.
- If you guess Fake News and the message comes from Fake News, you will receive 100 points.
- If you guess Fake News and the message comes from True News, you will receive 0 points.
- If you guess Equally Likely and the message comes from True News, you will receive 55 points.
- If you guess Equally Likely and the message comes from Fake News, you will receive 55 points.

Regardless of your guesses, a list of correct answers will be provided at the end of the study.

Figure 34: Receiver: Question page

Question

You are in Round 4 of 12.

The U.S. has seen a sharp rise in the share of undocumented immigrants over the past several years. Some people believe that undocumented immigrants are more likely to commit violent crime, while others believe that undocumented immigrants are less likely to commit violent crimes.

Texas is the only state that directly compares crime rates for US-born citizens to undocumented immigrants, and provided felony data from 2012-2018. During this time period, the felony violent crime rate was 213.0 per 100,000 U.S. citizens.

What was the felony violent crime rate per 100,000 undocumented immigrants?

Figure 35: Receiver: News page

News Assessment

You are in Round 4 of 12.

The previous question:

The U.S. has seen a sharp rise in the share of undocumented immigrants over the past several years. Some people believe that undocumented immigrants are more likely to commit violent crime, while others believe that undocumented immigrants are less likely to commit violent crimes.

Texas is the only state that directly compares crime rates for US-born citizens to undocumented immigrants, and provided felony data from 2012-2018. During this time period, the felony violent crime rate was 213.0 per 100,000 U.S. citizens.

What was the felony violent crime rate per 100,000 undocumented immigrants?

Your previous guess: 150.

The two messages say:

News A: *The answer is **greater than** your previous guess of 150.*

News B: *The answer is **less than** your previous guess of 150.*

Which news source do you think is True News, and which news source do you think is Fake News?

- ◯ News A is more likely to be True News; News B is more likely to be Fake News
- ◯ News B is more likely to be True News; News A is more likely to be Fake News
- ◯ The two news sources are similarly likely to be True News and Fake News

Figure 36: Instructions for choosing messages

Instructions: Choice of News

You will now be asked to choose a message for another participant that is like the message you previously saw.

If this question is selected for payment, the probability you win the bonus will depend on whether the other participant said that your chosen message was True News or Fake News.

- If the other participant said that your chosen message was True News, you will earn 100 points.
- If the other participant said that your chosen message was Fake News, you will earn 0 points.
- If the other participant said that your chosen message was similarly likely to be True or Fake, you will earn 50 points.

You will maximize the chance you win the bonus if you choose a message that the other participant thinks is True News.

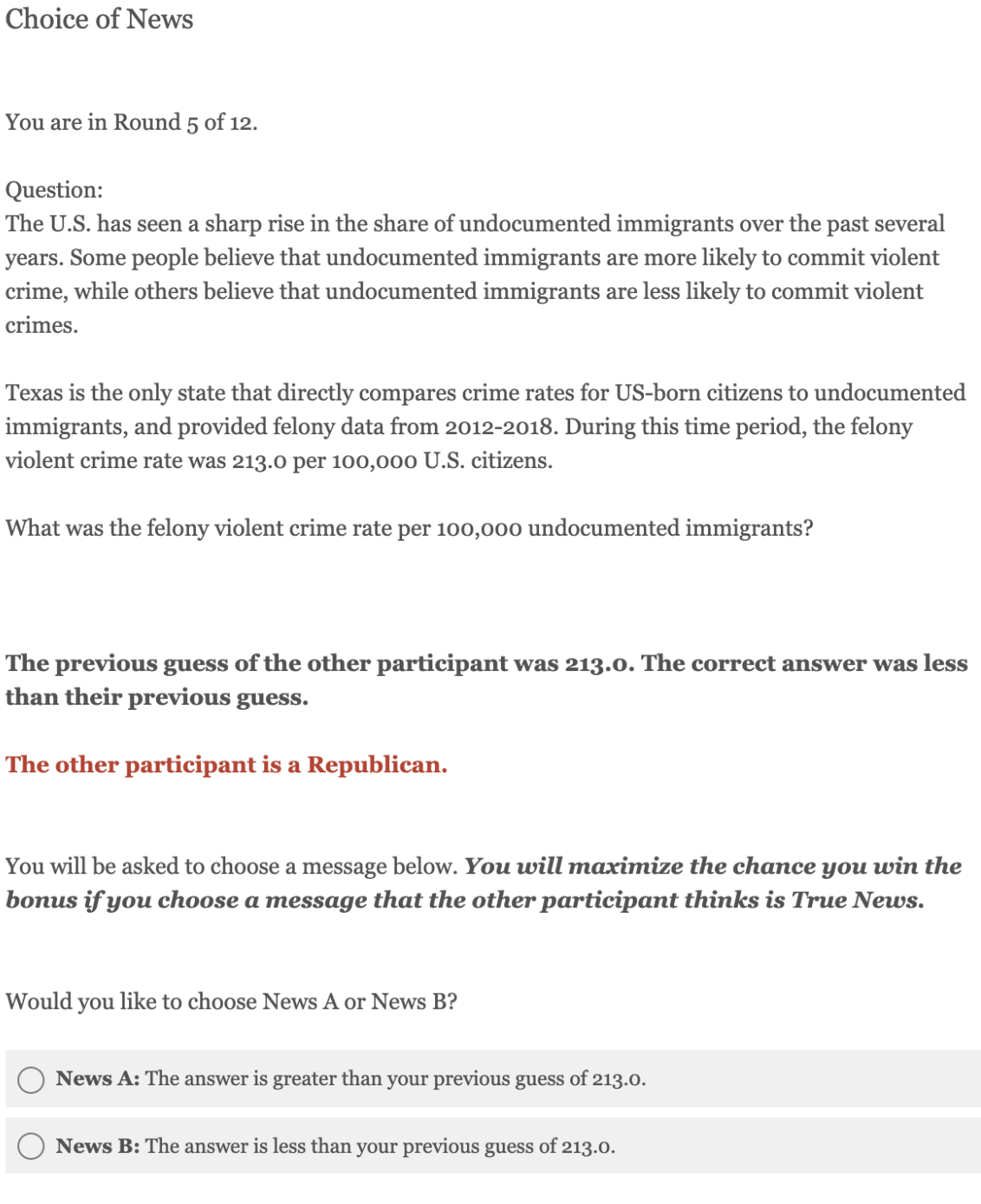
The other participant will see both messages, regardless of your choice. A list of correct answers will be provided to you and to the other participant at the end of the study.

You will be matched with different participants in different rounds. You will be equally likely to match with a Democrat or a Republican (including independents who prefer one party), and there is nothing to infer from a participants' answer about their party. However, you will sometimes be told whether they are a Democrat or a Republican.

You may have to wait for a few seconds in between rounds for the next page to load; please be patient.

Senders in the unincentivized treatment do not see the sentences about the bonus, and instead see: ***"These questions, unless otherwise specified, will not affect your chance to win a bonus payment."***

Figure 37: Sender: Choosing messages page



Choice of News

You are in Round 5 of 12.

Question:
The U.S. has seen a sharp rise in the share of undocumented immigrants over the past several years. Some people believe that undocumented immigrants are more likely to commit violent crime, while others believe that undocumented immigrants are less likely to commit violent crimes.

Texas is the only state that directly compares crime rates for US-born citizens to undocumented immigrants, and provided felony data from 2012-2018. During this time period, the felony violent crime rate was 213.0 per 100,000 U.S. citizens.

What was the felony violent crime rate per 100,000 undocumented immigrants?

The previous guess of the other participant was 213.0. The correct answer was less than their previous guess.

The other participant is a Republican.

You will be asked to choose a message below. ***You will maximize the chance you win the bonus if you choose a message that the other participant thinks is True News.***

Would you like to choose News A or News B?

◯ **News A:** The answer is greater than your previous guess of 213.0.

◯ **News B:** The answer is less than your previous guess of 213.0.

Senders in the unincentivized treatment do not see the sentence about the bonus.

Figure 38: Sender instructions: Demand for information

Instructions: Information About Other Participant

You will continue to choose messages for another participant.

This time, you will have the option to learn more about the other participant.

If this question is selected for payment, the probability you win the bonus will depend on whether the other participant said that your chosen message was True News or Fake News.

- If the other participant said that your chosen message was True News, you will earn 100 points.
- If the other participant said that your chosen message was Fake News, you will earn 0 points.
- If the other participant said that your chosen message was similarly likely to be True or Fake, you will earn 50 points.

You will maximize the chance you win the bonus if you choose a message that the other participant thinks is True News.

The other participant will see both messages, regardless of your choice. A list of correct answers will be provided to the other participant at the end of the study.

You will be matched with different participants in different rounds.

Figure 39: Sender: Information choice page

The previous guess of the other participant was 4.00. The correct answer was less than their previous guess.

You will be asked below whether you would like to learn more about the other participant before choosing a message to send.

You will maximize the chance you win the $100 bonus if you choose a message that the other participant thinks is True News.

News A: The answer is greater than your previous guess of 4.00.

News B: The answer is less than your previous guess of 4.00.

The other participant is equally likely to be a Democrat or a Republican. Before choosing a message, would you like to learn the **political party** of the other participant? If you do, you will have the opportunity to condition your message on their party.

You have the option to **either**

- Learn their party with **certainty** (100% chance) and receive **30 cents** if this question is selected for payment.

  OR

- Learn their party with **a smaller chance** (50% chance) and receive **35 cents** if this question is selected for payment.

◯ Learn the political party of the other participant with certainty and receive 30 cents

◯ Learn the political party of the other participant with a smaller chance and receive 35 cents

Figure 40: Sender: Can condition on receiver's party

Information About Other Participant

You are in Round 8 of 12.

Question:
The Trump administration campaigned on tough-on-crime policies. Some people believe that the Trump administration's policies were effective at reducing violent crime, while others believe that his rhetoric provoked more violence.

This question asks how violent crime rates changed during the Trump administration. In 2016 (before Trump became president), the violent crime rate was 386.6 per 100,000 Americans.

In 2020 (at the end of Trump's presidency), what was the violent crime rate (per 100,000 Americans)?

The previous guess of the other participant was 500.0. The correct answer was less than their previous guess.

You will learn the political party of the other participant. You will be asked to choose a message below.

Which news would you like to choose if the other participant is a **Democrat**?
Which news would you like to choose if the other participant is a **Republican**?

	News A: The answer is greater than your previous guess of 500.0.	**News B:** The answer is less than your previous guess of 500.0.
If the other participant is a **Democrat**	◯	◯
If the other participant is a **Republican**	◯	◯

Figure 41: Sender: Cannot condition on receiver's party

Information About Other Participant

You are in Round 9 of 12.

Question:
Some people believe that the Trump administration did a better job at increasing wages for most Americans, and some people believe that the Obama administration did a better job of wage growth.

In the last two years of the Obama administration (Jan 2015-Jan 2017), the median growth in Americans' wages was 3.28 percent on average.

What was the median growth in Americans' wages during the Trump administration on average?

The previous guess of the other participant was 4.00. The correct answer was less than their previous guess.

You did not learn the political party of the other participant. You will be asked to choose a message below.

Would you like to choose News A or News B?

◯ **News A:** The answer is greater than your previous guess of 4.00.

◯ **News B:** The answer is less than your previous guess of 4.00.

Figure 42: Attention check

Question

You are in Round 12 of 12.

In 1776 our fathers brought forth, upon this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

What is the year right now?

This is not a trick question and the first sentence is irrelevant; this is a check to make sure you are paying attention. If you get this question incorrect, you will not be eligible to receive a bonus payment.

Figure 43: Sender: Beliefs about receivers

Predictions About Other Participant

As you may have noticed, several of these questions were on topics that tend to be politically charged: racial discrimination, media bias, COVID stay-at-home orders, crime, undocumented immigrants, gun laws, unemployment, and wages. On each of these questions, Democrats and Republicans tend to have systematically different beliefs as to what the true answer is.

This question asks you whether people thought the news was more likely to be True News if it said that, as compared to people's original guess, the answer was more in the direction that Democrats believed or in the direction that Republicans believed.

When the two messages say:

News A: *The answer is* ***[more in the Democratic direction] than*** *your previous guess.*

News B: *The answer is* ***[more in the Republican direction] than*** *your previous guess.*

(Participants did not see the bracketed statements, but instead saw "greater than" or "less than" statements.)

What percent of the time did **Republicans** say that News A or News B was more likely to be True News? (Please have your answers add up to 100.)

% Republicans who said the two news sources are similarly likely to be True News and Fake News	0
% Republicans who said News A is more likely to be True News and News B is more likely to be Fake News	0
% Republicans who said News B is more likely to be True News and News A is more likely to be Fake News	0
Total	0

What percent of the time did **Democrats** say that News A or News B was more likely to be True News? (Please have your answers add up to 100.)

% Democrats who said News B is more likely to be True News and News A is more likely to be Fake News	0
% Democrats who said the two news sources are similarly likely to be True News and Fake News	0
% Democrats who said News A is more likely to be True News and News B is more likely to be Fake News	0
Total	0

If you are randomly selected to win a bonus payment and this question is randomly chosen, one of your answers on this question will be randomly selected to determine your payment. The details of the point system used to determine your chance of winning the bonus are a bit complicated, but are explained below if you are interested. It is in your best interest to guess an answer that is close to the correct answer. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by carefully and honestly answering these questions.**

Point system for your guess:

If a question is chosen for payment, you will receive between 0 and 100 points for your guess. The closer your guess is to the correct answer, the more likely it is that you'll win the prize.

If you guess the answer correctly, you will receive 100 points (the maximum).

If your guess is more than 10 away from the answer, you will receive 0 points.

If your guess is less than 10 away from the answer, you will receive points equal to 100 minus the squared distance to the correct answer.

For instance, if your guess is off by 7, you will receive 100 minus 7 times 7, or 51 points.