

DISCUSSION PAPER SERIES

IZA DP No. 17892

**Belief Updating About Moral Norms:  
Does Group Identity Matter?**

David L. Dickinson  
Marie Claire Villeval

MAY 2025

## DISCUSSION PAPER SERIES

IZA DP No. 17892

# Belief Updating About Moral Norms: Does Group Identity Matter?

**David L. Dickinson**

*Appalachian State University, Chapman University and IZA*

**Marie Claire Villeval**

*CNRS and IZA*

MAY 2025

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9  
53113 Bonn, Germany

Phone: +49-228-3894-0  
Email: [publications@iza.org](mailto:publications@iza.org)

[www.iza.org](http://www.iza.org)

## ABSTRACT

---

# Belief Updating About Moral Norms: Does Group Identity Matter?\*

We investigate how group identity affects belief updating about moral norms. Using a Belief Updating task, we found that individuals follow a cautious version of Bayesian updating. Group identity itself does not directly affect belief updating. However, when given an information signal about the truthfulness of a normative statement that is dissonant with one's perceived norm, individuals differ in their resistance to updating beliefs. This difference depends on whether the statement reflects moral norm judgments from people with the same or different political affiliation, and whether the signal supports or opposes honesty. This highlights the importance of understanding how one updates beliefs regarding moral norms, and how the group identity of those making normative judgments can be an important consideration.

**JEL Classification:** C91, D83, D91

**Keywords:** cheating, social norms, belief updating, group identity, online experiment

**Corresponding author:**

Marie Claire Villeval  
Université Jean-Monnet Saint-Etienne  
GATE  
35 rue Raulin  
69007-Lyon  
France  
E-mail: [villeval@gate.cnrs.fr](mailto:villeval@gate.cnrs.fr)

---

\* This research has benefited from financial support of the Walker College of Business at Appalachian State University. This work was performed within the framework of the LABEX CORTEX(ANR-11-LABX-0042) of l'Université Claude Bernard Lyon 1, within the program "Investissements d'Avenir" (2019-ANR-LABX-02) operated by the French National Research Agency (ANR).

# 1 Introduction

Human behavior is profoundly shaped by social norms, particularly by normative expectations regarding what most people consider appropriate conduct (Elster, 1989, 2009; Bicchieri, 2005, 2017). Social norms function as powerful behavioral rules because they help people coordinate, and deviations from these norms often trigger social sanctions, creating substantial pressure to conform. This pressure not only fosters adherence but also instills a strong aversion to rule-breaking, as individuals seek to avoid the disapproval and consequences associated with non-conformity (*e.g.*, Fehr et al., 2002; López-Pérez, 2008; Balafoutas et al., 2014; Garfield et al., 2023). This dynamic is particularly pronounced in moral contexts, where foundational lessons from family and educational settings discourage behaviors such as cheating from an early age.

Although social norms play a fundamental role in helping individuals predict others' behaviors and reactions in many settings, normative uncertainty is still frequently observed (*e.g.*, d'Adda et al., 2020; Bicchieri et al., 2022; Dimant and Gesche, 2023), raising critical questions about how individuals learn and adapt to social norms. In particular, some uncertainty arises because norms can vary substantially across social groups and cultures. Some groups may maintain stricter or more permissive norms, adapt their norms more rapidly, or enforce them differently. For example, enforcement mechanisms can range from tight to loose depending on culture (*e.g.*, Gelfand, 2018; Dimant et al., 2025). Moreover, even when a social norm is shared across groups ("lying is bad"), the frontier between what is considered appropriate or inappropriate may be relatively blurred, especially in the moral domain ("the gray area"). In this context, the signals individuals receive about the social acceptability of certain behaviors are crucial to reducing normative uncertainty. However, they may treat the signals differently depending on the fundamental value they communicate (honesty versus opportunism). Moreover, individuals may weigh the information value of signals depending on whether they are attached to the normative judgments from affinity group (*i.e.*, in-group) or non-affinity group (*i.e.*, out-group) members.

The first question our study addresses is how people treat signals to update their normative beliefs about socially appropriate or inappropriate behaviors regarding opportunities for cheating to increase earnings. Specifically, we examine whether individuals update their normative beliefs in the moral domain according to Bayes' rule and whether they process supporting and contradicting signals about the truthfulness of moral statements symmetrically. Are they equally likely to weigh signals advocating

stricter moral norms as they are to weigh signals that could justify cheating?

Our second question is whether individuals weigh signals about the social acceptability of cheating differently depending on the group identity of the information source. Previous research has extensively documented the influence of group identity on behavior in areas such as coordination, punishment, and social preferences (*e.g.*, [Akerlof and Kranton, 2000](#); [Chen and Li, 2009](#); [Chen et al., 2014](#)). In comparison, the interest in the role of group identity on belief formation is more recent ([Hill, 2017](#); [Bauer et al., 2023](#); [Dimant et al., 2024](#); [Dickinson, 2025](#)), although there are reasons to believe that the formation of beliefs is largely embedded in a social context defined in particular by group affiliations, as shown by the recent literature on echo chambers in social media (*e.g.*, [Williams, 2023](#)). While [Dickinson \(2025\)](#) was interested in how group identity influences preferences and beliefs about political policy issues, and [Bauer et al. \(2023\)](#) examined how group identity influenced beliefs about unemployment statistics in the context of the 2020 U.S. presidential elections, we explore this question in the domain of social norms and morality. In this domain, [Dimant et al. \(2024\)](#) were chiefly interested in how group identity influences the selection of information channels, while our study focuses on how individuals update beliefs after receiving exogenous signals on moral statements from given sources. We study whether norm-informative signals in the moral domain are perceived as more relevant when they are attached to the moral statements of an in-group source - individuals sharing the same group identity - compared to an out-group source in a setting where there is no objective difference in the quality of the signals across groups and no endogenous selection of the source.

To address this question, our study characterizes group identity in terms of political affiliation. While political affiliation is not predicted to correlate with lying behavior, based on [Dimant et al. \(2024\)](#), prior research suggested that political affiliation is a powerful source of group polarization, particularly in the United States (*e.g.*, [Bursztyn et al., 2020](#); [Klein, 2020](#); [Ross Arguedas et al., 2022](#); [Panizza et al., 2024](#)). Therefore, we hypothesized that when assessing normative statements, individuals will tend to place greater weight on information signals that are attached to norm-related statements about members of one's political affinity group compared to statements about an opposing ideological group's norms. This hypothesis is in line with a model where individuals derive utility from both their beliefs and their actions, and where the weight assigned to information may depend on whether the signal affirms or denies a moral statement, as well as on whether the source of this moral statement belongs to their affinity group or an out-group.

Finally, the third question our study addresses is whether exposure to new information about the acceptability or unacceptability of certain behaviors leads to changes in normative expectations and moral conduct.

These questions are critical because evolving ethical norms can fundamentally transform societies. If individuals do not update their beliefs about social norms in a Bayesian way, and if group identity significantly shapes this process, this could challenge the universality of ethical standards by highlighting the contextual and group-dependent nature of moral reasoning. This could also affect the evolution of norms. If individuals update their beliefs to a lesser extent when information is attached to an out-group, then, depending on the values of the respective groups, either cheaters may take longer to align with honesty norms, or group attachment could act as a catalyst to adopting stricter moral norms. Understanding how group identity mediates the learning and internalization of moral norms offers key insight into the dynamics of ethics.

To study the impact of the political group identity on belief updating about moral norms, we designed and pre-registered an experiment conducted online on the Prolific platform in the U.S.. 450 participants (half affiliated with the Democratic party and half with the Republican party) played an incentivized Coin Flip task before and after a Belief Updating task. In the Coin Flip task, participants were asked to flip a coin 10 times in private and report their total number of Heads flipped, with monetary rewards tied to the number of Heads reported. This task incentivizes the over-reporting of favorable (Heads) outcomes.

In the Belief Updating task, participants evaluated the truthfulness of four statements related to the social perception of the moral norm in the Coin Flip task. The statements differed based on their factual accuracy (true or false) and their normative implications, either supporting truthful reporting or endorsing over-reporting in the Coin Flip task. After an initial evaluation, participants could update their beliefs about each statement's truthfulness upon receiving signals that had a known 75% accuracy. These noisy signals reflected the judgments of a separate group of participants who were explained the Coin Flip task but did not perform it. The judgments addressed the acceptability of specific misreporting behaviors, such as claiming nine Heads when only flipping three or slightly inflating the number of Heads reported (*e.g.*, by three or fewer out of 10 flips). Moreover, we elicited participants' normative expectations about how others judged specific over-reporting behaviors in the Coin Flip task, both before and after the Belief Updating task.

The experiment consists of four between-subjects treatments that varied only the

conditions of the Belief Updating task. In the *No-Identity* treatment, the statements were not associated with any political affiliation. In the *In-Group* treatment, the four statements came from a politically aligned group (*i.e.*, Republicans for Republican participants and Democrats for Democrat participants). In the *Out-Group* treatment, the statements came from the opposing political group. Finally, in the *Mixed* treatment, statements varied across political identities within participants, that is, some statements reported the judgments of in-groups, and others the judgments of out-groups.

Importantly, the statements were equally false or true regardless of the group mentioned (*e.g.*, the statement that a significant percentage of “individuals X” think it is acceptable to over-report in the Coin Flip task remained true whether “individuals X” referred to Republicans or Democrats). This design ensured that the information quality was consistent across treatments for all participants. Consequently, any difference in belief updating across treatments cannot be attributed to information quality but rather to “source utility” as defined by [Bauer et al. \(2023\)](#).

Our results illustrate the complex interplay between normative expectations, belief updating about the social norm, and behavior. We found that True and False signals about the truthfulness of moral statements were weighted equally, although the updating process was more cautious than predicted by Bayes’ rule. Contrary to our predictions, we did not find evidence supporting our main hypothesis about the role of group identity in belief updating. Participants did not assign greater weight to signals about the statements of individuals sharing their political group affiliation compared to signals about the statements of out-group members or those without a group identity. However, incorporating the degree of dissonance between the signal and an individual’s initial normative expectations revealed a more nuanced dynamic. Signals with high dissonance were down-weighted somewhat more so when they related to in-group statements, though the in-group/out-group difference did not reach statistical significance unless considering *only* the subset of trials where the signal aligned with a truth-telling norm. A higher dissonance level of a signal aligning with a truth-telling norm implies the participant’s perceived norm leans towards lying. In contrast, when a signal aligned with a misreporting norm, participants who perceived an honesty social norm down-weighted the signal more if it was attached to an out-group statement.

Finally, individuals’ average expectations about the social norm in the Coin Flip task exhibited strong inertia, as post-updating measures mirrored pre-updating expectations despite the normative information provided during the Belief Updating task. This stickiness could explain why we did not observe a significant influence of the evolution

of normative expectations on changes in cheating behavior. Overall, an important implication of our findings relates to the nuanced dynamic of belief updating in moral norms. Specifically, when social norms are evolving toward higher ethical standards, group attachment and peer influence may unexpectedly hinder progress by reinforcing existing behavior and expectations among cheaters. Conversely, individuals who believe in a truth-telling norm are more likely to dismiss signals promoting a more lenient moral norm if the norm relates to an out-group’s morality. This suggests that group identity can play both as a barrier to moral improvement and a filter that reinforces existing moral expectations, shaping how individuals treat normative information.

The remainder of this paper is as follows. Section 2 specifies our contributions to the literature, and section 3 outlines the experimental design and procedures. Section 4 outlines a theoretical framework that combines utility derived from beliefs and action, which helps inform our preregistered hypotheses. Section 5 reports the results. Section 6 discusses these results and concludes.

## 2 Related Literature

Our study contributes to three strands of the literature. First, we advance the understanding of social norms in contexts of uncertainty by focusing on how individuals learn from norm-informative signals. Dimant et al. (2024) also investigated how social information influences lying behavior and perceptions of social norms, showing that normative social information had a stronger effect than empirical social information. Their primary focus was on how individuals select their source of information. Our study takes a different approach by imposing the source of information on participants and by introducing false *vs.* true moral statements. Moreover, we identify how participants updated their beliefs incrementally after receiving each signal, rather than eliciting beliefs only after exposure to all the signals. This allows us to capture the step-by-step process of belief updating and measure compliance with Bayes’ rule.

Second, our study contributes to the literature on individual biases in Bayesian updating (*e.g.*, Hill, 2017; Charness and Levin, 2005; Holt and Smith, 2009), with a particular focus on asymmetric updating (*e.g.*, Eil and Rao, 2011; Coutts, 2019; Barron, 2021; Dickinson, 2025). While the literature provided mixed evidence of asymmetry in belief updating related to ego-relevant performance or financial outcomes, we investigate a distinct domain: belief updating on normative moral statements. The specific context

of the belief-updating environment likely plays a critical role, as both Hill (2017) and Dickinson (2025) have demonstrated evidence of asymmetric belief updating in the realm of political ideology.

Our primary interest lies in examining how group identity influences belief formation and distortions in updating. Bauer et al. (2023) explored this dynamic in the context of beliefs about economic perspectives during the 2020 U.S. Presidential campaign. Their findings revealed a pronounced partisan gap. Individuals exhibited an aversion to information sourced from out-groups while assigning greater weight to information from an in-group source. This pattern was driven by a source-utility mechanism rather than differences in the quality of information, contributing to increasing political polarization. While we are also interested in analyzing whether individuals weigh signals differently depending on their source’s political identity, we did not consider information demand and focused on the learning of social norms rather than economic perspectives through exogenously assigned norm-informative signals. Moreover, our True and False signals have no group identity themselves, it is the moral statements that come from an in-group or an out-group source in some treatments.

Our third contribution is to the literature on lying behavior and group identity. While previous studies have examined how group identity influences individuals’ decisions to cheat (*e.g.*, Della Valle and Ploner, 2017; Banerjee et al., 2018; Aksoy and Palma, 2019; Benistant and Villeval, 2019), our approach in this study offers a distinct perspective by primarily focusing on how group identity shapes the way individuals learn social norms and update their moral beliefs. This shift in focus allows us to investigate the underlying cognitive and normative processes that precede moral decisions, contributing to a deeper understanding of the interplay between group identity, normative information, and moral reasoning.

### 3 Experimental Design and Procedures

The study design, procedures, hypotheses, and analysis plans were all preregistered on the Open Science Framework (<https://doi.org/10.17605/OSF.IO/ZMB85>). We administered an incentivized Coin Flip task before *and* after participants completed an incentivized Belief Updating task that focused on moral norms related to various behaviors in that same Coin Flip task. Normative expectations about others’ judgments of appropriateness of behavior in the Coin Flip task were assessed both before and

after the Belief Updating task. The statements used in the Belief Updating task were derived from a preliminary survey of moral norms administered to a distinct sample of participants.<sup>1</sup> First, we describe each task, then present the treatment variations, and finally the procedures.

### 3.1 The Tasks

**Coin Flip Task.** In this task, we asked participants to find a coin, flip the coin 10 times, and report the number of Heads flipped (as in, *e.g.*, Dickinson and McEvoy, 2021). We informed them that the bonus payment received for this task increased with each Heads reported. Therefore, this task presents a monetary temptation to cheat by over-reporting the number of Heads flipped.

The exact same task was administered both before and after the Belief Updating task presented in the next paragraph. It was common knowledge that the Coin Flip task would be administered twice, and participants were told that an equal-chance random draw at the end of the study would dictate which of the two administrations would be used to determine their bonus payment for that task.

**Belief Updating Task.** Here, we used the basic design in Hill (2017). Participants had to provide incentivized beliefs regarding the truthfulness of several factually true and false statements about morally acceptable and unacceptable behavior in the Coin Flip task. These statements came from a preliminary survey that we conducted on Prolific with a sample of 110 participants (n=55 Republicans, n=55 Democrats).<sup>2</sup> Each statement described the number of individuals surveyed who felt a particular type of over-reporting in the task was acceptable. As such, the statements in this Belief Updating task involved a moral norm of behavior in the same Coin Flip task the participant had completed.

Key to the task design is that, after an initial baseline assessment (the participant’s view of a statement’s truthfulness), a noisy signal regarding the statement’s truthfulness was presented that had a common-knowledge 75% signal strength. That is, the participant was given a signal stating that the statement is *True*, or stating that it is

---

<sup>1</sup>Appendix A contains full details on the preliminary survey used to generate the norms statements used in the main study, as well as full details on the main study survey.

<sup>2</sup>In this preliminary survey, participants were explained the incentivized Coin Flip task performed in the main experiment but did not play the game themselves. They only had to report their personal judgment about the appropriateness of ten different behaviors.

*False.* The participant was reminded of their previous belief and that such signals were accurate 3 out of 4 times. After the noisy signal, the participant’s belief regarding the truthfulness of the statement was re-elicited.<sup>3</sup> For a given statement, there was a total of 4 noisy signals presented, after each of which updated beliefs were elicited. Then, a new statement was given, and the procedure was repeated for that next statement. A total of 4 statements, which were randomized in order, were presented, producing a panel of 16 observations of belief updating per participant.<sup>4</sup> To elicit such beliefs, we used an incentive-compatible procedure, which was a variation of the Becker-DeGroot-Marschak mechanism (Becker et al., 1964).<sup>5</sup>

The four incentivized statements were designed to vary across two key criteria: *i*) whether they were factually true or false, and *ii*) and whether the statement, factually true or false, supported a truth-telling norm or a norm of over-reporting. Specifically, two of the statements were factually true, and two were false. Among these, one true and one false statement aligned with the moral interests of someone inclined to cheat, while the remaining two, one true and one false, aligned with the moral interests of someone inclined to truthfulness. Table 1 presents the statements used in the task.<sup>6</sup>

**Normative Expectations.** Both before participants received the instructions for the Belief Updating task (the *pre* measure), and after they performed that task (the *post* measure), we elicited their normative expectations regarding how others would judge certain behaviors in the Coin Flip task.<sup>7</sup> Specifically, on a [0,100] scale, for each of four scenarios describing over-reporting behaviors in various ways, they were asked “*how*

---

<sup>3</sup>In each trial, the participant was given a 40-second time constraint to submit the initial belief and a 20-second time constraint to submit responses after seeing the noisy signal (as in Hill, 2017).

<sup>4</sup>Practice trials, which were not incentivized, helped ensure a participant was familiar with the belief elicitation process and how the sequence of noisy signals would be presented. The practice statement was unrelated to ethical norms: it was a statement regarding the average temperature on Mars.

<sup>5</sup>Specifically, in each round the computer program randomly drew a number (Draw 1) between 1 and 100, each number being equally likely. If this number was lower than the participant’s response, the participant won the bonus if the statement was true and did not win if the statement was false. In other words, the higher the response, the more likely the participant won if the statement was true. If this number was higher than the participant’s response, the computer drew a second number (Draw 2) between 1 and 100, each number equally likely. If Draw 2 was lower than Draw 1, the participant won the bonus for that round. The participant did not win the bonus if Draw 2 was higher than Draw 1.

<sup>6</sup>In addition, both after completing the Coin Flip task and after completing the Belief Updating task, participants reported their emotional state using nine emotions (happy, excited, surprised, satisfied, angry, irritated, confused, regret, disgust) on a 7-item Likert scale. This was only motivated for exploratory analysis, and this data is not analyzed in this paper.

<sup>7</sup>Unlike the Belief Updating task, we did not incentivize the elicitation of these normative beliefs because doing so would have required introducing an additional incentivization mechanism. Since we had already implemented an incentive structure for belief elicitation in the updating task, we were concerned that introducing another mechanism for normative expectations might confuse participants.

Table 1: Factual moral norms statements (from a preliminary survey of  $n=100$  participants)

Statement number	Statement text	Truthfulness	Norm supported by the statement
Practice trials	The average temperature on Mars, as a whole, is -81 degrees Fahrenheit.	TRUE	No normative view
1	A significant percentage of [GROUP IDENTIFIER] surveyed think it is acceptable to over-report the number of HEADS actually flipped in this task.	TRUE	Over-reporting norm
2	None of the [GROUP IDENTIFIER] surveyed think it is acceptable to report flipping 9 HEADS if one actually flipped 3 HEADS (out of 10) in this task.	FALSE	Over-reporting norm
3	The majority of [GROUP IDENTIFIER] surveyed think that reporting more HEADS than what one actually flips in this task is not acceptable behavior.	TRUE	Truthful reporting norm
4	The majority of [GROUP IDENTIFIER] surveyed think it is acceptable to over-report the number of HEADS flipped by just a few ( <i>e.g.</i> , 3 or fewer extra HEADS reported out of 10 flips), but not by more than that.	FALSE	Truthful reporting norm

Notes: For each statement (1-4) the [GROUP IDENTIFIER] was either “Republicans”, “Democrats”, or “individuals”, depending on the experiment design cell (see Table 2 and section 3.2). See full survey details in Table C1 in Appendix C.

*acceptable or unacceptable you think “others” (i.e., a typical person) feel this behavior is.*” (1 = Not at all acceptable, 100 = Totally acceptable). The behaviors presented to the participants were similar to those reported in Table 1, and the measure we use for one’s normative expectations is the average response across the four scenarios.<sup>8</sup> The initial elicitation of normative expectations was done before the Belief Updating task to avoid any contamination by the signals and statements, and we consider the average acceptability report given to be that participant’s perceived baseline norm regarding misreporting in this task.

<sup>8</sup>Specifically: someone over-reports the number of Heads actually flipped; someone reports flipping 9 Heads even though 3 Heads were actually flipped; someone reports more Heads than actually flipped; and someone over-reports the number of Heads flipped by just a few (*e.g.*, 3 or fewer extra Heads reported out of 10 flips). Although the first and third statements obviously describe the same behavior and only differ in terms of phrasing, we chose to keep both to be as similar as possible to the statements used in the Belief Updating task.

## 3.2 Treatments

To test the key hypothesis that belief updating differs depending on the signal’s attachment to a source’s group identification, the normative statements in the Belief Updating task were described as being derived from the preliminary survey responses of Democrats, Republicans, or from generic “individuals” (*i.e.*, the pooled sample of respondents), depending on the treatment condition. This allowed us to vary the alignment between the group identity of the main study participants and that of the sample providing the moral norm. Crucially, the quality of information was orthogonal to the source: each statement was true or false independently of whether it was associated with Republicans, Democrats, or generic individuals.

This manipulation created three distinct between-subjects treatments: one where participants received statements from individuals with a political ideology affinity (Democrats or Republicans) from the preliminary survey (“*In-Group*” treatment), one where they received statements from the opposing political ideology group (“*Out-Group*” treatment) and one where they received statements from individuals with no reference to political group identification (“*No-Identity*” treatment).<sup>9</sup> In addition to these treatments which varied between-subjects participants’ affinity with the group identified in the statements, an additional treatment was included. In the “*Mixed*” treatment, the group identification varied randomly within participants across the four statements used in the Belief Updating task.

Participants’ political ideology was determined using their responses to Prolific’s profile screener question “In general, what is your political affiliation?”. Custom sampling capabilities of Prolific ensured that the study was made available only to self-reported Democrat and Republican participants.<sup>10</sup> In the final post-experimental questionnaire, participants were also asked to rate their political ideology on a 9-point scale where 1 indicated “very conservative”, and 9 indicated “very liberal”.

Table 2 summarizes the various treatments manipulating the moral norm statement association. The table also includes the number of participants per treatment, categorized by these participants’ own political ideology.

---

<sup>9</sup>For example, an *In-Group* treatment statement for a Republican participant stated: “A significant percentage of REPUBLICANS surveyed think it is acceptable to over-report the number of Heads actually filling in this task.” The *Out-Group* treatment would replace the word REPUBLICANS with DEMOCRATS, while the *No-Identity* treatment would replace it with “individuals”.

<sup>10</sup>We double-checked at the beginning of the survey that the participants fitted our custom-screening criterion, otherwise, they were not allowed to continue the study.

Table 2: Experimental treatments

<i>Moral norms statement association</i>				
<i>Prescreened political ideology</i>	<i>In-Group treatment</i>	<i>Out-Group treatment</i>	<i>No-Identity treatment</i>	<i>Mixed treatment</i>
Republicans	$N = 55$	$N = 55$	$N = 55$	$N = 60$
Democrats	$N = 55$	$N = 55$	$N = 55$	$N = 60$

Notes:  $N = 450$  total participants (each providing  $n=16$  trials of belief elicitation in the Updating task). “Mixed” corresponds to within-subjects in-group and out-group variations. Prescreened political ideology was based on the Prolific profile screener question “In general, what is your political affiliation?” (Republican or Democrat).

### 3.3 Procedures

We aimed to recruit a sample of 400-440 participants living in the U.S. on the Prolific platform, split evenly across the design cells shown in Table 2 (with any additional observations allowed by the budget allocated to help balance the sample size in each cell). The sample size was targeted based on an ex-ante power analysis using G\*Power software as well as budget. Such analysis suggests sufficient power (.80) to identify medium-sized effects (Cohen’s  $d = .60$ ) with  $\alpha = 0.05$  in means comparisons and smaller effect sizes using multi-variate analysis that can leverage the panel nature of the data set.<sup>11</sup> We actually collected 450 observations, each providing 16 trials of belief elicitation in the Belief Updating task.

The 450 participants include 225 Democrats and 225 Republicans. The demographics of the Democrats sample were: average age = 38.23 ( $\pm 14.70$ ) years, 69% female, 29% minority, 15% students, and 60% employed. The demographics of the Republicans sample were: average age = 44.32 ( $\pm 14.36$ ) years, 48% female, 16% minority, 15% students, and 65% employed.

The fixed earnings for participation in the Prolific study were \$2.25. The Coin Flip task bonus payment was an additional \$0.10 per Heads reported. Participants also earned a bonus payment for the Belief Updating task, such that total earnings could be as high as \$4.25 for the 10-15 minute study. On average, participants earned in total \$3.41 (st. dev.= \$0.20).

<sup>11</sup>Assuming that in-group and out-group effects are similar for Republican and Democrat participants, pooling the data from both groups allows us to identify smaller size effects (Cohen’s  $d = .40$ ). Moreover, in the *In-Group/Out-Group* treatments, with at least 200 observations, we have sufficient power to identify small size effects in a multiple regression test of a single coefficient with eight covariates ( $f^2$  effect size = .05). Our power is increased in any analysis extending to trial level data.

## 4 Theoretical Framework and Hypotheses

### 4.1 Theoretical Framework

To structure our predictions, we propose a theoretical framework where individuals derive utility from both holding certain beliefs about the social norm regarding the appropriate reporting behavior and from their action that may involve a trade-off between monetary gains and norm compliance. Choice variables are the individual’s beliefs and actions. The various weights described depend on individual preferences.

**Utility derived from beliefs:** Let  $\hat{b}$  denote the individual’s subjective belief about the social norm in the Coin Flip task. The prior belief about the appropriate level of reporting,  $\hat{b}_{\text{prior}}$ , is updated after receiving a signal ( $s$ ) about the truthfulness of a statement regarding the moral norms of one’s in-group versus out-group. These signals may be weighted differently depending on whether the signal indicates the statement is True or False and on the group identity of the moral norm statement:<sup>12</sup>

$$\begin{aligned} \hat{b} = & \omega_{\text{prior}} \hat{b}_{\text{prior}} + \mathbb{I}\{s \text{ is True}\} \left[ \omega_{IG}^T \mathbf{1}\{\text{In-group}\} + \omega_{OG}^T \mathbf{1}\{\text{Out-group}\} \right] s \\ & + \mathbb{I}\{s \text{ is False}\} \left[ \omega_{IG}^F \mathbf{1}\{\text{In-group}\} + \omega_{OG}^F \mathbf{1}\{\text{Out-group}\} \right] s. \end{aligned} \quad (1)$$

In this model:

- $\hat{b}_{\text{prior}}$  is the prior belief<sup>13</sup> and  $\omega_{\text{prior}}$  is the weight on the prior.
- $s$  is the new signal about the truthfulness of the statement.
- $\mathbb{I}\{s \text{ is True}\}$  equals 1 if the signal indicates the statement is true (and 0 otherwise), and similarly for  $\mathbb{I}\{s \text{ is False}\}$ .
- $\mathbf{1}\{\text{In-group}\}$  equals 1 if the signal is about an in-group statement, and  $\mathbf{1}\{\text{Out-group}\}$  equals 1 if it is about an out-group statement.
- $\omega_{IG}^T$  and  $\omega_{OG}^T$  capture the weights on signals True attached to in-group and out-group statements, respectively.

---

<sup>12</sup>Previous studies have suggested that individuals who derive utility from their ideological beliefs have distorted beliefs that can impact outcomes and behaviors (*e.g.*, Engelman et al., 2024; Drobner and Goerg, 2024; Dickinson, 2025).

<sup>13</sup>For the sake of parsimony, we assume that priors do not depend on the group affiliation of those making the moral normative statements.

- $\omega_{IG}^F$  and  $\omega_{OG}^F$  capture the weights on signals False attached to in-group and out-group statements, respectively.

Using the formulation in [Drobner and Goerg \(2024\)](#), we consider that the utility of holding a belief  $\hat{b}$  depends on three components: a direct utility component, a payoff associated with the accuracy of one’s subjective belief (relative to objectively accurate belief,  $b$ ), and a utility cost of any belief distortion, such that:

$$U_B = \alpha \hat{b} + \frac{1}{2} \left( 1 + 2b\hat{b} - \hat{b}^2 \right) P - \gamma(b - \hat{b})^2, \quad (2)$$

where:

- $\alpha \hat{b}$  is the direct utility from holding the belief  $\hat{b}$ , which may capture the intrinsic satisfaction from believing in a certain moral standard.
- The second term represents the payoff from belief accuracy relative to the true norm  $b$ , where  $P$  is the incentive for accurate beliefs. The precise specification results from the BDM elicitation procedure used in the Belief Updating task (see [Hill, 2017](#)).
- The last term represents a non-monetary utility cost of belief distortion, where  $\gamma$  penalizes deviations from the true norm  $b$ . It can be interpreted as the cognitive cost of constructing a story to persuade oneself that the social norm is lying when in fact, it is telling the truth.

Maximizing  $U_B$  with respect to  $\hat{b}$  yields the optimal belief:

$$\hat{b}^* = b + \frac{\alpha}{P + 2\gamma}. \quad (3)$$

Equations (1) to (3) show that group identification - through differential weighting of signals attached to in-group *vs.* out-group statements via the parameters  $\omega_{IG}$  and  $\omega_{OG}$  -, can affect belief updating. Also, the higher the direct utility component of the utility function,  $\alpha$ , relative to the cost of deviations from the true belief, the more inflated one’s beliefs will be regarding the norm relative to what is objectively true.

**Utility derived from action:** The utility from taking action  $a$  is assumed to depend on the updated belief about the social norm in the task,  $\hat{b}$ , as follows:

$$U_A(a; \hat{b}) = \mu a - \xi(a - \hat{b})^2, \quad (4)$$

where  $\mu > 0$  is the marginal benefit from the number of Heads reported (*e.g.*, action) and  $\xi > 0$  is the cost parameter for deviating from the belief about the norm.<sup>14, 15</sup>

To determine the optimal action  $a^*$ , we maximize  $U_A(a; \hat{b})$  with respect to  $a$ :

$$\frac{dU_A}{da} = \mu - 2\xi(a - \hat{b}) = 0, \quad (5)$$

$$\Rightarrow a^* = \hat{b} + \frac{\mu}{2\xi}. \quad (6)$$

This shows that the optimal reporting behavior depends on monetary incentives for over-reporting, the updated normative belief about the appropriate level of reporting  $\hat{b}$  (depending on group-weighted signals), and the cost of deviating from the social norm.

## 4.2 Hypotheses

A set of hypotheses was pre-registered in light of the literature and our theoretical framework. The first two hypotheses build on the previous research that has documented a statistically higher-than-expected number of favorable outcomes reported in repeated incentivized coin-flip tasks (*e.g.*, Abeler et al., 2014; Cohn et al., 2014, 2015; Conrads and Lotz, 2015; Garbarino et al., 2019; Dickinson and McEvoy, 2021; Dickinson and Masclet, 2023; Drupp et al., 2024). Previous work has also demonstrated that social norms, through peers' behavior or expectations, influence actions in environments where norm compliance conflicts with material payoffs (*e.g.*, Fortin et al., 2007; Gächter and Schulz, 2016; Bicchieri et al., 2022; Charroin et al., 2022; Dimant et al., 2024). This literature and our Equation (6) motivate the first two hypotheses. We anticipate that individuals will engage in cheating (*i.e.*, over-report favorable outcomes) when the monetary incentive ( $\mu$ ) is sufficiently high relative to the cost of deviating from the normative belief ( $\zeta$ ).

**Hypothesis 1 (Lying):** *Individuals report statistically significantly more than 50% of favorable outcomes (Heads) in the Coin Flip task.*

---

<sup>14</sup>The literature has shown that most individuals prefer to conform to a social norm by fear of social disapproval in case of deviation from others' normative expectations (*e.g.*, Brekke et al., 2003; Bicchieri, 2005; Krupka and Weber, 2013). The degree of conformity may not be uniform, notably if individuals assign different weights to the expectations and behavior of others, characterized in particular by their group identity or proximity (*e.g.*, Pickup et al., 2020; Bicchieri et al., 2022; Schneeberger and Krupka, 2024). Here, this is captured through the weight assigned to the signals depending on the source of the statements.

<sup>15</sup>Note that cheating could also entail personal moral costs and not only reputational costs; we omit these costs for the sake of parsimony.

**Hypothesis 2 (Compliance with Normative Expectations):** *Perceptions of how others view social acceptability of lying in the Coin Flip task predict the number of favorable outcomes reported (i.e., the more one thinks others view over-reporting as acceptable, the higher is the number of Heads one reports).*

The existence of normative uncertainty, and even pluralistic ignorance in some settings, is now well established in the literature. Introducing new information can help individuals update their understanding of prevailing social norms. However, laboratory research has shown that, from a general perspective, if individuals tend to update their beliefs in line with Bayes' rule, they often fail to fully weigh new signals as Bayes' rule would predict (*e.g.*, Charness and Levin, 2005; Holt and Smith, 2009; Coutts, 2019). This motivates Hypothesis 3.

**Hypothesis 3 (Belief Updating):** *Participants update their beliefs in a cautiously Bayesian way: they assign positive weight to both prior beliefs and new information signals when forming posterior beliefs, but the weights assigned are statistically lower than those predicted by Bayes' rule.*

Hypothesis 4 is derived from our modeling of belief-based utility, in particular Equations (1) and (2), indicating that individuals may place different values on signals, depending notably on whether the statements reflect in-group or out-group normative views.

**Hypothesis 4 (Group Identity):** *The estimated weight assigned to new information during belief updating regarding normative statements differs based on whether the information relates to individuals outside one's affinity group or to in-groups.*

The reasoning underlying Hypothesis 2 can be extended to our dynamic setting. Specifically, if individuals update their normative expectations regarding how appropriate others consider misreporting behaviors in the Coin Flip task, this adjustment is anticipated to influence their subsequent behavior in the second Coin Flip task, as stated in Hypothesis 5:

**Hypothesis 5 (Updating of Expectations and Evolution of Behavior):** *The difference between initial and posterior normative expectations about misreporting in*

the Coin Flip task predicts the difference between the initially reported Coin Flip task outcome and the subsequent outcome after completing the Updating task.<sup>16</sup>

## 5 Results

We first test Hypothesis 1 by examining whether the average number of Heads reported in the Coin Flip task is statistically greater than 5 out of 10 coin flips. Across all N=450 participants, we found that during the first administration (before the Belief Updating task), the average number of Heads reported was 5.40 (st. dev. = 1.66), and during the second administration (after the Belief Updating task), the average number of Heads reported was 5.28 (st. dev. = 1.77). Both means are statistically significantly greater than 5, using a one-sample Z-test ( $p < .001$  in both instances), and they are not significantly different from each other (matched pairs Wilcoxon signed-rank test,  $p = .216$ ). Hypothesis 1 is supported.

**Result 1 (Lying):** *Individuals reported statistically significantly more than 50% of favorable outcomes in the Coin Flip task.*

Hypothesis 2 examines the relationship between perceived moral norms and individual behavior. We regressed a measure of the participant’s normative expectations on the number of Heads reported by that participant in the Coin Flip task. We did this separately for the first administration of the Coin Flip task, using the original perceived norm as the key regressor, and its second administration, using the re-assessment of the norm after completing the Belief Updating task. Both *Pre-Mean Normative Expectation* and *Post-Mean Normative Expectation* were defined on the [0,100] interval as an average of the responses given to the four questions that described scenarios of over-reporting, as explained in Section 3. The averaged 0-100 scale measures serve as an indicator of the perceived social acceptability of over-reporting by a typical participant.

Table 3 presents simple linear regressions (columns (1) and (3)) and specifications with additional sociodemographic controls (columns (2) and (4)).<sup>17</sup> Note that through-

---

<sup>16</sup>Note that the pre-registered hypothesis stated: “The difference in one’s initial and posterior belief regarding truthfulness of a Coin Flip task norms statement will predict the difference between one’s initial reported Coin Flip task outcome and a subsequent Coin Flip task outcome assess after the Bayesian task.” While similar in spirit, we believe it makes more sense to consider the evolution of normative expectations rather than the evolution of the belief about the statements’ truthfulness.

<sup>17</sup>Socio-demographics data were not available from Prolific on all n=450 participants, given that profile entry data can expire or participants can choose not to share some information. Employment and student status

out the results section, our tables highlight statistical significance from the two-tailed test of the null hypothesis. However, our preregistered hypotheses imply that the more statistically powerful one-tailed test is appropriate in testing the significance of coefficients related to such preregistered hypotheses.

Table 3: Effect of perceived norms on coin flip reports

<i>Dependent variable:</i> # HEADS reported	1st Administration		2nd Administration	
	(1)	(2)	(3)	(4)
Pre - Mean Normative Expectation	.007** (.003)	.008** (.004)	-	-
Post - Mean Normative Expectation	-	-	.013*** (.003)	.011*** (.004)
Democrat (=1)	-	-.051 (.189)	-	-.163 (.196)
Age (years)	-	-.015** (.007)	-	-.030*** (.007)
Female (=1)	-	.094 (.190)	-	-.122 (.195)
Minority (=1)	-	-.276 (.222)	-	-.291 (.228)
Student (=1)	-	-.504* (.262)	-	-.564** (.270)
Employed (=1)	-	.091 (.189)	-	-.196 (.194)
Constant	5.180*** (.133)	5.816*** (.423)	4.854*** (.142)	6.635*** (.427)
Observations	450	368	450	368
R-squared	.010	.038	.030	.079

*Notes:* OLS regressions. The dependent variable is the number of Heads reported. Standard errors are in parentheses. \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2-tailed tests. Note that the more statistically powerful 1-tailed test is appropriate for any preregistered hypothesis test. The number of observations is lower in the models that include sociodemographic controls compared to those without them because Prolific did not provide all sociodemographic measures for all participants.

The regressions in Table 3 strongly support Hypothesis 2. Participants' mean normative expectations about the acceptability of over-reporting are associated with a

---

were the most common missing characteristics due to data expiration. We had n=194 Republicans (out of 250) who provided data to score an *Employed* indicator variable, and n=199 Democrats (out of 250). *Student status* was available on n=194 Republican and n=197 Democrat participants. *Minority status* was available on all participants, n=4 total participants (2 Republicans, 2 Democrats) chose not to report sex (so, missing *Female* indicator data), and n=5 participants (3 Republicans, 2 Democrats) did not have available data on *Age*.

higher number of Heads reported. This effect is both larger and more precisely estimated after the Belief Updating task (columns (3-4)). This result, in line with, *e.g.*, Dimant et al. (2024), motivates Result 2:

**Result 2 (Compliance with Normative Expectations):** *Perceptions of how others view social acceptability of over-reporting in the Coin Flip task correlates with the number of favorable outcomes reported in that task.*

To test Hypothesis 3, we estimated a belief-updating model using the panel data set produced by the Belief Updating task. This allows us to assess the degree to which participants’ belief updating aligns with Bayes’ rule (*i.e.*, appropriately weighting prior beliefs and new information from noisy signals). We employed the baseline log-odds model of Bayes’ rule commonly used in the literature (*e.g.*, Holt and Smith, 2009; Hill, 2017; Coutts, 2019). Following Coutts (2019), we suppressed the constant term to directly estimate the influence of *True* and *False* signals on participants’ belief updating. According to Bayes’ rule, receiving a *True* signal should increase posterior belief, whereas receiving a *False* signal should decrease posterior belief in the likelihood that the statement is “true”. The belief updating model is specified as follows:<sup>18</sup>

$$\begin{aligned} \text{logit}(\hat{p}_t) = & \delta \cdot \text{logit}(\hat{p}_{t-1}) + \beta_1 \cdot \mathbf{I}\{s_t = 1\} \cdot \ln(LR_1) \\ & + \beta_0 \cdot \mathbf{I}\{s_t = 0\} \cdot \ln(LR_0) + e_{it} \end{aligned} \quad (7)$$

Equation (7) estimates the posterior belief at time  $t$ , expressed as the log-odds of the statement being true,  $\text{logit}(\hat{p}_t) = \ln\left(\frac{\hat{p}_t}{1-\hat{p}_t}\right)$ , is a function of the prior log-odds ( $\text{logit}(\hat{p}_{t-1})$ ) and the new evidence provided by noisy signals. The terms  $\ln(LR_1)$  and  $\ln(LR_0)$  represent the log-likelihood ratios for *True* ( $s_t = 1$ ) and *False* ( $s_t = 0$ ) signals, respectively.  $\beta_1$  and  $\beta_0$  capture the weight participants place on these signals when updating their beliefs. The indicator functions  $\mathbf{I}\{s_t = 1\}$  and  $\mathbf{I}\{s_t = 0\}$  ensure that the model separately estimates the effect of *True* and *False* signals.  $\delta$  measures the influence of prior beliefs on posterior beliefs. Recall that we set the signal strength in our design at  $\frac{3}{4}$ . Therefore, participants should update their beliefs regarding the statement’s truth upward by  $\ln\left(\frac{3/4}{1/4}\right) = \ln(3)$  when receiving a *True* signal, and downward by

---

<sup>18</sup>Using log-odds, the regression is an approximation ( $\text{logit}(\hat{p}) = \hat{b}$ ) of the utility-based Equation (1) that expresses the updated belief as a weighted average of the prior and the signals, with  $\delta \approx \omega_{\text{prior}}$ , and  $\beta s \approx \omega^T$  and  $\omega^F$ . We ignore for the moment in the regression the group identity of the information source.

$\ln\left(\frac{1/4}{3/4}\right) = \ln\left(\frac{1}{3}\right)$  when receiving a *False* signal.

Table 4: Determinants of belief updating

<i>Dependent variable:</i> Posterior beliefs = lnOdds(t)	(1)	(2)
lnOdds(t-1) = Prior beliefs ( $\delta$ )	.773*** (.020)	.775*** (.022)
<i>True</i> Signal * lnLR ( $\beta_1$ )	.613*** (.038)	.586*** (.076)
<i>False</i> Signal * lnLR ( $\beta_0$ )	.621*** (.039)	.667*** (.082)
Democrat (=1)	–	.028 (.035)
Age (years)	–	–.000 (.001)
Female (=1)	–	–.017 (.037)
Minority (=1)	–	.108** (.045)
Student (=1)	–	–.015 (.051)
Employed (=1)	–	.005 (.035)
Tests of Bayesian model outcomes	<i>Model(1)tests</i>	<i>Model(2)tests</i>
Test: $H_0 : \delta = 1$ (Fully weighted priors)	$F(1, 449) = 124.17^{***}$	$F(1, 449) = 100.25^{***}$
Test: $H_0 : \beta_1 = 1$ (Fully weighted True signals)	$F(1, 449) = 103.25^{***}$	$F(1, 449) = 29.47^{***}$
Test: $H_0 : \beta_0 = 1$ (Fully weighted False signals)	$F(1, 449) = 93.87^{***}$	$F(1, 449) = 16.55^{***}$
Test: $H_0 : \beta_1 = \beta_0$ (Equal True & False signal weights)	$F(1, 449) = .08$	$F(1, 449) = .36$
Observations	7,200	5,888
R-squared	.646	.650

Notes: The dependent variable is the posterior belief in trial  $t$ . Robust standard errors, clustered at the individual level, are in parentheses. \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2-tailed tests. Note that the more statistically powerful 1-tailed test is appropriate for any preregistered hypothesis test.

Table 4 presents the estimates of this model using the full dataset (pooled across treatments), both without socio-demographic controls (column (1)) and with socio-demographic controls (column (2)). Given the panel structure of the dataset, with 16 belief-updating observations per participant, we estimated the models with robust standard errors clustered at the individual level.

The regressions in Table 4 provide support for Hypothesis 3. Participants positively weighted both prior beliefs and new evidence when forming posterior beliefs. However, the estimated coefficients for  $\delta$ ,  $\beta_0$ , and  $\beta_1$  are all significantly less than 1, indicating cautious Bayesian updating, as labeled by Hill (2017). The tests at the bottom of Table 4 show no significant differences in the extent to which participants weighted *True vs. False* signals during the task. This supports Result 3:

**Result 3 (Belief Updating):** *Participants updated their beliefs in a cautiously Bayesian way: while they assigned positive weight to both prior beliefs and new signals when forming posterior beliefs, these weights were statistically lower than those predicted by Bayes’ rule.*

Hypothesis 4 about the impact of group identity on belief updating constitutes the main hypothesis of this paper. To test it, we first coded trials based on the treatment assignment and the participant’s binary political group identification. A trial could be either *In-Group* = 1 (the participant shares the same political identity, *e.g.*, Democrat or Republican, with the group referenced in the trial), *Out-Group* = 1 (the participant’s political identity differs from the group referenced in the trial), or *No-Identity* = 1 (the trial serves as the reference group where no political identity is mentioned, *i.e.*, *In-Group* = 0 and *Out-Group* = 0). We estimated the following model, in which  $\eta_1 \cdot (InGroup_t \cdot \ln(LR_t))$  measures whether receiving a signal about an in-group normative statement affects the weight participants assign to the signal in their belief updating process;  $\eta_2 \cdot (OutGroup_t \cdot \ln(LR_t))$  measures whether receiving a signal about an out-group normative statement changes the weight participants assign to the signal when updating their beliefs. Since we found no significant difference between the effects of *True vs. False* signals (see the  $\beta_1$  and  $\beta_0$  coefficient estimates in Table 4), in this model we collapse these differences and focus only on the effects of the signal’s strength (as measured by  $\ln(LR_t)$ ) and its interaction with group identity.<sup>19</sup>

$$\begin{aligned} \text{logit}(\hat{p}_t) = & \text{constant} + \delta \cdot \text{logit}(\hat{p}_{t-1}) + \beta \cdot \ln(LR_t) + \gamma_1 \cdot InGroup_t + \gamma_2 \cdot OutGroup_t \\ & + \eta_1 \cdot (InGroup_t \cdot \ln(LR_t)) + \eta_2 \cdot (OutGroup_t \cdot \ln(LR_t)) + e_{it} \end{aligned} \tag{8}$$

Instead of only relying on the participants’ reported political affiliation (Democrat or Republican), we also elicited a more granular measure of ideological strength at the

---

<sup>19</sup>As for Equation (7), this logit linear regression is an approximation to the belief updating process described in the utility-based Equation (1). Since our empirical estimation found no significant differences between the effects of True and False signals, the model in Equation (1) simplifies to:

$$\hat{b} = \omega_{\text{prior}} \hat{b}_{\text{prior}} + [\omega_{IG} \mathbf{1}\{\text{In-group}\} + \omega_{OG} \mathbf{1}\{\text{Out-group}\}] s,$$

where  $\omega_{IG}$  and  $\omega_{OG}$  represent the effective weights on the signal attached to in-group and out-group statements, respectively. Taking the logit transformation (letting  $\hat{b} = \text{logit}(\hat{p})$ ) in Equation (8),  $\delta \approx \omega_{\text{prior}}$ ,  $\beta \approx \omega_s$ ,  $\beta + \eta_1 \approx \omega_{IG}$ ,  $\beta + \eta_2 \approx \omega_{OG}$ .

beginning of the study, *Liberal Score*  $\in [1, 9]$  (with 1 = Very Conservative, 5 = Middle Of The Road, 9 = Very Liberal). This allowed us to code variables to proxy the degree of in-groupness or out-groupness between the participant and the normative-statement source in that trial, because those who nominally identified with one political affiliation or the other may feel a stronger or weaker attachment to that affiliation. With this 9-point scale, we constructed *In-GroupDegree*  $\in [1, 9]$  to represent the degree of alignment between the in-groupness of a statement and one’s political party identification, and *Out-GroupDegree*  $\in [1, 9]$  to represent the degree of misalignment between the out-groupness of a statement and one’s party. Appendix D gives the details of the construction of these variables and replicates all the regression analyses of the main text, replacing the binary *In-Group* and *Out-Group* variables with the continuous *In-GroupDegree* and *Out-GroupDegree* variables. The results are always qualitatively similar.

Table 5 reports the estimates that the statement is true from a set of specifications using the *In-Group* and *Out-Group* binary indicators (see Table D1 for those using the continuous measures). In all regressions, the key variable to test Hypothesis 4 is the interaction terms (the  $\theta$  coefficients) between the group variables and the log-likelihood ratio (lnLR) variable (highlighted in bold fonts). In columns (1) to (4) of both tables, we pooled observations from the *In-Group*, *Out-Group* and *No-Identity* treatments, excluding observations from the *Mixed* treatment. Columns (5) to (8) present estimates for the within-subjects models using only observations from the *Mixed* treatment.<sup>20</sup> Regressions in columns (1) and (5) include all participants from the specified treatments without sociodemographic controls, while those in columns (2) and (6) include sociodemographic controls. Finally, columns (3), (4), (7), and (8) present separate estimates for Democrat and Republican participants.

Table 5 (as Table D1 ) shows no evidence that participants weighted signals differently based on whether these signals were attached to an *In-Group* or an *Out-Group* moral norm statement, compared to a signal attached to a non-group-specific statement.

---

<sup>20</sup>In the *Mixed* treatment, statements were either *In-Group* or *Out-Group* (*i.e.*, no generic norm statements were included). Consequently, both group indicator variables cannot be included in these models as they are in models with data pooled across the other treatments.

Table 5: Bayes model estimates by In-Group or Out-Group norms statements and by group identity

Dep. var.:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Posterior belief	Between-subjects treatments All	Between-subjects treatments All	Between-subjects treatments Democrats	Between-subjects treatments Republicans	Mixed treatment All	Mixed treatment All	Mixed treatment Democrats	Mixed treatment Republicans
lnOdds(t-1) =	.770***	.765***	.789***	.734***	.782***	.801***	.847***	.762***
Prior belief ( $\delta$ )	(.024)	(.026)	(.031)	(.044)	(.041)	(.044)	(.040)	(.069)
lnLR (signals) ( $\beta$ )	<b>.591***</b>	<b>.606***</b>	<b>.604***</b>	<b>.607***</b>	<b>.592***</b>	<b>.569**</b>	<b>.564***</b>	<b>.563***</b>
IG trial ( $\gamma_1$ )	(.060)	(.066)	(.077)	(.107)	(.079)	(.089)	(.086)	(.147)
OG trial ( $\gamma_2$ )	-.025	-.007	.014	-.040	-.048	-.045	-.046	.011
	(.047)	(.051)	(.070)	(.075)	(.064)	(.068)	(.082)	(.104)
	.007	.032	.152**	-.120	-	-	-	-
	(.046)	(.053)	(.073)	(.083)	-	-	-	-
IG trial * lnLR ( $\eta_1$ )	<b>.095</b>	<b>.090</b>	<b>.130</b>	<b>.052</b>	<b>.035</b>	<b>.038</b>	<b>.135</b>	<b>-.041</b>
OG trial * lnLR ( $\eta_2$ )	(.096)	(.105)	(.120)	(.171)	(.069)	(.074)	(.095)	(.119)
	<b>-.011</b>	<b>.011</b>	<b>.91</b>	<b>-.081</b>	-	-	-	-
	(.079)	(.094)	(.133)	(.126)	-	-	-	-
Age (years)	-	-.002	-.003	-.000	-	.003	.001	.005*
	-	(.002)	(.002)	(.002)	-	(.002)	(.002)	(.003)
Female (=1)	-	-.041	-.070	-.039	-	.055	.053	.019
	-	(.046)	(.069)	(.063)	-	(.055)	(.075)	(.081)
Minority (=1)	-	.136**	.142*	.122*	-	.104*	.076	.107
	-	(.060)	(.082)	(.068)	-	(.058)	(.062)	(.101)
Student (=1)	-	-.031	-.119	.020	-	-.007	-.109	.105
	-	(.062)	(.092)	(.070)	-	(.071)	(.074)	(.126)
Employed (=1)	-	-.025	-.017	-.070	-	.055	.046	.069
	-	(.041)	(.058)	(.058)	-	(.059)	(.064)	(.077)
Constant	.016	.084	.123	.101	-.020	-.250*	-.099	-.456**
	(.033)	(.074)	(.112)	(.098)	(.039)	(.136)	(.122)	(.211)
Observations	5,280	4,240	2,176	2,064	1,920	1,648	800	848
R-squared	.636	.632	.675	.587	.676	.701	.784	.632

Notes: The dependent variable is the posterior belief that the statement is true in trial t. Robust standard errors, clustered at the participant level, are in parentheses. \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2-tailed tests. Note that the more statistically powerful 1-tailed test is appropriate for any preregistered hypothesis test. IG is for the in-Group, and OG for Out-Group. The between-subjects treatments (columns (1) to (4)) include the In-Group, Out-Group, and No-Identity treatments (they exclude the Mixed treatments). Mixed treatment trials were always either IG=1 or OG=1 trials, but the between-subjects treatment trials could also be IG=0 and OG=0 trials when the norm source was No-Identity.

That said, any proper test of Hypothesis 4 should take into account whether the norm information signal was consonant or dissonant with the participant’s baseline perception of the social norm. By task design, the mix of *True* and *False* signals supporting a lying norm or a truthful reporting norm means that participants likely encountered signals both more and less aligned with their normative expectations. Any differences in effects based on whether a signal was consonant or dissonant with these normative expectations may cancel each other out in the previous analysis that ignored this possibility. To address this, we coded a variable to capture the degree of signal dissonance with participants’ normative expectations before they started the Belief Updating task. We did not preregister the construction of this variable; therefore, this analysis is exploratory.

To account for signal dissonance, we considered each statement, the signal, and the participant’s initial normative expectation regarding the misreporting behavior described in this statement. For example, a *True* signal for Statement 1 reinforced the belief in a lying norm.<sup>21</sup> Such a signal would be dissonant for a participant whose *Pre-Normative Expectation* was low, indicating that they believed others would generally view the described lying behaviors as highly unacceptable, as reflected in lower scores on the [0,100] response scale average. In this case, we calculated the *Signal Norm Dissonance Degree* as  $100 - \text{Pre-Normative Expectation}$ . For example, an individual with *Pre-Normative Expectation* = 23 would have a *Signal Norm Dissonance Degree* = 77, indicating a relatively high degree of dissonance between the signal for Statement 1 and the individual’s initial normative expectation.<sup>22</sup> On the other hand, if a *False* signal was received for Statement 1, then the *Signal Norm Dissonance Degree* was set equal to the participant’s *Pre-Normative Expectation* because the *False* signal aligns with a truthful reporting norm.

This approach was applied consistently across all trials to score the dissonance degree of each signal for each statement and each participant. Importantly, the construction of the *Signal Norm Dissonance Degree* is independent of participants’ political ideology. It strictly measures the signal’s alignment (or misalignment) between the signal and the participant’s initial perception of the social norm.

---

<sup>21</sup>Recall that Statement 1 was: “A significant percentage of [GROUP IDENTIFIER] surveyed think it is acceptable to over-report the number of HEADS actually flipped in this task.” This statement is true and supports the perception of a norm favoring lying.

<sup>22</sup>Conversely, someone who believed that others generally consider the lying behavior relatively acceptable might have *Pre-Normative Expectation* = 82, for example, resulting in a *Signal Norm Dissonance Degree* = 18, reflecting a low degree of dissonance with the signal.

We modified the original specification of the posterior belief at time  $t$  from Equation (8), pooling the log-likelihood ratios for the *True* and *False* signal types in the estimation to obtain the baseline specification. We augmented the model by including the *Signal Norm Dissonance Degree* as an additional variable, along with an interaction term between this degree and the log-likelihood ratio. The coefficient  $\psi$  represents the direct impact of the signal’s dissonance on the posterior belief, and  $\eta$  measures how the degree of dissonance moderates the weight participants assigned to the signal during the updating process. We obtain:

$$\begin{aligned} \text{logit}(\hat{p}_t) = & \text{constant} + \delta \cdot \text{logit}(\hat{p}_{t-1}) + \beta \cdot \ln(\text{LR}_t) + \gamma_1 \cdot \text{InGroup}_t + \gamma_2 \cdot \text{OutGroup}_t \\ & + \eta_1 \cdot (\text{InGroup}_t \cdot \ln(\text{LR}_t)) + \eta_2 \cdot (\text{OutGroup}_t \cdot \ln(\text{LR}_t)) \\ & + \psi \cdot \text{SignalNormDissonanceDegree}_t + \theta \cdot (\text{SignalNormDissonanceDegree}_t \cdot \ln(\text{LR}_t)) + e_{it} \end{aligned} \quad (9)$$

Table 6 presents the estimates of this model, pooling data from the three between-subjects treatments (*In-Group*, *Out-Group* and *No-Identity*) in column (1), as well as for each treatment analyzed separately in columns (2) to (4), and for the mixed treatment in column (5) (see Table D2 for the continuous measures of groupness). The key coefficient estimate is the interaction term *Signal Norm Dissonance Degree \* ln(LR)*, highlighted in bold for emphasis.

Table 6 (as Table D2) shows a consistently significant negative impact of signal dissonance on the weighting of new information, indicating that signals more dissonant with the initially perceived moral norm were weighted less heavily during the updating process. Figure 1 illustrates how the signal’s dissonance degree is estimated to reduce the weight placed on the information signal when attached to an out-group and in-group statement (*i.e.*, estimated effects from models (2) and (3) in Table 6). However, a test on the slope difference from the pooled treatments estimation (*i.e.*, the interaction effect slope) concludes that this difference is not statistically significant at conventional levels ( $p > .10$ ) in the fully specified model with sociodemographic controls, and marginally significant (t-test on the coefficient estimate,  $p < .10$ ) in the model without controls that has more observations.<sup>23</sup>

---

<sup>23</sup>In Table 6, the separate estimates of the effects of the signal norm dissonance in the In-Group treatment (column (2)) and the Out-Group treatment (column (3)) are presented side-by-side to facilitate the interpretation. However, to test the significance of the difference between the coefficients of this variable in the two treatments, we pooled the data from the In-Group and Out-Group treatments, and included an *In-Group* indicator with interaction variables. The test reported in Figure 1 is on the significance of the triple-interaction

Table 6: Bayes model estimates, focusing on the degree of dissonance between the signal and the initial normative expectations

Dep. var.:	(1)	(2)	(3)	(4)	(5)
Posterior belief	Between-subjects treatments	Out-Group treatment	In-Group treatment	No Identity treatment	Mixed treatment
lnOdds(t-1) = Prior belief ( $\delta$ )	.746*** (.027)	.761*** (.047)	.687*** (.048)	.786*** (.047)	.793*** (.046)
<b>lnLR (signals) (<math>\beta</math>)</b>	<b>.848***</b> <b>(.086)</b>	<b>.764***</b> <b>(.118)</b>	<b>1.075***</b> <b>(.146)</b>	<b>.807***</b> <b>(.119)</b>	<b>.688***</b> <b>(.119)</b>
IG Trial (=1) ( $\gamma_1$ )	-.003 (.052)	-	-	-	-.043 (.069)
OG Trial (=1) ( $\gamma_2$ )	.034 (.055)	-	-	-	-
IG Trial * lnLR ( $\eta_1$ )	.096 (.105)	-	-	-	.042 (.074)
OG Trial * lnLR ( $\eta_2$ )	-.014 (.094)	-	-	-	-
Signal Norm Dissonance Degree $\in [0, 100]$ ( $\psi$ )	-.002** (.001)	.002 (.001)	-.006*** (.002)	-.002 (.002)	-.002 (.001)
<b>Signal Norm Dissonance Degree * lnLR (<math>\theta</math>)</b>	<b>-.005***</b> <b>(.001)</b>	<b>-.003*</b> <b>(.001)</b>	<b>-.007***</b> <b>(.002)</b>	<b>-.004**</b> <b>(.002)</b>	<b>-.002**</b> <b>(.001)</b>
Age (years)	-.002 (.002)	-.002 (.003)	-.001 (.002)	-.000 (.003)	-.003 (.002)
Female (=1)	-.040 (.047)	.058 (.080)	-.062 (.083)	-.111 (.087)	.053 (.056)
Minority (=1)	.143** (.062)	.177 (.123)	.236** (.115)	.050 (.076)	.104* (.058)
Student (=1)	-.032 (.064)	.011 (.115)	-.072 (.124)	-.010 (.085)	-.003 (.071)
Employed (=1)	-.028 (.042)	.039 (.086)	.056 (.077)	-.132** (.061)	.056 (.060)
Constant	.208** (.102)	-.078 (.169)	.328* (.187)	.237* (.139)	-.161 (.137)
Observations	4,240	1,264	1,440	1,536	1,648
R-squared	.636	.640	.604	.669	.702

Notes: The dependent variable is the posterior belief that the statement is true in trial  $t$ . Robust standard errors, clustered at the participant level, are in parentheses. IG is for the in-Group, and OG for Out-Group. The between-subjects treatments (columns (1)) include the In-Group, Out-Group, and No-Identity treatments (it excludes the Mixed treatment). \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2-tailed tests.

Overall, this analysis, summarized in Result 4, does not support Hypothesis 4 regardless of whether the dissonance between the information signal and initial normative expectations is explicitly accounted for. Later, we report additional exploratory analyses examining whether the influence of a signal's degree of norm dissonance on its weighting in Bayesian judgments depends on whether the information signal supports misreporting or truth-telling.

**Result 4 (Group Identity):** *The estimated weight assigned to new information duration, Signal Norm Dissonance Degree \* In-Group \* ln(LR), to assess whether the moderating effect of the signal's dissonance degree differs by the group identity of the statement.*

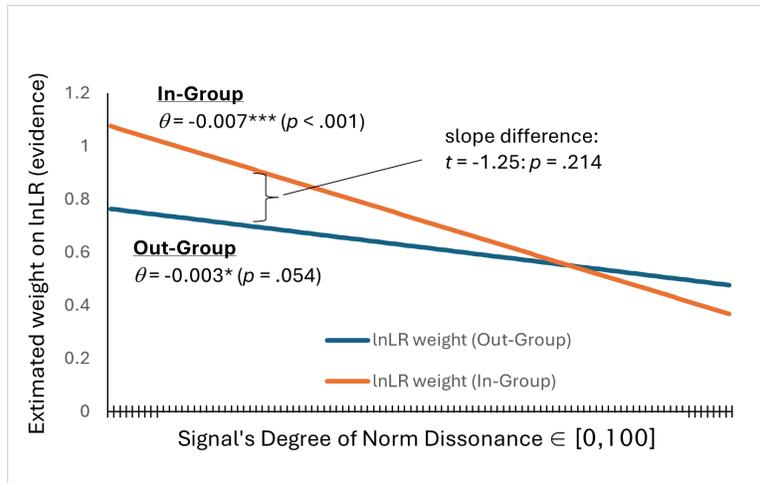


Figure 1: Group identity moderation effect on signal weights.

Notes: Significance of the slope difference was evaluated by pooling the Out-Group and In-Group treatments data, adding an In-Group indicator variable with interaction term, and examining the significance of the triple interaction term (Signal Norm Dissonance Degree \* In-Group \*  $\ln(LR)$ ) from the estimation results (model estimated as specified in Table 6). The slope difference is marginally significant  $p = .083$  in a model that does not control for socio-demographics ( $n=3520$  observations), compared to the model yielding the results above, which included sociodemographic controls but had fewer observations due to missing socio-demographics on some Prolific participants ( $n=2704$ ).

ing belief updating does not significantly differ based on the whether the signal is about an in-group or out-group normative statement, even when accounting for the dissonance between the signal and participants' initial normative expectations.

Finally, Hypothesis 5 states that the change in an individual's normative expectations, calculated as the difference between their mean normative expectations elicited before and after the Updating task (*i.e.*, *Post-Mean Normative Expectation* – *Pre-Mean Normative Expectation*), correlates with the change in the number of Heads reported across the two Coin Flip tasks. We test this hypothesis with one observation per participant reflecting the difference in outcomes.

While Result 2 supported Hypothesis 2, suggesting that perceived norms predict the number of Heads reported, the results in Table C2 in the Appendix C do not show that a change in mean normative expectations predicts a change in behavior in the Coin Flip task. The lack of support for Hypothesis 5 may be attributed to the minimal change observed in both the number of Heads reported and in the mean normative expectations between the two measurements. As indicated earlier in this section, two-tailed

Wilcoxon signed-rank tests on the matched-pairs data show no significant differences either between the average number of Heads reported before the Belief Updating task (5.40) and after (5.28) ( $p = 0.216$ ), or between the average *Post-Mean Normative Expectations* (33.50) and *Pre-Mean Normative Expectations* (33.61) ( $p = 0.220$ ). As visible in Figure B2 in the Appendix B, there is a clear mode at *Post-Pre Mean Normative Expectations Change* = 0, indicating very little to no change in expectations. This analysis supports our final result:

**Result 5 (Updating of Expectations and Evolution of Behavior):** *There is no evidence that the difference between an individual’s prior and posterior normative expectations about lying predicts changes in reporting behavior.*

**Exploratory Analysis** – While the analysis of Result 4 found no statistically significant difference in the moderating effect of *Signal Norm Dissonance Degree* on evidence weight when updating beliefs regarding in-group and out-group norms, we now examine whether the group identity of the statement influences the weight assigned to a signal (True or False) depending on whether this signal supports a truth-telling or lying norm.

We scored an indicator variable equal to 1 for cases where a new information signal supports a truth-telling norm, *Truth-Supporting Signal*, and another indicator for instances where the signal supports a lying norm, *Lie-Supporting Signal*. We then estimated models analogous to those in Table 6 for the separate subsets of *Truth-Supporting* and *Lie-Supporting* signal trials. Full results are presented in Tables C3 and C4 in the Appendix C, respectively (with Tables D3 and D4 for the continuous measures), while Figures 2 and 3 summarize the key forecasts, mirroring those in Figure 1.

Figure 2 shows that the qualitative pattern observed in Figure 1 holds when focusing only on the subset of trials when the signal aligns with the truth-telling norm. Participants are more likely to discount a more dissonant signal supporting a truth-telling norm when it is attached to an in-group normative statement. In this case, the difference in slopes from a pooled-treatment estimation (*i.e.*, the interaction effect) between the in-group and out-group sources is statistically significant (t-test of the coefficient estimate,  $p = .024$ ).

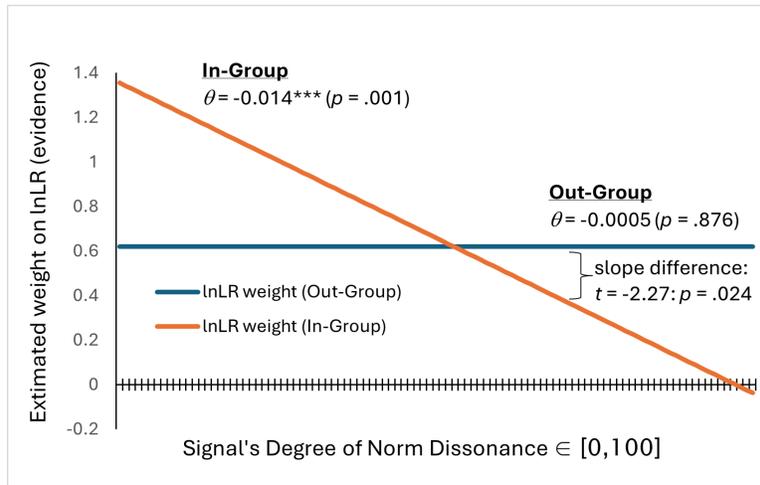


Figure 2: Group identity moderation effect on signal weights—WHEN SIGNAL ALIGNS WITH A TRUTH-TELLING NORM

Notes: Forecasts based only on trials where signal aligns with a truth-telling norm (n=1,338). Significance of the slope difference was evaluated by pooling the Out-Group and In-Group treatments data, adding an In-Group indicator variable with interaction term, and examining the significance of the triple interaction term (*Signal Norm Dissonance Degree \* In-Group \* ln(LR)*) from the estimation results (model estimated as specified in Table C3 in the Appendix). The slope difference is significant  $p = .008$  in a model that does not control for socio-demographics (n=1,745 observations), compared to the model yielding the results above, which includes sociodemographic controls but has fewer observations due to missing socio-demographics on some Prolific participants (n=1,338).

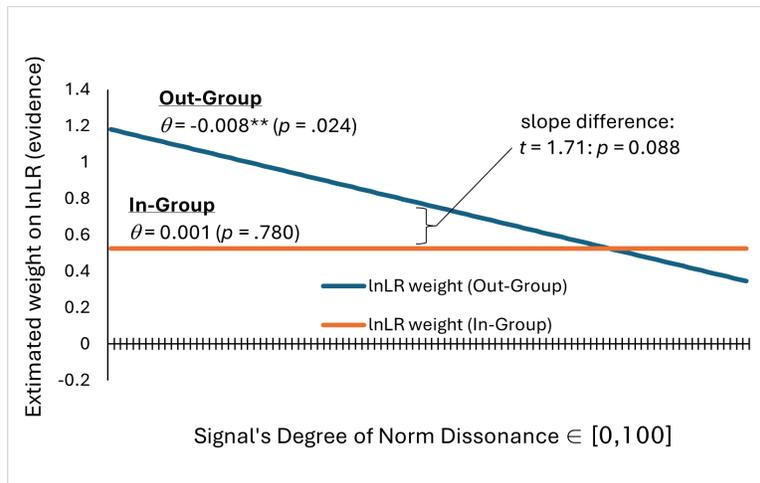


Figure 3: Group identity moderation effect on signal weights—WHEN INFORMATION SIGNAL ALIGNS WITH A LYING NORM

Notes: Forecasts based only on trials where signal aligns with a misreporting norm (n=1,366). Significance of the slope difference was evaluated by pooling the Out-Group and In-Group treatments data, adding an In-Group indicator variable with interaction term, and examining the significance of the triple interaction term (*Signal Norm Dissonance Degree \* In-Group \* ln(LR)*) from the estimation results (model estimated as specified in Table C4 in the Appendix). The slope difference on the interaction term of the pooled treatments model is marginally significant here (t-test test,  $p = .088$ ) in a model that controls for socio-demographics (n=1,366). In the model that does not include socio-demographics (n=1,775 observations), the difference is statistically insignificant at conventional levels ( $p = .132$ ).

In contrast, Figure 3 reveals that in trials when the signal aligns with a lying norm, the decision weight placed on evidence is moderated negatively by the signal’s degree of dissonance with the participant’s honesty norm, but only when the signal is attached to an out-group normative statement. That is, participants are more likely to discount a more dissonant signal, but only when it is attached to an out-group normative statement. The difference in slopes from a pooled-treatment estimation between the in-group and out-group sources does not reach significance at a standard level ( $p = .088$ ).

These findings are intriguing, given what a dissonant information signal means in each instance. When the signal aligns with a truth-telling norm, a higher *Signal Norm Dissonance Degree* indicates that the participant believes in a lying norm. In this scenario, participants place *less* weight on the truth-promoting signal when it is attached to an in-group normative statement. This suggests that individuals who perceive misreporting as the norm are reticent to update their beliefs when confronted with a contradictory signal about their own group. One possible interpretation is that this is a way to manage cognitive dissonance self-servingly by avoiding inconvenient information that challenges their perceived norm. This contrasts with the result in Table C4 and Figure 3 which indicates that participants discount a dissonant signal only when it challenges their perceived truth-telling norm *and* is attached to an out-group statement.

## 6 Discussion and Conclusion

In this study, we examined how group identity shapes belief updating when individuals receive signals about moral norms. We combined a repeated Coin Flip task, where over-reporting increases earnings, a Belief Updating task with statements about others’ moral judgments, and measures of prior and posterior normative expectations about how others judge misreporting. In line with the previous literature, we found evidence of lying in the Coin Flip task, whose extent varied with individuals’ perceptions of how others assess the acceptability of over-reporting. Overall, individuals updated their beliefs cautiously, weighting signals less than predicted by Bayes’ rule - a finding that may help explain the stickiness of social norms frequently observed (*e.g.*, Young, 2015; Bicchieri, 2017; Andreoni et al., 2021; Andrighetto and Vriens, 2022) and pluralistic ignorance (*e.g.*, Bursztyn et al., 2020) when individuals may privately reject a norm but continue to conform to it because they believe others support it.

Our core hypothesis that signals attached to in-group *vs.* out-group moral norms

statements would differently affect belief updating regarding the social acceptability of cheating was not supported. Political group identity did not significantly influence how signals altered normative beliefs. This is consistent with [Dimant et al. \(2024\)](#) who found that political group identity influenced the selection of the source of information but not expectations, but contrasts with [Bauer et al. \(2023\)](#) and [Dickinson \(2025\)](#), who both identified a group identity effect on belief formation. The discrepancy between these findings, although all studies kept constant the signals' and statements' informative quality across political affiliations, suggests that the effect of political identity on belief formation depends on the environment (social norms and lying in one case, economic perspectives in the other). Individuals may expect that various groups have different views on economic perspectives but not different perceptions of the social norm in the domain of cheating.

However, we should not conclude definitively on the absence of any effect of group identity on belief updating about the social norm. Exploratory analysis shed light on this issue as an area for future research. Indeed, when a new information signal was more dissonant with the participant's perceived social honesty norm, the resulting down-weighting of the signal in one's updating was greater and varied depending on two factors: whether the information signal was attached to an in-group *vs.* out-group normative statement, and whether the signal promoted honesty *vs.* dishonesty as the norm.

The asymmetry was striking. When a signal that aligned with an in-group honesty norm was highly dissonant with a participant's baseline perceived norm, it was down-weighted more than an equivalent signal attached to an out-group normative statement. Conversely, a signal promoting dishonesty and dissonant with one's baseline perceived norm was only down-weighted when it related to an out-group normative statement. Thus, the more one perceived the social norm as favoring honesty, the more a signal endorsing dishonesty by out-groups was discounted. Conversely, if one perceived the norm as favoring misreporting, a signal promoting truthful reporting by in-groups was less influential on beliefs than a signal promoting truthful reporting by out-groups.

One interpretation is that individuals endorsing lenient moral norms cannot avoid contradictory information in our setting, but they can minimize the informational value of such inconvenient signals, especially when they are attached to valued normative sources. This asymmetry in belief updating, where peer and identity factors lead to a preference for less ethical norms, is a clear area for future research.

Finally, we found no evidence that changes in average posterior normative expect-

tations regarding morally acceptable behavior (elicited independently of political affiliation) influenced the evolution of cheating behavior. Exclusive exposure to norm-informative signals attached to in-group or out-group normative statements did not alter participants' perception of the average social norm. This may reflect strong inertia in normative expectations, even in a context where signals were informative about the social norm.

We acknowledge several limitations. First, moral decision-making in the real world occurs in more diverse and complex social contexts than the stylized coin flip task used in this experiment. Second, there may be a spurious connection between the norms we elicited and behavior; for example, individuals might report a higher perceived cheating norm immediately after cheating- a challenge likely present in field data as well, given that perceptions of norms are always embedded in everyday behavior. Third, it might be interesting to explore a setting in which the information signals themselves emanate from sources with a diverse group identity. Finally, our study focuses on cheating behavior, a context where moral rules tend to transcend political group boundaries. This was a deliberate design choice to ensure that the quality of signals and statements did not differ across treatment cells. Further research could explore belief updating about social norms in settings where groups hold divergent normative views, which would shed light on the complex interplay between belief updating about social norms and group identity.

## References

- Abeler, J., A. Becker, and A. Falk (2014). Representative evidence on lying costs. *Journal of Public Economics* 113, 96–104.
- Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics* 115(3), 715–753.
- Aksoy, B. and M. A. Palma (2019). The effects of scarcity on cheating and in-group favoritism. *Journal of Economic Behavior & Organization* 165, 100–117.
- Andreoni, J., N. Nikiforakis, and S. Siegenthaler (2021). Predicting social tipping and norm change in controlled experiments. *Proceedings of the National Academy of Sciences of the United States* 118(16).
- Andrighetto, G. and E. Vriens (2022). A research agenda for the study of social norm change. *Philosophical Transactions of the Royal Society A* 380(2227), 20200411.
- Balafoutas, L., N. Nikiforakis, and B. Rockenbach (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences of the United States* 111(45), 15924–15927.
- Banerjee, R., N. Datta Gupta, and M. C. Villeval (2018). The spillover effects of affirmative action on competitiveness and unethical behavior. *European Economic Review* 101, 567–604.
- Barron, K. (2021). Belief updating: does the ‘good-news, bad-news’ asymmetry extend to purely financial domains? *Experimental Economics* 24(1), 31–58.
- Bauer, K., Y. Chen, F. Hett, and M. Kosfeld (2023). Group identity and belief formation: A decomposition of political polarization. *CESifo Working Paper* 108595.
- Becker, G. M., M. H. DeGroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science* 9(3), 226–232.
- Benistant, J. and M. C. Villeval (2019). Unethical behavior and group identity in contests. *Journal of Economic Psychology* 72, 128–155.
- Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
- Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford: Oxford University Press.
- Bicchieri, C., E. Dimant, M. Gelfand, and S. Sonderegger (2022). Social norms and behavior change: The interdisciplinary research frontier. *Journal of Economic Behavior & Organization*.

- Bicchieri, C., E. Dimant, S. Gächter, and D. Nosenzo (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior* 132, 59–72.
- Brekke, K. A., S. Kverndokk, and K. Nyborg (2003). An economic model of moral motivation. *Journal of Public Economics* 87(9-10), 1967–1983.
- Bursztyn, L., G. Egorov, and S. Fiorin (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review* 110(11), 3522–3548.
- Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review* 110(10), 2997–3029.
- Charness, G. and D. Levin (2005). When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. *American Economic Review* 95(4), 1300–1309.
- Charroin, L., B. Fortin, and M. C. Villeval (2022). Peer effects, self-selection and dishonesty. *Journal of Economic Behavior & Organization* 200, 618–637.
- Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review* 99(1), 431–457.
- Chen, Y., S. X. Li, T. X. Liu, and M. Shih (2014). Which hat to wear? Impact of natural identities on coordination and cooperation. *Games and Economic Behavior* 84, 58–86.
- Cohn, A., E. Fehr, and M.-A. Marechal (2014). Business culture and dishonesty in the banking industry. *Nature* 516, 86–89.
- Cohn, A., M.-A. Marechal, and T. Noll (2015). Bad boys: How criminal identity salience affects rule violation. *Review of Economic Studies* 82, 1289–1308.
- Conrads, J. and S. Lotz (2015). The effect of communication channels on dishonest behavior. *Journal of Behavioral and Experimental Economics* 58, 88–93.
- Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics* 22(2), 369–395.
- d’Adda, G., M. Dufwenberg, F. Passarelli, and G. Tabellini (2020). Social norms with private values: Theory and experiments. *Games and Economic Behavior* 124, 288–304.
- Della Valle, N. and M. Ploner (2017). Reacting to unfairness: Group identity and dishonest behavior. *Games* 8(3), 28.
- Dickinson, D. L. (2025). Political ideology, emotion response, and confirmation bias. *Economic Inquiry* 63(1), 181–205.

- Dickinson, D. L. and D. Masclet (2023). Unethical decision making and sleep restriction: Experimental evidence. *Games and Economic Behavior* 141, 484–502.
- Dickinson, D. L. and D. M. McEvoy (2021). Further from the truth: The impact of moving from in-person to online settings on dishonest behavior. *Journal of Behavioral and Experimental Economics* 90, 101649.
- Dimant, E., F. Galeotti, and M. Villeval (2024). Motivated information acquisition and social norm formation. *European Economic Review* 167.
- Dimant, E., M. Gelfand, A. Hochleitner, and S. Sonderegger (2025). Strategic behavior with tight, loose, and polarized norms. *Management Science* 71(3), 2245–2263.
- Dimant, E. and T. Gesche (2023). Nudging enforcers: How norm perceptions and motives for lying shape sanctions. *PNAS Nexus*, pgad224.
- Drobner, C. and S. J. Goerg (2024). Motivated belief updating and rationalization of information. *Management Science* 70(7), 4583–4592.
- Drupp, M. A., M. Khadjavi, and R. Voss (2024). The truth-telling of truth-seekers: Evidence from online experiments with scientists. *CESifo Working Paper 10897*.
- Eil, D. and J. M. Rao (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics* 3(2), 114–138.
- Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives* 3(4), 99–117.
- Elster, J. (2009). Social norms and the explanation of behavior. In P. Hedstrom and P. Bearman (Eds.), *The Oxford Handbook of Analytical Sociology*, pp. 195–217. Oxford University Press.
- Engelmann, J. B., M. Lebreton, N. A. Salem-Garcia, P. Schwardmann, and J. J. van der Weele (2024). Anticipatory anxiety and wishful thinking. *American Economic Review* 114(4), 926–960.
- Fehr, E., U. Fischbacher, , and S. Gaechter (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature* 13(1), 1–25.
- Fortin, B., G. Lacroix, and M.-C. Villeval (2007). Tax evasion and social interactions. *Journal of Public Economics* 91(11), 2089–2112.
- Gächter, S. and J. F. Schulz (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature* 531(7595), 496—499.
- Garbarino, E., R. Slonim, and M. C. Villeval (2019). Loss aversion and lying behavior. *Journal of Economic Behavior & Organization* 158, 379–393.

- Garfield, Z. H., E. J. Ringen, W. Buckner, D. Medupe, R. W. Wrangham, and L. Glowacki (2023). Norm violations and punishments across human societies. *Evolutionary Human Sciences* 5, 5:e11.
- Gelfand, M. J. (2018). *Rule makers, rule breakers: How tight and loose cultures wire our world*. New-York: Simon & Schuster.
- Hill, S. J. (2017). Learning together slowly: Bayesian learning about political facts. *The Journal of Politics* 79(4), 1403–1418.
- Holt, C. and A. Smith (2009). An update on bayesian updating. *Journal of Economic Behavior & Organization* 69(2), 125–134.
- Klein, E. (2020). *Why We're Polarized*. Avid Reader Press / Simon & Schuster.
- Krupka, E. L. and R. A. Weber (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association* 11(3), 495–524.
- López-Pérez, R. (2008). Aversion to norm-breaking: A model. *Games and Economic Behavior* 64(1), 237–267.
- Panizza, F., E. Dimant, E. O. Kimbrough, and A. Vostroknutov (2024). Measuring norm pluralism and perceived polarization in u.s. politics. *Working Paper*. Available at SSRN: <https://dx.doi.org/10.2139/ssrn.4779225>.
- Pickup, M. A., E. O. Kimbrough, and E. A. de Rooij (2020). Identity and the self-reinforcing effects of norm compliance. *Southern Economic Journal* 86(3), 1222–1240.
- Ross Arguedas, A., C. Robertson, R. Fletcher, and R. Nielsen (2022). Echo chambers, filter bubbles, and polarisation: a literature review. *Reuters Institute for the Study of Journalism*.
- Schneeberger, A. and E. L. Krupka (2024). Determinants of norm compliance: Moral similarity and group identification. *SSRN Working Paper* <https://ssrn.com/abstract=3969227>.
- Williams, C. (2023, 10). Echo chambers: Social learning under unobserved heterogeneity. *The Economic Journal* 134(658), 837–855.
- Young, H. P. (2015). The evolution of social norms. *Annual Review of Economics* 7(1), 359–387.

# ONLINE APPENDIX

## A Online Appendix: Instructions

### A.1 Pilot survey (to generate data on moral norms by political ideology)

#### **Informed Consent:**

You are being asked to complete this online survey that will ask about your views on whether certain behaviors are acceptable or not. This is a very short survey that we estimate to take 5 minutes or less. We are not disclosing the criteria used to determine your eligibility for this study, but the inclusion criteria will be told to you at the end of the study.

Participation in this online survey is completely voluntary, your responses to this survey will remain completely confidential, the data will be securely stored, your name will not be recorded anywhere on this survey. The only identifier we will record will be your Prolific ID, which we as researchers cannot link to personally identifiable data of yours.

This survey is estimated to take 5 minutes to complete and your payment for successful survey completion will be \$1.00.

There is an attention-check questions within the survey to ensure that you are being attentive and reading each question prior to responding. Successful survey completion includes passing the attention check questions. **Failure to answer the attention-check questions correctly may jeopardize receiving your payment** for this Prolific study.

There are no known risks associated with this study beyond those associated with everyday life. Although this study will not benefit you personally, its results will help our understanding of how people view certain behaviors.

For additional information related to this questionnaire, contact Dr. David Dickinson, Department of Economics, Appalachian State University, at dickinsondl@appstate.edu. Appalachian State University's Institutional Review Board has determined this study to be exempt from review by the IRB Administration (828-262-4060).

- I Consent** (and I understand there is not complete disclosure regarding inclusion criteria for this study until after I have completed the study. **I wish to continue**)
- I do not consent** to participating and **do not wish to continue**

The following questions are **screening validation questions** to make sure we get the desired sample we advertised for this survey

What is your current age (in years)?

18 23 29 34 40 45 50 56 61 67 72

Years of age	
--------------	--

----- page break -----

What sex were you assigned at birth, such as on an original birth certificate?

- Female
- Male

----- page break -----

In general, what is your political affiliation?

- Republican
- Democrat

----- page break -----

As described earlier, we are interested in factors that influence the decisions you might make. In order for the results of this survey to be valid, **it is essential that you read all the instructions and questions carefully**. So we know that you have read these instructions, please place the slider below on the answer to (33+12)=? Thank you for taking the time to read these instructions.

0 10 20 30 40 50 60 70 80 90 100

My response	
-------------	--

----- page break -----

**Imagine the following task:**

A researcher asks a study participant to flip a coin ten times in private such the researcher cannot see the outcome of the flips (e.g., the participant is in a separate room, or is doing the study online). The participant is asked to report the results of the 10 flips (i.e., how many HEADS versus TAILS were flipped).

Of course, the coin should flip 5 HEADS and 5 TAILS, on average, though there is a possibility of flipping mostly HEADS or mostly TAILS for any one participant.

**Incentives-->** the study will pay the participant additional compensation for each HEADS reported, such that the more HEADS reported implies a higher payoff in the study.

**In the table below, please respond to whether you each task scenario describes behavior in this task that you view as ACCEPTABLE or UNACCEPTABLE.**

Note: we are not asking you to do the coin flip task. Rather, we are describing possibilities from participants in such a study and want to know your views on the acceptability (or not) of the

behavior described. There are not right or wrong answers, and your responses will not affect your compensation for participating in this study.

	This behavior is <b>ACCEPTABLE</b>	This behavior is <b>UNACCEPTABLE</b>
The participant <b>flips 3 HEADS</b> and <b>reports 3 HEADS</b>	<input type="radio"/>	<input type="radio"/>
The participant <b>flips 3 HEADS</b> but <b>reports 5 HEADS</b>	<input type="radio"/>	<input type="radio"/>
The participant <b>flips 3 HEADS</b> but <b>reports 7 HEADS</b>	<input type="radio"/>	<input type="radio"/>
The participant <b>flips 3 HEADS</b> but <b>reports 9 HEADS</b>	<input type="radio"/>	<input type="radio"/>
The participant <b>flips 5 HEADS</b> and <b>reports 5 HEADS</b>	<input type="radio"/>	<input type="radio"/>
The participant <b>flips 5 HEADS</b> but <b>reports 7 HEADS</b>	<input type="radio"/>	<input type="radio"/>
The participant <b>flips 5 HEADS</b> but <b>reports 9 HEADS</b>	<input type="radio"/>	<input type="radio"/>
The participant <b>flips 7 HEADS</b> but <b>reports 7 HEADS</b>	<input type="radio"/>	<input type="radio"/>
The participant <b>flips 7 HEADS</b> but <b>reports 9 HEADS</b>	<input type="radio"/>	<input type="radio"/>
The participant <b>flips 4 HEADS</b> but <b>reports 10 HEADS</b>	<input type="radio"/>	<input type="radio"/>

----- page break -----

Thank you for your responses.

**Inclusion criteria for this study:** Eligible participants were custom-screened to be US-based individuals who self-reported either their political affiliation as either "Democrat" or "Republican" in their Prolific profile data.

To finalize this survey, **please click "FINISH SURVEY" below.**  
(please do this to make sure your Completion Code registers for Prolific study payment)

## A.2 Main Study Survey (Coin Flip task and Updating task)

*[Condensed for space] [Commentary to aid reader is given in blue font within square brackets]*

### **Informed Consent:**

You are being asked to complete this online survey that includes reporting your political ideology, answering some mood/emotion questions, and completing two incentivized decision tasks (one of this examines your beliefs in the midst of uncertainty). The study is only open to individuals with a specific political ideology in their Prolific profile. You will not be told the specifics of the political ideology screening until after completion of the study.

Participation in this online survey is completely voluntary, your responses to this survey will remain completely confidential, the data will be securely stored, your name will not be recorded anywhere on this survey. The only identifier we will record will be your Prolific ID, which we as researchers cannot link to personally identifiable data of yours.

This survey is estimated to take 13 minutes to complete and your payment for successful survey completion will be \$2.25. Plus, **you will have the chance to earn bonuses of up to an additional \$2.00 based** on your responses in the incentivized decision tasks within the survey.

There may be multiple attention-check questions within the survey to ensure that you are being attentive and reading each question prior to responding. Successful survey completion includes passing the attention check question. **Failure to answer attention-check questions correctly may jeopardize receiving your payment** for this Prolific study.

There are no known risks associated with this study beyond those associated with everyday life. Although this study will not benefit you personally, its results will help our understanding of how people make decisions about politics.

For additional information related to this questionnaire, contact Dr. David Dickinson, Department of Economics, Appalachian State University, at dickinsondl@appstate.edu. Appalachian State University's Institutional Review Board has determined this study to be exempt from review by the IRB Administration (828-262-4060).

- I Consent** and wish to continue with this study
- I do not consent** to participating and **do not wish to continue**

---

**Before you start, please switch off phone/ e-mail/ music so that you can focus on this study.** Thank you!

----- page break -----

Please carefully enter your Prolific ID (or double check if it has auto-filled) \_\_\_\_\_

----- page break -----

The following questions are **screening validation questions** to make sure we get the desired sample we advertised for this survey

In general, what is your political affiliation?

- Republican
- Democrat

**A couple of quick questions regarding your political views.**

In terms of politics, do you consider yourself conservative, liberal, or middle-of-the-road?

- VERY CONSERVATIVE**
- Quite conservative
- Conservative
- Somewhat conservative
- MIDDLE OF THE ROAD**
- Somewhat liberal
- Liberal
- Quite liberal
- VERY LIBERAL**

----- page break -----

**Bonus Payment Task (Coin Flip)**

This next question asks you to complete a task that you will also complete later in the

experiment (same task twice). **A bonus payment in this study will be based on your report in just one of tasks** (either this "coin flip" task or the "coin flip" task you will complete later in this survey), but not both. There is a 50% chance your "coin flip" bonus payment may be paid based on your report from either coin flip task, so please respond here as if your response will determine your "coin flip" bonus payoff for this study, because it may.

---

**COIN FLIP TASK**

**INSTRUCTIONS:** This next question asks you to flip a coin 10 times and report the results (making note of the order of Heads and Tails outcomes). You will be paid based on the outcome of the coin flips, so ***please read the instructions on the next page carefully. Before advancing to the next page, first please get/find a quarter (or any coin with a Heads and Tails side to it) that you can flip and something to write with and then click the button below. Please use an actual coin and not a virtual coin flipper. Please do not flip the coin before advancing to the next page.***

---

I have read the instructions above, I have a quarter/coin and something to write with, and I am ready to start flipping once I click this button

---

page break

**This task offers you the chance for additional compensation, so please read carefully.**

Please take your coin and **Flip the coin 10 times** and record what you get for each flip along with the total number of HEADS you flip. Please only perform the 10 coin flips once. **Your payoff for this task will be \$0.10 (i.e., 10 cents) for each HEADS outcome from these coin flips**, and this payment will be paid to you as a BONUS within 3 days of completing the survey. For example, if you flip HEADS 3 times, your additional compensation (a BONUS payment in addition to the fixed payment you were promised completing the HIT) will be another \$0.30. Similarly, if you flip HEADS 6 times you will receive an additional \$0.60, etc.

**Please tell us below how many times you flipped HEADS out of your 10 total coin flips, and we will add 10 cents for each HEADS flipped to your HIT payoff as a bonus.**

0 1 2 3 4 5 6 7 8 9 10

---

Number of HEADS outcomes from flipping coin 10 times



---

page break

**Now, we'd like to get your baseline ratings on some mood/emotion states.**

Please rate how strongly you feel each of these emotions *right now*.

**Right now I feel.....**

	Zero level of this emotion (1)	(2)	(3)	Mid-Range level of this emotion (4)	(5)	(6)	Maximum level of this emotion (7)
Happy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Excited	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Surprised	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Satisfied	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Angry	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Irritated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Confused	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Regret	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Disgust	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

----- page break -----

Before you complete the next task in this study, we would first like to know your beliefs regarding how acceptable or unacceptable *others* view certain behaviors in the Coin Flip task.

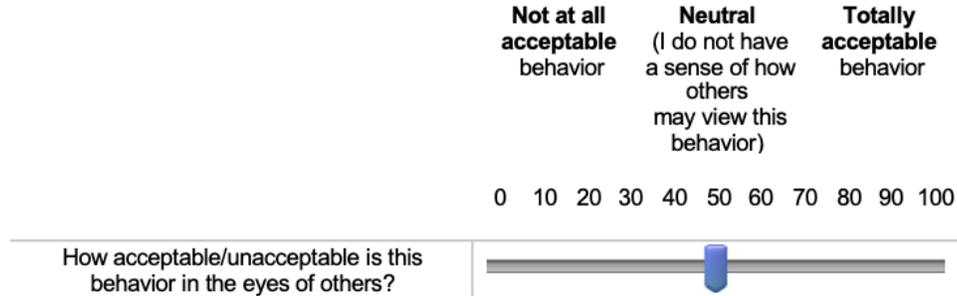
Remember, the Coin Flip task asks one to flip a coin in private 10 times and report the total number of HEADS flipped. One is paid for each HEADS reported, such that reporting more HEADS means a higher payoff for the individual.

Your beliefs will be assessed on a slider scale (0-100) for each behavior described. Lower scores on this slider mean you feel the behavior described would not be considered acceptable by most other people, and higher scores mean you think others would view the behavior as acceptable. **Remember, we are not asking whether YOU feel the behavior is acceptable or not, but rather we are asking you about how you think "OTHERS" (i.e., a typical person) view the acceptability of the behavior.**

----- page break -----

**BEHAVIOR** -----> Imagine that **someone over-reports the number of HEADS actually flipped** in this task.

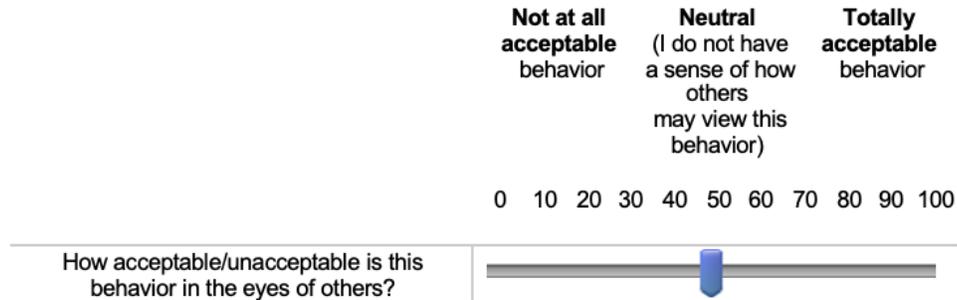
Using the slider below, please tell us how acceptable or unacceptable you think "others" (i.e., a typical person) feel this behavior is.



----- page break -----

**BEHAVIOR** -----> Imagine that **someone reports flipping 9 HEADS even though 3 HEADS** (out of 10) **were actually flipped** in this task

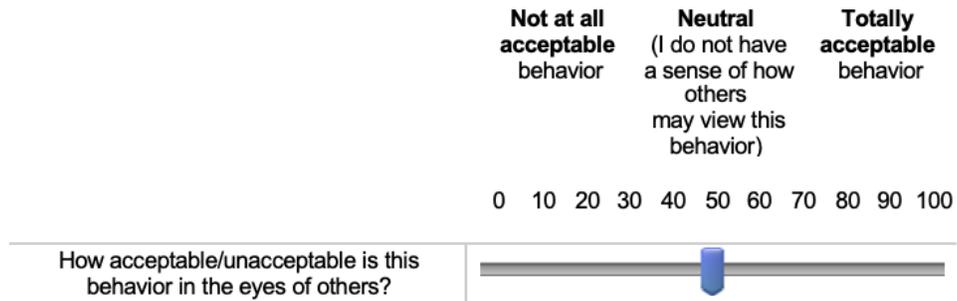
Using the slider below, please tell us how acceptable or unacceptable you think "others" (i.e., a typical person) feel this behavior is.



----- page break -----

**BEHAVIOR** -----> Imagine that **someone reports more HEADS than what he/she actually flipped** in this task.

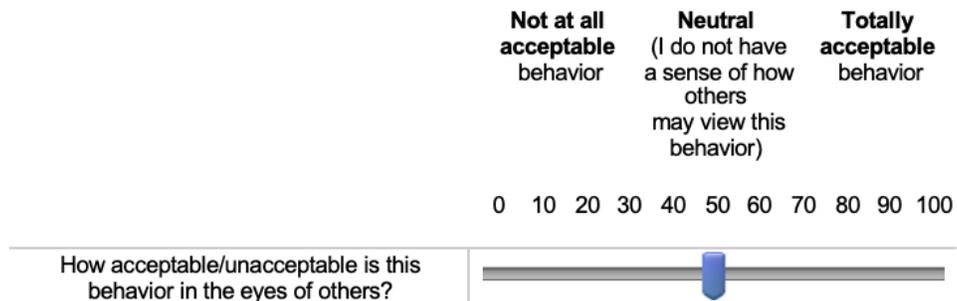
Using the slider below, please tell us how acceptable or unacceptable you think "others" (i.e., a typical person) feel this behavior is.



----- page break -----

**BEHAVIOR** -----> Imagine that **someone over-reports the number of HEADS flipped by just a few** (e.g., 3 or fewer extra HEADS reported out of 10 flips), but not by more than that.

Using the slider below, please tell us how acceptable or unacceptable you think "others" (i.e., a typical person) feel this behavior is.



----- page break -----

**STATEMENT EVALUATION TASK: INSTRUCTIONS**

In this next task, we will ask you to evaluate the truthfulness of 4 different factual statements over the course of 20 rounds (5 rounds per statement). "Factual" does not mean the statement is necessarily true, rather it means it is objectively either "TRUE" or "FALSE" as a matter of fact.

**Each statement will describe something about how REPUBLICANS view the acceptability or unacceptability of certain behaviors in the Coin Flip task that you**

completed earlier, based on a sample of survey data we obtained.

You will have a chance to earn up to \$1.00 in bonus payments (5 cents per round) based on your responses in this task (this bonus payment is independent of the "coin flip" task bonus payment).

**Here's how the task works**

We will first present to you a statement that may be true or false. We will then ask you to indicate how likely you believe the statement is true on a scale of 0 to 100. On this scale, "0" means you are certain the statement is *false* but "100" means you are certain the statement is *true*. A response of "50" means you are totally uncertain about the statement's truth. If you believe a statement is likely true but you are less certain, then a response like "65" or "83" (as examples) can be entered--the closer to "100" meaning you believe more strongly the statement is true. On the other hand, if you believe a statement is likely false (not true) but you are not totally sure, then a response like "14" or "41" (as examples) can be entered--the closer to "0" meaning you believe more strongly the statement is false.

The next page will tell you about how your response in each round determines whether or not you win that round.

----- page break -----

**INSTRUCTIONS FOR THE DECISION TASK (continued)**

Winning a \$0.05 bonus in each round depends on your response, and you will *maximize* your chance of the highest bonus by giving a response (belief) in each round that truly reflects how likely you think the statement is "true".

-----

**Here is how your response generates a bonus in each round of this task.**  
***(you can skip these shaded details if you are not interested in the underlying process).***

In each round, the computer will draw a random number from 0 to 100. Each number from 0 to 100 is equally likely to be drawn by the computer. We'll call this number Draw 1. How you win or lose that round of the task depends on what number the computer draws for Draw 1 and your response:

1) If Draw 1 is less than your response, you win if the statement is true and do not win if the statement is false. For example, if you enter a response of 99, you are very likely to win a bonus if the statement is true and very likely to not win if the statement is false. The higher your response, the more likely you win if the statement is true. Similarly, the lower your responses, the more likely you win if the statement if false.

2) If Draw 1 is greater than your response, then the computer will draw a second random number from 0 to 100. As before, each number from 0 to 100 is equally likely to be drawn by

the computer. We'll call this random number Draw 2. If Draw 2 is less than Draw 1, then you win the bonus for that round. If Draw 2 is greater than Draw 1, then you do not win the round.

---

Again, **the bonus payment process is designed so that *you have the best chance for earning a 5 cent bonus each round by being as accurate as possible with your response* (which can earn you up to a total bonus payment of 20 rounds times 5 cents, or \$1.00).** The random numbers and payment calculations will happen behind the scenes after you have finished the study. You will not see the draws in any round.

*Finally, **you will have a time limit on each page to submit your belief responses** (40 seconds when evaluating a statement for the first time, and 20 seconds for each subsequent evaluation of the same statement), so please stay attentive and input your response in a timely manner.*

---

page break

---

**INSTRUCTIONS FOR THE DECISION TASK (continued)**

We will ask for your belief whether a statement is true for each of 20 rounds. Each factual statement is presented more than once.

**When we repeat a statement, the computer will provide you with a signal about whether the correct answer is "TRUE" or "FALSE" for that statement.** The computer will present you a signal "TRUE" or "FALSE." Part of the task is that three out of every four signals are correct, on average. That is, if the statement is factually true, the computer will signal "TRUE" three out of four times and "FALSE" one out of four times. If the statement is factually false, the computer will signal "FALSE" three out of four times and "TRUE" one out of four times. A new independent signal is produced every time you are presented with the same statement in a new round. You will not know, however, whether or not the signal you see in any given round is correct.

**Remember, the signals you receive will be accurate three out of four times, on average.**

You may use the information from the signal to change your response (i.e., your belief about whether the statement is true) in that round from what you had said earlier.

When we give you more than one signal about the same question, and when each new signal is presented, we will store and remind you of your belief response regarding the truthfulness of that statement from the previous round so that you do not have to keep track in your head. After you have completed the survey, we will calculate how many rounds you won the bonus and pay you your total bonus payment separately from your main Prolific compensation for this task. We anticipate bonus payments to be paid within 4 working days of you completing

the study.

----- page break -----

As described earlier, we are interested in factors that influence the decisions you might make. In order for the results of this survey to be valid, **it is essential that you read all the instructions and questions carefully**. So we know that you have read these instructions, please place the slider below on the answer to  $(33+12)=?$  Thank you for taking the time to read these instructions.

0 10 20 30 40 50 60 70 80 90 100



----- page break -----

**The next pages involve a practice statement** for the belief-task you will perform, and responses for this practice set of rounds will not count towards bonus earnings. The practice will help you get familiar with the decision making environment you will face asking about several other factual statements. We will include a timer on the practice pages for you to get used to the timed nature of your task each round, but these practice rounds are actually not timed (i.e., the practice rounds will let you go longer without automatically moving you to the next page, unlike in the real rounds)

**READY to evaluate the practice statement?**

----- page break -----

**FACTUAL (True or False) STATEMENT:**

The average temperature on Mars, as a whole, is -81 degrees Fahrenheit.

**How likely do you believe that the statement is true?**

[Here, we elicit the baseline priors for the statement]

<u>0 =</u>		<u>50 =</u>		<u>100 =</u>
CERTAINLY FALSE	Likely FALSE	I'm not sure	Likely TRUE	CERTAINLY TRUE

0 10 20 30 40 50 60 70 80 90 100



**FACTUAL (True or False) STATEMENT:**

The average temperature on Mars, as a whole, is -81 degrees Fahrenheit.

Your last response was a [piped text of prior round belief% **belief** the statement is true

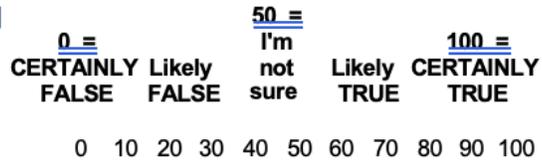
**Computer Signal:**

The computer has produced a signal for you. Remember, three out of four times this signal will be accurate and one out of four times it will be inaccurate, on average.

**--The signal from the computer this round is that the statement is "TRUE."**

**How likely do you believe that the statement is true?**

[Now that noisy signal has been shown, participant can update beliefs]



[This represents ONE trial of signal→belief elicitation. The exact same process from (i.e., factual statement and last response belief reminder, then noisy signal and re-elicitation of belief) continues for 3 more trials for this practice statement. Thus concludes the practice trials]

**This concludes your practice.**

The statements and rounds that come on the next page will count towards the calculation of your bonus earnings.

**Instructions Quick Summary:**

- **First Round for each new Statement:** You will presented with a factual statement, and we'll ask your belief regarding its truthfulness.
- **Follow-up Rounds that repeat the same statement:** You will be provided a signal from the computer whether the statement is "TRUE" or "FALSE". These signals are

designed to be accurate 3 out of 4 times, on average. You are then asked to again submit your belief regarding the statement's truthfulness.

- **Bonus payments on the task are maximized when you response [sic] what your actual belief is** in each round (which may or may not change after seeing computer signals regarding the statement's truthfulness)
- **Rounds are timed**, so please submit answers before the timer runs down for each round.

(advance past this page when you are ready to start the real rounds)

----- page break -----

[What follows is a sample presentation of one of the 4 coin-flip norms statements, which were randomized in order for all participants. This sample presentation displays that the norm provided in the statement is from a Republican sample of individuals—depending on whether the study participant were a pre-screened Democrat or Republican, this would represent a statement in the “Democrat shown Republican Norms” (i.e., out-group) or “Republican shown Republican Norms” (i.e., in-group) treatment, respectively]

**READY to evaluate a new statement?**

(pages are timed after this and so you should be ready and attentive because your responses impact the bonus you'll receive from this task.....and ***remember, you stand to earn the most in this task by submitting a response each round that reflects your true belief of how likely the statement is true on the [0,100] scale***)

----- page break -----

**Recall:** you have 40 seconds to submit your responses on this page.

Consider the "coin flip" task you completed in this study, where earnings depend on the number of HEADS reported. The task is done in private, and so no one can verify a participant's reported outcome, and one's payoff is higher in this task the more HEADS he/she reports flipping.

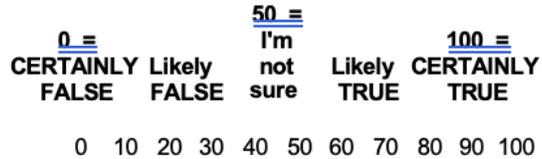
**FACTUAL (True or False) STATEMENT:**

A significant percentage of REPUBLICANS surveyed think it is acceptable to *over-report* the

number of HEADS actually flipped in this task.

How likely do you believe that the statement (about what others think is acceptable or unacceptable behavior) is true?

[Here, we elicit the baseline priors for the statement]



My belief regarding whether the statement is true

----- page break -----

**Recall:** you have 20 seconds to submit your responses on this page.

**FACTUAL (True or False) STATEMENT:**

A significant percentage of REPUBLICANS surveyed think it is acceptable to *over-report* the number of HEADS actually flipped in this task.

Your last response was a [piped text of prior round belief]% belief the statement is true

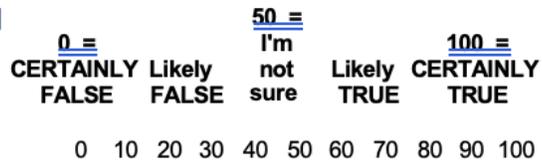
**Computer Signal:**

The computer has produced a signal for you. Remember, three out of four times this signal will be accurate and one out of four times it will be inaccurate, on average.

**--The signal from the computer this round is that the statement is "FALSE."**

-----  
How likely do you believe that the statement (about what others think is acceptable or unacceptable behavior) is true?

[Now that noisy signal has been shown, participant can update beliefs]



My belief regarding whether the statement is true

[This represents ONE trial of signal→belief elicitation. The exact same process from (i.e., factual statement and last response belief reminder, then noisy signal and re-elicitation of belief) continues for 3 more trials for this statement. Then, the set of trials for the next statement follows, etc. After the sets of trials on all 4 statements (not include initial Practice statement trials) is completed, the Bayesian task is finished]

----- page break -----

**Now, we'd like to get one last assessment of your current mood/emotion states.**

Please rate how strongly you feel each of these emotions *right now*.

**Right now I feel.....**

	Zero level of this emotion (1)	(2)	(3)	Mid-Range level of this emotion (4)	(5)	(6)	Maximum level of this emotion (7)
Happy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Excited	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Surprised	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Satisfied	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Angry	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Irritated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Confused	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Regret	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Disgust	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

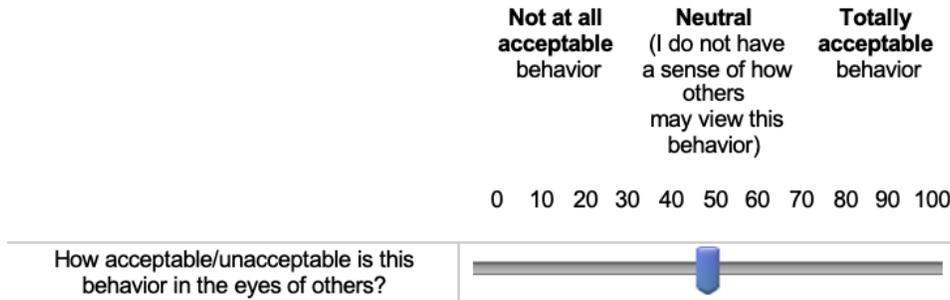
[this is a re-elicitation of perceived norms that were first elicited prior to the Bayes task]

We would first like to ask one more time about your beliefs regarding how acceptable or unacceptable *others* view certain behaviors in the Coin Flip task.

**Remember, we are not asking whether YOU feel the behavior is acceptable or not, but rather we are asking you about how your think "OTHERS" (i.e., a typical person) view the acceptability of the behavior.**

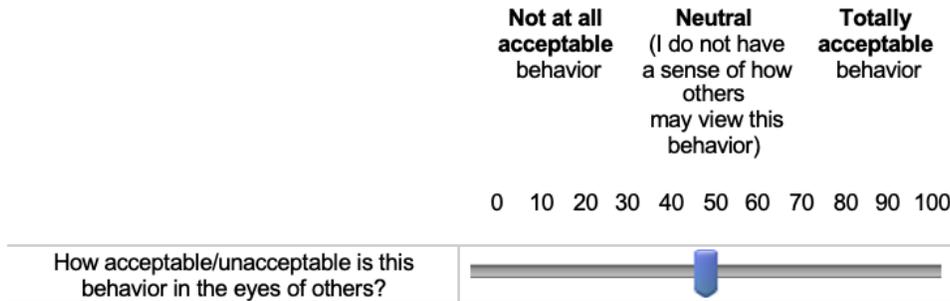
**BEHAVIOR** -----> Imagine that **someone over-reports the number of HEADS actually flipped** in this task.

Using the slider below, please tell us how acceptable or unacceptable you think "others" (i.e., a typical person) feel this behavior is.



**BEHAVIOR** -----> Imagine that **someone reports flipping 9 HEADS even though 3 HEADS (out of 10) were actually flipped** in this task

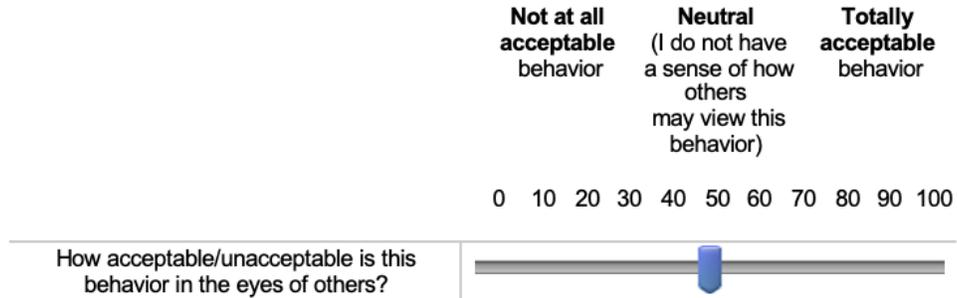
Using the slider below, please tell us how acceptable or unacceptable you think "others" (i.e., a typical person) feel this behavior is.



page break

**BEHAVIOR** -----> Imagine that **someone reports more HEADS than what he/she actually flipped** in this task.

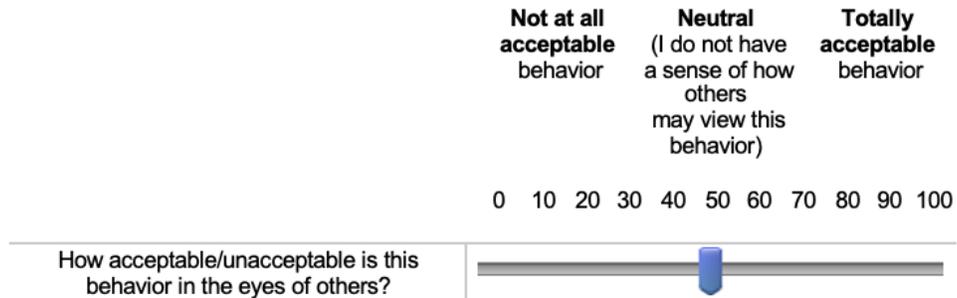
Using the slider below, please tell us how acceptable or unacceptable you think "others" (i.e., a typical person) feel this behavior is.



page break

**BEHAVIOR** -----> Imagine that **someone over-reports the number of HEADS flipped by just a few** (e.g., 3 or fewer extra HEADS reported out of 10 flips), but not by more than that.

Using the slider below, please tell us how acceptable or unacceptable you think "others" (i.e., a typical person) feel this behavior is.



page break

**Bonus Payment Task (Coin Flip)**

This next question asks you to complete the same coin flip task as you did earlier in the experiment. The instructions below are the same as before, so please read them carefully, but note that **your "coin flip" bonus payment will be determined by your report in just one of coin flip tasks** (either this one or the one completed earlier in this survey), but not both. There

is a 50% chance you may be paid based on your report from either task, so please respond here as if your response will determine your bonus payoff for this study, because it may.

---

### **COIN FLIP TASK**

**INSTRUCTIONS:** Here you are asked to flip a coin 10 times and report the results (making note of the order of Heads and Tails outcomes). You will be paid based on the outcome of the coin flips, so ***please read the instructions on the next page carefully. Before advancing to the next page, first please get/find a quarter (or any coin with a Heads and Tails side to it) that you can flip and something to write with and then click the button below. Please use an actual coin and not a virtual coin flipper. Please do not flip the coin before advancing to the next page.***

I have read the instructions above, I have a quarter/coin and something to write with, and I am ready to start flipping once I click this button

---

page break

---

### **One last item before the coin flip task**

Here is a reminder of your final belief-response to each of the 4 statements you just evaluated regarding acceptable or unacceptable behavior. Please take a moment to review the statement and your belief regarding the truthfulness of each statement.

**STATEMENT:** A significant percentage of Republicans surveyed think it is acceptable to over-report the number of HEADS actually flipped in this task.

-->Your final response to this statement was a **[piped text to remind participant of her final belief of this statement's truthfulness]% belief that this statement is TRUE.**

**STATEMENT:** None of the Republicans surveyed think it is acceptable to report flipping 9 HEADS if one actually flipped 3 HEADS (out of 10) in this task.

-->Your final response to this statement was a **[piped text to remind participant of her final belief of this statement's truthfulness]% belief that this statement is TRUE.**

**STATEMENT:** Most Republicans (i.e., more than half) think that reporting more HEADS than what one actually flips in this task is not acceptable behavior.

-->Your final response to this statement was a **[piped text to remind participant of her final belief of this statement's truthfulness]% belief that this statement is TRUE.**

**STATEMENT:** The majority of Republicans think it is acceptable to over-report the number of HEADS flipped ***by just a few*** (e.g., 3 or fewer extra HEADS reported out of 10 flips), but not by more than that.

-->Your final response to this statement was a **[piped text to remind participant of her final**

belief of this statement's truthfulness] % belief that this statement is TRUE.

---

The next page will be the coin flip task for you to complete.

I have reviewed these and am ready to do the coin flip task

---

page break

---

**This task offers you the chance for additional compensation, so please read carefully.**

Please take your coin and **Flip the coin 10 times** and record what you get for each flip along with the total number of HEADS you flip. Please only perform the 10 coin flips once. **Your payoff for this task will be \$0.10 (i.e., 10 cents) for each HEADS outcome from these coin flips**, and this payment will be paid to you as a BONUS within 3 days of completing the survey. For example, if you flip HEADS 3 times, your additional compensation (a BONUS payment in addition to the fixed payment you were promised completing the HIT) will be another \$0.30. Similarly, if you flip HEADS 6 times you will receive an additional \$0.60, etc.

**Please tell us below how many times you flipped HEADS out of your 10 total coin flips, and we will add 10 cents for each HEADS flipped to your HIT payoff when you receive your payment.**

0 1 2 3 4 5 6 7 8 9 10

Number of HEADS outcomes from flipping  
coin 10 times



---

page break

---

To finalize this survey, **please click "FINISH SURVEY"** below and advance the page (otherwise, Prolific completion code may not register properly and/or you will not get your completion code).

This study required that you either be a self-identified Republican or Democrat on your Prolific profile information used for custom screening our sample (and your response in this survey matched what you had previously indicated in your Prolific profile information).

**Note:** you should generally get paid your promised fixed payment within 48 hours of completing the study, and we aim to separately process bonus payments within 3-4 days. Please realize we cannot respond to individual inquiries about bonus payments, or we will be inundated with emails. **We will determine your total bonus payment based on the procedure that was outlined in the decision tasks used for the bonus payment.** I believe Prolific will notify you when you receive the bonus payment, which will reflect the sum of the bonus payments earned

in all rounds for all statements in the decision task. Thanks for understanding, and thank you again for your participation in our study.

**FINISH SURVEY**

## B Online Appendix: Figures

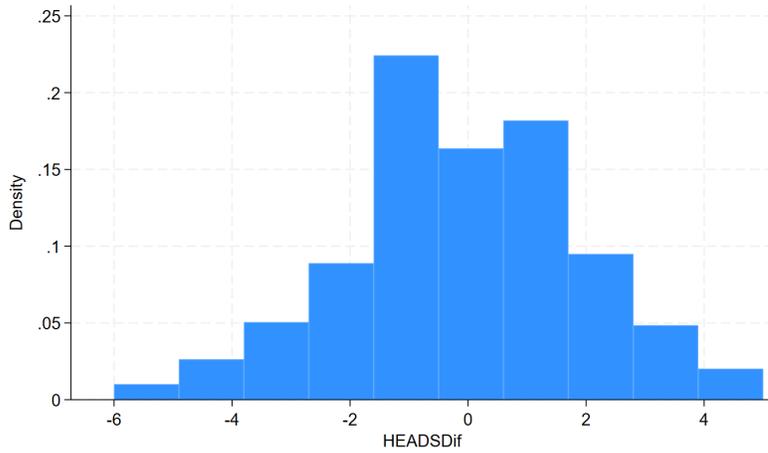


Figure B1: Density of changes in mean HEADS reports

Notes: The figure represents the change in the number of HEADS reported in the two Coin Flip task decisions, calculated as the difference between their mean HEADS report before versus after the Belief Updating task.

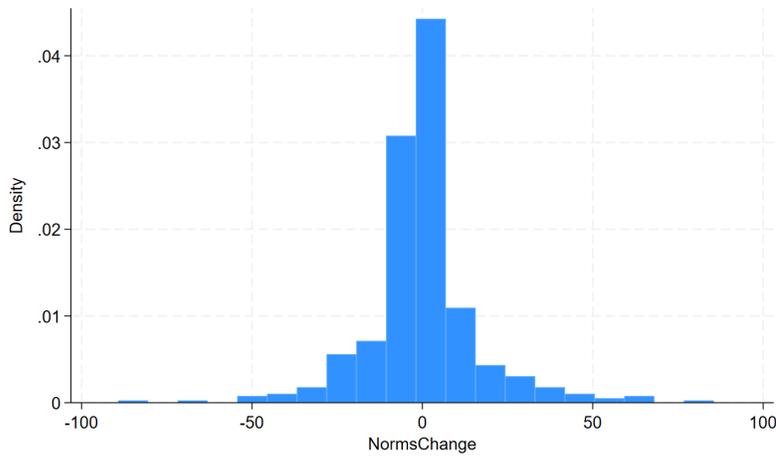


Figure B2: Density of changes in mean normative expectations

Notes: The figure represents the change in an individual's normative expectations, calculated as the difference between their mean normative expectations elicited before the Updating task and after this task (*i.e.*, *Post-Mean Normative Expectations* – *Pre-Mean Normative Expectations*).

## C Online Appendix: Tables

Table C1: Factual moral norms statements (from the preliminary survey of n=100 participants)

Statement number	Statement text	Truthfulness & normative support Data documentation
1	A significant percentage of [GROUP IDENTIFIER] surveyed think it is acceptable to over-report the number of HEADS actually flipped in this task.	<b>True:</b> Percentage saying “acceptable” at some level: <b>Overall</b> (n=110): 19.09%, <b>Republicans</b> (n=55): 14.55%, <b>Democrats</b> (n=55): 23.64%.
2	None of the [GROUP IDENTIFIER] surveyed think it is acceptable to report flipping 9 HEADS if one actually flipped 3 HEADS (out of 10) in this task.	<b>False:</b> Percentage saying “acceptable”: <b>Overall</b> (n=110): 3.64%, <b>Republicans</b> (n=55): 3.64%, <b>Democrats</b> (n=55): 3.64%.
3	The majority of [GROUP IDENTIFIER] surveyed think that reporting more HEADS than what one actually flips in this task is not acceptable behavior.	<b>True:</b> Percentage saying “unacceptable” at any level: <b>Overall</b> (n=110): 81.91%, <b>Republicans</b> (n=55): 85.45%, <b>Democrats</b> (n=55): 76.36%.
4	The majority of [GROUP IDENTIFIER] surveyed think it is acceptable to over-report the number of HEADS flipped by just a few ( <i>e.g.</i> , 3 or fewer extra HEADS reported out of 10 flips), but not by more than that.	<b>False:</b> Percentage saying “acceptable”: <b>Overall</b> (n=110): 15.45%, <b>Republicans</b> (n=55): 9.09%, <b>Democrats</b> (n=55): 21.82%.

*Notes:* Statement 1 percentages are tested against the null hypothesis that they are no different from zero. The null hypothesis is rejected using the one-sample Z-test ( $p < .01$  in all instances). Documenting Statement 2 is false requires no test given its phrasing and the fact that some considered the behavior described as acceptable. Statements 3 and 4 percentages are tested against the null hypothesis that they are no different from 50% against the one-sided alternatives. The null hypotheses are rejected in favor of the one-sided alternatives (one-sample Z-tests:  $p < .01$  in all instances).

Table C2: Changes in mean normative expectations and in the number of Heads reported in the Coin Flip task

<i>Dep. variable:</i> Change in number of Heads reported	(1) Pooled treatments	(2) Pooled treatments	(3) In-Group treatment	(4) Out-Group treatment	(5) No-Identity treatment	(6) Mixed treatment
Mean Normative Expectation Change	-.004 (.005)	.003 (.006)	-.013 (.016)	.012 (.011)	.011 (.012)	-.004 (.011)
Age (years)	-	-.015** (.007)	-.019 (.018)	-.016 (.015)	-.000 (.015)	-.012 (.015)
Female (=1)	-	-.230 (.210)	.323 (.487)	-.121 (.404)	-.380 (.437)	-.856** (.396)
Minority (=1)	-	-.024 (.245)	.275 (.623)	-.681 (.450)	.306 (.487)	.032 (.455)
Student (=1)	-	-.028 (.291)	-.512 (.773)	-.100 (.507)	.552 (.560)	-.022 (.583)
Employed (=1)	-	-.288 (.208)	.542 (.489)	-.470 (.456)	-.620 (.403)	-.714* (.385)
Democrat (=1)	-	-.078 (.209)	-.044 (.501)	-.160 (.412)	-.064 (.405)	.217 (.409)
Constant	-.120 (.087)	.900** (.437)	.263 (1.091)	1.289 (.902)	.475 (.798)	1.177 (.876)
Observations	450	368	90	79	96	103
R-squared	.001	.019	.052	.078	.060	.084

Notes: Standard errors are in parentheses. The dependent variable is the difference in the number of Heads reported between the first and the second administration of the Coin Flip task. There is one observation per individual. *Mean Normative Expectation Change* is calculated as the difference between the mean normative expectation elicited before the Belief Updating task and the mean normative expectation elicited after, that is, *Post-Mean Normative Expectation – Pre-Mean Normative Expectation*. \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2-tailed test.

Table C3: Bayes model estimates, focusing on the degree of dissonance between the signal and the initial normative expectations - WHEN THE INFORMATION SIGNAL ALIGNS WITH A TRUTH-TELLING NORM

<i>Dep. variable:</i>	(1)	(2)	(3)	(4)	(5)
Posterior belief	Between-subjects treatments	Out-Group treatment	In-Group treatment	No Identity treatment	Mixed treatment
lnOdds(t-1) = Prior belief ( $\delta$ )	.700*** (.035)	.773*** (.054)	.604*** (.059)	.738*** (.062)	.785*** (.059)
<b>lnLR (signals) (<math>\beta</math>)</b>	<b>.884***</b> <b>(.120)</b>	<b>.621**</b> <b>(.160)</b>	<b>1.356***</b> <b>(.193)</b>	<b>.719***</b> <b>(.171)</b>	<b>.683***</b> <b>(.182)</b>
IG Trial (=1)( $\gamma_1$ )	.054 (.083)	-	-	-	-.041 (.088)
OG Trial (=1)( $\gamma_2$ )	-.091 (.080)	-	-	-	-
IG Trial * lnLR ( $\eta_1$ )	.125 (.116)	-	-	-	.182 (.120)
OG Trial * lnLR ( $\eta_2$ )	-.074 (.106)	-	-	-	-
Signal Norm Dissonance	-.002 (.002)	.000 (.002)	-.008** (.004)	-.002 (.003)	-.000 (.002)
Degree $\in [0, 100]$ ( $\psi$ )					
<b>Signal Norm Dissonance Degree * lnLR (<math>\theta</math>)</b>	<b>-.005**</b> <b>(.002)</b>	<b>-.001</b> <b>(.003)</b>	<b>-.014***</b> <b>(.004)</b>	<b>-.001</b> <b>(.004)</b>	<b>-.003</b> <b>(.004)</b>
Age (years)	-.000 (.002)	-.002 (.003)	.000 (.004)	.001 (.004)	.001 (.003)
Female (=1)	-.035 (.070)	-.136 (.093)	-.034 (.121)	.060 (.122)	.046 (.078)
Minority (=1)	.167** (.083)	.189 (.124)	.373** (.168)	.020 (.126)	.095 (.079)
Student (=1)	.083 (.101)	.122 (.125)	.035 (.242)	.049 (.135)	-.095 (.103)
Employed (=1)	-.104 (.066)	-.046 (.129)	.003 (.121)	-.184* (.100)	-.003 (.074)
Constant	.221 (.139)	.074 (.264)	.374 (.261)	.168 (.217)	-.104 (.175)
Observations	2,086	641	697	748	837
R-squared	.635	.641	.643	.645	.721

*Notes:* The dependent variable is the posterior belief that the statement is true in trial  $t$  when the signal promotes truth-telling. Robust standard errors, clustered at the participant level, are in parentheses. IG is for the in-Group, and OG for Out-Group. The between-subjects treatments (columns (1)) include the In-Group, Out-Group, and No-Identity treatments (it excludes the Mixed treatment). \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2- tailed tests.

Table C4: Bayes model estimates, focusing on the degree of dissonance between the signal and the initial normative expectations - WHEN THE INFORMATION SIGNAL ALIGNS WITH A LYING NORM

<i>Dep. variable:</i>	(1)	(2)	(3)	(4)	(5)
Posterior belief	Between-subjects treatments	Out-Group treatment	In-Group treatment	No Identity treatment	Mixed treatment
lnOdds( $t-1$ ) = Prior belief ( $\delta$ )	.779*** (.027)	.742*** (.056)	.764*** (.044)	.816*** (.046)	.785*** (.044)
<b>lnLR (signals)(<math>\beta</math>)</b>	<b>.871***</b> <b>(.158)</b>	<b>1.180***</b> <b>(.261)</b>	<b>.524**</b> <b>(.245)</b>	<b>1.038***</b> <b>(.242)</b>	<b>.565***</b> <b>(.192)</b>
IG Trial (=1) ( $\gamma_1$ )	-.062 (.079)	-	-	-	-.054 (.095)
OG Trial (=1) ( $\gamma_2$ )	.151** (.075)	-	-	-	-
IG Trial * lnLR ( $\eta_1$ )	.088 (.119)	-	-	-	-.101 (.099)
OG Trial * lnLR ( $\eta_2$ )	.089 (.112)	-	-	-	-
Signal Norm Dissonance	-.001 (.001)	.000 (.002)	-.003 (.003)	-.002 (.002)	-.003* (.002)
Degree $\in [0, 100]$ ( $\psi$ )					
<b>Signal Norm Dissonance</b>	<b>-.005**</b>	<b>-.008**</b>	<b>.001</b>	<b>-.008**</b>	<b>-.000</b>
<b>Degree * lnLR (<math>\theta</math>)</b>	<b>(.002)</b>	<b>(.004)</b>	<b>(.004)</b>	<b>(.004)</b>	<b>(.003)</b>
Age (years)	-.003 (.003)	-.002 (.004)	-.004 (.005)	.000 (.004)	.005 (.003)
Female (=1)	-.048 (.069)	.265** (.120)	-.090 (.125)	-.258** (.114)	.087 (.098)
Minority (=1)	.113 (.080)	.159 (.170)	.113 (.135)	.087 (.107)	.118 (.103)
Student (=1)	-.145* (.078)	-.092 (.180)	-.202 (.127)	-.059 (.103)	.145 (.138)
Employed (=1)	.053 (.069)	.125 (.103)	.114 (.150)	-.062 (.092)	.140 (.105)
Constant	.168 (.185)	-.109 (.195)	.262 (.438)	.230 (.206)	-.300 (.214)
Observations	2,154	623	743	788	811
R-squared	.637	.649	.572	.691	.671

*Notes:* The dependent variable is the posterior belief that the statement is true in trial  $t$  when the signal promotes misreporting. Robust standard errors, clustered at the participant level, are in parentheses. IG is for the in-Group, and OG for Out-Group. The between-subjects treatments (columns (1)) include the In-Group, Out-Group, and No-Identity treatments (it excludes the Mixed treatment). \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2- tailed tests.

## D Online Appendix: Tables with continuous *In-GroupDegree* and *Out-GroupDegree* variables

We elicited a granular measure of ideological strength at the beginning of the study, *Liberal Score*  $\in [1, 9]$  (with 1 = Very Conservative, 5 = Middle Of The Road, 9 = Very Liberal). With this 9-point scale, we constructed *In-GroupDegree*  $\in [1, 9]$  to represent the degree of alignment between an in-group statement source and one’s political party identification, and *Out-GroupDegree*  $\in [1, 9]$  to represent the degree of misalignment between an out-group statement source and one’s political party.

Specifically, for in-group treatment trials *In-GroupDegree* = *Liberal Score* for Democrat participants seeing signals about Democrats, whereas *In-GroupDegree* = (10-*Liberal Score*) for Republican participants seeing signals about Republicans. *In-GroupDegree* was then set to zero for out-group treatment trials, as well as for generic norms treatment trials. A similar process for out-group treatment trials constructed *Out-GroupDegree* = *Liberal Score* where a Democrat participant viewed a Republican statement, whereas *Out-GroupDegree* = (10-*Liberal Score*) for Republican participants viewing Democrat statements (and *Out-GroupDegree* = 0 for trials in the in-group or generic norms treatments).<sup>24</sup>

The tables in this appendix section replicate those in the main text, replacing the binary InGroup and OutGroup variables with In-GroupDegree and Out-GroupDegree variables. The results are qualitatively similar.

---

<sup>24</sup>The average *Liberal Score* is 5.91 ( $\pm$  2.11 st. dev.) in our Democrat samples (pooled across all treatments) and 3.90 in our Republican samples ( $\pm$  2.50 st. dev.), which ensures variation in our construct of political *In-Group* or *Out-Group* identification.

Table D1: Bayes model estimates by In-Group or Out-Group norms statements and by degree of in- and out-groupness

Dep. variable:	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
	Between-subjects treatments All	Between-subjects treatments	Between-subjects treatments All	Between-subjects treatments	Between-subjects treatments Democrats	Between-subjects treatments Republicans	Between-subjects treatments Democrats	Between-subjects treatments Republicans	Mixed treatment All	Mixed treatment	Mixed treatment Democrats	Mixed treatment All	Mixed treatment Democrats	Mixed treatment Republicans	Mixed treatment Democrats	Mixed treatment Republicans
InOdds(t-1) =	.770***	.764***	.787***	.735***	.781***	.802***	.846***	.763***								
Prior belief ( $\delta$ )	(.024)	(.026)	(.031)	(.044)	(.041)	(.045)	(.040)	(.069)								
InLR (signals) ( $\beta$ )	<b>.602***</b>	<b>.609***</b>	<b>.630***</b>	<b>.592***</b>	<b>.600***</b>	<b>.590***</b>	<b>.597***</b>	<b>.573***</b>								
	<b>(.057)</b>	<b>(.063)</b>	<b>(.077)</b>	<b>(.099)</b>	<b>(.072)</b>	<b>(.082)</b>	<b>(.092)</b>	<b>(.130)</b>								
IGDegree $\in [1, 9]$	-0.008	-0.006	-0.004	-0.008	-0.010	-0.010	-0.007	-0.004								
	(.007)	(.008)	(.011)	(.011)	(.010)	(.010)	(.013)	(.013)								
OGDegree $\in [1, 9]$	.003	.008	.026**	-0.014	-	-	-	-								
	(.007)	(.008)	(.012)	(.010)												
<b>IGDegree*InLR</b>	<b>.016</b>	<b>.017</b>	<b>.020</b>	<b>.013</b>	<b>.003</b>	<b>-.001</b>	<b>.011</b>	<b>-.010</b>								
	<b>(.015)</b>	<b>(.017)</b>	<b>(.018)</b>	<b>(.029)</b>	<b>(.010)</b>	<b>(.009)</b>	<b>(.011)</b>	<b>(.014)</b>								
<b>OGDegree*InLR</b>	<b>-.006</b>	<b>-.001</b>	<b>.004</b>	<b>-.007</b>	-	-	-	-								
	<b>(.010)</b>	<b>(.012)</b>	<b>(.015)</b>	<b>(.017)</b>												
Age (years)	-	-.002	-.003	-.000	-	.003	.002	.006*								
		(.002)	(.002)	(.002)		(.002)	(.002)	(.003)								
Female (=1)	-	-.037	-.069	-.038	-	.053	.040	.017								
		(.046)	(.069)	(.063)		(.055)	(.080)	(.073)								
Minority (=1)	-	.134**	.138*	.117*	-	.102*	.078	.108								
		(.059)	(.080)	(.068)		(.058)	(.063)	(.101)								
Student (=1)	-	-.030	-.125	.026	-	.000	-.097	.102								
		(.061)	(.090)	(.070)		(.068)	(.076)	(.118)								
Employed (=1)	-	-.025	-.028	-.064	-	.053	.044	.066								
		(.041)	(.058)	(.060)		(.059)	(.065)	(.076)								
Constant	.018	.085	.147	.080	-.014	-.250*	-.112	-.445**								
	(.029)	(.072)	(.111)	(.099)	(.036)	(.136)	(.128)	(.208)								
Observations	5,280	4,240	2,176	2,064	1,920	1,648	800	848								
R-squared	.636	.632	.675	.587	.676	.701	.783	.632								

Notes: This table corresponds to Table 5 in the main text that uses instead binary variables for In-Group and Out-Group trials. The dependent variable is the posterior belief that the statement is true in trial  $t$ . Robust standard errors, clustered at the participant level, are in parentheses. \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2-tailed tests. Note that the more statistically powerful 1-tailed test is appropriate for any preregistered hypothesis test. IG is for the in-Group, and OG for Out-Group. The between-subjects treatments (columns (1) to (4)) include the In-Group, Out-Group, and No-Identity treatments (they exclude the Mixed treatments). Mixed treatment trials were always either IG=1 or OG=1 trials, but the between-subjects treatment trials could also be IG=0 and OG=0 trials when the norm source was No-Identity.

Table D2: Bayes model estimates, focusing on the degree of dissonance between the signal and the initial normative expectations, and by degree of in- and out-groupness

Dep. var.:	(1)	(2)	(3)	(4)	(5)
Posterior belief	Between-subjects treatments	Out-Group treatment	In-Group treatment	No Identity treatment	Mixed treatment
lnOdds(t-1) = Prior belief ( $\delta$ )	.746*** (.027)	.758*** (.047)	.685*** (.047)	.786*** (.047)	.793*** (.046)
<b>lnLR (signals) (<math>\beta</math>)</b>	<b>.853***</b> <b>(.088)</b>	<b>.879***</b> <b>(.274)</b>	<b>.944***</b> <b>(.271)</b>	<b>.807***</b> <b>(.119)</b>	<b>.628***</b> <b>(.188)</b>
IGDegree $\in$ [1, 9]	-.005 (.008)	-	-.034** (.016)	-	-.011 (.012)
OGDegree $\in$ [1, 9]	.008 (.008)	.039* (.020)	-	-	-.002 (.013)
IGDegree * lnLR	.018 (.017)	-	.025 (.036)	-	.011 (.032)
OGDegree * lnLR	-.001 (.012)	-.018 (.032)	-	-	.015 (.036)
Signal Norm Dissonance Degree $\in$ [0, 100] ( $\psi$ )	-.002** (.001)	.002 (.001)	-.006*** (.002)	-.002 (.002)	-.002 (.001)
<b>Signal Norm Dissonance Degree * lnLR (<math>\theta</math>)</b>	<b>-.005***</b> <b>(.001)</b>	<b>-.003*</b> <b>(.001)</b>	<b>-.007***</b> <b>(.002)</b>	<b>-.004**</b> <b>(.002)</b>	<b>-.002**</b> <b>(.001)</b>
Age (years)	-.002 (.002)	-.003 (.003)	-.001 (.002)	-.000 (.003)	.003 (.002)
Female (=1)	-.036 (.047)	.081 (.083)	-.018 (.083)	-.111 (.087)	.054 (.059)
Minority (=1)	.141** (.061)	.186 (.122)	.236** (.114)	.050 (.076)	.102* (.059)
Student (=1)	-.031 (.063)	-.021 (.111)	-.009 (.123)	-.010 (.085)	.007 (.072)
Employed (=1)	-.029 (.042)	.034 (.084)	.084 (.077)	-.132** (.061)	.056 (.060)
Constant	.209** (.100)	-.305 (.196)	.440** (.203)	.237* (.139)	-.159 (.149)
Observations	4,240	1,264	1,440	1,536	1,648
R-squared	.636	.642	.606	.669	.702

Notes: This table corresponds to Table 6 in the main text that uses instead binary variables for In-Group and Out-Group trials. The dependent variable is the posterior belief that the statement is true in trial  $t$ . Robust standard errors, clustered at the participant level, are in parentheses. IG is for in-Group, and OG for Out-Group. The between-subjects treatments (columns (1)) include the In-Group, Out-Group, and No-Identity treatments (it excludes the Mixed treatment). \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2-tailed tests.

Table D3: Bayes model estimates, focusing on the degree of dissonance between the signal and the initial normative expectations, and by degree of in- and out-groupness - WHEN THE INFORMATION SIGNAL ALIGNS WITH A TRUTH-TELLING NORM

<i>Dep. variable:</i>	(1)	(2)	(3)	(4)	(5)
Posterior belief	Between-subjects treatments	Out-Group treatment	In-Group treatment	No Identity treatment	Mixed treatment
lnOdds(t-1) = Prior belief ( $\delta$ )	.700*** (.035)	.770*** (.053)	.597*** (.056)	.738*** (.062)	.787*** (.060)
<b>lnLR (signals)(<math>\beta</math>)</b>	<b>.858***</b> <b>(.120)</b>	<b>.659**</b> <b>(.304)</b>	<b>1.115***</b> <b>(.318)</b>	<b>.719***</b> <b>(.171)</b>	<b>.706***</b> <b>(.240)</b>
IGDegree $\in$ [1, 9]	.001 (.014)	-	-.075*** (.026)	-	-.019 (.018)
OGDegree $\in$ [1, 9]	-.008 (.011)	.040** (.019)	-	-	-.013 (.017)
IGDegree * lnLR	.030* (.017)	-	.043 (.038)	-	.019 (.035)
OGDegree * lnLR	-.008 (.013)	-.005 (.031)	-	-	-.000 (.035)
Signal Norm Dissonance	-.003	.001	-.010**	-.002	-.000
Degree $\in$ [0, 100] ( $\psi$ )	(.002)	(.002)	(.004)	(.003)	(.002)
<b>Signal Norm Dissonance</b>	<b>-.005**</b>	<b>-.001</b>	<b>-.013***</b>	<b>-.001</b>	<b>-.003</b>
<b>Degree * lnLR (<math>\theta</math>)</b>	<b>(.002)</b>	<b>(.003)</b>	<b>(.004)</b>	<b>(.004)</b>	<b>(.004)</b>
Age (years)	-.000 (.002)	-.003 (.003)	.000 (.004)	.001 (.004)	.002 (.003)
Female (=1)	-.036 (.069)	-.114 (.088)	.059 (.116)	.060 (.122)	.037 (.080)
Minority (=1)	.166** (.083)	.197 (.124)	.377** (.166)	.020 (.126)	.090 (.078)
Student (=1)	.077 (.099)	.130 (.122)	.178 (.235)	.049 (.135)	-.071 (.105)
Employed (=1)	-.110* (.066)	-.047 (.123)	.066 (.123)	-.184* (.100)	-.006 (.072)
Constant	.229 (.139)	-.166 (.293)	.703** (.275)	.168 (.217)	-.018 (.187)
Observations	2,086	641	697	748	837
R-squared	.635	.642	.648	.645	.721

Notes: This table corresponds to Table C3 in the Appendix C that uses instead binary variables for In-Group and Out-Group trials. The dependent variable is the posterior belief that the statement is true in trial  $t$  when the signal promotes truth-telling. Robust standard errors, clustered at the participant level, are in parentheses. IG is for the in-Group, and OG for Out-Group. The between-subjects treatments (columns (1)) include the In-Group, Out-Group, and No-Identity treatments (it excludes the Mixed treatment). \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2- tailed tests.

Table D4: Bayes model estimates, focusing on the degree of dissonance between the signal and the initial normative expectations, and by degree of in- and out-groupness - WHEN THE INFORMATION SIGNAL ALIGNS WITH A LYING NORM

<i>Dep. variable:</i>	(1)	(2)	(3)	(4)	(5)
Posterior belief	Between-subjects treatments	Out-Group treatment	In-Group treatment	No Identity treatment	Mixed treatment
lnOdds(t-1) = Prior belief ( $\delta$ )	.780*** (.027)	.736*** (.058)	.764*** (.045)	.816*** (.046)	.785*** (.044)
<b>lnLR (signals) (<math>\beta</math>)</b>	<b>.896*** (.157)</b>	<b>1.372*** (.381)</b>	<b>.520 (.354)</b>	<b>1.038*** (.242)</b>	<b>.471* (.264)</b>
IGDegree $\in$ [1, 9]	-.012 (.013)	-	-.008 (.024)	-	.007 (.022)
OGDegree $\in$ [1, 9]	.024** (.012)	.038 (.031)	-	-	.023 (.024)
IGDegree * lnLR	.013 (.021)	-	.001 (.042)	-	-.003 (.033)
OGDegree * lnLR	.007 (.014)	-.030 (.041)	-	-	.023 (.038)
Signal Norm Dissonance	-.001	.000	-.003	-.002	-.004*
Degree $\in$ [0, 100] ( $\psi$ )	.001	.002	.003	.002	.002
<b>Signal Norm Dissonance Degree * lnLR (<math>\theta</math>)</b>	<b>-.005** (.002)</b>	<b>-.008** (.004)</b>	<b>.001 (.004)</b>	<b>-.008** (.004)</b>	<b>-.000 (.003)</b>
Age (years)	-.003 (.003)	-.003 (.005)	-.004 (.005)	.000 (.004)	.005 (.003)
Female (=1)	-.038 (.068)	.289** (.131)	-.079 (.124)	-.258** (.114)	.094 (.100)
Minority (=1)	.114 (.078)	.171 (.171)	.116 (.136)	.087 (.107)	.127 (.107)
Student (=1)	-.138* (.079)	-.079 (.177)	-.187 (.142)	-.059 (.103)	.132 (.149)
Employed (=1)	.055 (.069)	.120 (.103)	.120 (.150)	-.062 (.092)	.148 (.106)
Constant	.159 (.182)	-.325 (.248)	.274 (.448)	.230 (.206)	-.404* (.233)
Observations	2,154	623	743	788	811
R-squared	.637	.650	.572	.691	.673

Notes: This table corresponds to Table C4 in the Appendix C that uses instead binary variables for In-Group and Out-Group trials. The dependent variable is the posterior belief that the statement is true in trial  $t$  when the signal promotes misreporting. Robust standard errors, clustered at the participant level, are in parentheses. IG is for the in-Group, and OG for Out-Group. The between-subjects treatments (columns (1)) include the In-Group, Out-Group, and No-Identity treatments (it excludes the Mixed treatment). \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .10$  for the 2- tailed tests.