Do in-group biases lead to overconfidence in performance? Experimental evidence.*

Lia Q. Flores† and Miguel A. Fonseca‡

April 25, 2024

Abstract

Is the phenomenon of people overestimating their skill relative to their peers (overplacement) exacerbated by group affiliation? Social identity theory predicts people evaluate in-group members more positively than out-group members, and we hypothesized that this differential treatment may result in greater overplacement when interacting with an out-group member. We tested this hypothesis with 301 US voters affiliated with either the Republican or Democratic party in the run-up to the 2020 Presidential election, a time when political identities were salient and highly polarized. We found there is a higher tendency for overplacement when faced with an out-group opponent than with an in-group opponent. Decomposition analysis suggests this difference is due to underestimating the opponent, as opposed to overestimating one's own performance to a higher degree. Moreover, any tendency to incur in overplacement is mitigated when faced with an opponent with the same political identity relative to one with a neutral one. Group affiliation biases initial priors, but not how they are updated.

Keywords: Overconfidence, belief updating, motivated beliefs, overplacement, social identity, competition.
JEL classification: D8, D91, Z1, C9

*We thank Scott Vincent, Hannes Titeca and Philipp Chapkovski for their support regarding the experimental software. We also thank Julian Jamison, Lutfi Rahimi and Hannes Titeca, as well as participants at the 2021 SWDTP Conference, the 2021 PGR Economics Conference at the U. of Exeter, the 2021 ESA World Conference, the WiP seminar at cefUP (U. Porto) in September 2021 and the seminar at NIPE (U. Minho) in January 2022 for their valuable comments and suggestions. Lia Flores gratefully acknowledges the financial support from Portuguese public funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., in the framework of the PhD scholarship SFRH/BD/136976/2018. Additionally, research at CEF.UP is also financed by Portuguese public funds through FCT, in the framework of the project UIDB/04105/2020. The usual disclaimer applies.

†University of Porto. E-mail: *up201203503@fep.up.pt*. Postal address: University of Porto School of Economics and Management (FEP), Rua Dr. Roberto Frias. Porto, Portugal.

‡University of Exeter and NIPE, Universidade do Minho. E-mail: *m.a.fonseca@exeter.ac.uk*. Postal address: University of Exeter Business School, Streatham Court, Rennes Drive, Exeter EX4 4ST, United Kingdom.

1 Introduction

People often have misplaced beliefs about their own ability. Usually described as an error of judgement (Plous, 1993; Moore and Healy, 2008; Johnson and Fowler, 2011), overconfidence has been attributed to failures of individual and collective judgment such as wars (Tuchman, 1984; Johnson, 2004), unprofitable investment decisions by CEOs (Malmendier and Tate, 2005), or excess entry in markets (March and Shapira, 1987; Camerer and Lovallo, 1999).[1]

The economic analysis of overconfidence takes an individual perspective. Individuals may be overconfident as a by-product of them valuing self-confidence (Bénabou and Tirole, 2002) or self-image (Kőszegi, 2006). Santos-Pinto and Sobel (2005) propose that individuals may differ in their production technology, in which many different skills combine to determine ability in a given task. If individuals compare themselves to others on the basis of their own production function, overconfidence may result. Other theoretical frameworks consider the case in which individuals learn about their own ability over time; overconfidence emerges if they acquire information about their own ability in a self-serving way (Zábojník, 2004).

Although extremely insightful, this program of research by construction cannot account for the evidence of systematic differences in overconfidence across groups. For instance, evidence suggests that East Asian individuals are less overconfident than Americans (Kitayama et al., 1997; Heine et al., 2001; Heine and Hamamura, 2007). Moore et al. (2018) in addition find that individuals from collectivist cultures are less prone to overplacement than those from individualistic cultures. They argue that the need for self-enhancement biases individuals to prefer a positive self-image over a negative image, which in turn could lead to overconfidence. Schulz and Thöni (2016) also detect group-level differences in overconfidence across social groups within a cultural setting. Swiss undergraduate students in political science and business disciplines are overconfident in their own ability, while some STEM and humanities students tend to be underconfident.

These group differences suggest social identity may play an important role in explaining overconfidence. They add an interesting social layer to Moore and Healy's (2008) Bayesian and information-based argument, suggesting that groups (or their membership), either through socialization or through in-group bias, determine the degree of one's overconfidence relative to others. Cheng et al. (2021) demonstrate experimental evidence for the former channel, in that overconfidence is transmitted in social contexts. Importantly, they find this transmission

[1] Trivers (2000) argues that self-deception (of which overconfidence is a specific instance) can be evolutionarily advantageous in that the ability to self-deceive can make one more effective at deceiving others. Kaplan and Ruffle (2004) propose a two-player "king of the hill" game in which overconfidence of one player changes the unique equilibrium to their benefit. Johnson and Fowler (2011) show that overconfidence can emerge as an evolutionarily stable trait in a population in which randomly paired individuals compete for resources in a winner-takes-all contest where there is incomplete information about the opponent's ability. Recent experimental evidence by Schwardmann and van der Weele (2019) and Solda et al. (2020) further support the claim that overconfidence can be advantageous in social interactions.

is most likely to happen between in-group members.

In this paper, we study experimentally the extent to which a strong, highly salient in-group affiliation can lead individuals to believe that their performance is better than the performance of others – an aspect of overconfidence known as *overplacement* (Larrick, Burson, and Soll, 2007; Moore and Healy, 2008). We also examine how participants revise their overplacement upon receiving new performance-relevant information: by revising their beliefs about their own performance, or by revising their beliefs about the performance of their comparison group. Finally, we study whether participants engage in information avoidance regarding their objective performance on the task, and whether information avoidance (if any) varies as a function of overplacement. Acknowledging Moore and Healy's (2008) insight, subjects are asked to evaluate their performance in an easy task which would maximise our chances of finding overplacement and allow us to focus on assessing whether differences in overplacement would emerge from in-group affiliation.

Specifically, we recruited 301 U.S. voters registered with either the Republican or Democratic party in the week running up to the 2020 U.S. Presidential election, a time when American voters' political identities were very salient and highly polarized. We asked our participants to perform a logical reasoning test in a competitive setting – a task which, at least theoretically, should be orthogonal to their political identity. If their performance was better than the average performance of a separate sample of participants, they would receive a bonus payment. After completing the test, but before knowing their own score or whether they had won the bonus, we elicited participants' beliefs about their own performance and about the average performance of their competitors. We then informed participants about whether they won or lost the bonus and asked them to revise their beliefs. Finally, we gave participants the option to find out at zero cost the actual number of correct questions they got in the task, after informing them of the average score by their competitors. The main treatment variable was the political identity of the competitors: in-group members (i.e. belonging to the same political party), out-group members (i.e. belonging to the other main political party), or neutral (i.e. U.S. citizens).

We find that participants are more likely to overplace their performance relative to out-group members than to in-group members, and the extent of their overplacement is also larger. Interestingly, our data suggest the likelihood of engaging in overplacement is reduced by in-group identification, while the magnitude of overplacement is augmented by out-group derogation. This is consistent with Cacault and Grieder's (2019) findings that overconfidence relative to members of the group their participants were randomly matched with is reduced when a shared (artificially induced) identity is present within the group as opposed to none. However, in their study, participants are always set in a competition together with their group against people that belong to an external group (not necessarily an out-group) thus there is less incentive to be competitive relative to the members of the group the participant belongs

to. Moreover, in Brookins et al. (2014) underestimating others from one's (randomly formed) group is reduced when social identity is induced or when their group is set in competition with an out-group.

Upon good (bad) news of performing better (worse) than their competitors and thus (not) earning the bonus, belief updating reveals behavior that is consistent with wanting to look good (or not bad) relative to others, consistent with Murad and Starmer (2021); however, the effect is not magnified depending on the identity of the opponent. Differences in gender, political party and affiliation intensity are not systematically correlated with overplacement. Our data replicate the link between lower ability and higher overconfidence first identified by Kruger and Dunning (1999). However, we only observe this link in overplacement levels, not belief revisions upon receiving new information.

Overall, the effect of social identity is robust: in-group biases lead to overplacement in initial beliefs. However, the way in which participants update their beliefs following news is unaffected by group affiliation. Given that we used real identities in a context where they were extremely salient leads us to conjecture that social identity processes might not be important determinants of overconfidence, especially in contexts in which there are multiple opportunities to learn about one's performance/ability over time. Further study of socialization processes, such as those described by Cheng et al. (2021) is a promising avenue of future research.

When we broke up our sample and performed exploratory analysis to see if our findings were matched at the political party level, we found the same directional effects in both samples, although they were more often statistically significant for Republicans and not for Democrats – this is likely due to small sample once we condition on political affiliation and the fact that Democrats were better at the task than Republicans, which may have limited the extent to which we could detect an effect.

The remainder of the paper is structured as follows. Section 2 describes the experimental design and hypotheses, and Section 3 summarizes the results. Section 4 concludes the paper.

2 Experimental Design and Procedures

In this section we outline the pre-registered procedures, sample selection and design for our experiment (https://aspredicted.org/blind.php?x=v2ms2b). Copies of the experimental materials are included in Appendix E. After consenting to data collection, participants were presented with the instructions on screen. The instructions explained that participants would have to do a logical reasoning test, and that their performance on the test would impact their earnings. Specifically, participants would compete against a sample of participants that had completed the same task at an earlier point in time; if their score was better than the average score in the competitor sample they would receive a bonus of £2 ($2.64 at the time) and £0 otherwise. Participants had five minutes in which to read the instructions and go through a

sample example.

2.1 The Real-Effort Task

The logical reasoning test consisted of 10 items. Each item was a sequence of five figures. Participants had to predict the figure that would come next in the sequence from a set of five options. Participants had to complete each item in 45 seconds or less; if 45 seconds elapsed without an answer, the software marked the item as answered incorrectly and proceeded with the experiment, either to the next item or the next stage if that was the last item.

The type of task we selected tests problem-solving abilities, not general knowledge. Thus, performance in this task was independent of participants' background, namely education and/or occupation, except via aptitude itself (see meta-analysis by Klauer and Phye, 2008). Importantly, the task was theoretically orthogonal to the identity that we made salient since we do not expect political views to be related to logical reasoning. This design feature facilitates replication to other contexts and participant pools. We are interested in overplacement; Moore and Healy (2008) theorize and find experimentally that average overplacement should be larger the easier the task. As a result, we used a simple logical reasoning task.

2.2 Making Political Identity Salient

After completing the logical reasoning test, participants then answered a number of questions that made their political identity salient. Firstly, participants reported their general political affiliation on a seven-point scale: Strong Democrat, Democrat, lean towards Democrat, Independent/Other/None, lean towards Republican, Republican and Strong Republican. Participants stated how they voted in the previous elections and how they would vote in the upcoming Presidential election if it happened the following day, from three options: Republican, Democrat, Neither. Depending on their answer to the last question, either an image of Donald Trump, Joe Biden or a question mark was shown on screen for five seconds.

Secondly, participants were asked about their perception on their political opponents' opinion regarding topics that have been polarizing Americans in the last few years (Hawkins et al., 2018; Yudkin et al., 2019). These include immigration policies and Islamophobia, sexual harassment, police and racism, welfare state and firearm possession. We also asked participants to select the adjective that in their opinion best characterized a Democrat, Republican or American, out of a selection of three positive adjectives (caring, honest, reasonable) and three negative adjectives (brainwashed, hateful, racist). Finally, participants were asked about their own engagement and participation in US politics.

We used a natural identity as opposed to an artificial identity because the former generates a stronger behavioral response than the latter (see review on social identity theory by Li, 2020; and Weisel and Böhm, 2015). Since individuals have multiple identities (Turner et al., 1987),

each of which become relevant depending on the context, we needed to make political identity salient. We did so through both our questionnaire and the timing of the experiment. We note that people, in the main, self-select into real identities. To counteract the issue of selection, we used a task which is orthogonal to the identity we made salient which minimizes any selection effects that are specific to political identity. Moreover, we find it unlikely that there are different degrees of information (or more or less accurate signals) about others' potential performance in this specific task depending on sharing or not the same political views.

2.3 Experimental Manipulation

The software randomly assigned participants to a condition[2] and, after completing the identity saliency stage of the experiment, it informed each participant about the political identity of the group against which they were competing for the bonus payment.[3] The political identity of the opponent was the main experimental manipulation: in-group (a voter affiliated with the same political party as the participant), out-group (a voter affiliated with the opposing political party as the participant), and neutral identity, (a voter described as "American"). The screen also included the symbol of the relevant political party, or the US flag (see Figure 14 in the Appendix).

The performance of each participant was compared with the average performance of a separate sample of 80 participants (40 Democrats and 40 Republicans) who had answered the same logical reasoning test at an earlier date. In the in-group/out-group treatments, the benchmark was the performance of the 40 Democrats (8.0 out of 10) or the 40 Republicans (7.4 out of 10). For the neutral identity treatment, we took the performance of all the 80 participants (7.7 out of 10).

Using a separate sample allowed us to determine winners and losers immediately and against the same benchmark (conditional on treatment assignment). Without it, the opponents' performance would vary as a function of the progress in the data collection (see section A.1 for details on the auxiliary sample). The approach we adopted guaranteed no deception.[4]

2.4 Eliciting Beliefs About Performance

Before informing participants whether they had out-performed their opponents or not, we asked participants to state their belief about their own performance (the number of correct answers in the logical reasoning test) as well as the performance of their opponents. We

[2]The assignment to treatment was pre-determined to follow a certain fixed order; however, since time of arrival or starting moment of participants was unpredictable, we view treatment allocation "as if" random.

[3]We intentionally only revealed the opponent's identity after participants completed the logical reasoning test. Making identity salient prior to the task could have led participants to perform differently (see Shih et al., 1999 and Hoff and Pandey, 2006).

[4]As a robustness check, we repeated the analysis using the main sample as the reference; our results did not change - see section D.7 in the Appendix.

incentivized beliefs by compensating participants on the accuracy of each of the four guesses, such that they would earn £0.25 if their guess was within +/- 0.5 from the truth.[5]

We then informed participants about whether they had won or lost and asked participants to revise their beliefs about their own performance and the performance of their opponent. The second belief elicitation was incentivized in the same way as the first.

2.5 Eliciting Demand for Information

After completing the second belief elicitation round, participants were told about the actual performance of their opponent, and could find out their actual score in the logical reasoning test at no cost by clicking a "Reveal my score" button on-screen (as an alternative, participants could click on the "Don't reveal my score" button, that would take them directly to the next stage of the experiment). After making that decision, participants responded to a post-experimental survey, which included a Big Five personality questionnaire (Rammstedt and John, 2007) and a set of open-ended questions concerning the rationale for their decisions, as well as an incentivized real-effort task.[6] The experimental software then informed participants of their final payment and the experiment ended.

2.6 Participant pool, sample selection and incentives

The experiment was conducted online and the software used to program the experiment was oTree (Chen et al., 2016; Konrad, 2018). We collected the auxiliary sample data on the 22$^{\text{nd}}$ and 23$^{\text{rd}}$ of October, 2020. The main data collection happened between the 24$^{\text{th}}$ of October and the 3$^{\text{rd}}$ of November, 2020 (the Presidential Election day) to maximize the salience of political identity. All participants were recruited via Prolific (https://www.prolific.co).

At the recruitment stage we pre-screened participants on the basis of their political party affiliation and on their voting behavior: participants had to have voted at least once in the USA presidential elections. Based on pre-registered criteria, we excluded participants from the sample if (1) they did not complete the study within the time limit; (2) in the main test, they did not answer to a question and the software moved on to next one in more than 3 out of 10 questions; (3) they spent 3 seconds or less in more than 3 out of 10 questions; (4) they answered inconsistently relative to political affiliation in Prolific's pre-screening questionnaire and in our study;[7] (5) they were assessed as having weak political affiliation either categorized

[5] As the opponents' score was the average of performances ranging from 0 to 10, it was unlikely to be an integer.

[6] The real-effort task was intended as a second measure of ability. Hereafter interchangeably referred to as vowel-counting task, it consisted of counting the number of vowels in a quote for as many quotes as possible in one minute. Participants were only able to move on to the next quote once they had correctly answered to the current one. See a discussion by Charness et al. (2018) on different real-effort task designs.

[7] Participants that presented inconsistent answers were excluded in Prolific and replaced by new ones whenever they had previously reported to Prolific as being affiliated to one party – either Republican or Democrat

as "lean" or "swing" voters.[8]

We took a number of steps in our experimental protocol to ensure randomisation to treatment. Sessions for the main sample were conducted in small chunks both for democrats and republicans in order to increase control over the online data collection process and ensure submissions were valid. We conducted six sessions in two consecutive days for only democrats, followed by eight sessions for only republicans during four days, and then two additional sessions were conducted, one for each political party, to reach the pre-registered sample target of 50 participants for each treatment and political party. Conditioned by time of arrival of participants, sessions were launched such that for each political party there was an even spread across different times of the day, including peak times in Prolific activity in EST and PST timezones.

Our final sample was 301 participants, 151 Democrats and 150 Republicans. We randomly allocated 50 Democrats to the in-group and out-group conditions and 51 Democrats to the neutral condition. We allocated 50 Republicans to in-group, out-group and neutral conditions. Based on self-reported data available through Prolific, participants were on average 37 years old (s.d.=11.4, min=19, max=72), 54% were female, 16% were students, and 50% (15%) were employed full-time (part-time). All participants were registered voters; 91% of participants were born in the USA and 60% spoke English as their first language; 45% of the sample identified themselves as "Strong Democrats" or "Strong Republicans" while the remaining, identified themselves as "Democrats" or "Republicans".[9]

Participants were paid for their participation, performance in the experiment, and accuracy of guesses: £1.20 for completing the study; £2 if their score was higher than that of the opponent they were matched with, and £0 if otherwise; £0.25 for each accurate guess in the assessment of performance; and £0.02 for each correct answer in the vowel-counting task.[10] The study took on average 11.6 minutes to complete and average earnings were £2.40.[11]

- and subsequently reported being affiliated to the other party or to "Independent/None/Other" in the experiment in the question "Generally speaking, you usually think of yourself as a...".

[8] "Lean" voters were those that reported being "lean towards Democrat" or "lean towards Republican" to the question "Generally speaking, you think of yourself as a...". "Swing" voters were those that did not select the same political party in "Generally speaking, you think of yourself as a..." and either "Which party did you vote for in the previous election?" or "If the presidential election were held tomorrow, which party would you vote for?". 23% of Democrats and 55% of Republicans among those that were neither rejected nor returned (by their own account) were "lean" and/or "swing" voters and thus excluded from the analysis.

[9] For differences between the Democratic and Republican samples, please see Appendix A.2.

[10] While participants were pre-screened to only include those who had voted before in a presidential election in the U.S.A, the payment currency was in GBP, as this was the only currency used by Prolific at the time of data collection.

[11] The study was predicted to take less than 12 minutes to complete and the participation fee was determined by Prolific's minimum payment of £5/hour. With few exceptions, 5 minutes was the standard time limit per page across the study – the penalty for going over time was exclusion from the study, such that participants would not be able to move on to the next page or end the study.

2.7 Experimental Hypotheses

Our main outcome measure of interest is overplacement. We use the definition of overplacement proposed by Moore and Healy (2008). Equation (1) defines participant i's overplacement (if $O_{ij} > 0$) or underplacement (if $O_{ij} < 0$) of their performance relative to participant j.

$$O_{ij} = (b_i - b_j) - (x_i - x_j), \tag{1}$$

The variables x_i and x_j are the actual scores by i and j, and b_i and b_j refer to i's beliefs over those scores.

In our experiment, we elicited beliefs from participants twice: before and after participants found out whether they had won or lost the bonus. We denote the pre-announcement beliefs as b_{i1}, b_{j1}, and the post-announcement beliefs as b_{i2}, b_{j2}.

We hypothesize that group identity magnifies the tendency to overplace one's own performance. People act in favor of members of their in-group in detriment of the out-group and hold a sense of esteem or pride for belonging to that same group (Tajfel and Turner, 1979; Chen and Li, 2009; Akerlof, 2016; Cadsby et al., 2016; Chakravarty et al., 2016, Li, 2020; see Balliet et al., 2014 for a meta-analysis). Such pride in belonging to a group or "positive social identity" is a source of self-esteem (Trepte and Loy, 2017). Consequently, people tend to accentuate the similarities (differences) between themselves and the other members of their in-group (out-group) (Hogg and Abrams, 1988). The pursuit of self-esteem in which one strives to hold a good image of oneself may lead to overconfidence (Baumeister, Heatherton and Tice, 1993; Heatherton and Ambady, 1993). Overconfidence in one's abilities may translate into overconfidence in relative ability, i.e., overplacement (Zell et al., 2020). Given one's self-image alone or "personal identity" creates overplacement, we expect that valuing the image of oneself as member of an esteemed group or "social identity" also generates overplacement (the differentiation between the two identities and their interaction is discussed in Turner et al., 1987). However, to the best of our knowledge, no empirical study looks at the interaction between overplacement and social identity except Cacault and Grieder (2019) and Brookins et al. (2014) although in their studies group identity is artificially induced in their experiments and there is no competition between the participant and their in-group.

Hypothesis 1. *The likelihood of individuals overplacing their performance is larger when comparing themselves to the out-group rather than to the in-group.*

Hypothesis 2. *Conditional on there being overplacement, its magnitude is larger when individuals compare themselves to the out-group rather than to the in-group.*

The same processes that lead to biased initial beliefs could lead to imperfect adjustments in beliefs upon receiving information pertaining to performance. Potential deviations to Bayesian

updating, namely an "ostrich effect" in which inconvenient information (that does not match convenient information or good news) is avoided, forgotten or distorted have been examined by Benoît et al. (2015), Benoît and Dubra (2011), Burks et al., (2013), Merkle and Weber (2011), Möbius et al. (2011), Gottlieb (2011), Eil and Rao (2011), Karlsson et al. (2009) among others. Potentially, in a competitive context, the drive to distort beliefs to protect one's self-image could be enhanced when faced with an opponent identified as rival. This leads to the next set of hypotheses.

Hypothesis 3. *Being informed of having won, the likelihood of overplacement in revised beliefs is larger when individuals compare themselves to the out-group rather than to the in-group.*

Hypothesis 4. *Being informed of having won, any upward update of belief of own performance ($b_{i2} - b_{i1}$) is larger when playing against the out-group than against the in-group.*

Hypothesis 5. *Being informed of having won, any downward update of belief of the opponents' average performance ($b_{j2} - b_{j1}$) is larger when playing against the out-group than against the in-group.*

Hypothesis 6. *Being informed of having lost, any downward update of belief of own performance ($b_{i2} - b_{i1}$) is smaller when playing against the out-group than against the in-group.*

Hypothesis 7. *Being informed of having lost, any upward update of belief of the opponents' average performance ($b_{j2} - b_{j1}$) is smaller when playing against the out-group than against the in-group.*

3 Results

Throughout the text, our unit of analysis will be the participant, whether pertaining to performance, belief about performance, or a function of the two outcomes. In a nutshell, illustrated by Table 1 below, facing an out-group member aggravates overplacement relative to facing an in-group member both in probability and magnitude - this is explored in Results 1 and 2 that refer to Hypothesis 1 and 2, respectively. Moreover, facing bad news of performing worse than the opponent, the difference in behavior conditional on their identity is somewhat reduced, explored in Result 4, an unexpected outcome to testing Hypothesis 7. Regarding other behaviors, no significant difference was found between those facing out-group members and in-group members (Hypothesis 3 to 6 were not confirmed, giving way to null results).

Unfortunately, it is not possible for us to use subjects' own performance in the task as an additional control, as it introduces endogeneity issues in the estimation. Having overplacement as the dependent variable, measured as $(b_i - b_j) - (x_i - x_j)$, while having one of its

components, performance in the logical reasoning task, x_i, as an independent variable originates simultaneity between the left-hand side and right-hand side of the equation. We take an alternative approach by using a proxy for ability instead (performance in a vowel-counting task) which correlates significantly with performance in the main task ($r = 0.48$). The results from this approach have the same sign and statistical significance as the analysis run with standard t-tests. As such, hereafter, we report tests of mean treatment differences in text and corresponding regressions in the Appendix (Table A16 provides more detail to displayed Table 1). Note, nevertheless, that our sample size ($N = 301$) means the asymptotic properties of the test are valid (Sheskin, 2011)[12]. Unless otherwise noted, p-values relate to two-sided tests. To facilitate exposition of the material, we report test statistics and p-values pertaining to hypothesis tests conducted in Sections 3.2 and 3.3 in Tables A6, A7, A8, and A9 in the appendix.

3.1 Sample and Manipulation Checks

The sample characteristics were balanced across treatments.[13] To check whether our experimental manipulation concerning identity was effective, we compare the proportion of participants in each treatment who used a positive adjective to characterize their opponent (three available options out of six). 100% of in-group members were perceived positively, characterized with a positive adjective, suggesting in-group favoritism (56% "reasonable", 27% "caring", 17% "honest"). By contrast, 77% of out-group members were perceived negatively, characterized with a negative adjective, suggesting out-group derogation (43% "brainwashed", 19% "hateful", 15% "racist"), while only a small minority used positive terms (12% "caring", 7% "reasonable" and 4% "honest"). In addition, 67% of neutral-group members were perceived positively (27% "caring", 21% "honest", 19% "reasonable"), with a small proportion of negative adjectives (23% "brainwashed", 6% "racist", 4% "hateful"), in between in-group and out-group members. Our manipulation check suggests that our identity manipulation was successful.

Table 2 summarizes information on performance in the task, pre-announcement beliefs, and their difference. The average performance in the task was 7.16 out of 10. Performance in the out-group treatment was lower than in-group and neutral-group treat-

[12] Our results do not change if we use the Mann-Whitney test as shown in Section D.3 of the Appendix; moreover, results are robust to using the main sample to calculate the opponent's score instead of the auxiliary sample as explored in section D.7 in the appendix.

[13] We performed Kolmogorov-Smirnov (KS) tests on the score in the logical reasoning test, socio-demographic variables, and survey responses. Table A2 in the Appendix shows summary statistics on the full set of covariates among other descriptive variables. No significant differences were found between in-group, out-group and neutral-group treatments ($p > 0.10$ for all variables except openness with $p = 0.036$ and $p = 0.014$ when comparing out-group with neutral-group and in-group with neutral-group, respectively). For differences between treatments considering sub-samples of either political party, see summary statistics presented in Section A.2 of the Appendix.

Table 1: Hypotheses testing with regressions

	Pre-announcement		Post-announcement				
			Won			Lost	
DV:	O^{+}_{ij1}	O_{ij1} if $O_{ij1} > 0$	O^{+}_{ij2}	$\Delta(b_i - x_i)$	$\Delta(b_j - x_j)$	$\Delta(b_i - x_i)$	$\Delta(b_j - x_j)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Social Identity	0.170**	1.182***	-0.110	-0.094	-0.406	-0.226	-1.143***
	(0.071)	(0.334)	(0.129)	(0.199)	(0.268)	(0.422)	(0.343)
Vowel Score	-0.027	-0.213**	-0.019	0.110	0.154*	0.051	0.078
	(0.020)	(0.087)	(0.033)	(0.075)	(0.089)	(0.099)	(0.083)
Democrat	-0.096	-0.734*	-0.028	-0.305	0.157	0.685*	0.470
	(0.083)	(0.424)	(0.144)	(0.306)	(0.306)	(0.374)	(0.387)
Additional controls	✓	✓	✓	✓	✓	✓	✓
Constant	1.060	51.977***	1.163	9.601	-31.234*	-14.395	15.392
	(2.864)	(13.857)	(5.599)	(9.770)	(17.833)	(19.673)	(11.494)
N	200	124	81	81	81	119	119
R-squared	0.120	0.299	0.217	0.149	0.250	0.109	0.185

Notes: The dependent variable in regression (1) is $\mathbb{1}(O_{ij1} > 0)$ and in (3) it is $\mathbb{1}(O_{ij2} > 0)$, where O_{ij} is defined by equation (1). The dependent variable in regression (2) is the magnitude of O_{ij} conditional on $O_{ij} > 0$ happening. In regressions (4) and (6), the dependent variable is $(b_{i2} - b_{i1})$ which is equivalent to $\Delta(b_i - x_i)$, and in (5) and (7), it is $(b_{j2} - b_{j1})$ which is equivalent to $\Delta(b_j - x_j)$. LPM estimates for regressions (1) and (3) and OLS for the remaining. Social identity is a dummy =1 if out-group and =0 if in-group. Vowel Score is the performance in the vowel-counting task, a proxy for x_i (we couldn't include the latter directly due to endogeneity via simultaneity relative to the dependent variables). Additional controls include Strong affiliation (dummy), Female (dummy), Full-time (dummy), Age, Prolific Score (proxy for quality) and the Big Five personality traits, namely, Extraversion, Agreeableness, Conscientiousness, Neuroticism, Openness. Robust standard errors in parentheses.

ments (in-group vs. out-group, $t(198) = 1.764, p = 0.079$; out-group vs. neutral-group, $t(199) = -2.116, p = 0.036$), while there was no difference between in-group and neutral-group ($t(199) = -0.409, p = 0.683$).

These small differences in performance did not translate directly into differences in success rates in the contest: participants in the in-group and out-group treatments had success rates of 40% (in-group vs. out-group: $p = 1.000$, Fisher's exact test, henceforth FET), while the success rate of participants in the neutral-group was 58% (in-group vs. neutral-group: $p = 0.011$; out-group vs. neutral-group, $p = 0.017$, FET).

3.2 Performance and Initial Beliefs

We begin by testing Hypothesis 1.

Result 1 (Overplacement Frequency)**.** *Prior to announcement of results, the frequency of overplacement of performance is larger when competing against an out-group member than an in-group member.*

Result 1 comes about by comparing the proportion of individuals in each treatment who overplaced their performance.[14] It is significantly lower in the in-group treatment than in the out-group treatment and the neutral-group treatment. There is no significant difference between out-group and neutral-group. The similarity between out-group and neutral-group treatments and the difference between them and in-group suggest an in-group bias in the extensive margin.

Observation 1 ((Over)placement bias)**.** *Prior to announcement of results, the positive bias in the belief regarding relative performance is larger when competing against an out-group member than an in-group member. This is driven by underestimation of out-group members' performance.*

We now turn to participants' potential biases in judging their own and their opponent's performances. Looking at the average bias in estimation of own performance (the difference between estimated and actual performance), we see that the sample mean is not significantly different from zero. Breaking down the sample into the three treatments, the average estimation bias is only significantly different from zero in the out-group treatment. Interestingly, the percentage of individuals who overestimated their own performance is slightly higher in the in-group and out-group treatments than in the neutral-group treatment, although the differences are not significant.

We now turn to participants' beliefs about their opponents' performance. We observe general underestimation of the opponents' performance in the sample. However, there are important differences across treatments: underestimation is significantly larger in the out-group

[14] Table A6 in the appendix summarizes all statistical tests pertaining to Result 1 and Observation 1.

and neutral-group treatments than in the in-group treatment. The proportion of individuals who underestimated the performance of their opponent is significantly lower in the in-group treatment (54%) than the out-group treatment (70%) or neutral-group (75%). There is no difference between out-group and neutral group treatments (statistical tests in Table A6 in appendix).

The evidence so far suggests some degree of overplacement in the sample. The positive bias in relative placement is there both in the sample as a whole, as well as in each treatment. Overplacement was smallest in the in-group treatment and largest in the out-group treatment (statistical tests in Table A6 in appendix).

Result 2 (Overplacement magnitude)**.** *Conditional on overplacement, its magnitude is larger when competing against an out-group member than an in-group member. This is driven by underestimation of out-group members' performance.*

Result 2 is based on the sub-sample of those who effectively overplaced their performance, corresponding to 55-70% of participants in each treatment (see Table 3). The magnitude of overplacement was again smaller in the in-group treatment than in the out-group treatment, but not significantly smaller than the neutral-group. There is a significant difference between out-group and neutral-group, suggesting an out-group bias in the intensive margin.[15]

Still among overplacers, while the positive bias in the estimation of own performance is not significantly different between out-group and in-group treatments, the negative bias relative to the opponent is larger in the out-group treatment. Moreover, the bias in estimation of own performance is smallest in the neutral-group treatment but not significantly so relative to the out-group. As for underestimating the opponent, the neutral-group's bias lies in between the out-group's and in-group's with statistical significance on both sides (statistical tests in Table A7).

It is worthwhile to note that there is an asymmetry in the way overplacement manifests itself in our sample. We find evidence of out-group derogation in average overplacement when comparing out-group to neutral-group, but not in the relative frequency of overplacement (i.e., the extensive margin). This suggests that the difference in mean overplacement is driven by the intensity with which overplacers are biased against the out-group. In contrast, we see differences in mean overplacement and extensive margin of overplacement when comparing in-group to neutral-group.

In short, overplacement is more frequent regarding the out-group than regarding the in-group. The fact that out-group and neutral-group proportions of overplacement are similar suggests an in-group bias. Moreover, among overplacers, the degree of overplacement is larger regarding the out-group than regarding the in-group and the fact that in-group and neutral-group average levels of overplacement are similar suggests an out-group bias. Note that the

[15] Table A7 in the appendix summarizes all statistical tests pertaining to Result 2

Table 2: Task performance and pre-announcement beliefs, overestimation and overplacement

		Treatment		
	All	In-group	Out-group	Neutral-group
Own performance (x_i)	7.16 (2.25)	7.30 (2.12)	6.74 (2.36)	7.43 (2.23)
Success rate (% of ($x_i > x_j$))	47%	40%	41%	58%
Belief about own performance (b_{i1})	7.33 (1.96)	7.56 (1.84)	7.25 (1.64)	7.17 (2.33)
Bias in estimation of own performance ($b_{i1} - x_i$)	0.17 (2.53)	0.26 (2.33)	0.51 (2.52)	-0.26 (2.69)
% overestimating own performance	38%	41%	42%	31%
Belief about others' performance (b_{j1})	6.62 (1.85)	7.36 (1.33)	6.21 (2.06)	6.30 (1.86)
Bias in estimation of others' performance ($b_{j1} - x_j$)	-1.08 (1.87)	-0.34 (1.39)	-1.49 (2.08)	-1.40 (1.86)
% underestimating others' performance	66%	54%	70%	75%
(Over)placement $(b_{i1} - b_{j1}) - (x_i - x_j)$	1.25 (2.54)	0.60 (1.97)	2.00 (2.99)	1.15 (2.37)
% overplacing performance	65%	55%	69%	70%
N	301	100	100	101

Notes: Standard deviations in parentheses.

latter is specific to overplacement and is not mirrored to the left of null-overplacement or underplacement: conditional on underplacement, underplacement magnitudes are not statistically different between treatments.

3.3 Revision of beliefs following negative news

We now turn to the effect of announcing the outcome of the contest. Losers nominally underestimate their performance when evaluating it for the second time but not significantly; in fact, the mean bias in the estimation of own performance is not significantly different from zero in any treatment.[16]

Upon discovering that they had lost the contest, participants revise their beliefs about their own performance ($b_{i2} - b_{i1}$) significantly downwards across all treatments, without significant differences between treatments, in line with Result 16 (statistical tests in Table A8).

Result 3 (Revision of beliefs about own performance by losers)**.** *Being informed of having lost, there are no significant differences in downward updates of belief of own performance ($b_{i2} - b_{i1}$) as a function of the opponents' identity.*

[16]Table A8 summarizes all statistical tests related to Result .

Table 3: Task performance and pre-announcement beliefs, overestimation and overplacement: overplacers sub-sample

		Treatment		
	All	In-group	Out-group	Neutral-group
Own performance	6.35	6.29	5.83	6.90
(x_i)	(2.22)	(2.11)	(2.20)	(2.24)
Success rate (% of $(x_i > x_j)$)	32%	20%	23%	49%
Belief about own performance	7.37	7.89	6.88	7.44
(b_{i1})	(1.93)	(1.78)	(1.65)	(2.18)
Bias in estimation of own performance	1.02	1.60	1.06	0.54
$(b_{i1} - x_i)$	(2.54)	(2.05)	(2.77)	(2.59)
% overestimating own performance	55%	69%	57%	44%
Belief about others' performance	6.14	7.24	5.41	6.00
(b_{j1})	(1.86)	(1.36)	(1.84)	(1.83)
Bias in estimation of others' performance	-1.56	-0.40	-2.32	-1.70
$(b_{j1} - x_j)$	(1.88)	(1.41)	(1.83)	(1.83)
% underestimating others' performance	78%	58%	88%	83%
Overplacement	2.58	2.00	3.38	2.24
$(b_{i1} - x_{j1}) - (x_i - x_j)$	(2.09)	(1.43)	(2.55)	(1.78)
N	195	55	69	71

Notes: Standard deviations in parentheses.

Losers update their beliefs about their opponents' performance upwards with the exception of the case where the opponents are in-group members. Thus, in conformity with Result 4, the update is larger for the out-group treatment than the in-group one. We therefore reject Hypothesis 7. Moreover, the difference between out-group and neutral-group is not significant (statistical tests in Table A8).

Result 4 (Revision of beliefs about others' performance by losers)**.** *Being informed of having lost, participants update their belief about the opponents' performance ($b_{j2} - b_{j1}$) to a greater extent when the opponents are out-group members than when they are in-group members.*

In response to the announcement, losers revise their overplacement significantly downwards in all treatments, and as a result do not significantly hold a bias on relative performance when revising their beliefs except in the out-group treatment. As expected, the out-group treatment overplaces more frequently than the in-group, holding a larger bias in relative performance (statistical tests in Table A9).

3.4 Revision of beliefs following positive news

Among winners, participants' behavior does not differ depending on their opponent's social identity whether it is overplacement levels or belief updating regarding own and others' performance. Hence, our data does not support Hypotheses 3 to 5. Detailed analysis are in the Appendix Section D.1.

3.5 Exploratory analysis: sub-sample and individual differences

We now revisit our main outcome variables of interest and explore whether the results differ between Democrats and Republicans. This particular sub-sample comparison is important since we detected a baseline difference in ability in the logical reasoning task. We repeat the exercise with respect to other observable characteristic. We do so with two caveats: first, we did not have any expectations of differences along any observable characteristic (including political affiliation). Second, we may not be appropriately powered to detect differences when we analyze sub-samples. Indeed, our sample in some cases is not sufficiently large to satisfy the asymptotic properties of the t-test, which forces us to report results from Mann-Whitney tests (hereafter, referred to as MW test).

We begin by looking at the Democrat/Republican sub-samples. There was an equal number (50) of Democrat and Republican participants in each condition, except the neutral condition where we had 51 Democrat participants. Democrats were significantly better than Republicans at the task (7.85 vs 6.45 out of 10 questions across all three treatments, $t(299) = 5.67, p < 0.001$). This means that Democrats have less scope for overplacement, due to a ceiling effect. It therefore makes sense to examine our results at a sub-sample level.

Table 4: Beliefs revision and overplacement conditional on having lost.

		Treatment		
	All	In-group	Out-group	Neutral-group
Own performance (x_i)	5.53 (1.82)	6.02 (1.73)	5.25 (1.92)	5.24 (1.68)
Revised belief about own performance (b_{i2})	5.22 (2.13)	5.50 (2.05)	5.10 (1.84)	5.00 (2.59)
Bias in revised estimation of own performance ($b_{i2} - x_i$)	-0.31 (2.53)	-0.52 (2.25)	-0.15 (2.59)	-0.24 (2.85)
% overestimating own performance pre-announcement	59%	58%	63%	55%
% overestimating own performance post-announcement	35%	32%	36%	38%
Change in beliefs about own performance ($b_{i2} - b_{i1}$)	-1.48 (2.00)	-1.62 (2.06)	-1.61 (1.81)	-1.10 (2.15)
Revised belief about others' performance (b_{j2})	7.16 (1.80)	7.40 (1.56)	6.88 (1.61)	7.19 (2.30)
Bias in revised estimation of others' performance ($b_{j2} - x_j$)	-0.59 (1.80)	-0.31 (1.56)	-0.93 (1.59)	-0.51 (2.30)
% underestimating others' performance pre-announcement	66%	52%	80%	69%
% underestimating others' performance post-announcement	52%	47%	61%	45%
Change in beliefs about others' performance ($b_{j2} - b_{j1}$)	0.71 (1.79)	0.05 (1.42)	1.12 (1.97)	1.07 (1.77)
Revised (Over)placement $(b_{i2} - b_{j2}) - (x_i - x_j)$	0.28 (1.92)	-0.21 (1.60)	0.77 (2.12)	0.27 (1.93)
% overplacing performance	45%	32%	56%	48%
Change in (Over)placement $(b_{i2} - b_{j2}) - (b_{i1} - b_{j1})$	-2.19 (2.16)	-1.67 (1.75)	-2.73 (2.29)	-2.17 (2.37)
N	161	60	59	42

Notes: Standard deviations in parentheses.

The percentage of overplacers is indeed smaller for Democrats (54%) than Republicans (75%). However, we find no systematic difference in treatment effects between the two samples.

Result 1 is replicated in each sub-sample, in that we observe the same directional effect for Republicans and Democrats, though the significance level of the effect differs across sub-samples. Both Democrats and Republicans are more likely to be overplacers when facing an out-group individual than an in-group individual; however, that difference is only significant for Democrats (Democrats: 62% vs. 44%, $p = 0.054$, Republicans, 76% vs. 66%, $p = 0.189$). Individuals of both political affiliations are more likely to be overplacers when facing a neutral individual than an in-group individual, although again the difference is only significant for Republicans (Democrats: 57% vs. 44%, $p = 0.137$, Republicans, 84% vs. 66%, $p = 0.032$); are not significantly more or less likely to be overplacers when facing an out-group than when facing a neutral-group (Democrats: 62% vs. 57%, $p = 0.373$, Republicans: 76% vs. 84%, $p = 0.227$).

The directional effect of Result 2 is also replicated in each sub-sample, although again statistical significance differs across sub-samples. The magnitude of overplacement among overplacers is higher when facing out-group individuals than in-group ones (Democrats: 0.77, $p = 0.225$; Republicans: 1.92, $p = 0.002$, MW test); not significant when comparing in-group to neutral-group (Democrats: -0.13, $p = 0.143$; Republicans: 0.49, $p = 0.131$, MW test); and higher when facing out-group members than neutral-group ones (Democrats: 0.90, $p = 0.362$; Republicans: 1.43, $p = 0.002$).

Result 3 is partially replicated at a sub-sample level. Conditional on having lost, Democrats revise their beliefs about their own performance downwards by the same amount when comparing the in-group and out-group treatments ($p = 0.157$, MW test), when facing an in-group member than a neutral individual ($p = 0.124$, MW test), but by more when facing an out-group member than a neutral individual ($p = 0.025$, MW test). In contrast, we find no significant differences when performing the same analysis in the Republican sample ($p \geq 0.489$. MW test).

Result 4 is also replicated at the sub-sample level. We find a significant difference in the revision of beliefs about others' performance between in-group and out-group following a loss in both Democrats ($p = 0.032$, MW test) and Republicans ($p = 0.006$, MW test). We find no significant differences in any other comparison except for in-group vs neutral among Republicans ($p = 0.015$, all other comparisons $p \geq 0.164$, MW test). The econometric analysis in Table A11 corroborates this analysis.

We also explored the effect of observable characteristics such as sex, employment status, age, the acceptance score on the Prolific platform, which is a proxy for quality (Prolific Score), the big-five personality characteristics, and a proxy for ability from the vowel-counting task (Vowel Score). Details about the model we estimated and estimates can be found in Appendix D.2. With few exceptions, most individual characteristics are never predictive of behavior

in the experiment. We do not find any variable that systematically correlates well with overestimation of own and/or other performance, and/or overplacement, pre- and/or post-announcement. An exception is found for ability: in line with Kruger and Dunning (1999), higher ability is often linked with lower overconfidence, and consistently in the same direction (in particular, smaller likelihood of overestimation of own performance and overplacement; while for belief updating the trend is not so clear).

3.6 Information Avoidance

We conclude our analysis by looking at the frequency with which individuals avoided finding out information about their objective performance after having revised their beliefs. Table 5 outlines the relative frequencies of individuals who chose not to reveal their personal score at the end of the experiment. We find no systematic relationship between treatment assignment and information avoidance ($p = 0.652$, Fisher's exact test (FET)), conditional on having won ($p = 1.000$, FET), or having lost ($p = 0.426$, FET). Pooling across treatments, the frequency of information avoiders among losers is not statistically different from that of winners (10% vs. 6%: $p = 0.302$, FET).

Observation 2. *We find very low rates of information avoidance, which is uncorrelated with treatment assignment or outcome in the competition.*

Table 5: Rates of information avoidance

	Treatment			
	In-group	Out-group	Neutral-group	All
Win	5%	7%	7%	6%
Loss	7%	10%	14%	10%
All	6%	9%	10%	8%

4 Conclusion

We explore how social identity affects the way people judge their ability relative to others in a competitive setting. We focus on overplacement, the tendency to judge one's performance above that of someone else's. Our study is motivated by observations that the degree of over-confidence in performance seems to be different across different cultural settings (Kitayama et al., 1997; Heine et al., 2000; Heine and Hamamura, 2007) and across different occupational groups within the same cultural context (Schulz and Thöni, 2016). This suggests that there are social processes afoot which may augment inherent biases in human judgement.

These differences across social groups suggest that there may be an overconfidence norm that emerges in certain groups but not in others. Cheng et al. (2020) find strong evidence of

social transmission of overconfidence, especially among in-group members. Could, however, overconfidence emerge simply as a by-product of in-group bias? Our experimental evidence shows that while people tend to overplace their performance in easy tasks, they do so even more when comparing themselves to out-group targets. Decomposing this effect, it suggests that in-group love prevents many people from being overconfident of their relative ability (extensive margin) while those that fall into the trap would have their bias further worsened by out-group derogation (intensive margin).

However, the way participants react to new information is in contrast with what we hypothesize which is inspired by the literature on motivated reasoning and the "ostrich effect". In the experiment, participants have the opportunity to update their beliefs after being informed of having won or lost relative to their in-group or out-group opponents. Moreover, they can reveal how good their performance was.

Firstly, we observe that upon receiving bad news of having lost in the contest, those facing out-group targets adjust their beliefs about others' performance upwards, recognizing that they had underestimated their opponents, and they do so to a larger extent than those facing in-group targets. Secondly, among winners, overplacement in revised beliefs is magnified but the identity of the opponent doesn't make it any worse. Thirdly, those who win the contest against in-group members tend to start by underestimating themselves relative to others and then correct it with the confirmation of their success.

This implies that beliefs, earlier differentiated by the opponents' identity, may converge to a situation of no bias. Hence, our results show that social identity of opponents doesn't prevent new information from getting participants' beliefs closer to the truth. Our paper has a message of hope, aligned with Coppock's (2023) argument of parallel persuasion: while motivated reasoning may affect beliefs about others, compromising dialogue, cooperation and harmony, new information will adjust beliefs a little bit closer to the truth as opposed to intensifying biases.

How robust are our results? Individual characteristics do not seem to play a confounding role, namely age, gender or any of the Big Five personality traits. Nevertheless, we find some evidence for the link between ability and overconfidence (Kruger and Dunning, 1999) but no significant differences across treatments. Interestingly, the self-reported intensity of political affiliation does not play a significant role which suggests that the identity of the opponent is more important than the intensity of one's affiliation for analyzing the effect of social identity on the construction of beliefs. While there were some statistically significant differences between Democrats and Republicans in the way they overplaced performance, they were not systematic across treatments or win/loss announcements, suggesting they may be spurious.

We found some differences along the political affiliation dimension, in that Democrat participants scored better than Republicans in the real effort task, which created a stronger

ceiling effect for the former than for the latter. Broadly speaking, the directional effect of the findings of the full sample was replicated in each sub-sample, although for the Democrat sub-sample, treatment effects, although large, were not always statistically significant. We believe this lack of significance was due to small samples. Understanding whether there are fundamental differences along political dimension in terms of overplacement (or overconfidence more generally) is an interesting question, worthy of further study.

As an alternative explanation to social identity hypotheses, Moore and Healy (2008) suggest that overplacement in easy tasks like ours is a consequence of "as if" rational behavior with predictable Bayesian updating and the fact that the information about others is likely to be less perfect than the information that is held about oneself. Thus, regression to the mean would imply that: "Consequently, when performance is exceptionally high, people will underestimate their own performances, underestimate others even more so, and thus believe that they are better than others." (Moore and Healy 2009, p.504). As an extension of the argument, if information about the out-group is less accurate than information about the in-group, this could explain how underestimation of the opponent—even under financial incentives for accuracy—is worse in the out-group treatment. However, this line of argument requires participants to have differing degrees of information about cognitive reasoning skills (i.e., identifying visual patterns) along political party lines, which we find unlikely.

Notwithstanding, to further corroborate the social identity argument, future research could explore in which direction results change in an environment where the task is difficult and, as predicted in Moore and Healy's theory (2008), participants tend to underestimate their relative ability. If participants tend to underestimate their relative ability even further when facing an out-group member as opposed to an in-group member, the Bayesian information-based argument shadows the social identity argument. By contrast, if participants tend to underestimate their relative ability less when facing an out-group member as opposed to an in-group member, this supports the argument that there is resistance to believing rivals are good based on esteem over one's group affiliation, thus extending our results to a larger array of contexts.

Could in-group bias be mistaken by people simply inferring each others' logical reasoning abilities from how "wise" they think others' political opinions and general behaviours are? Firstly, there is extensive literature arguing how people are more likely to accept vague information that favours their own group than information that favours the rival group (Kunda, 1990; Taber and Lodge, 2006; Bolsen, Druckman and Cook, 2014; Adida, Gottlieb, Kramon and McClendon, 2017). In a world in which available information is diversified and easily accessible, if information is being "consumed" in a passive way where ego or esteem over one's social identity has no apparent role, picking up on casual conversations from one's entourage being it selected social groups or selected social media is still a manifestation of in-group bias which would be avoided with small effort to hold opinions as close to the truth as possible

as opposed to confirming prior beliefs. Thus, in-group bias would make the inference from disapproved political opinions to lack of objective logical reasoning easier.

Secondly, given the inference is reasonable, it is important to recognize that the two American parties' political stands have changed throughout history quickly enough for what one considers "wise" opinions and behaviours to shift from being associated to one party to the other. Thus, it is not evident that one party holds the "wise" views without debate. Finally, while it is true that in our sample Democrats fare slightly better than Republicans in the logical reasoning task, overplacement only picks up on the excess relative to the observed performances. Moreover, when using the main sample of 301 participants as reference point instead of the auxiliary sample of 80, in-group bias persists.

As mentioned, we didn't find the expected differences in the way individuals update their beliefs conditional on the opponent's identity. We must add that we cannot fully reject our hypotheses since ex-post power analyses indicate that the required sample to detect a significant effect for Hypotheses 3-6 was well beyond what is feasible. Moreover, it remains possible that the persisting overplacement in the out-group treatment (and only in that treatment) after bad news is due to an ego-driven resistance to make rival others look good, even if their revision towards erasing the bias is the largest.

Overall, our results confirm and extend recent research that finds an important group aspect to overconfidence: existing in-group biases translate into overplacement of performance, intensified by any out-group bias. However, individuals in isolation tend to update their beliefs following objective information in a rationalizable manner. Therefore, even if the reference to the identity of opponents is no more than a label and one is not being actively pressured to conform to a certain way of thinking by in-group members, in-group biases are still at work. While participants answered on their computers or tablets and very likely on their own, especially given the social-gathering restrictions in force at the time of our data collection, many human activities and behaviors are inserted in some social context. Hence, our results may be a lower bound for the behavioral response when corporate culture or other institutional norms are salient. Exploring how such contexts affect overconfidence and belief updating is an interesting avenue for future research.

On a final note, the form of social identity we use is based on existing political differences in the US, a natural identity. We believe requirements are met for it to be relevant and thus consequential, as opposed to minimal groups, namely a significantly more positive perception over the in-group relative to the out-group. Since people are not randomly assigned to political identity and the effect of identity on overplacement may be domain specific, we invite future investigations of this research question with other identities.

References

[1] Adida, C., Gottlieb, J., Kramon, E., & McClendon, G. (2017). Reducing or reinforcing in-group preferences? An experiment on information and ethnic voting.Quarterly Journal of Political Science, 12(4), 437-477.

[2] Alicke, M. D., & Govorun, O. (2005). The Better-than-Average Effect. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), The Self in Social Judgement (pp. 85–106). New York: Psychology Press.

[3] Akerlof, R. (2016). "We Thinking" and Its Consequences. American Economic Review, 106(5), 415-19.

[4] Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1993). When ego threats lead to self-regulation failure: negative consequences of high self-esteem. Journal of Personality and Social Psychology, 64(1), 141.

[5] Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. The Quarterly Journal of Economics, 117(3), 871-915.

[6] Benoît, J. P. & Dubra, J. (2011). Apparent overconfidence. Econometrica, 79(5), 1591–1625.

[7] Benoît, J. P., Dubra, J., & Moore, D. A. (2015). Does the better-than-average effect show that people are overconfident?: Two experiments. Journal of the European Economic Association, 13(2), 293-329.

[8] Bolsen, T., Druckman, J. N., & Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. Political Behavior, 36(2), 235-262.

[9] Brookins, P., Lucas, A., & Ryvkin, D. (2014). Reducing within-group overconfidence through group identity and between-group confidence judgments. Journal of Economic Psychology, 44, 1-12.

[10] Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2013). Overconfidence and social signalling. Review of Economic Studies, 80(3), 949-983.

[11] Cacault, M. P., & Grieder, M. (2019). How group identification distorts beliefs. Journal of Economic Behavior & Organization, 164, 63-76.

[12] Cadsby, C. B., Du, N., & Song, F. (2016). In-group favoritism and moral decision-making. Journal of Economic Behavior & Organization, 128, 59-71.

[13] Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. American Economic Review, 89(1), 306-318.

[14] Chakravarty, S., Fonseca, M. A., Ghosh, S., & Marjit, S. (2016). Religious fragmentation, social identity and cooperation: Evidence from an artefactual field experiment in India. European Economic Review, 90, 265-279.

[15] Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. Journal of Economic Behavior & Organization, 149, 74-87.

[16] Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9, 88-97.

[17] Chen, Y., & Li, S. X. (2009). Group identity and social preferences. American Economic Review, 99(1), 431-57.

[18] Cheng, J. T., Anderson, C., Tenney, E. R., Brion, S., Moore, D. A., & Logg, J. M. (2021). The social transmission of overconfidence. Journal of Experimental Psychology: General, 150(1), 157.

[19] Coppock, A. (2023). Persuasion in parallel: How information changes minds about politics. University of Chicago Press.

[20] Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. American Economic Journal: Microeconomics, 3(2), 114-38.

[21] Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. Journal of Economic Perspectives, 30(3), 133-40.

[22] Gottlieb, D. (2011). Will you never learn? self deception and biases in information processing. Unpublished Manuscript. Wharton School, University of Pennsylvania.

[23] Hawkins, S., Míriam, J.-T., Yudkin, D., & Dixon, T. (2018). Hidden Tribes: A Study of America's Polarized Landscape. New York: More in Common.

[24] Heatherton, T. F., & Ambady, N. (1993). Self-esteem, self-prediction, and living up to commitments. In Self-esteem (pp. 131-145). Springer, Boston, MA.

[25] Heine, S. J., & Hamamura, T. (2007). In search of East Asian self-enhancement. Personality and Social Psychology Review, 11(1), 4-27.

[26] Heine, S. J., Kitayama, S., & Lehman, D. R. (2000). Cultural differences in self-evaluation: Japanese readily accept negative self-relevant information. Journal of Cross-Cultural Psychology, 32(4), 434-443.

[27] Hoff, K., & Pandey, P. (2006). Discrimination, social identity, and durable inequalities. American Economic Review, 96(2), 206-211.

[28] Hogg, M. A., & Abrams, D. (1988). Social identifications: A social psychology of intergroup relations and group processes. Taylor Frances/Routledge.

[29] Horrace, W. C., & Oaxaca, R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. Economics Letters, 90(3), 321-327.

[30] Jakobsson, N., Levin, M., & Kotsadam, A. (2013). Gender and overconfidence: effects of context, gendered stereotypes, and peer group. Advances in Applied Sociology, 3(02), 137.

[31] Johnson, D. D. (2004). Overconfidence and war. Harvard University Press.

[32] Johnson, D. D., & Fowler, J. H. (2011). The evolution of overconfidence. Nature, 477(7364), 317-320.

[33] Kaplan, T. R., & Ruffle, B. J. (2004). The Self-serving Bias and Beliefs about Rationality. Economic Inquiry, 42(2), 237-246.

[34] Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information. Journal of Risk and Uncertainty, 38(2), 95-115.

[35] Kitayama, S., Markus, H. R., Matsumoto, H., & Norasakkunkit, V. (1997). Individual and collective processes in the construction of the self: self-enhancement in the United States and self-criticism in Japan. Journal of Personality and Social Psychology, 72(6), 1245.

[36] Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. Review of Educational Research, 78(1), 85-123.

[37] Konrad, M. (2019). oTree: Implementing experiments with dynamically determined data quantity. Journal of Behavioral and Experimental Finance, 21, 58-60.

[38] Kőszegi, B. (2006). Ego utility, overconfidence, and task choice. Journal of the European Economic Association, 4(4), 673-707.

[39] Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology, 77(6), 1121.

[40] Kunda, Z. (1990). The case for motivated reasoning. Psychological Bulletin, 108(3), 480.

[41] Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). Organizational Behavior and Human Decision Processes, 102(1), 76-94.

[42] Li, S. X. (2020). Group identity, ingroup favoritism, and discrimination. Handbook of Labor, Human Resources and Population Economics, 1-28.

[43] Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. The Journal of Finance, 60(6), 2661-2700.

[44] March, J. G., & Shapira, Z. (1987). Managerial perspectives on risk and risk taking. Management Science, 33(11), 1404-1418.

[45] Merkle, C., & Weber, M. (2011). True overconfidence: The inability of rational information processing to account for apparent overconfidence. Organizational Behavior and Human Decision Processes, 116(2), 262-271.

[46] Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). Managing self-confidence: Theory and experimental evidence (No. w17014). National Bureau of Economic Research.

[47] Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. Psychological Review, 115(2), 502.

[48] Moore, D. A., Dev, A. S., Goncharova, E. Y., Vazire, S., & Tullett, A. (2018). Overconfidence across cultures. Collabra: Psychology, 4(1).

[49] Murad, Z., & Starmer, C. (2021). Confidence snowballing and relative performance feedback. Journal of Economic Behavior & Organization, 190, 550-572.

[50] Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. The Quarterly Journal of Economics, 122(3), 1067-1101.

[51] Plous, S. (1993). The psychology of judgment and decision making. Mcgraw-Hill Book Company.

[52] Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality, 41(1), 203–212.

[53] Santos-Pinto, L., & Sobel, J. (2005). A model of positive self-image in subjective assessments. American Economic Review, 95(5), 1386-1402.

[54] Schulz, J. F., & Thöni, C. (2016). Overconfidence and career choice. PloS one, 11(1), e0145126.

[55] Schwardmann, P., & Van der Weele, J. (2019). Deception and self-deception. Nature human behaviour, 3(10), 1055-1061.

[56] Sheskin, D.J. 2011. Handbook of Parametric and Nonparametric Statistical Procedures, Fifth Edition, Routledge.

[57] Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. Psychological Science, 10(1), 80-83.

[58] Solda, A., Ke, C., Page, L., & Von Hippel, W. (2020). Strategically delusional. Experimental Economics, 23(3), 604-631.

[59] Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. American Journal of Political Science, 50(3), 755-769.

[60] Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), The social psychology of intergroup relations (pp. 33-47). Monterey, CA: Brooks/Cole.

[61] Tuchman, B. W. (1988). The march of folly: From Troy to Vietnam. NY: Knopf.

[62] Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). Rediscovering the social group: A self-categorization theory. Basil Blackwell.

[63] Trepte, S., & Loy, L. S. (2017). Social identity theory and self-categorization theory. The International Encyclopedia of Media Effects, 1-13.

[64] Trivers, R. (2000). The elements of a scientific theory of self-deception. Annals of the New York Academy of Sciences, 907(1), 114-131.

[65] Weisel, O., & Böhm, R. (2015). "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. Journal of Experimental Social Psychology, 60, 110-120.

[66] Yudkin, D., Hawkins, S., & Dixon, T. (2019). The Perception Gap: How False Impressions are Pulling Americans Apart. New York: More in Common.

[67] Zábojník, J. (2004). A model of rational bias in self-assessments. Economic Theory, 23(2), 259-282.

[68] Zell, E., Strickhouser, J. E., Sedikides, C., & Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. Psychological Bulletin, 146(2), 118.

This Appendix contains summary statistics (Section A); statistical tests reported in the main text (Section B); an exposition on deviations from the pre-registration (Section C); alternative model specifications and robustness checks to results in section D; and instruction sets and screenshots for the experimental interface (Section E).

A Summary Statistics

Subsection A.1 provides additional details on the auxiliary sample, while subsection A.2 provides additional details on the main sample.

A.1 Details on the auxiliary sample

Recruitment of the auxiliary sample of 80 participants - 40 Democrats and 40 Republicans - occurred on the 22-23/11/2020 with 1 morning session and 1 afternoon session (in EST time) for each political party.

Participants answered the same logical reasoning test (hereafter, LRT) as the main sample. Time constraints were the same as were financial incentives for performance, now compared to the average of the auxiliary sample itself. Participants also answered the 3-question political affiliation survey as part of their participation in the study.

The experiment was expected to last 5 minutes, thus participation fee was £0.05, proportional to the compensation in the main sample[17]. Data from this sample was used for one purpose only - to calculate the average scores of the main sample's opponents, whether identified as "Americans" (7.7/10), "Democrats" (8.0/10) or "Republicans" (7.4/10).

Note that "lean" and "swing" voters were not excluded from the auxiliary sample, as opposed to the main sample, following the assumption that people perceive the less extreme as being part of the group to which "the average" Democrat, Republican or American belongs to. "Lean" and/or "swing" voters comprised 28% of the sample of Democrats and 45% of the sample of Republicans.

Table A1 compiles summary statistics about the auxiliary sample. Note that 45% of the sub-sample of Democrats identified themselves as "Strong Democrats", 35% as "Democrats" and 20% as "lean towards Democrats". Relative to the sub-sample of Republicans, 17.5% of the sample identified themselves as "Strong Republicans", 50% as "Republicans" and 32.5% as "lean towards Republicans".

[17] With the bonus, average payoff was £1.3; average time taken was 4.49 minutes.

Table A1: Summary statistics of auxiliary sample

	All	Democrats	Republicans
Score in the LRT (x_i)	7.66	7.95	7.38
	(2.37)	(2.10)	(2.61)
Age	34.24	32.23	36.25
	(9.49)	(9.29)	(9.37)
Female	58%	60%	55%
Student	16%	23%	10%
Swing or Lean	36%	28%	45%
Strong Affiliation	31%	45%	18%
English is first language	74%	73%	75%
Born in the USA	98%	95%	100%
Resident in the USA	95%	95%	95%
Full-time	55%	50%	60%
Part-time	14%	13%	15%
N	80	40	40

Standard deviations in parentheses.

A.2 Details on the main sample by treatment and political party

Here we present additional details that describe the main sample, namely by treatment in Table A2 and by political affiliation in Table A3. Moreover, we compare treatment groups within each political party in Tables A4 and A5 for Democrats and Republicans respectively.

B Statistical tests reported in the main text

Here we summarize the statistical tests and corresponding p-values reported in the main text. We report a table for each subsection of the results.

Table A2: Summary statistics of main sample

	All	Treatment In-group	Out-group	Neutral-group
Score in the LRT (x_i)	7.16 (2.25)	7.30 (2.12)	6.74 (2.36)	7.43 (2.23)
Age	36.52 (11.37)	36.51 (11.88)	36.55 (10.93)	36.50 (11.41)
Female	54%	52%	51%	59%
Student	16%	18%	14%	17%
Strong Affiliation	45%	42%	45%	47%
English is first language	60%	59%	58%	62%
Born in the USA	95%	94%	94%	96%
Resident in the USA	92%	93%	90%	92%
Full-time	50%	51%	51%	47%
Part-time	15%	16%	15%	13%
Extraversion	2.67 (1.01)	2.71 (1.00)	2.69 (0.99)	2.62 (1.05)
Agreeableness	3.54 (0.92)	3.63 (0.84)	3.56 (0.89)	3.42 (1.01)
Conscientiousness	3.83 (0.87)	3.99 (0.87)	3.76 (0.89)	3.76 (0.83)
Neuroticism	2.85 (1.13)	2.89 (1.19)	2.72 (1.04)	2.94 (1.15)
Openness	3.60 (0.97)	3.84 (0.96)	3.52 (0.86)	3.45 (1.05)
Negative adjective	37%	0%	77%	33%
Engagement in US politics	4.06 (0.66)	4.05 (0.71)	4.06 (0.61)	4.07 (0.67)
Vowel Score	3.49 (2.05)	3.74 (2.14)	3.17 (1.84)	3.54 (2.14)
N	301	100	100	101

Standard deviations in parentheses.

Other details: (1) Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness constitute the Big Five personality traits, measured in a 1-5 Likert scale, each trait associated to two questions; (2) Engagement in US politics refers to the mean of responses to 4 questions in a 1-5 Likert scale, that concerns voting, following the news and main political events. (3) Vowel Score refers to the score in the real-effort task of correctly counting as many vowels as possible in a limited time frame;

Table A3: Summary statistics of main sample by political party

	All	Democrats	Republicans
Score in the LRT (x_i)	7.16	7.85	6.45
	(2.25)	(1.73)	(2.49)
Age	36.52	33.54	39.52
	(11.37)	(10.86)	(11.13)
Female	54%	66%	42%
Student	16%	28%	5%
Strong Affiliation	45%	50%	39%
English is first language	60%	70%	50%
Born in the USA	95%	91%	98%
Resident in the USA	92%	94%	89%
Full-time	50%	42%	57%
Part-time	15%	17%	12%
Extraversion	2.67	2.63	2.71
	(1.01)	(1.04)	(0.99)
Agreeableness	3.54	3.45	3.63
	(0.92)	(0.87)	(0.95)
Conscientiousness	3.83	3.63	4.04
	(0.87)	(0.86)	(0.83)
Neuroticism	2.85	3.20	2.49
	(1.13)	(1.14)	(1.00)
Openness	3.60	3.77	3.43
	(0.97)	(0.97)	(0.94)
Negative adjective	37%	54%	19%
Engagement in US politics	4.06	4.00	4.12
	(0.66)	(0.67)	(0.65)
Vowel Score	3.49	4.05	2.92
	(2.05)	(1.93)	(2.03)
N	301	151	150

Standard deviations in parentheses.

Other details: (1) Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness constitute the Big Five personality traits, measured in a 1-5 Likert scale, each trait associated to two questions; (2) Engagement in US politics refers to the mean of responses to 4 questions in a 1-5 Likert scale, that concerns voting, following the news and main political events. (3) Vowel Score refers to the score in the real-effort task of correctly counting as many vowels as possible in a limited time frame;

Table A4: Summary statistics for Democrats' sub-sample

	All	Treatment In-group	Out-group	Neutral-group
Score in the LRT (x_i)	7.85 (1.73)	7.92 (1.52)	7.56 (1.75)	8.08 (1.90)
Age	33.54 (10.86)	32.22 (10.59)	33.90 (11.40)	34.49 (10.65)
Female	66%	66%	56%	76%
Student	28%	34%	24%	25%
Strong Affiliation	50%	50%	50%	51%
English is first language	70%	70%	66%	73%
Born in the USA	91%	92%	90%	92%
Resident in the USA	94%	94%	94%	94%
Full-time	42%	40%	42%	43%
Part-time	17%	22%	16%	14%
Extraversion	2.63 (1.04)	2.60 (0.99)	2.48 (0.95)	2.79 (1.15)
Agreeableness	3.45 (0.87)	3.61 (0.88)	3.45 (0.83)	3.28 (0.89)
Conscientiousness	3.63 (0.86)	3.79 (0.93)	3.53 (0.84)	3.57 (0.81)
Neuroticism	3.20 (1.15)	3.42 (1.16)	2.93 (1.07)	3.25 (1.17)
Openness	3.77 (0.97)	4.02 (0.99)	3.68 (0.84)	3.61 (1.05)
Negative adjective	54%	0%	100%	61%
Engagement in US politics	4.00 (0.67)	4.06 (0.71)	4.01 (0.61)	3.94 (0.70)
Vowel Score	4.05 (1.93)	4.42 (1.99)	3.80 (1.75)	3.92 (2.01)
N	151	50	50	51

Standard deviations in parentheses.

Other details: (1) Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness constitute the Big Five personality traits, measured in a 1-5 likert scale, each trait associated to two questions; (2) Engagement in US politics refers to the mean of responses to 4 questions in a 1-5 likert scale, that concerns voting, following the news and main political events. (3) Vowel Score refers to the score in the real-effort task of correctly counting as many vowels as possible in a limited time frame;

Table A5: Summary statistics for Republicans' sub-sample

	All	Treatment In-group	Out-group	Neutral-group
Score in the LRT (x_i)	6.45 (2.49)	6.68 (2.45)	5.92 (2.61)	6.76 (2.37)
Age	39.52 (11.13)	40.80 (11.63)	39.20 (9.84)	38.56 (11.90)
Female	42%	38%	46%	42%
Student	5%	2%	4%	8%
Strong Affiliation	39%	34%	40%	42%
English is first language	50%	48%	50%	52%
Born in the USA	98%	96%	98%	100%
Resident in the USA	89%	92%	86%	90%
Full-time	57%	62%	60%	50%
Part-time	12%	10%	14%	12%
Extraversion	2.71 (0.99)	2.81 (1.01)	2.89 (0.99)	2.44 (0.92)
Agreeableness	3.63 (0.95)	3.65 (0.80)	3.67 (0.94)	3.56 (1.10)
Conscientiousness	4.04 (0.83)	4.19 (0.76)	3.98 (0.89)	3.95 (0.82)
Neuroticism	2.49 (1.00)	2.36 (0.98)	2.50 (0.97)	2.61 (1.05)
Openness	3.43 (0.94)	3.65 (0.89)	3.35 (0.86)	3.28 (1.03)
Negative adjective	0.19 (0.40)	0.00 (0.00)	0.54 (0.50)	0.04 (0.20)
Engagement in US politics	4.12 (0.65)	4.04 (0.72)	4.11 (0.62)	4.21 (0.62)
Vowel Score	2.92 (2.03)	3.06 (2.09)	2.54 (1.72)	3.16 (2.23)
N	150	50	50	50

Standard deviations in parentheses.

Other details: (1) Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness constitute the Big Five personality traits, measured in a 1-5 Likert scale, each trait associated to two questions; (2) Engagement in US politics refers to the mean of responses to 4 questions in a 1-5 Likert scale, that concerns voting, following the news and main political events. (3) Vowel Score refers to the score in the real-effort task of correctly counting as many vowels as possible in a limited time frame;

Hypothesis	Test	D.F., Statistic	p-value
Over-placement frequency			
In-group = Out-group	FET	-	0.029
In-group = Neutral-group	FET	-	0.018
Out-group = Neutral-group	FET	-	0.418
Overestimation of own performance			
Sample mean = 0	t-test	300, 1.163	0.246
In-group mean = 0	t-test	99, 1.118	0.266
Out-group mean = 0	t-test	99, 2.023	0.046
Neutral-group mean = 0	t-test	100, -0.963	0.338
% In-group = % Out-group	FET	-	1.000
% In-group = % Neutral-group	FET	-	0.143
% Out-group = % Neutral-group	FET	-	0.108
Underestimation of others' performance			
Sample mean = 0	t-test	300, -9.995	< 0.001
In-group = Out-group	t-test	198, 4.593	< 0.001
In-group = Neutral-group	t-test	199, 4.581	< 0.001
Out-group = Neutral-group	t-test	199, -0.312	0.755
% In-group = % Out-group	FET	-	0.029
% In-group = % Neutral-group	FET	-	0.002
% Out-group = % Neutral-group	FET	-	0.432
Overplacement of performance			
Sample mean = 0	t-test	300, 8.533	< 0.001
In-group mean = 0	t-test	99, 3.048	0.003
Out-group mean = 0	t-test	99, 6.678	< 0.001
Neutral-group mean = 0	t-test	100, 4.861	< 0.001
In-group = Out-group	t-test	198, -3.911	< 0.001
In-group = Neutral-group	t-test	198, -1.779	0.077
Out-group = Neutral-group	t-test	199, 2.245	0.026

Table A6: Statistical tests: Performance and initial beliefs, Result 1 & Observation 1

Hypothesis	Test	D.F., Statistic	p-value
Over-placement magnitude			
In-group = Out-group	t-test	122, -3.589	< 0.001
In-group = Neutral-group	t-test	124, -0.79	0.433
Out-group = Neutral-group	t-test	138, -3.10	0.002
Overestimation of own performance | overplacers			
In-group = Out-group	t-test	122, 1.210	0.229
In-group = Neutral-group	t-test	124, 2.50	0.014
Out-group = Neutral-group	t-test	138, -1.15	0.251
Underestimation of others' performance | overplacers			
In-group = Out-group	t-test	122, 6.42	< 0.001
In-group = Neutral-group	t-test	124, 4.35	< 0.001
Out-group = Neutral-group	t-test	138, 2.02	0.045
Underplacement magnitude | underplacers			
In-group = Out-group	t-test	61, -0.85	0.398
In-group = Neutral-group	t-test	64, 0.13	0.890
Out-group = Neutral-group	t-test	55, 0.65	0.521

Table A7: Statistical tests: Performance and initial beliefs, Result 2

Hypothesis	Test	D.F., Statistic	p-value
Under-estimation of own performance | bad news			
Sample mean = 0	t-test	160, -1.56	0.121
In-group mean = 0	t-test	59, -1.78	0.081
Out-group mean = 0	t-test	58, -0.45	0.652
Neutral-group mean = 0	t-test	41, -0.54	0.592
Belief revision of own performance | bad news			
In-group mean = 0	t-test	59, -6.08	< 0.001
Out-group mean = 0	t-test	58, -6.83	< 0.001
Neutral-group mean = 0	t-test	41, -3.30	0.002
In-group = Out-group	t-test	117, -0.02	0.986
In-group = Neutral-group	t-test	100, -1.24	0.219
Out-group = Neutral-group	t-test	99, 1.31	0.196
Belief revision of others' performance | bad news			
In-group mean = 0	t-test	59, 0.27	0.786
Out-group mean = 0	t-test	58, 4.35	< 0.001
Neutral-group mean = 0	t-test	41, 3.92	< 0.001
In-group = Out-group	t-test	117, -3.39	< 0.001
In-group = Neutral-group	t-test	100, -3.23	0.002
Out-group = Neutral-group	t-test	99, -0.12	0.902

Table A8: Statistical tests: Revision of beliefs following bad news, Result 3

Hypothesis	Test	D.F., Statistic	p-value
Overplacement revision | bad news			
In-group mean = 0	t-test	59, -7.36	< 0.001
Out-group mean = 0	t-test	58, -9.16	< 0.001
Neutral-group mean = 0	t-test	41, -5.93	< 0.001
Bias in relative performance | bad news			
In-group mean = 0	t-test	59, -1.00	0.322
Out-group mean = 0	t-test	58, 2.80	0.007
Neutral-group mean = 0	t-test	41, -0.91	0.367
Overplacement | bad news			
In-group = Out-group	FET	-	0.010
In-group = Neutral-group	FET	-	0.147
Out-group = Neutral-group	FET	-	0.147
Bias in relative performance | bad news			
In-group = Out-group	t-test	117, -2.85	0.005
In-group = Neutral-group	t-test	100, -1.36	0.176
Out-group = Neutral-group	t-test	99, 1.22	0.227

Table A9: Statistical tests: Revision of beliefs following bad news, Result 4

C Deviations from pre-register

Here we report and justify deviations in the adopted methodology relative to the pre-register.

For hypotheses 1 and 3, the pre-registered dependent variable in the regressions were not adequate, not addressing the hypotheses. Instead of the probability of overplacement occurring, the dependent variable was pre-registered as the probability of the relative performance gap being believed to be positive ($b_i - b_j > 0$). Even after controlling for the observed relative performance ($x_i - x_j$), the coefficient of the key regressor, social identity, is not measuring its effect on overplacement.

While hypotheses were paraphrased, their meaning did not change relative to the pre-registered script.

Finally, an additional participant was recruited by mistake, a Democrat in the neutral-group treatment. Thus, the participant pool of the main sample sums up to 301 instead of 300.

D Complementary data analysis and robustness checks

In this section, we report the remainder of the pre-registered analysis besides other details. In sub-section D.1 we present our analysis in detail regarding Hypotheses 3 to 5. In sub-section D.2 we analyse the effect of different variables in the relationship between social identity and

overplacement and its components. In sub-section D.3, Mann-Whitney U tests are presented, complementing the t-tests presented in the main text. Additionally, in subsection D.4, regression estimates are reported as a robustness check to results 1 to 4. Note that this analysis focuses on the comparison of out-group and in-group treatments only, in line with the Hypotheses. Finally, in sub-section D.8, we present analogues of the tables in the main text which report performance, pre-announcement beliefs and overplacement (both unconditionally and conditional on being an overplacer) for the Democrat and Republican sub-samples.

In sub-section D.5, we assess the relationship between the two measures of ability available to our study, the Vowel-counting task and the score in the main task, the logical reasoning test. In sub-section D.6, we share additional evidence on how accurate participants are in their beliefs of own performance. Finally, in section D.7, we show tests for the main hypotheses considering an alternative reference for the opponent's score based on the main sample itself instead of the auxiliary sample - results don't change.

D.1 Winners and Hypotheses 3 to 5

While overplacement is significant among known-to-be winners ($t(139) = 14.88, p < 0.001$), there are important differences in overplacement levels post-announcement: winners do not significantly overplace their performance in the in-group treatment ($t(39) = 1.60, p = 0.117$), but they do so in both the out-group ($t(40) = 2.55, p = 0.015$) and neutral-group ($t(58) = 5.73, p < 0.001$) treatments. Even so, as stated in Result 5, the relative frequency of overplacement is not significantly different across all three treatments (in-group vs. out-group: $p = 0.657$, in-group vs. neutral-group: $p = 0.095$, out-group vs. neutral-group: $p = 0.291$).

Result 5 (Revised Overplacement Frequency)**.** *Being informed of having won, the frequency of overplacement in revised beliefs is not different as a function of the identity of the opponent.*

When beliefs are elicited a second time, winners on average underestimate their performance ($t(139) = -5.36, p < 0.001$). This is a combination of a natural ceiling effect (the average score for winners was 9 out of 10, which limits the scope for overestimation), and the tendency for people to underestimate their own performance when doing an easy task (Moore and Healy, 2008).

We next examine how winners revised their beliefs about their own performance post-announcement ($b_{i2} - b_{i1}$), which is equivalent to them revising their overestimation, as well as their beliefs about the performance of their opponents ($b_{j2} - b_{j1}$).

Result 6 (Revision beliefs about own performance by winners)**.** *Being informed of having won, there are no significant differences in upward updates of belief of own performance ($b_{i2} - b_{i1}$) as a function of the opponents' identity.*

We see a small yet significant revision upwards in beliefs about own performance across all treatments following the winning announcement (in-group: $t(39) = 2.01, p = 0.051$; out-group: $t(40) = 3.54, p = 0.001$; neutral-group: $t(58) = 3.58, p < 0.001$). That revision of beliefs is not significantly different across treatments (in-group vs. out-group, $t(79) = -0.09, p = 0.930$; in-group vs. neutral group, $t(97) = -0.96, p = 0.341$; out-group vs. neutral-group, $t(98) = 1.01, p = 0.315$), namely between out-group and and in-group as stated in Result 6.

Result 7 (Revision beliefs about others' performance by winners)**.** *Being informed of having won, there are no significant differences in downward updates of belief about others' performance ($b_{j2} - b_{j1}$) as a function of the opponents' identity.*

In addition, we see a downwards revision in beliefs about others' performance across all treatments, although that change is only significant for the in-group case (in-group: $t(39) = -3.39, p = 0.002$; out-group: $t(40) = -1.67, p = 0.103$; neutral-group: $t(58) = -0.23, p = 0.823$). Moreover, the difference between in-group and out-group treatments is not significant ($t(79) = -1.31, p = 0.193$), in line with Result 7.

These two effects combined translate to an increase in overplacement by winners across all treatments (in-group: $t(39) = 5.42, p < 0.001$, out-group: $t(40) = 3.81, p < 0.001$, neutral-group: $t(58) = 4.60, p < 0.001$). In other words, winners do not get better at judging their relative ability.

D.2 Exploratory analysis on individual heterogeneity

In this section, among other variables, we assess any potential effect of political affiliation on the results on a strictly empirical basis. Firstly, we notice that Democrat's average score is higher than Republican's across every treatment (in-group: t(98)=-3.0365, p=0.0031; out-group: t(98)=-3.6898, p=0.0004; neutral-group: t(99) = -3.0908, p=0.0026). Moreover, the likelihood of overplacement happening is not significantly lower in the in-group relative to the neutral-group among democrats (t(99)=-1.2905, p=0.1999) but it is so among republicans (t(98)=-2.1035, p=0.0380). The likelihood of overplacement happening is similar in the out-group and neutral-group treatments across both democrats (t(99)=-0.5211, p=0.6034) and republicans (t(98)=0.9949, p=0.3222). Likewise, the likelihood of overplacement happening is not significantly lower in the in-group relative to the out-group among democrats (t(98)=-1.8149, p=0.0726) and republicans (t(98)=-1.0975, p=0.2751). Note, however, that this analysis by sub-samples lacks power.

Moving to regression estimations, Tables A11, A12 and A13 summarize the evidence of any potential interaction effects.

We start by looking at beliefs pre-announcement. Regression (1) of Table A11 shows that Democrats significantly overestimate their performance by less than Republicans in the

Table A10: Beliefs revision and overplacement conditional on having won.

		Treatment		
	All	In-group	Out-group	Neutral-group
Own performance (x_i)	9.02 (0.78)	9.23 (0.77)	8.88 (0.71)	8.98 (0.82)
Revised belief about own performance (b_{i2})	8.46 (1.20)	8.55 (1.24)	8.37 (1.22)	8.47 (1.16)
Bias in revised estimation of own performance ($b_{i2} - x_i$)	-0.56 (1.23)	-0.68 (1.42)	-0.51 (1.16)	-0.51 (1.15)
% overestimating own performance pre-announcement	14%	15%	12%	14%
% overestimating own performance post-announcement	16%	18%	15%	17%
Change in beliefs about own performance ($b_{i2} - b_{i1}$)	0.42 (0.99)	0.33 (1.02)	0.34 (0.62)	0.54 (1.16)
Revised belief about others' performance (b_{j2})	6.55 (1.24)	6.75 (1.21)	6.56 (1.38)	6.39 (1.16)
Bias in revised estimation of others' performance ($b_{j2} - x_j$)	-1.11 (1.27)	-0.94 (1.27)	-0.99 (1.40)	-1.31 (1.16)
% underestimating others' performance pre-announcement	66%	58%	56%	80%
% underestimating others' performance post-announcement	77%	70%	73%	85%
Change in beliefs about others' performance ($b_{j2} - b_{j1}$)	-0.28 (1.10)	-0.63 (1.17)	-0.29 (1.12)	-0.03 (0.98)
Revised (Over)placement $(b_{i2} - b_{j2}) - (x_i - x_j)$	0.55 (1.11)	0.26 (1.03)	0.47 (1.19)	0.80 (1.07)
% overplacing performance	61%	53%	59%	69%
Change in (over)placement $(b_{i2} - b_{j2}) - (b_{i1} - b_{j1})$	0.70 (1.04)	0.95 (1.11)	0.63 (1.07)	0.57 (0.95)
N	140	40	41	59

Notes: Standard deviations in parentheses.

out-group and in-group treatments (in-group: $b = -1.23, F(1, 284) = 6.98, p = 0.009$; out-group: $b = -0.81, F(1, 284) = 2.77, p = 0.097$; neutral-group: $b = -1.71, F(1, 284) = 12.30, p < 0.001$). Regression (2) shows that Democrats underestimate their competitors' performance to a larger extent than Republicans, with statistical significance, only in the in-group treatment (in-group: $b = -1.00, F(1, 284) = 11.60, p < 0.001$; out-group: $b = 0.57, F(1, 284) = 1.88, p = 0.172$; neutral-group: $b = -0.25, F(1, 284) = 0.43, p = 0.513$). Estimates from regression (3) reveal that Democrats overplace their performance by less than Republicans in all treatments but not significantly in the in-group treatment (in-group: $b = -0.23, F(1, 284) = 0.28, p = 0.600$; out-group: $b = -1.38, F(1, 284) = 6.29, p = 0.013$; neutral-group: $b = -1.46, F(1, 284) = 11.23, p < 0.001$). Focusing on other determinants of overplacement, while a lower Prolific Score is linked with higher overestimation of own performance, and larger age is linked with less underestimation of the opponent's, these effects do not determine overplacement significantly (Prolific Score: $b = -0.19, F(1, 284) = 2.07, p = 0.151$; Age: $b = -0.02, F(1, 284) = 1.73, p = 0.190$). Higher conscientiousness correlates significantly with more underestimation of the opponent's performance which is marginally reflected in more overplacement ($b = 0.32, F(1, 284) = 3.73, p = 0.055$). By contrast, a higher Vowel Score is strongly correlated with less overplacement ($b = -0.27, F(1, 284) = 14.12, p < 0.001$) via significantly less overestimation of own performance. This is consistent with Kruger and Dunning (1999) that links lower ability with higher overconfidence.

Observation 3. *Democrats tend to overplace less than Republicans but not significantly so in the in-group treatment. Higher ability consistently correlates with lower overplacement.*

We now move to the update of beliefs post-announcement, and we start with winners. We repeat the analysis of regressions (1-3) by assessing potential differences between Democrats and Republicans. Regression (4) looks at the change in the belief about one's own performance and we find that Democrats increase their guess on own performance by less than Republicans in the out-group treatment only, with no significant differences between Democrats and Republicans in the other treatments (in-group: $b = -0.08, F(1, 123) = 0.04, p = 0.843$; out-group: $b = -0.84, F(1, 123) = 7.77, p = 0.013$; neutral-group: $b = 0.22, F(1, 123) = 0.95, p = 0.332$). Regression (5) looks at the change in the belief about others' performance. We do not find significant differences between Republicans and Democrats in any treatment (in-group: $b = 0.45, F(1, 123) = 1.48, p = 0.226$; out-group: $b = -0.24, F(1, 123) = 0.21, p = 0.645$; neutral-group: $b = 0.00, F(1, 123) = 0.00, p = 0.990$). Regression (6) takes the revision or change in overplacement as the dependent variable. We observe no significant differences between Democrats and Republicans across treatments (in-group: $b = -0.52, F(1, 123) = 2.64, p = 0.107$; out-group: $b = -0.60, F(1, 123) = 2.03, p = 0.157$; neutral-group: $b = 0.22, F(1, 123) = 0.87, p = 0.354$). Considering other determinants, upon winning, older individuals update about their opponent's performance more negatively although this

is not reflected in the revision of overplacement ($b = 0.01, F(1, 123) = 0.43, p = 0.514$). For the latter, Conscientiousness, Prolific Score and Affiliation intensity have only marginally negative effects (Conscientiousness: $b = -0.22, F(1, 123) = 3.38, p = 0.069$; Prolific Score: $b = -0.25, F(1, 123) = 3.40, p = 0.068$; Affiliation intensity: $b = -0.30, F(1, 123) = 3.05, p = 0.083$, respectively). By contrast, we find that women increase their overplacement significantly more than men following positive news ($b = 0.66, F(1, 123) = 18.94, p < 0.001$), driven by a significantly larger upward update of the belief on own performance.

Observation 4. *We do not find consistent differences between Republicans and Democrats in the update of beliefs that concern overplacement following positive news. Female participants are associated to magnifying overplacement more than men.*

We now look at losers' post-announcement beliefs. Based on Regression (7), differences between Democrats and Republicans regarding the update of belief of own performance following a loss are not consistent across treatments with a larger contraction in the in-group and a smaller contraction in the neutral-group although the latter only marginally significant (in-group: $b = -1.11, F(1, 144) = 4.15, p = 0.044$; out-group: $b = 0.43, F(1, 144) = 0.91, p = 0.34$; neutral-group: $b = 1.13, F(1, 144) = 3.01, p = 0.085$). Regression (8) shows that Democrats' upward update of the opponents' performance is marginally larger than that of Republicans in the in-group treatment with no significant differences in the other treatments (in-group: $b = 0.75, F(1, 144) = 2.88, p = 0.092$; out-group: $b = 0.50, F(1, 144) = 0.79, p = 0.377$; neutral-group: $b = -0.33, F(1, 144) = 0.41, p = 0.523$). According to Regression (9), overplacement revision is significantly different between Democrats and Republicans in the neutral-group only, such that the (downward) mitigation of overplacement is smaller for Democrats (in-group: $b = 0.35, F(1, 144) = 0.48, p = 0.490$; out-group: $b = -0.07, F(1, 144) = 0.01, p = 0.906$; neutral-group: $b = 1.46, F(1, 144) = 4.12, p = 0.044$). Age and, to a less extent, intensity of affiliation and extraversion are positively correlated with the revision of overplacement, making its mitigation less pronounced (Age: $b = 0.03, F(1, 144) = 5.55, p = 0.020$; Affiliation Intensity: $b = 0.56, F(1, 144) = 2.83, p = 0.095$; Extraversion: $b = 0.36, F(1, 144) = 3.71, p = 0.056$). Interestingly, the smaller downwards update of the belief of own performance of the more conscientious is compensated by the larger upwards update of the belief about others' performance, thus having no significant effect on the revision of overplacement ($b = -0.09, F(1, 144) = 0.16, p = 0.689$).

Observation 5. *We do not find consistent differences between Republicans and Democrats in the update of beliefs that concern overplacement following negative news. Age is associated to an upwards revision of overplacement.*

Next, we present an analysis of heterogeneous effects by looking at the likelihood of overestimating or overplacing performance, both pre- and post-announcement, and explore

Table A11: Individual-level determinants of inaccurate beliefs and overplacement

	Pre-announcement			Post-announcement					
				Won			Lost		
DV:	$(b_{i1} - x_i)$	$(b_{j1} - x_j)$	O_{ij1}	$\Delta(b_i - x_i)$	$\Delta(b_j - x_j)$	ΔO_{ij}	$\Delta(b_i - x_i)$	$\Delta(b_j - x_j)$	ΔO_{ij}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
In-group	0.29	1.48***	-1.19***	0.06	-0.92***	0.98***	-0.71	-1.63***	0.92
	(0.53)	(0.36)	(0.42)	(0.29)	(0.33)	(0.27)	(0.64)	(0.52)	(0.57)
Out-group	0.18	-0.50	0.68	0.65**	-0.11	0.76*	-0.15	-0.33	0.18
	(0.60)	(0.41)	(0.51)	(0.28)	(0.53)	(0.41)	(0.51)	(0.51)	(0.57)
Democrat	-1.71***	-0.25	-1.46***	0.22	0.00	0.22	1.13*	-0.33	1.46**
	(0.49)	(0.38)	(0.44)	(0.23)	(0.21)	(0.24)	(0.65)	(0.52)	(0.72)
Democrat ×	0.48	-0.76*	1.23**	-0.30	0.44	-0.74*	-0.02	1.09*	-1.11
In-group	(0.67)	(0.45)	(0.57)	(0.40)	(0.39)	(0.38)	(0.85)	(0.65)	(0.87)
Democrat ×	0.90	0.82	0.08	-1.06***	-0.24	-0.82*	-0.70	0.83	-1.53*
Out-group	(0.70)	(0.56)	(0.69)	(0.38)	(0.57)	(0.48)	(0.77)	(0.79)	(0.91)
Strong Affiliation	0.00	0.31	-0.31	0.03	0.34**	-0.30*	0.04	-0.53*	0.56*
	(0.28)	(0.20)	(0.27)	(0.16)	(0.16)	(0.17)	(0.34)	(0.29)	(0.33)
Female	-0.25	0.06	-0.31	0.42***	-0.24	0.66***	0.30	-0.14	0.44
	(0.28)	(0.22)	(0.27)	(0.15)	(0.17)	(0.15)	(0.34)	(0.30)	(0.35)
Full-time	-0.21	-0.22	0.02	-0.03	0.08	-0.11	0.25	0.29	-0.04
	(0.28)	(0.21)	(0.28)	(0.16)	(0.18)	(0.16)	(0.35)	(0.30)	(0.37)
Age	0.00	0.02**	-0.02	-0.01	-0.02**	0.01	0.03**	-0.01	0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Prolific Score	-0.27**	-0.08	-0.19	-0.07	0.17	-0.25*	0.01	-0.15	0.16
	(0.12)	(0.10)	(0.13)	(0.06)	(0.11)	(0.13)	(0.16)	(0.10)	(0.16)
Extraversion	-0.02	0.15	-0.17	-0.02	-0.15	0.13	0.12	-0.24*	0.36*
	(0.14)	(0.10)	(0.15)	(0.11)	(0.10)	(0.10)	(0.20)	(0.14)	(0.19)
Agreeableness	-0.01	0.12	-0.12	0.08	-0.04	0.13	-0.04	0.14	-0.18
	(0.16)	(0.12)	(0.15)	(0.14)	(0.12)	(0.10)	(0.15)	(0.14)	(0.18)
Neuroticism	0.05	0.06	-0.01	-0.09	-0.14	0.05	0.20	-0.00	0.20
	(0.13)	(0.11)	(0.14)	(0.09)	(0.10)	(0.08)	(0.14)	(0.16)	(0.18)
Conscientiousness	-0.04	-0.36***	0.32*	-0.13	0.09	-0.22*	0.31*	0.40**	-0.09
	(0.17)	(0.13)	(0.16)	(0.10)	(0.10)	(0.12)	(0.18)	(0.17)	(0.21)
Openness	0.08	0.09	-0.01	-0.04	0.06	-0.10	-0.12	-0.08	-0.04
	(0.13)	(0.11)	(0.13)	(0.10)	(0.10)	(0.10)	(0.13)	(0.13)	(0.15)
Vowel Score	-0.27***	0.00	-0.27***	0.04	0.06	-0.02	0.02	0.04	-0.01
	(0.07)	(0.06)	(0.07)	(0.04)	(0.05)	(0.04)	(0.09)	(0.07)	(0.10)
Constant	28.17**	5.55	22.62*	8.37	-16.70	25.08*	-5.40	15.25	-20.65
	(12.13)	(10.04)	(13.17)	(5.52)	(10.59)	(13.31)	(15.62)	(9.89)	(16.10)
N	301	301	301	140	140	140	161	161	161
R^2	0.20	0.17	0.23	0.14	0.17	0.28	0.14	0.18	0.20

Notes: Robust standard errors in parentheses.

whether individual-level characteristics played a role in determining inaccurate beliefs about own/others' performance and/or overplacement[18].

We estimated a series of models of the form:

$$y_i = \beta_0 + \beta_1 I + \beta_2 O + \beta_3 I \times D + \beta_4 O \times D + X\gamma + \varepsilon_i \tag{2}$$

where I, O and D are dummies for in-group and out-group treatments, respectively (the omitted category is neutral-group), D is a dummy for Democrat voters, and X is a vector of characteristics, including an intensity of political affiliation dummy (Strong Affiliation), a female dummy (Female), the Democrat voter dummy (Democrat), a dummy for full-time employment (Full-time), age, the acceptance score on the Prolific platform, which is a proxy for quality (Prolific Score), the big-five personality characteristics, and a proxy for ability from the vowel-counting task (Vowel Score). Vowel Score is significantly correlated with the score on the logical reasoning test; and allows us to proxy ability without including x_i as a regressor, which would lead to endogeneity issues.[19] Detailed analysis including estimation tables are reported in Appendix D.2.

Table A12 reports on estimates from a linear probability model outlined in equation (2).[20]

Like the previous analysis, we start by looking at beliefs pre-announcement. Regression (1) shows that Democrats are less likely to overestimate their performance than Republicans in the in-group and neutral-group treatments (in-group: $b = -0.34, F(1, 284) = 5.81, p = 0.017$; out-group: $b = -0.40, F(1, 284) = 2.23, p = 0.137$; neutral-group: $b = -0.38, F(1, 284) = 17.85, p < 0.001$). Regression (2) shows that there are no differences between Democrats and Republicans in the likelihood of underestimating others in none of treatments (in-group: $b = -0.01, F(1, 284) = 0.01, p = 0.917$; out-group: $b = -0.08, F(1, 284) = 0.72, p = 0.398$; neutral-group: $b = 0.14, F(1, 284) = 2.28, p = 0.132$).

Moreover, Regression (3) shows that Democrats are less likely to overplace themselves than Republicans in the neutral-group while we find no difference in the other treatments (in-group: $b = -0.12, F(1, 284) = 1.18, p = 0.279$; out-group: $b = -0.05, F(1, 284) = 0.23, p = 0.634$; neutral-group: $b = -0.19, F(1, 284) = 4.05, p = 0.045$). Female participants are less likely

[18] When the dependent variable refers to the likelihood of overconfidence i.e. a binary variable, the analysis considers a linear probability model, robust standard errors and confirms that the predicted probabilities fall within the unit interval, addressing the common issues (Horrace and Oaxaca, 2006).

[19] The Pearson correlation between performance in the vowel-counting task and performance in the main task is 0.48, $p < 0.001$. We also find evidence for a significant positive relationship between these two variables using OLS regressions. See Appendix D.5 for details.

[20] Using a linear probability model (i.e., OLS) to estimate a binary variable can lead to bias if the linear probability model predicted probabilities fall outside the unit interval; moreover, OLS imposes an assumption of homoskedasticity (Horrace and Oaxaca, 2006). We resolve the latter problem by estimating robust standard errors. The former concern is unlikely to apply to our case, as our predicted probabilities generated by our estimations lie within the unit interval.

to overplace than men but the effect is only marginal ($b = -0.11, F(1, 284) = 3.61, p = 0.059$). Higher conscientiousness correlates strongly with a higher probability of underestimation of the opponent's performance which is significantly reflected in more overplacement ($b = 0.09, F(1, 284) = 6.43, p = 0.012$). Still consistent with the literature on the Dunning-Kruger effect, a higher Vowel Score is strongly associated to a smaller likelihood of overplacement ($b = -0.03, F(1, 284) = 4.42, p = 0.036$) via significantly smaller likelihood of overestimation of own performance.

Observation 6. *Democrats are less likely to overestimate their performance and less likely to overplace themselves than Republicans in the neutral-group treatment only. Lower ability and higher conscientiousness consistently correlate with higher overplacement likelihood.*

Turning to beliefs after a win announcement, regressions (4) and (5) reveal a difference between Democrats and Republicans in the likelihood of revising own performance upwards but only in the out-group treatment (in-group: $b = -0.13, F(1, 123) = 0.73, p = 0.394$; out-group: $b = -0.56, F(1, 123) = 11.18, p = 0.001$; neutral-group: $b = 0.00, F(1, 123) = 0.00, p = 0.972$) and no difference in the likelihood of revising others' performance downwards (in-group: $b = -0.06, F(1, 123) = 0.12, p = 0.73$; out-group: $b = 0.07, F(1, 123) = 0.17, p = 0.683$; neutral-group: $b = 0.12, F(1, 123) = 1.56, p = 0.214$). Moreover, regression (6) suggests no differences in the likelihood of increasing overplacement post-announcement (in-group: $b = -0.18, F(1, 123) = 1.36, p = 0.245$; out-group: $b = -0.16, F(1, 123) = 0.84, p = 0.360$; neutral-group: $b = -0.00, F(1, 123) = 0.00, p = 0.978$).

We find significant coefficients Female ($b = 0.28, F(1, 123) = 11.95, p < 0.001$) via both larger likelihood of upward update of own and downward update of others' performances' beliefs[21], Prolific Score ($b = -0.086, F(1, 123) = 5.29, p = 0.023$) and, albeit marginally, Extraversion ($b = 0.08, F(1, 123) = 2.77, p = 0.099$).

Observation 7. *Following a win announcement, women have a larger likelihood of increasing overplacement than men, while the opposite is found for participants with a higher Prolific Score. By contrast, no significant differences are found between Democrats and Republicans within each treatment.*

Finally, we look at beliefs after a loss announcement. Regressions (7), (8) and (9) show that Democrats are more likely to update upwards their belief of own performance (in-group: $b = -0.00, F(1, 144) = 0.00, p = 0.955$; out-group: $b = 0.03, F(1, 144) = 0.26, p = 0.611$;

[21] Note that men are often associated to overconfidence more than women. Niederle and Vesterlund (2007) find that men are more willing to engage in competitive tournaments than women, which is likely explained by higher overconfidence. Jakobsson et al. (2013) suggest that overconfidence varies by gender depending on the existing stereotypes on the performed task and gender representation in the set environment. Our data suggests that the role of gender on overconfidence is nuanced. While at the start female participants tend to overplace less than men, they are more likely to magnify it following positive news.

neutral-group: $b = 0.23, F(1, 144) = 3.85, p = 0.052$), less likely to update downwards their belief of others' performance (in-group: $b = -0.05, F(1, 144) = 0.20, p = 0.655$; out-group: $b = -0.02, F(1, 144) = 0.08, p = 0.783$; neutral-group: $b = -0.22, F(1, 144) = 4.77, p = 0.031$) and more likely to increase overplacement (in-group: $b = -0.01, F(1, 144) = 0.01, p = 0.905$; out-group: $b = 0.03, F(1, 144) = 1.01, p = 0.318$; neutral-group: $b = 0.22, F(1, 144) = 4.13, p = 0.044$), but only in the neutral-group treatment. Moreover, we find that the likelihood of increasing overplacement following a loss announcement is positively correlated with Prolific Score, but only marginally ($b = 0.02, F(1, 144) = 2.84, p = 0.094$). Note that Conscientiousness and Vowel Score, and, albeit marginally, Age, have significant coefficients on the likelihood of updating downwards the belief of others' performance but have no significant effect on the likelihood of increasing overplacement.

Observation 8. *Following a loss announcement, differences between Democrats and Republicans are only found when participants face a neutral-group member.*

We complement this analysis by assessing how the likelihood of overplacement and its subcomponents post-announcement are affected by individual heterogeneity. Nevertheless, in Table A13 we repeat regressions (1)-(3) of the former table that concern pre-announcement likelihoods, to ease comparison with the remaining regressions (4)-(9) that concern post-announcement likelihoods.

Following a win announcement, differences in the likelihood of overestimation between Democrats and Republicans are inconsistent across treatments (in-group: $b = -0.29, F(1, 123) = 7.01, p = 0.009$; out-group: $b = 0.21, F(1, 123) = 5.78, p = 0.018$; neutral-group: $b = -0.15, F(1, 123) = 1.98, p = 0.162$). Moreover, there are no differences between Democrats and Republicans in the likelihood of underestimating others (in-group: $b = -0.06, F(1, 123) = 0.12, p = 0.730$; out-group: $b = 0.07, F(1, 123) = 0.17, p = 0.683$; neutral-group: $b = 0.12, F(1, 123) = 1.56, p = 0.214$) nor in the likelihood of overplacing performance (in-group: $b = -0.18, F(1, 123) = 1.36, p = 0.245$; out-group: $b = -0.16, F(1, 123) = 0.84, p = 0.360$; neutral-group: $b = -0.00, F(1, 123) = 0.00, p = 0.978$). Neuroticism ($b = 0.12, F(1, 123) = 7.82, p = 0.006$) and Conscientiousness ($b = 0.16, F(1, 123) = 10.38, p = 0.002$), and to a small extent Age ($b = 0.01, F(1, 123) = 4.72, p = 0.032$), are positively and significantly correlated with the likelihood of overplacement, while Strong Affiliation ($b = -0.20, F(1, 123) = 226.19, p < 0.001$) and, albeit marginally, Agreeableness ($b = -0.09, F(1, 123) = 3.85, p = 0.052$) are negatively correlated with it.

Observation 9. *Following a win announcement, larger Neuroticism and Conscientiousness and less intense political affiliation are correlated with a larger likelihood of overplacing performance. By contrast, no significant differences are found between Democrats and Republicans within each treatment.*

Table A12: Individual-level determinants of likelihood of overestimation and overplacement, pre-announcement levels and post-announcement changes

	Pre-announcement			Post-announcement					
				Won			Lost		
DV:	$(b_{i1}-x_i)^+$	$(b_{j1}-x_j)^-$	O^+_{ij1}	$\Delta^+(b_i-x_i)$	$\Delta^-(b_j-x_j)$	Δ^+O_{ij}	$\Delta^+(b_i-x_i)$	$\Delta^-(b_j-x_j)$	Δ^+O_{ij}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
In-group	0.04	-0.17*	-0.21**	0.10	0.38***	0.37***	-0.01	0.06	0.09*
	(0.10)	(0.09)	(0.08)	(0.13)	(0.12)	(0.13)	(0.06)	(0.11)	(0.05)
Out-group	-0.02	0.07	-0.10	0.42**	0.15	0.18	0.01	-0.07	-0.00
	(0.10)	(0.09)	(0.08)	(0.17)	(0.15)	(0.16)	(0.05)	(0.10)	(0.01)
Democrat	-0.38***	0.14	-0.19**	0.00	0.12	-0.00	0.23*	-0.22**	0.22**
	(0.09)	(0.09)	(0.09)	(0.12)	(0.10)	(0.13)	(0.12)	(0.10)	(0.11)
Democrat ×	0.14	-0.15	0.07	-0.14	-0.18	-0.17	-0.23*	0.17	-0.23*
In-group	(0.13)	(0.13)	(0.13)	(0.19)	(0.18)	(0.19)	(0.13)	(0.14)	(0.12)
Democrat ×	0.23*	-0.22*	0.14	-0.56***	-0.06	-0.15	-0.20	0.20	-0.19*
Out-group	(0.13)	(0.13)	(0.13)	(0.21)	(0.18)	(0.21)	(0.14)	(0.13)	(0.11)
Strong Affiliation	0.03	-0.11*	-0.08	-0.03	-0.11	-0.13	-0.03	0.05	-0.02
	(0.06)	(0.05)	(0.05)	(0.08)	(0.08)	(0.08)	(0.04)	(0.06)	(0.03)
Female	0.03	-0.00	-0.11*	0.18**	0.18**	0.28***	-0.05	0.02	-0.04
	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.08)	(0.04)	(0.06)	(0.04)
Full-time	0.01	0.10*	0.03	-0.03	-0.06	-0.08	-0.04	0.03	-0.01
	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.08)	(0.04)	(0.06)	(0.04)
Age	0.00	-0.00*	0.00	-0.00	0.00	0.00	0.00	0.01*	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Prolific Score	-0.01	0.03	-0.01	-0.12***	-0.01	-0.09**	-0.00	0.02	0.02*
	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.04)	(0.01)	(0.02)	(0.01)
Extraversion	-0.03	-0.05	-0.00	0.02	0.07*	0.08*	0.03	0.03	0.04
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(0.03)	(0.03)
Agreeableness	0.01	-0.01	-0.02	0.05	-0.01	0.03	-0.03	0.01	-0.03
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.02)	(0.03)	(0.02)
Neuroticism	0.01	-0.00	0.05	-0.02	-0.00	-0.01	0.00	0.02	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.02)	(0.03)	(0.02)
Conscientiousness	0.01	0.10***	0.09**	-0.08	0.01	-0.08	0.05	-0.08**	0.00
	(0.04)	(0.03)	(0.03)	(0.05)	(0.05)	(0.06)	(0.03)	(0.04)	(0.03)
Openness	0.01	-0.00	-0.00	-0.04	-0.03	-0.05	-0.01	0.00	-0.02
	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.02)	(0.03)	(0.02)
Vowel Score	-0.05***	0.02	-0.03**	0.04*	-0.01	0.01	-0.00	-0.03**	-0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Constant	1.76	-2.72	1.54	12.17***	0.98	9.04**	0.39	-2.33	-1.55*
	(2.39)	(2.27)	(2.09)	(3.89)	(3.28)	(3.73)	(1.26)	(1.87)	(0.92)
N	301	301	301	140	140	140	161	161	161
R^2	0.16	0.14	0.13	0.20	0.15	0.20	0.13	0.14	0.16

Notes: The dependent variable in regressions (1) is $\mathbb{1}(b_{i1}-x_i>0)$. The dependent variable in regressions (2) is $\mathbb{1}(b_{j1}-x_j<0)$. The dependent variable in regressions (3) is $\mathbb{1}(O_{ij1}>0)$, where O_{ij1} is defined by equation (1). The dependent variable in regressions (4) and (7) is $\mathbb{1}(b_{i2}-b_{i1}>0)$. The dependent variable in regressions (5) and (8) is $\mathbb{1}(b_{j2}-b_{j1}<0)$. The dependent variable in regressions (6) and (9) is $\mathbb{1}(O_{ij2}-O_{ij1}>0)$.
LPM estimates. Robust standard errors in parentheses.

Following a loss announcement, Democrats are more likely to overestimate their performance than Republicans in the neutral-group, while no differences are found in the other two treatments (in-group: $b = -0.00, F(1,144) = 0.00, p = 0.955$; out-group: $b = 0.03, F(1,144) = 0.26, p = 0.611$; neutral-group: $b = 0.23, F(1,144) = 3.85, p = 0.052$). Furthermore, Democrats are more likely to underestimate others in the neutral-group only (in-group: $b = -0.12, F(1,144) = 0.63, p = 0.429$; out-group: $b = 0.08, F(1,144) = 0.34, p = 0.558$; neutral-group: $b = 0.31, F(1,144) = 3.09, p = 0.081$). Not surprisingly, it is only in the neutral-group treatment that the likelihood of overplacement in revised beliefs is larger for Democrats than for Republicans (in-group: $b = -0.14, F(1,144) = 1.13, p = 0.290$; out-group: $b = 0.09, F(1,144) = 0.40, p = 0.530$; neutral-group: $b = -0.29, F(1,144) = 3.34, p = 0.070$). Vowel Score is negatively correlated ($b = -0.05, F(1,144) = 6.16, p = 0.014$) with the likelihood of overplacement following a loss announcement, while Conscientiousness is positively yet marginally correlated with it ($b = -0.08, F(1,144) = 2.98, p = 0.087$).

Observation 10. *Following a loss announcement, exclusively when facing a neutral-group member, Democrats are significantly more likely to overestimate their own performance and underestimate others, thus more likely to overplace than Republicans. Higher ability is negatively correlated with revised overplacement.*

D.3 Mann-Whitney U tests

Here we report the non-parametric results relative to Hypotheses 2 and 4 to 7, using the Mann-Whitney's U test or or Wilcoxon Rank Sum Test (hereafter, MWT), in line with the pre-registered methodology. We compare them to the parametric t-tests already reported in the main text. Since hypotheses are directional, we considered one-sided p-values. Table A14 shows that results do not change.

Nevertheless, note that the two tests are not measuring the same thing. While the t-test assesses whether means are the same, the MWT evaluates whether the two distributions are the same. If we assume that the two distributions — of out-group and in-group sub-samples — have the same shape and spread, then we can also test for the equality of medians. Because the distribution of U statistics is symmetric, the one-tailed p-value is obtained by dividing the two-tailed one by 2. By contrast, in the t-test, one-sided p-values differ depending on the direction that is being considered. For that reason, in Table A14, reported values refer to the hypothesized direction. If medians are significantly different, the MWT one-sided p-value is not informative about the direction and it is necessary to look at the median values to test the hypotheses. This is the case of Hypothesis 7 in which the opposite of the hypothesized is observed with statistical significance, which explains the large p-value for the directional t-test and small p-value for the MWT.

Table A13: Individual-level determinants of likelihood of overestimation and overplacement, pre- and post-announcement levels

	Pre-announcement			Post-announcement					
				Won			Lost		
DV:	$(b_{i1}-x_i)^+$	$(b_{j1}-x_j)^-$	O^+_{ij1}	$(b_{i2}-x_i)^+$	$(b_{j2}-x_j)^-$	O^+_{ij2}	$(b_{i2}-x_i)^+$	$(b_{j2}-x_j)^-$	O^+_{ij2}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
In-group	0.04	-0.17*	-0.21**	0.05	-0.08	0.00	-0.09	0.16	-0.15
	(0.10)	(0.09)	(0.08)	(0.14)	(0.15)	(0.14)	(0.14)	(0.14)	(0.14)
Out-group	-0.02	0.07	-0.10	-0.26**	0.04	-0.31	-0.10	0.21	-0.02
	(0.10)	(0.09)	(0.08)	(0.10)	(0.16)	(0.20)	(0.13)	(0.13)	(0.14)
Democrat	-0.38***	0.14	-0.19**	-0.15	0.18*	0.02	-0.22	0.31*	-0.29*
	(0.09)	(0.09)	(0.09)	(0.11)	(0.10)	(0.13)	(0.15)	(0.18)	(0.16)
Democrat ×	0.14	-0.15	0.07	-0.14	-0.10	-0.36**	0.19	-0.43*	0.15
In-group	(0.13)	(0.13)	(0.13)	(0.15)	(0.19)	(0.18)	(0.20)	(0.22)	(0.20)
Democrat ×	0.23*	-0.22*	0.14	0.37***	-0.23	0.30	0.27	-0.22	0.38*
Out-group	(0.13)	(0.13)	(0.13)	(0.13)	(0.18)	(0.24)	(0.20)	(0.22)	(0.21)
Strong Affiliation	0.03	-0.11*	-0.08	-0.04	-0.05	-0.20**	0.10	0.01	0.05
	(0.06)	(0.05)	(0.05)	(0.07)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
Female	0.03	-0.00	-0.11*	-0.01	0.11	-0.03	-0.02	-0.13	-0.03
	(0.06)	(0.06)	(0.06)	(0.07)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)
Full-time	0.01	0.10*	0.03	-0.01	0.08	0.06	0.03	-0.10	-0.03
	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
Age	0.00	-0.00*	0.00	-0.00	0.00	0.01**	0.01**	-0.01	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Prolific Score	-0.01	0.03	-0.01	-0.05	0.00	-0.03	-0.04	0.04	-0.03
	(0.02)	(0.02)	(0.02)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Extraversion	-0.03	-0.05	-0.00	0.00	-0.01	0.07	0.03	0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
Agreeableness	0.01	-0.01	-0.02	-0.01	-0.04	-0.09*	-0.05	-0.00	-0.06
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)
Neuroticism	0.01	-0.00	0.05	0.01	0.03	0.12***	0.03	-0.05	0.06
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
Conscientiousness	0.01	0.10***	0.09**	0.05	0.08	0.16***	0.03	0.05	0.08*
	(0.04)	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Openness	0.01	-0.00	-0.00	-0.03	0.01	-0.02	0.00	-0.03	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Vowel Score	-0.05***	0.02	-0.03**	-0.02	-0.01	-0.03	-0.03	0.02	-0.05**
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	1.76	-2.72	1.54	5.66	0.13	2.99	4.41	-3.36	3.40
	(2.39)	(2.27)	(2.09)	(4.80)	(3.94)	(3.62)	(2.87)	(2.97)	(2.91)
N	301	301	301	140	140	140	161	161	161
R^2	0.16	0.14	0.13	0.16	0.10	0.23	0.14	0.10	0.17

Notes: The dependent variable in regression (1) is $\mathbb{1}(b_{i1}-x_i>0)$, and in (4) and (7) it is $\mathbb{1}(b_{i2}-x_i>0)$. The dependent variable in regression (2) is $\mathbb{1}(b_j-x_j<0)$, and in (5) and (8) it is $\mathbb{1}(b_{j2}-x_j<0)$. The dependent variable in regression (3) is $\mathbb{1}(O_{ij1}>0)$, and in (6) and (9) it is $\mathbb{1}(O_{ij2}>0)$, where O_{ij} is defined by equation (1).
LPM estimates. Robust standard errors in parentheses.

Table A14: Comparing parametric to non-parametric test results

	Out-group (O)		In-group (I)		p-value	
	$\overline{X}(s.d.)$	*Median*	$\overline{X}(s.d.)$	*Median*	t-test	MWT
Hypothesis 2: $O > I$	3.38 (2.55)	3	2.00 (1.43)	1.4	**<0.001**	**0.001**
N (overplacers)	69		55			
Hypothesis 4: $O > I$	0.34 (0.62)	0	0.33 (1.02)	0	0.465	0.431
Hypothesis 5: $O < I$	-0.29 (1.12)	0	-0.63 (1.17)	0	0.904	0.066
N (winners)	41		40			
Hypothesis 6: $O > I$	-1.61 (1.81)	-1	-1.62 (2.06)	-1	0.493	0.370
Hypothesis 7: $O < I$	1.12 (1.97)	1	0.05 (1.42)	0	**<0.001**	**<0.001**
N (losers)	59		60			

all p-values are one-sided; p-values less than 0.05 are in bold;

D.4 Other econometric analysis

To complement the main analysis, in this section, we start by presenting regression estimates as an alternative to mean tests. No control variables are considered in Table A15 while the full set of control variables is considered in Table A16. In both tables, each of the 7 columns regards one of the 7 hypotheses. Results remain consistent. As argued in the main text, using a linear probability model (i.e., OLS) to estimate a binary variable can lead to bias if the predicted probabilities fall outside the unit interval, and OLS imposes an assumption of heteroskedasticity (Horrace and Oaxaca, 2006). We resolve the latter problem by estimating robust standard errors. The former concern is unlikely to apply to our case, as our predicted probabilities generated by our estimations lie within the unit interval.

Notwithstanding, we also report Probit and OLS estimates with different specifications, in line with the pre-registered methodology, still focusing on treatments out-group and in-group. Table A17 confirms 1 with average marginal effects from Probit estimates. Table A18 confirms 2 with OLS estimates, robust standard errors and interaction terms in the set of control variables. Table A19 reports a pre-registered analysis that we later deemed less adequate to test Hypothesis 2. Results would nevertheless point in the same direction.

Tables A20 and A21 decompose the dependent variable, overplacement, in overestimation of own performance and underestimation of other's performance, respectively - among overplacers. The coefficient of Social Identity in Table A20 is consistently not statistically significant thus does not have a role in the bias regarding own performance. By contrast, facing an out-group makes any underestimation of other's performance significantly worse.

Table A22 presents average marginal effects from Probit estimates and confirms Result 5 that refers to overplacement in revised beliefs among winners. Tables A23, A24, A25 and A26 confirm results 6 to 4.

Table A15: Hypotheses testing with regressions

	Pre-announcement		Post-announcement				
			Won			Lost	
DV:	O^{+}_{ij1}	O_{ij1} if $O_{ij1} > 0$	O^{+}_{ij2}	$\Delta(b_i - x_i)$	$\Delta(b_j - x_j)$	$\Delta(b_i - x_i)$	$\Delta(b_j - x_j)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Social Identity	0.140**	1.379***	-0.060	0.016	0.335	0.006	1.069***
	(0.068)	(0.362)	(0.112)	(0.188)	(0.255)	(0.355)	(0.316)
Constant	0.690***	3.383***	0.585***	0.341**	-0.293***	-1.610***	1.119***
	(0.046)	(0.306)	(0.078)	(0.096)	(0.175)	(0.236)	(0.257)
N	200	124	81	81	81	119	119
R-squared	0.021	0.096	0.004	0.000	0.021	0.000	0.090

Notes: The dependent variable in regression (1) is $\mathbb{1}(O_{ij1} > 0)$ and in (3) it is $\mathbb{1}(O_{ij2} > 0)$, where O_{ij} is defined by equation (1). The dependent variable in regression (2) is the magnitude of O_{ij} conditional on $O_{ij} > 0$ happening. In regressions (4) and (6), the dependent variable is $(b_{i2} - b_{i1})$ which is equivalent to $\Delta(b_i - x_i)$, and in (5) and (7), it is $(b_{j2} - b_{j1})$ which is equivalent to $\Delta(b_j - x_j)$. LPM estimates for regressions (1) and (3) and OLS for the remaining. Social identity is a dummy =1 if out-group and =0 if in-group. Robust standard errors in parentheses.

Table A16: Hypotheses testing with regressions - control variables added

	Pre-announcement		Post-announcement				
			Won			Lost	
DV:	O^+_{ij1}	O_{ij1} if $O_{ij1} > 0$	O^+_{ij2}	$\Delta(b_i - x_i)$	$\Delta(b_j - x_j)$	$\Delta(b_i - x_i)$	$\Delta(b_j - x_j)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Social Identity	0.170**	1.182***	0.110	0.094	0.406	-0.226	1.143***
	(0.0713)	(0.334)	(0.129)	(0.199)	(0.268)	(0.422)	(0.343)
Vowel Score	-0.027	-0.213**	-0.019	0.110	0.154*	0.051	0.078
	(0.020)	(0.087)	(0.033)	(0.075)	(0.089)	(0.099)	(0.083)
Democrat	-0.096	-0.734*	-0.028	-0.305	0.157	0.685*	0.470
	(0.083)	(0.424)	(0.144)	(0.306)	(0.306)	(0.374)	(0.387)
Strong Affiliation	-0.079	-0.138	-0.186	-0.246	0.164	-0.008	-0.316
	(0.071)	(0.407)	(0.115)	(0.197)	(0.245)	(0.385)	(0.328)
Female	-0.064	0.411	0.126	0.308	-0.220	0.374	0.148
	(0.072)	(0.377)	(0.114)	(0.199)	(0.272)	(0.375)	(0.330)
Full-time	0.035	-0.616	0.019	0.036	-0.005	0.303	0.401
	(0.070)	(0.403)	(0.121)	(0.208)	(0.222)	(0.426)	(0.345)
Age	0.000	-0.024	0.007	-0.006	-0.028**	0.019	-0.009
	(0.003)	(0.017)	(0.005)	(0.010)	(0.012)	(0.016)	(0.016)
Prolific Score	-0.009	-0.467***	-0.015	-0.087	0.313*	0.101	-0.149
	(0.029)	(0.138)	(0.057)	(0.096)	(0.179)	(0.199)	(0.117)
Extraversion	-0.014	-0.305	0.049	-0.039	-0.251*	0.156	-0.320*
	(0.040)	(0.197)	(0.055)	(0.132)	(0.128)	(0.224)	(0.170)
Agreeableness	0.012	0.062	-0.086	0.036	0.068	0.052	0.143
	(0.043)	(0.202)	(0.063)	(0.106)	(0.118)	(0.192)	(0.157)
Conscientiousness	0.128***	0.235	0.233***	-0.170	0.090	0.266	0.318
	(0.045)	(0.261)	(0.069)	(0.131)	(0.153)	(0.252)	(0.239)
Neuroticism	0.043	-0.153	0.073	-0.125	-0.297*	0.214	-0.100
	(0.040)	(0.203)	(0.061)	(0.130)	(0.151)	(0.173)	(0.176)
Openness	0.017	-0.022	-0.013	0.098	0.226	-0.267	-0.079
	(0.040)	(0.225)	(0.068)	(0.136)	(0.169)	(0.189)	(0.170)
Constant	1.059	51.977***	1.163	9.601	-31.234*	-14.395	15.392
	(2.864)	(13.857)	(5.599)	(9.770)	(17.833)	(19.673)	(11.494)
N	200	124	81	81	81	119	119
R-squared	0.120	0.299	0.217	0.149	0.250	0.109	0.185

Notes: The dependent variable in regression (1) is $\mathbb{1}(O_{ij1} > 0)$ and in (3) it is $\mathbb{1}(O_{ij2} > 0)$, where O_{ij} is defined by equation (1). The dependent variable in regression (2) is the magnitude of O_{ij} conditional on $O_{ij} > 0$ happening. In regressions (4) and (6), the dependent variable is $(b_{i2} - b_{i1})$ which is equivalent to $\Delta(b_i - x_i)$, and in (5) and (7), it is $(b_{j2} - b_{j1})$ which is equivalent to $\Delta(b_j - x_j)$. LPM estimates for regressions (1) and (3) and OLS for the remaining. Social identity is a dummy =1 if out-group and =0 if in-group. Vowel Score is the performance in the vowel-counting task, a proxy for x_i. Robust standard errors in parentheses.

In both Tables A17 and A22, while interaction terms between treatment and affiliation were included in the regressions, their average marginal effects are naturally not reported. We observe that the probability of overplacement changes in a significantly different way between out-group (dummy=1) and in-group due to participants being Democrats (dummy=1) or Republicans. We thus explore this matter further. Remaining aware of the small sample sizes within sub-samples, frequencies (or predictive margins) are nevertheless reported in Tables A27 and A28. There is no statistically significant difference in the likelihood of overplacement for Result 1. By contrast, the enhancing effect (positive coefficient) of being a Democrat as opposed to a Republican on the effect of social identity on revised overplacement among winners is statistically significant (p-values< 0.05).

Additionally, regarding any role of political affiliation in the effect of social identity on the magnitude of overplacement given overplacement, it is not significant (Table A18). Going through the remaining tables, we conclude that the differences between Democrats and Republicans in their effect of social identity on overplacement outcomes, including its subcomponents, is sporadic, thus potentially spurious.

Regarding the effect of any individual heterogeneity, reported in section 3.5 in the main text and sub-section D.2 in the appendix, we confirm that no variable systematically determines behavior except ability (in line with the Kruger-Dunning effect). Moreover, by restricting the sample to treatments out-group and in-group, the effect is smaller than that previously presented in the main text.

Table A17: Effect of social identity (=1 if in-group; =0 if out-group) on the probability of initial overplacement: $Y = 1$ if $(b_{i1} - b_{j1}) - (x_i - x_j) > 0$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Identity	0.14**	0.14**	0.12*	0.12*	0.12*	0.17**	0.17***	0.13*
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Democrat		-0.18***	-0.15**	-0.14*	-0.13	-0.12	-0.08	-0.13*
		(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)
Vowel Score			-0.03	-0.03	-0.03	-0.03	-0.02	-0.03
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Female				-0.03	-0.03	-0.05	-0.06	
				(0.07)	(0.07)	(0.07)	(0.07)	
Age				0.00	0.00	-0.00	0.00	
				(0.00)	(0.00)	(0.00)	(0.00)	
Full-time					0.06	0.02	0.03	
					(0.07)	(0.07)	(0.07)	
Prolific Score					0.01	-0.01	-0.01	
					(0.03)	(0.03)	(0.03)	
Extraversion						-0.02	-0.02	
						(0.04)	(0.04)	
Agreeableness						0.02	0.02	
						(0.04)	(0.04)	
Conscientiousness						0.13***	0.13***	
						(0.04)	(0.04)	
Neuroticism						0.04	0.04	
						(0.04)	(0.04)	
Openness						0.01	0.02	
						(0.04)	(0.04)	
Engagement in US politics							0.00	0.02
							(0.06)	(0.06)
Strong Affiliation							-0.09	-0.07
							(0.07)	(0.07)
English is first language							-0.09	
							(0.07)	
N	200	200	200	200	200	200	200	200
Significance (p-value)								
Social Identity × Democrat	N.A.	0.659	0.695	0.738	0.739	0.606	0.676	0.703

Average marginal effects based on probit results.

In the regression specifications (2)-(8), an interaction term between the dummies Social Identity and Democrat was included. The p-values for the significance test on whether the interaction term's effect is significantly different from zero are reported in the lower part of the table.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A18: Effect of Social Identity (=1 if in-group; =0 if out-group) on the magnitude of overplacement, conditional on overplacement happening: $Y = (b_{i1} - b_{j1}) - (x_i - x_j) | (b_{i1} - b_{j1}) - (x_i - x_j) > 0$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Identity	1.38***	1.92***	1.75***	1.68***	1.67***	1.69***	1.56***	1.74***
	(0.36)	(0.51)	(0.49)	(0.49)	(0.47)	(0.49)	(0.47)	(0.47)
Democrat		-1.61***	1.14*	-1.26**	-1.33**	-1.26**	-1.18**	-1.03**
		(0.58)	(0.57)	(0.54)	(0.53)	(0.57)	(0.58)	(0.60)
Democrat ×		1.15*	1.22*	1.16	1.19*	1.30*	1.24	1.17*
Social Identity		(0.68)	(0.68)	(0.70)	(0.67)	(0.75)	(0.75)	(0.68)
Vowel Score			-0.26***	-0.28***	-0.21**	-0.22**	-0.19**	-0.26***
			(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Female				0.24	0.33	0.38	0.48	
				(0.36)	(0.36)	(0.39)	(0.38)	
Age				-0.02	-0.01	-0.02	-0.03*	
				(0.02)	(0.02)	(0.02)	(0.02)	
Full-time					-0.58	-0.73*	-0.76*	
					(0.35)	(0.41)	(0.39)	
Prolific Score					-0.41***	-0.46***	-0.45***	
					(0.15)	(0.14)	(0.14)	
Extraversion						-0.30*	-0.39*	
						(0.17)	(0.20)	
Agreeableness						-0.02	-0.03	
						(0.21)	(0.21)	
Conscientiousness						0.28	0.25	
						(0.27)	(0.28)	
Neuroticism						-0.21	-0.19	
						(0.20)	(0.20)	
Openness						-0.04	-0.11	
						(0.22)	(0.23)	
Engagement in US politics							0.70**	0.48*
							(0.33)	(0.28)
Strong Affiliation							-0.41	-0.51
							(0.40)	(0.40)
English is first language							-0.50	
							(0.37)	
Constant	3.38***	4.11***	4.65***	5.25***	46.27***	51.91***	49.20***	2.87**
	(0.31)	(0.43)	(0.49)	(0.83)	(14.87)	(13.76)	(14.04)	(1.17)
N	124	124	124	124	124	124	124	124
R^2	0.10	0.17	0.22	0.22	0.29	0.32	0.35	0.24

OLS coefficient estimations.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A19: Effect of social identity (=1 if in-group; =0 if out-group) on the magnitude of perceived relative performance, given overplacement happened: $Y = (b_{i1} - b_{j1})|(b_{i1} - b_{j1}) - (x_i - x_j) > 0$, with $(x_i - x_j)$ as control variable

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Relative performance $(x_i - x_j)$	0.33***	0.35***	0.33***	0.32***	0.34***	0.32***	0.33***	0.33***
	(0.07)	(0.08)	(0.09)	(0.09)	(0.11)	(0.10)	(0.10)	(0.09)
Social Identity	-1.01***	-1.04***	-1.05***	-1.01**	-1.03**	-0.97**	-0.96**	-1.06***
	(0.26)	(0.38)	(0.38)	(0.39)	(0.39)	(0.39)	(0.40)	(0.39)
Democrat		-0.17	-0.22	-0.23	-0.26	-0.15	-0.17	-0.20
		(0.47)	(0.47)	(0.47)	(0.51)	(0.53)	(0.53)	(0.48)
Democrat ×		-0.04	-0.01	-0.02	0.01	-0.02	0.00	0.00
Social Identity		(0.55)	(0.56)	(0.57)	(0.58)	(0.62)	(0.63)	(0.56)
Vowel Score			0.05	0.05	0.05	0.04	0.04	0.05
			(0.08)	(0.09)	(0.09)	(0.09)	(0.10)	(0.08)
Female				-0.15	-0.13	-0.06	-0.03	
				(0.30)	(0.31)	(0.29)	(0.28)	
Age				-0.01	-0.01	-0.02	-0.02	
				(0.01)	(0.01)	(0.01)	(0.02)	
Full-time					-0.06	-0.14	-0.17	
					(0.34)	(0.38)	(0.37)	
Prolific Score					-0.04	-0.08	-0.08	
					(0.14)	(0.12)	(0.12)	
Extraversion						-0.41***	-0.43***	
						(0.14)	(0.16)	
Agreeableness						0.06	0.06	
						(0.17)	(0.17)	
Conscientiousness						0.22	0.21	
						(0.23)	(0.24)	
Neuroticism						-0.20	-0.21	
						(0.16)	(0.17)	
Openness						-0.06	-0.07	
						(0.19)	(0.20)	
Engagement in US politics							0.09	-0.01
							(0.29)	(0.27)
Strong Affiliation							0.00	-0.13
							(0.34)	(0.31)
English is first language							-0.11	
							(0.34)	
Constant	2.11***	2.21***	2.06**	2.54**	6.64	11.57	11.90	2.16*
	(0.22)	(0.35)	(0.48)	(0.74)	(14.02)	(12.29)	(12.22)	(1.19)
N	124	124	124	124	124	124	124	124
R^2	0.23	0.23	0.24	0.24	0.24	0.30	0.30	0.24

OLS coefficient estimations.

The coefficient estimate of the key independent variable, Social Identity, measures its effect on $(b_{i1} - b_{j1})$. Since we control for $(x_i - x_j)$, this effect should be similar to that on overplacement, $(b_{i1} - b_{j1} - x_i - x_j)$. See table A18.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A20: Effect of social identity (=1 if in-group; =0 if out-group) on the magnitude of overestimation, given overplacement happened: $Y = (b_{i1} - x_i)|(b_{i1} - b_{j1}) - (x_i - x_j) > 0$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Identity	0.54	0.35	0.57	0.57	0.57	0.51	0.74	0.68
	(0.43)	(0.67)	(0.64)	(0.65)	(0.61)	(0.65)	(0.57)	(0.62)
Democrat		-1.51**	-0.81	-0.80	-0.95	-1.01	-0.97	-0.74
		(0.62)	(0.59)	(0.59)	(0.60)	(0.65)	(0.62)	(0.61)
Democrat ×		-0.28	-0.39	0.40	0.46	0.66	0.64	-0.32
Social Identity		(0.78)	(0.76)	(0.77)	(0.75)	(0.89)	(0.85)	(0.75)
Vowel Score			-0.39***	-0.39***	-0.31***	-0.32***	-0.23*	-0.38***
			(0.11)	(0.11)	(0.10)	(0.11)	(0.12)	(0.11)
Female				0.17	0.27	0.25	0.46	
				(0.43)	(0.41)	(0.45)	(0.42)	
Age				0.01	0.01	0.01	0.00	
				(0.02)	(0.02)	(0.02)	(0.02)	
Full-time					-0.83**	-0.96**	-1.13***	
					(0.42)	(0.48)	(0.42)	
Prolific Score					-0.46***	-0.47***	-0.45***	
					(0.16)	(0.17)	(0.14)	
Extraversion						-0.02	-0.25	
						(0.20)	(0.22)	
Agreeableness						-0.22	-0.23	
						(0.25)	(0.24)	
Conscientiousness						0.12	0.05	
						(0.29)	(0.28)	
Neuroticism						-0.05	-0.10	
						(0.20)	(0.21)	
Openness						0.04	-0.05	
						(0.24)	(0.24)	
Engagement in US politics							0.99***	0.81**
							(0.36)	(0.34)
Strong Affiliation							-0.11	-0.14
							(0.45)	(0.45)
English is first language							-1.22***	
							(0.41)	
Constant	1.06***	1.74***	2.54***	2.06***	48.14***	49.54***	45.94***	-0.79
	(0.33)	(0.53)	(0.58)	(0.95)	(15.47)	(16.35)	(14.10)	(1.38)
N	124	124	124	124	124	124	124	124
R^2	0.01	0.09	0.17	0.17	0.24	0.25	0.34	0.20

OLS coefficient estimations.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A21: Effect of social identity (=1 if in-group; =0 if out-group) on the magnitude of underestimation of others' performance, given overplacement happened: $Y = (b_{j1} - x_j)|(b_{i1} - b_{j1}) - (x_i - x_j) > 0$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Identity	1.92***	2.27***	2.35***	2.24***	2.24***	2.20***	2.30***	2.42***
	(0.29)	(0.43)	(0.43)	(0.44)	(0.44)	(0.43)	(0.40)	(0.42)
Democrat		0.10**	0.32	0.46	0.39	0.25	0.22	0.29
		(0.44)	(0.45)	(0.44)	(0.46)	(0.49)	(0.47)	(0.47)
Democrat ×		-0.86	-0.83	-0.76	-0.73	-0.65	-0.59	-0.86
Social Identity		(0.55)	(0.55)	(0.56)	(0.56)	(0.61)	(0.60)	(0.54)
Vowel Score			-0.13	-0.11	-0.10	-0.10	-0.04	-0.11
			(0.08)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)
Female				-0.07	-0.06	-0.13	-0.02	
				(0.32)	(0.32)	(0.34)	(0.33)	
Age				0.03**	0.03*	0.03**	0.03**	
				(0.01)	(0.01)	(0.01)	(0.01)	
Full-time					-0.26	-0.23	-0.37	
					(0.31)	(0.35)	(0.32)	
Prolific Score					-0.05	-0.01	-0.00	
					(0.17)	(0.16)	(0.15)	
Extraversion						0.28**	0.14	
						(0.14)	(0.14)	
Agreeableness						-0.20	-0.20	
						(0.17)	(0.17)	
Conscientiousness						-0.15	-0.20	
						(0.22)	(0.20)	
Neuroticism						0.16	0.09	
						(0.15)	(0.16)	
Openness						0.08	0.06	
						(0.19)	(0.18)	
Engagement in US politics							0.29	0.34
							(0.27)	(0.26)
Strong Affiliation							0.30	0.36
							(0.34)	(0.31)
English is first language							-0.72**	
							(0.33)	
Constant	-2.32***	-2.37***	-2.10***	-3.19***	1.87	-2.36	-3.26	-3.67***
	(0.22)	(0.33)	(0.39)	(0.70)	(16.46)	(16.27)	(15.47)	(1.11)
N	124	124	124	124	124	124	124	124
R^2	0.25	0.27	0.28	0.31	0.31	0.34	0.39	0.31

OLS coefficient estimations.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A22: Effect of social identity (=1 if in-group; =0 if out-group) on the probability of overplacement in revised beliefs among winners: $Y = 1$ if $(b_{i2} - b_{j2}) - (x_i - x_j) > 0 | x_i > x_j$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Identity	-0.06	-0.10	-0.10	-0.08	-0.08	-0.09	-0.08	-0.10
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)	(0.09)	(0.11)
Democrat		-0.14	-0.14	-0.12	-0.12	-0.01	0.09	-0.14
		(0.12)	(0.12)	(0.13)	(0.13)	(0.11)	(0.10)	(0.12)
Vowel Score			-0.02	-0.02	-0.02	-0.03	-0.01	-0.02
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Female				0.10	0.09	0.13	0.14	
				(0.11)	(0.11)	(0.10)	(0.09)	
Age				0.00	0.00	0.00	0.01***	
				(0.01)	(0.01)	(0.01)	(0.00)	
Full-time					0.05	-0.05	-0.01	
					(0.11)	(0.11)	(0.10)	
Prolific Score					0.04	0.02	0.03	
					(0.06)	(0.05)	(0.04)	
Extraversion						0.06	0.11**	
						(0.05)	(0.05)	
Agreeableness						-0.07	-0.07	
						(0.06)	(0.05)	
Conscientiousness						0.24***	0.28***	
						(0.05)	(0.06)	
Neuroticism						0.08	0.10**	
						(0.05)	(0.05)	
Openness						-0.02	0.02	
						(0.06)	(0.06)	
Engagement in US politics							-0.11*	-0.06
							(0.07)	(0.08)
Strong Affiliation							-0.24***	-0.12
							(0.08)	(0.11)
English is first language							-0.29***	
							(0.11)	
N	81	81	81	81	81	81	81	81
Significance (p-value)								
Social Identity × Democrat	N.A.	0.033	0.030	0.026	0.025	0.005	0.001	0.021

Average marginal effects based on probit results.

In the regression specifications (2)-(8), an interaction term between the dummies Social Identity and Democrat was included. The p-values for the significance test on whether the interaction term's effect is significantly different from zero are reported in the lower part of the table.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A23: Effect of Social Identity (=1 if in-group; =0 if out-group) on winners' update of belief of own performance: $Y = (b_{i2} - b_{i1})|x_i > x_j$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Identity	-0.02	-0.51	-0.53	-0.51	-0.58	-0.70*	-0.65	-0.48
	(0.19)	(0.37)	(0.37)	(0.36)	(0.37)	(0.39)	(0.42)	(0.38)
Democrat		-0.61**	-0.61**	-0.66**	-0.71**	-0.86***	-0.79**	-0.55*
		(0.26)	(0.27)	(0.28)	(0.28)	(0.29)	(0.34)	(0.29)
Democrat ×		0.69*	0.70*	0.65	0.74*	0.94**	0.88*	0.64
Social Identity		(0.41)	(0.41)	(0.41)	(0.42)	(0.46)	(0.47)	(0.41)
Vowel Score			0.09	0.08	0.10*	0.12	0.12	0.09
			(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.06)
Female				0.23	0.25	0.28	0.30	
				(0.16)	(0.16)	(0.19)	(0.20)	
Age				-0.00	-0.00	-0.01	-0.01	
				(0.01)	(0.01)	(0.01)	(0.01)	
Full-time					-0.01	0.04	0.07	
					(0.18)	(0.21)	(0.21)	
Prolific Score					-0.15**	-0.12	-0.11	
					(0.07)	(0.09)	(0.09)	
Extraversion						-0.12	-0.11	
						(0.14)	(0.13)	
Agreeableness						0.07	0.04	
						(0.10)	(0.11)	
Conscientiousness						-0.20	-0.20	
						(0.12)	(0.13)	
Neuroticism						-0.18	-0.16	
						(0.14)	(0.14)	
Openness						0.07	0.07	
						(0.13)	(0.13)	
Engagement in US politics							0.08	0.03
							(0.15)	(0.12)
Strong Affiliation							-0.20	-0.23
							(0.20)	(0.19)
English is first language							-0.02	
							(0.21)	
Constant	0.34**	0.80	0.44	0.56	15.01**	13.43	12.70	-0.13
	(0.10)	(0.24)	(0.36)	(0.45)	(6.83)	(8.88)	(9.33)	(0.36)
N	81	81	81	81	81	81	81	81
R^2	0.00	0.05	0.09	0.11	0.13	0.19	0.20	0.10

OLS coefficient estimations.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A24: Effect of social identity (=1 if in-group; =0 if out-group) on winners' update of belief about others' performance: $Y = (b_{j2} - b_{j1})|x_i > x_j$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Identity	-0.33	-0.61	-0.63	-0.72	-0.63	-0.84	-0.83	-0.67
	(0.26)	(0.63)	(0.65)	(0.60)	(0.58)	(0.59)	(0.61)	(0.65)
Democrat		-0.13	-0.14	-0.24	-0.16	-0.19	-0.25	-0.18
		(0.56)	(0.59)	(0.53)	(0.53)	(0.53)	(0.56)	(0.60)
Democrat ×		0.52	0.53	0.53	-0.42	0.68	0.67	0.56
Social Identity		(0.67)	(0.68)	(0.64)	(0.62)	(0.66)	(0.70)	(0.69)
Vowel Score			0.14	0.14	0.11	0.17*	0.16*	0.14
			(0.09)	(0.09)	(0.08)	(0.09)	(0.09)	(0.10)
Female				-0.37*	-0.40*	-0.22	-0.24	
				(0.21)	(0.22)	(0.26)	(0.27)	
Age				-0.02	-0.02	-0.03**	-0.03**	
				(0.01)	(0.01)	(0.01)	(0.01)	
Full-time					0.16	0.05	0.01	
					(0.23)	(0.23)	(0.22)	
Prolific Score					0.27	0.28	0.27	
					(0.20)	(0.18)	(0.17)	
Extraversion						-0.27*	-0.27**	
						(0.14)	(0.13)	
Agreeableness						0.06	0.09	
						(0.12)	(0.13)	
Conscientiousness						0.07	0.07	
						(0.14)	(0.15)	
Neuroticism						-0.31*	-0.33**	
						(0.16)	(0.15)	
Openness						0.22	0.21	
						(0.16)	(0.16)	
Engagement in US politics							-0.11	-0.16
							(0.19)	(0.15)
Strong Affiliation							0.28	0.09
							(0.27)	(0.27)
English is first language							0.22	
							(0.29)	
Constant	-0.29*	-0.20	-0.76	0.18	-26.40	-27.68	-25.92	-0.12
	(0.18)	(0.54)	(0.63)	(0.67)	(20.14)	(18.08)	(17.35)	(0.73)
N	81	81	81	81	81	81	81	81
R^2	0.02	0.04	0.09	0.13	0.17	0.26	0.28	0.09

OLS coefficient estimations.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A25: Effect of social identity (=1 if in-group; =0 if out-group) on losers' update of belief of own performance: $Y = (b_{i2} - b_{i1})|x_i < x_j$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Identity	-0.01	-0.46	-0.47	-0.51	-0.46	-0.46	-0.50	-0.45
	(0.36)	(0.58)	(0.58)	(0.57)	(0.58)	(0.59)	(0.62)	(0.60)
Democrat		0.20	0.12	0.13	0.21	0.39	0.26	0.10
		(0.43)	(0.44)	(0.44)	(0.47)	(0.49)	(0.46)	(0.44)
Democrat ×			0.80	0.72	0.71	0.65	0.60	0.69
0.72								
Social Identity		(0.68)	(0.67)	(0.66)	(0.67)	(0.68)	(0.70)	(0.70)
Vowel Score			0.07	0.06	0.03	0.04	-0.02	0.07
			(0.10)	(0.10)	(0.09)	(0.10)	(0.10)	(0.10)
Female				0.57	0.55	0.38	0.36	
				(0.36)	(0.37)	(0.38)	(0.38)	
Age				0.02	0.02	0.02	0.01	
				(0.01)	(0.01)	(0.02)	(0.02)	
Full-time					0.30	0.28	0.29	
					(0.37)	(0.43)	(0.42)	
Prolific Score					0.13	0.10	0.09	
					(0.19)	(0.20)	(0.20)	
Extraversion						0.16	0.21	
						(0.22)	(0.23)	
Agreeableness						0.03	0.02	
						(0.19)	(0.19)	
Conscientiousness						0.26	0.29	
						(0.24)	(0.27)	
Neuroticism						0.19	0.23	
						(0.17)	(0.18)	
Openness						-0.27	-0.23	
						(0.19)	(0.19)	
Engagement in US politics							-0.26	-0.25
							(0.30)	(0.30)
Strong Affiliation							0.14	0.20
							(0.40)	(0.40)
English is first language							0.59	
							(0.36)	
Constant	-1.62***	-1.68***	-1.84***	-2.81***	-15.27	-13.65	-12.40	-0.89
	(0.24)	(0.32)	(0.41)	(0.72)	(18.64)	(19.82)	(19.75)	(1.24)
N	119	119	119	119	119	119	119	119
R^2	0.00	0.04	0.04	0.08	0.09	0.11	0.14	0.05

OLS coefficient estimations.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A26: Effect of social identity (=1 if in-group; =0 if out-group) on losers' update of belief about others' performance: $Y = (b_{j2} - b_{j1})|x_i < x_j$

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Identity	-1.07***	-1.23***	-1.24***	-1.22***	-1.20***	-1.14***	-1.20***	-1.30***
	(0.32)	(0.43)	(0.43)	(0.44)	(0.44)	(0.43)	(0.44)	(0.43)
Democrat		0.29	0.19	0.13	0.25	0.35	0.35	0.21
		(0.56)	(0.55)	(0.52)	(0.54)	(0.55)	(0.55)	(0.56)
Democrat ×		0.21	0.11	0.09	0.09	0.08	0.16	0.18
Social Identity		(0.67)	(0.68)	(0.68)	(0.68)	(0.69)	(0.67)	(0.66)
Vowel Score			0.08	0.07	0.09	0.08	0.06	0.09
			(0.08)	(0.08)	(0.09)	(0.08)	(0.10)	(0.08)
Female				0.21	0.22	0.19	0.17	
				(0.32)	(0.32)	(0.32)	(0.34)	
Age				-0.01	-0.00	-0.01	-0.01	
				(0.02)	(0.02)	(0.02)	(0.02)	
Full-time					0.39	0.33	0.40	
					(0.33)	(0.36)	(0.35)	
Prolific Score					-0.09	-0.15	-0.14	
					(0.12)	(0.11)	(0.12)	
Extraversion						-0.34**	-0.30*	
						(0.17)	(0.17)	
Agreeableness						0.13	0.13	
						(0.15)	(0.16)	
Conscientiousness						0.30	0.30	
						(0.24)	(0.26)	
Neuroticism						-0.15	-0.09	
						(0.18)	(0.19)	
Openness						-0.08	-0.08	
						(0.17)	(0.18)	
Engagement in US politics							0.09	0.16
							(0.31)	(0.27)
Strong Affiliation							-0.32	-0.38
							(0.35)	(0.35)
English is first language							0.22	
							(0.36)	
Constant	1.11	1.03	0.84	0.96	9.32	15.43	14.41	0.33
	(0.26)	(0.31)	(0.39)	(0.77)	(11.95)	(11.06)	(11.75)	(1.15)
N	119	119	119	119	119	119	119	119
R^2	0.09	0.10	0.11	0.11	0.13	0.18	0.19	0.12

OLS coefficient estimations.

Heteroskedastic-robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

Table A27: Pre-announcement overplacement levels by treatment and political party

	All		Democrats		Republicans	
	In-group	Out-group	In-group	Out-group	In-group	Out-group
	55%	69%	44%	62%	66%	76%
N	100	100	50	50	50	50

Table A28: Post-announcement overplacement levels by treatment and political party

	All		Democrats		Republicans	
	In-group	Out-group	In-group	Out-group	In-group	Out-group
	53%	59%	32%	61%	71%	50%
N(winners)	40	41	19	31	21	10
	32%	56%	19%	58%	45%	55%
N(losers)	60	59	31	19	29	40

D.5 Logical reasoning test and Vowel-counting task

In the main text, we used the score in the vowel-counting effort task as a measure of ability. Using the score in the logical reasoning task instead (LRT; x_i) would create an endogeneity problem since there would be simultaneity (dependent variable: $(b_i - bj) - (x_i - x_j)$). Additionally, there is support in the literature for a negative correlation between ability and overconfidence, referred to as the Dunning-Kruger effect (based on the seminal paper by Kruger and Dunning, 1999). Thus, to avoid both endogeneity and the omitted variables bias, it is important to include in the set of control variables a measure of ability that depends on similar factors as those that influence the score in the LRT but not overconfidence about performance in the LRT in the belief-elicitation stage.

Recall that, in the Counting-vowels task, participants were asked to count the number of vowels in as many quotes as possible in 1 minute; the task was set at the end of the experiment, after both the LRT and belief-elicitation stage.

Intuitively, performance in the LRT depends on logical reasoning abilities or, generally, "intelligence", while performance in the vowel-counting task is not as demanding. By contrast, performance in the vowel-counting task depends on attention and patience which also affect performance in the logical reasoning test. Finally, attention and patience are not so relevant for overconfidence, since guessing 4 values is not significantly tedious - thus, factors affecting performance in the vowel-counting task do not affect belief formation or overconfidence.

Empirically, we observe a high correlation of 48% between the two scores. Moreover, Table A29 shows that the score in the Vowel-counting task is a good predictor of the score in the LRT with a coefficient estimate of around 1 (p-value<0.01).

Table A29: Vowel-counting task's ability to predict performance in the LRT

Variables	(1)	(2)
Vowel Score	0.528***	1.088***
	(0.057)	(0.186)
Vowel Score squared		-0.075***
		(0.021)
Constant	5.316***	4.590***
	(0.262)	(0.375)
N	301	301
R^2	0.23	0.26

OLS coefficient estimations.
Heteroskedastic-robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

D.6 Ability to predict

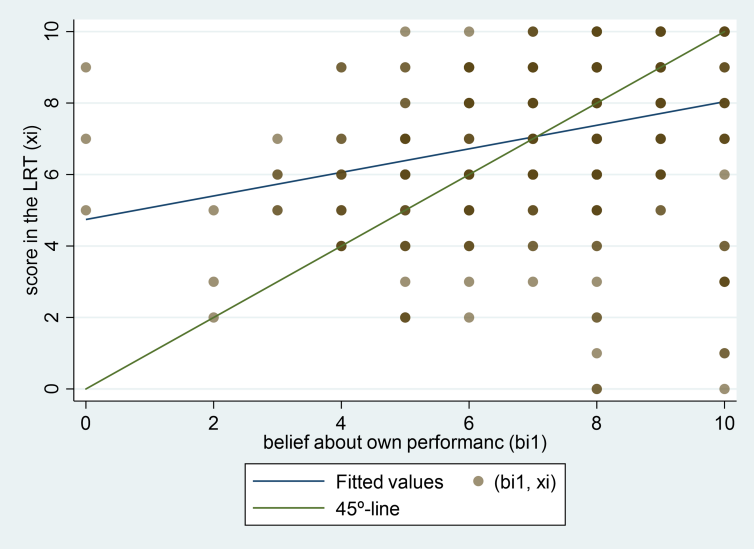
How good are subjects at predicting their own score? The large constant estimates presented in Table A30 suggest subjects aren't too good at predicting. Even though we observe a correlation of 29% between the two variables, belief and observed performance, note that 54% of subjects only got their performance wrong by one question. Moreover, the constant's coefficient being positive and > 1 is in line with the Dunning-Kruger effect, with higher ability coming together with lower overconfidence and vice-versa.

Table A30: Predicting performance

Variables	Score in Logical Reasoning Test (x_i)
Belief of own performance (b_{i1})	0.33***
	(0.07)
Constant	4.74***
	(0.50)
N	301
R-squared	0.082

OLS regression.
Heteroskedastic-robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Relationship between performance and the respective belief



D.7 Opponent's score calculated from main sample

Table A31: Test results to alternative specification of the opponent's score (x_j)

Hypotheses		Results with x_j from auxiliary sample	Results with x_j from main sample
1	The likelihood of individuals overplacing their performance is larger when comparing themselves to the out-group rather than to the in-group.	OUT>IN**	OUT>IN**
2	Conditional on there being overplacement, its magnitude is larger when individuals compare themselves to the out-group rather than to the in-group	OUT>IN*** because $b_{i1} = x_i$ & $b_{j1} < x_j$***	OUT>IN*** because $b_{i1} = x_i$ & $b_{j1} < x_j$***
3	Being informed of having won, the likelihood of overplacement in revised beliefs is larger when individuals compare themselves to the out-group rather than to the ingroup.	OUT=IN	OUT=IN
4	Being informed of having won, any upward update of belief of own performance ($b_{i2} - b_{i1}$) is larger when playing against the out-group than against the in-group.	OUT=IN	OUT=IN
5	Being informed of having won, any downward update of belief of the opponents' average performance ($b_{j2} - b_{j1}$) is larger when playing against the out-group than against the in-group.	OUT=IN	OUT=IN
6	Being informed of having lost, any downward update of belief of own performance ($b_{i2} - b_{i1}$) is smaller when playing against the out-group than against the in-group.	OUT=IN	OUT=IN
7	Being informed of having lost, any upward update of belief of the opponents' average performance ($b_{j2} - b_{j1}$) is smaller when playing against the out-group than against the in-group.	OUT>IN*** (opposite of hypothesized)	OUT>IN*** (opposite of hypothesized))

*** p<0.01, ** p<0.05, * p<0.1.

By contrast with signs > and <, the = sign is used when the difference is not statistically significant. "OUT" refers to the value obtained in the out-group treatment and "IN" refers to the value obtained in the in-group treatment.

Table A32: Task performance and pre-announcement beliefs, overestimation and overplacement – Democrat sub-sample

	All	Treatment In-group	Out-group	Neutral-group
Own performance (x_i)	7.85 (1.73)	7.92 (1.52)	7.56 (1.75)	8.08 (1.90)
Success rate (% of ($x_i > x_j$))	56%	38%	62%	69%
Belief about own performance (b_{i1})	7.23 (1.90)	7.34 (1.53)	7.50 (1.57)	6.84 (2.43)
Bias in estimation of own performance ($b_{i1} - x_i$)	-0.63 (1.84)	-0.58 (1.79)	-0.06 (1.60)	-1.24 (1.97)
% overestimating own performance	23%	26%	32%	10%
Belief about others' performance (b_{j1})	6.58 (1.70)	7.22 (1.06)	6.24 (2.05)	6.27 (1.69)
Bias in estimation of others' performance ($b_{j1} - x_j$)	-1.12 (1.66)	-0.78 (1.06)	-1.16 (2.05)	-1.43 (1.69)
% underestimating others' performance	67%	54%	66%	80%
(Over)placement ($b_{i1} - b_{j1}$) − ($x_i - x_j$)	0.49 (2.14)	0.20 (1.67)	1.10 (2.52)	0.19 (2.07)
% overplacing performance	54%	44%	62%	57%
N	151	50	50	51

Notes: Standard deviations in parentheses.

D.8 Democrat vs. Republican subsamples

In this section, we replicate the information in Tables 2, 3, and 4 conditioning on the political affiliation of the participants: Democrat or Republican.

E Instructions

The link to the experiment was shared by email, sent by Prolific to participants, with the content shown in Figure 2. Once participants clicked on the button with the respective hyperlink, they would find themselves in the consent page (Figure 3) and data collection would start (we used Heroku to save answers in a dataset with one line per participant, https://id.heroku.com/login). The experiment was programmed using oTree and the following set of screenshots illustrates what participants visualised. In order not to distort images, namely in the logical reasoning test, participants were only allowed to participate using their computers or tablets, but not their phones - an option available on Prolific. In the last page of the experiment (Figure 28), clicking the button would take participants back to Prolific's website, confirming completion.

Table A33: Task performance and pre-announcement beliefs, overestimation and overplacement – Republican sub-sample

		Treatment		
	All	In-group	Out-group	Neutral-group
Own performance	6.45	6.68	5.92	6.76
(x_i)	(2.49)	(2.45)	(2.61)	(2.37)
Success rate (% of $(x_i > x_j)$)	37%	42%	20%	48%
Belief about own performance	7.43	7.78	7.00	7.50
(b_{i1})	(2.02)	(2.09)	(1.69)	(2.20)
Bias in estimation of own performance	0.97	1.10	1.08	0.74
$(b_{i1} - x_i)$	(2.85)	(2.51)	(3.10)	(2.96)
% overestimating own performance	53%	56%	52%	52%
Belief about others' performance	6.67	7.50	6.18	6.32
(b_{j1})	(1.99)	(1.55)	(2.09)	(2.04)
Bias in estimation of others' performance	-1.03	0.10	-1.82	-1.38
$(b_{j1} - x_j)$	(2.07)	(1.55)	(2.09)	(-2.04)
% underestimating others' performance	66%	54%	74%	70%
(Over)placement	2.01	1.00	2.90	2.12
$(b_{i1} - b_{j1}) - (x_i - x_j)$	(2.68)	(2.17)	(3.18)	(2.27)
% overplacing performance	75%	66%	76%	84%
N	150	50	50	50

Notes: Standard deviations in parentheses.

Table A34: Task performance and pre-announcement beliefs, overestimation and overplacement: Democrat overplacers sub-sample

		Treatment		
	All	In-group	Out-group	Neutral-group
Own performance (x_i)	7.15 (1.64)	6.95 (1.17)	6.71 (1.57)	7.76 (1.86)
Success rate (% of ($x_i > x_j$))	43%	14%	42%	66%
Belief about own performance (b_{i1})	7.33 (1.67)	7.82 (1.18)	6.94 (1.55)	7.38 (2.03)
Bias in estimation of own performance ($b_{i1} - x_i$)	0.18 (1.57)	0.86 (1.13)	0.23 (1.76)	-0.38 (1.47)
% overestimating own performance	39%	59%	45%	17%
Belief about others' performance (b_{j1})	5.88 (1.57)	7.14 (0.94)	5.13 (1.61)	5.72 (1.33)
Bias in estimation of others' performance ($b_{j1} - x_j$)	-1.79 (1.46)	-0.86 (0.94)	-2.27 (1.61)	-1.98 (1.33)
% underestimating others' performance	84%	55%	94%	97%
Overplacement $(b_{i1} - x_{j1}) - (x_i - x_j)$	1.97 (1.61)	1.73 (1.08)	2.50 (2.13)	1.60 (1.11)
N	82	22	31	29

Notes: Standard deviations in parentheses.

Table A35: Task performance and pre-announcement beliefs, overestimation and overplacement: Republican overplacers sub-sample

		Treatment		
	All	In-group	Out-group	Neutral-group
Own performance	5.77	5.85	5.11	6.31
(x_i)	(2.41)	(2.46)	(2.38)	(2.31)
Success rate (% of $(x_i > x_j)$)	24%	24%	8%	38%
Belief about own performance	7.40	7.94	6.84	7.48
(b_{i1})	(2.10)	(2.11)	(1.75)	(2.31)
Bias in estimation of own performance	1.63	2.09	1.74	1.17
$(b_{i1} - x_i)$	(2.92)	(2.38)	(3.24)	(3.00)
% overestimating own performance	67%	76%	66%	62%
Belief about others' performance	6.33	7.30	5.63	6.19
(b_{j1})	(2.03)	(1.59)	(2.01)	(2.10)
Bias in estimation of others' performance	-1.39	-0.10	-2.37	-1.51
$(b_{j1} - x_j)$	(2.12)	(1.59)	(2.01)	(2.10)
% underestimating others' performance	73%	61%	84%	74%
Overplacement	3.01	2.19	4.11	2.68
$(b_{i1} - x_{j1}) - (x_i - x_j)$	(2.28)	(1.62)	(2.65)	(2.02)
N	113	33	38	42

Notes: Standard deviations in parentheses.

Table A36: Beliefs revision and overplacement conditional on having lost – Democrat subsample.

		Treatment		
	All	In-group	Out-group	Neutral-group
Own performance (x_i)	6.26 (1.26)	6.94 (1.00)	5.63 (1.07)	5.69 (1.30)
Revised belief about own performance (b_{i2})	5.35 (1.76)	5.94 (1.48)	4.95 (1.47)	4.69 (2.24)
Bias in revised estimation of own performance ($b_{i2} - x_i$)	-0.91 (1.89)	-1.00 (1.61)	-0.68 (1.89)	-1.00 (2.42)
% overestimating own performance pre-announcement	42%	42%	58%	25%
% overestimating own performance post-announcement	26%	23%	32%	25%
Change in beliefs about own performance ($b_{i2} - b_{i1}$)	-1.00 (1.49)	-1.13 (1.26)	-1.47 (1.22)	-0.18 (1.91)
Revised belief about others' performance (b_{j2})	7.14 (1.56)	7.58 (1.26)	6.53 (1.31)	7.00 (2.10)
Bias in revised estimation of others' performance ($b_{j2} - x_j$)	-0.62 (1.50)	-0.42 (1.26)	-0.87 (1.31)	-0.70 (2.10)
% underestimating others' performance pre-announcement	65%	48%	84%	75%
% underestimating others' performance post-announcement	53%	42%	68%	56%
Change in beliefs about others' performance ($b_{j2} - b_{j1}$)	0.70 (1.54)	0.29 (1.24)	1.32 (2.06)	0.75 (1.13)
Revised (Over)placement $(b_{i2} - b_{j2}) - (x_i - x_j)$	-0.29 (1.43)	-0.58 (1.23)	0.19 (1.40)	-0.30 (1.75)
% overplacing performance	33%	19%	58%	31%
Change in (Over)placement $(b_{i2} - b_{j2}) - (b_{i1} - b_{j1})$	-1.70 (1.89)	-1.42 (1.52)	-2.79 (1.81)	-0.94 (2.14)
N	66	31	19	16

Notes: Standard deviations in parentheses.

Table A37: Beliefs revision and overplacement conditional on having lost – Republican sub-sample.

		Treatment		
	All	In-group	Out-group	Neutral-group
Own performance (x_i)	5.03 (1.98)	5.03 (1.82)	5.08 (2.20)	4.96 (1.84)
Revised belief about own performance (b_{i2})	5.14 (2.36)	5.03 (2.47)	5.18 (2.00)	5.19 (2.81)
Bias in revised estimation of own performance ($b_{i2} - x_i$)	0.11 (2.83)	0.00 (2.71)	0.10 (2.84)	0.23 (3.04)
% overestimating own performance pre-announcement	71%	76%	65%	73%
% overestimating own performance post-announcement	41%	41%	38%	46%
Change in beliefs about own performance ($b_{i2} - b_{i1}$)	-1.81 (2.23)	-2.14 (2.59)	-1.68 (2.04)	-1.65 (2.13)
Revised belief about others' performance (b_{j2})	7.17 (1.96)	7.21 (1.84)	7.05 (1.72)	7.31 (2.45)
Bias in revised estimation of others' performance ($b_{j2} - x_j$)	-0.57 (1.99)	-0.19 (1.84)	-0.95 (1.72)	-0.39 (2.45)
% underestimating others' performance pre-announcement	67%	55%	78%	65%
% underestimating others' performance post-announcement	51%	52%	58%	38%
Change in beliefs about others' performance ($b_{j2} - b_{j1}$)	0.72 (1.96)	-0.21 (1.57)	1.03 (1.95)	1.27 (2.07)
Revised (Over)placement $(b_{i2} - b_{j2}) - (x_i - x_j)$	0.67 (2.12)	0.19 (1.86)	1.05 (2.35)	0.62 (1.98)
% overplacing performance	53%	45%	55%	58%
Change in (Over)placement $(b_{i2} - b_{j2}) - (b_{i1} - b_{j1})$	-2.53 (2.28)	-1.93 (1.96)	-2.70 (2.50)	-2.92 (2.21)
N	95	29	40	26

Notes: Standard deviations in parentheses.

Moreover, in all pages with the headline "Time left to complete this page", there was a time countdown starting on 5 minutes (instructions, surveys and main belief elicitation pages), 1 minute (vowel-counting effort task), 45 seconds (each of the 10 questions in the logical reasoning test), 10 seconds (score revelation page) or 5 seconds (exposure to preferred candidate image). Going over the given time would have different consequences. Participants were informed beforehand that going over the time limit could exclude them from the study - if they timed out, they would see the page in Figure 5 and would not be able to proceed with the study nor submit a completion code to Prolific. This happened in all 5-minute countdowns in which the expected time completion was under 1 minute to nudge participants to focus on the study and avoid distractions. Moreover, participants would not be able to move on to the next page without answering all the questions or, equivalently, filling in all the fields. By contrast, going over time in the logical reasoning test would result in an automatic submission equivalent to a wrong answer and participants would be led on to the next question or task. Similarly, not clicking on the "Next" button and going over time in the score-revelation page would move participants on to the next page, but with no associated penalty. Relative to the preferred-candidate image page, participants would be "stuck" in it for the fixed few seconds, after which they would be automatically directed to the following page. Similarly, in the vowel-counting task, it would only end after the 1 minute time-frame, during which participants had the opportunity to go through as many within-pages as their correct answers. Finally, in the consent page there was no countdown with the intention of giving participants the opportunity to prepare themselves for the following timed tasks or to delay participation to a more convenient time in which they could be focused, under no penalty.

Although participants were US citizens, incentives were in pounds because that was the currency Prolific payed with at the time of the experiment (Fall 2020). Moreover, the 10 questions that composed the logical reasoning test (Figures 7 to 10) were an adaptation of the test found (https://www.assessment-training.com/Training/Free#/test/245?mode=free&pid=0), including new questions with the intention of keeping the test at an easy level (in line with Moore and Healy's (2008) predictions). The correct answers are [3,2,3,2,5,4,4,3,5,4] considering 1 to be the position most at left and 5, most at right.

For the Big Five Personality Test[22], displayed in Figure 21, Extraversion was calculated as the average of the reverted 1st and 6th answers; Agreeableness was the average of the reverted 7th and 2nd answers; Conscientiousness was the average of the reverted 3rd and 8th answers; Neuroticism was the average of the reverted 4th and the 9th answers; Openness was the average of the 5th and the 10th answers. For example, a more extroverted person would score lower in the 1st question ("is reserved") and higher in the 6th question ("is outgoing, sociable") in the five-point Likert scale.

[22]Rammstedt & John, 2007; code adapted from the following source https://github.com/chapkovski/bigfive/blob/master/bigfive/models.py

Relative to differences between treatments (out-group, in-group, neutral-group), there are none until participants arrive at the perception-of-the-opponent page (Figure 13), one of the set of 4 pages that intend to make identity salient (Figures 11-15) - participants would only visualize one of these 3 pages depending on their identity as Democrat or Republican and their treatment. The difference is in the bold words, stating the opponent's identity as either "Republican", "Democrat" or "American" (although participants did not know at this point that they were being asked about their opponent), and the symbolic figure of the party or national flag, in accordance with the opponent's identity. Similarly, in the pages relative to belief elicitation, revision and confirmation (Figures 16, 17, 18 and 19), differences between treatments lie only on the reference to the opponent's identity.

Figure 2: Content of email shared by Prolific with link to start study

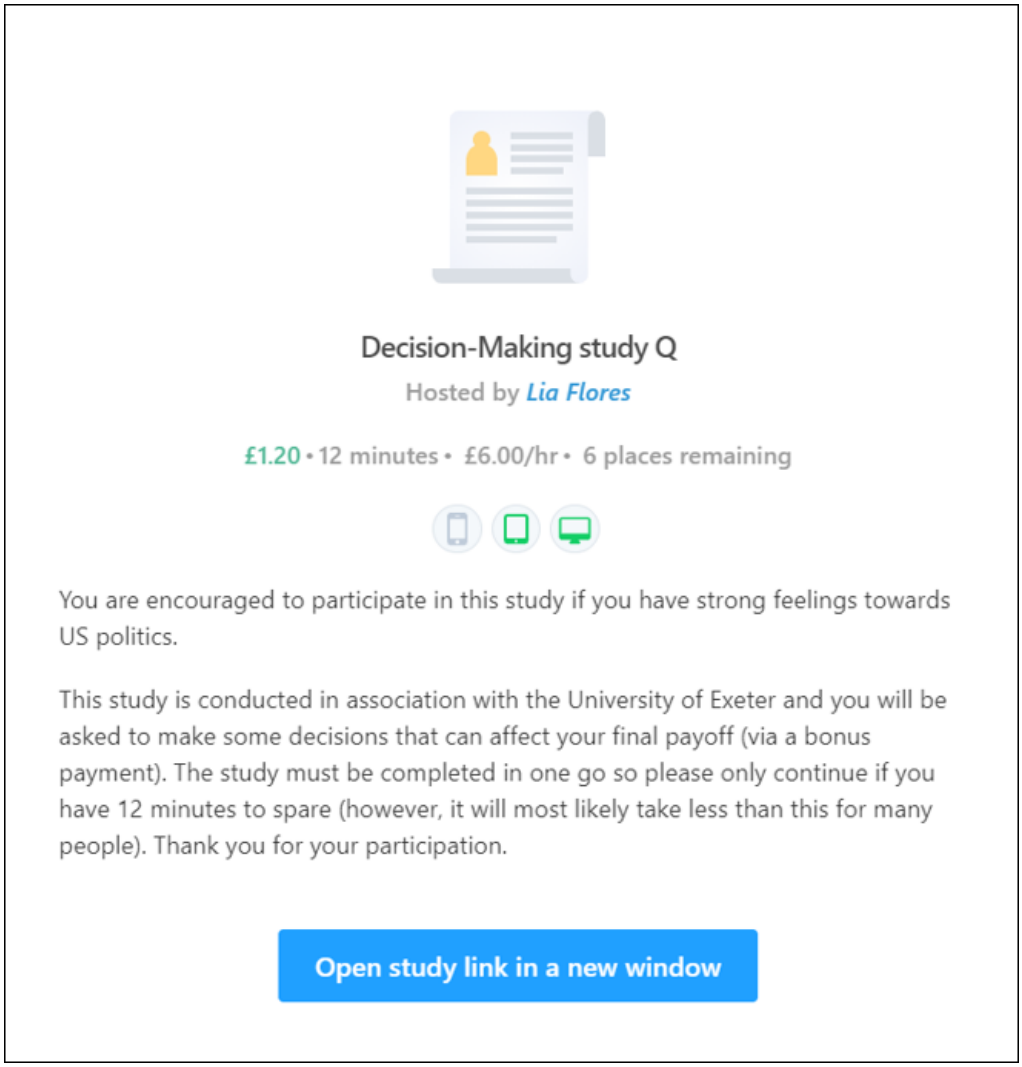
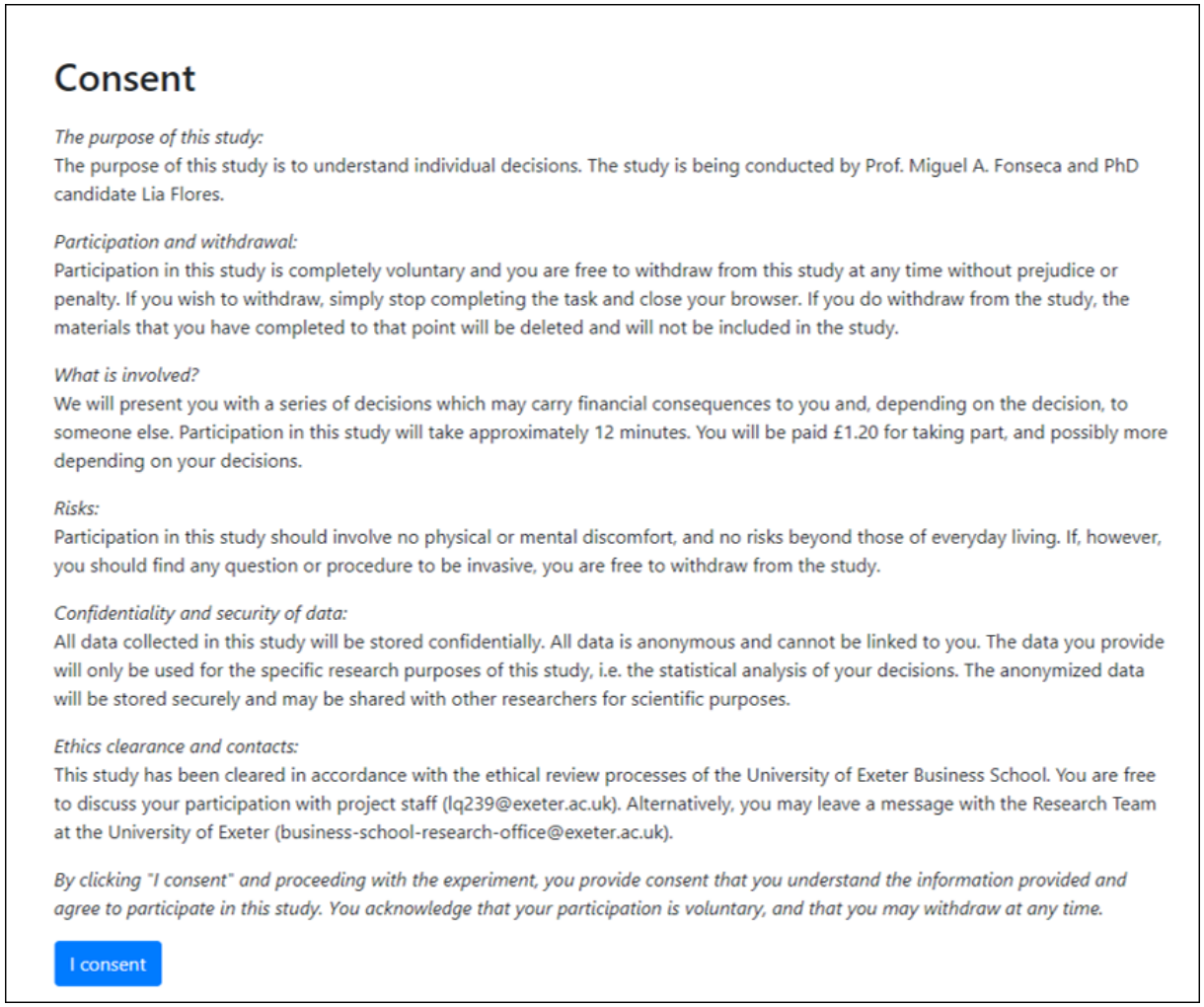


Figure 3: Consent page



Consent

The purpose of this study:
The purpose of this study is to understand individual decisions. The study is being conducted by Prof. Miguel A. Fonseca and PhD candidate Lia Flores.

Participation and withdrawal:
Participation in this study is completely voluntary and you are free to withdraw from this study at any time without prejudice or penalty. If you wish to withdraw, simply stop completing the task and close your browser. If you do withdraw from the study, the materials that you have completed to that point will be deleted and will not be included in the study.

What is involved?
We will present you with a series of decisions which may carry financial consequences to you and, depending on the decision, to someone else. Participation in this study will take approximately 12 minutes. You will be paid £1.20 for taking part, and possibly more depending on your decisions.

Risks:
Participation in this study should involve no physical or mental discomfort, and no risks beyond those of everyday living. If, however, you should find any question or procedure to be invasive, you are free to withdraw from the study.

Confidentiality and security of data:
All data collected in this study will be stored confidentially. All data is anonymous and cannot be linked to you. The data you provide will only be used for the specific research purposes of this study, i.e. the statistical analysis of your decisions. The anonymized data will be stored securely and may be shared with other researchers for scientific purposes.

Ethics clearance and contacts:
This study has been cleared in accordance with the ethical review processes of the University of Exeter Business School. You are free to discuss your participation with project staff (lq239@exeter.ac.uk). Alternatively, you may leave a message with the Research Team at the University of Exeter (business-school-research-office@exeter.ac.uk).

By clicking "I consent" and proceeding with the experiment, you provide consent that you understand the information provided and agree to participate in this study. You acknowledge that your participation is voluntary, and that you may withdraw at any time.

I consent

Figure 4: Page with overview of the experiment and general instructions

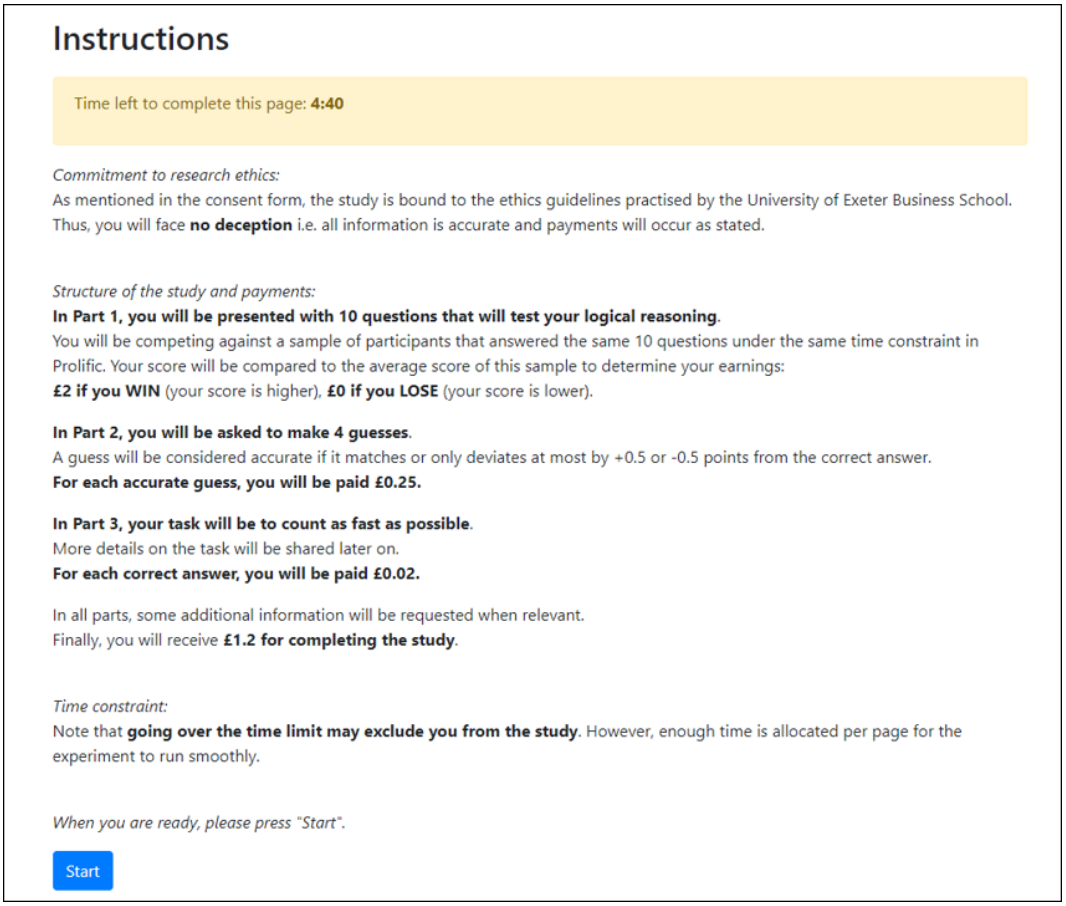


Figure 5: Time out page - visualised when participants went over time in the 5-minute long pages

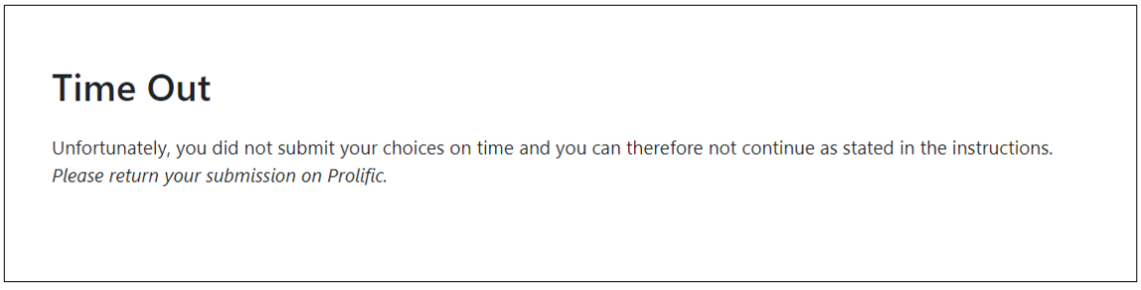


Figure 6: Instructions for the logical reasoning test with an example

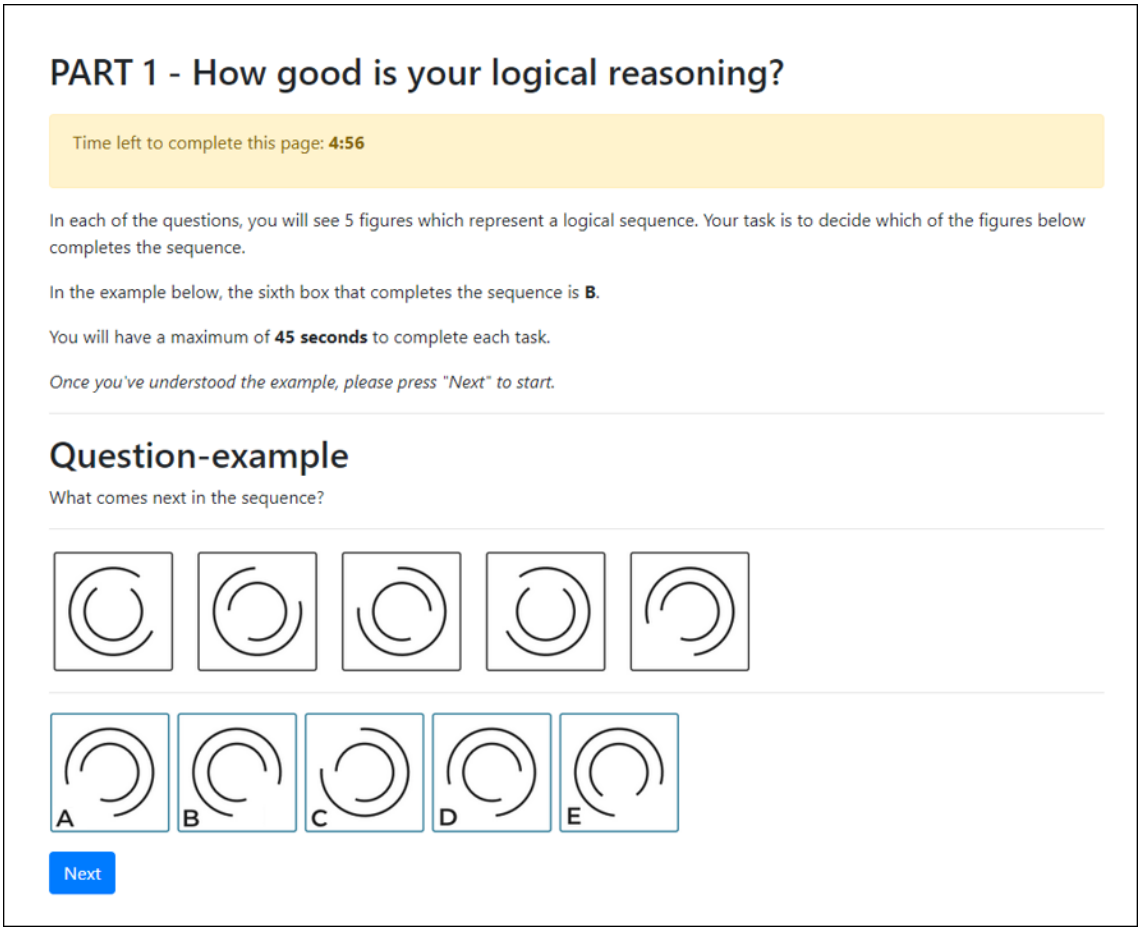


Figure 7: Logical reasoning test - Question 1 out of 10

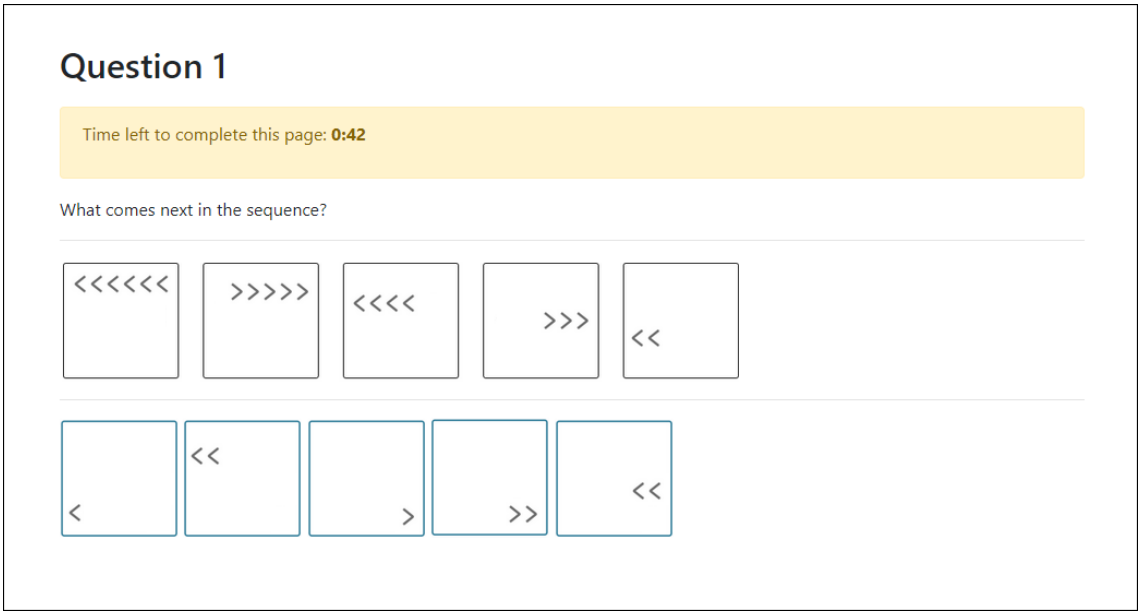


Figure 8: Logical reasoning test - Questions 2 to 4

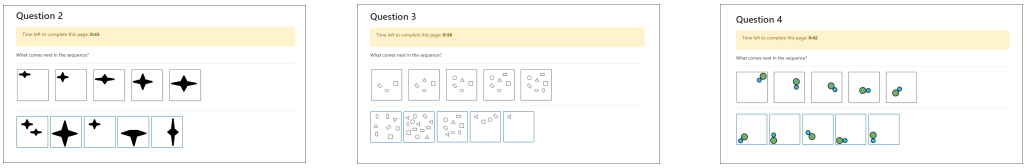


Figure 9: Logical reasoning test - Questions 5 to 7

Figure 10: Logical reasoning test - Questions 8 to 10

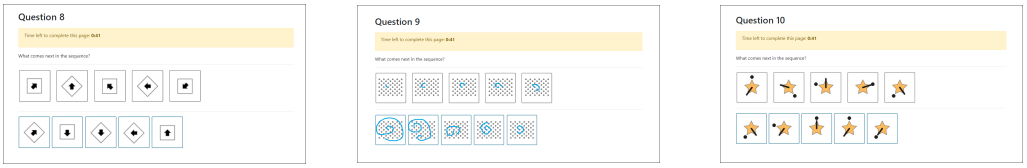


Figure 11: Identity salience pages - political affiliation

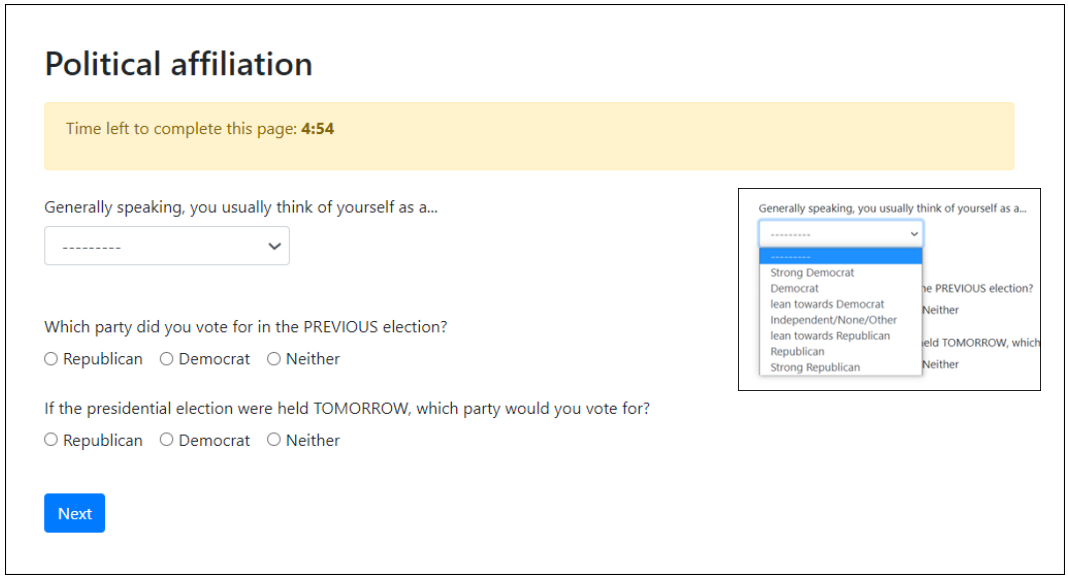
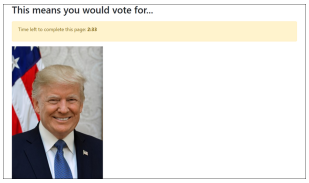


Figure 12: Identity salience pages - exposure to one of the 3 images below, concerning participants' preferred candidate, depending on the answer to the last question in the former page (Republican/Trump, Democrat/Biden, Neither/"?")



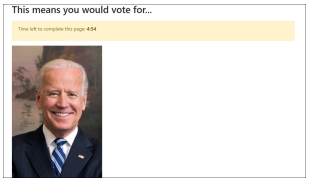


Figure 13: Identity salience pages - participants' perception of their opponent when the latter is a Republican

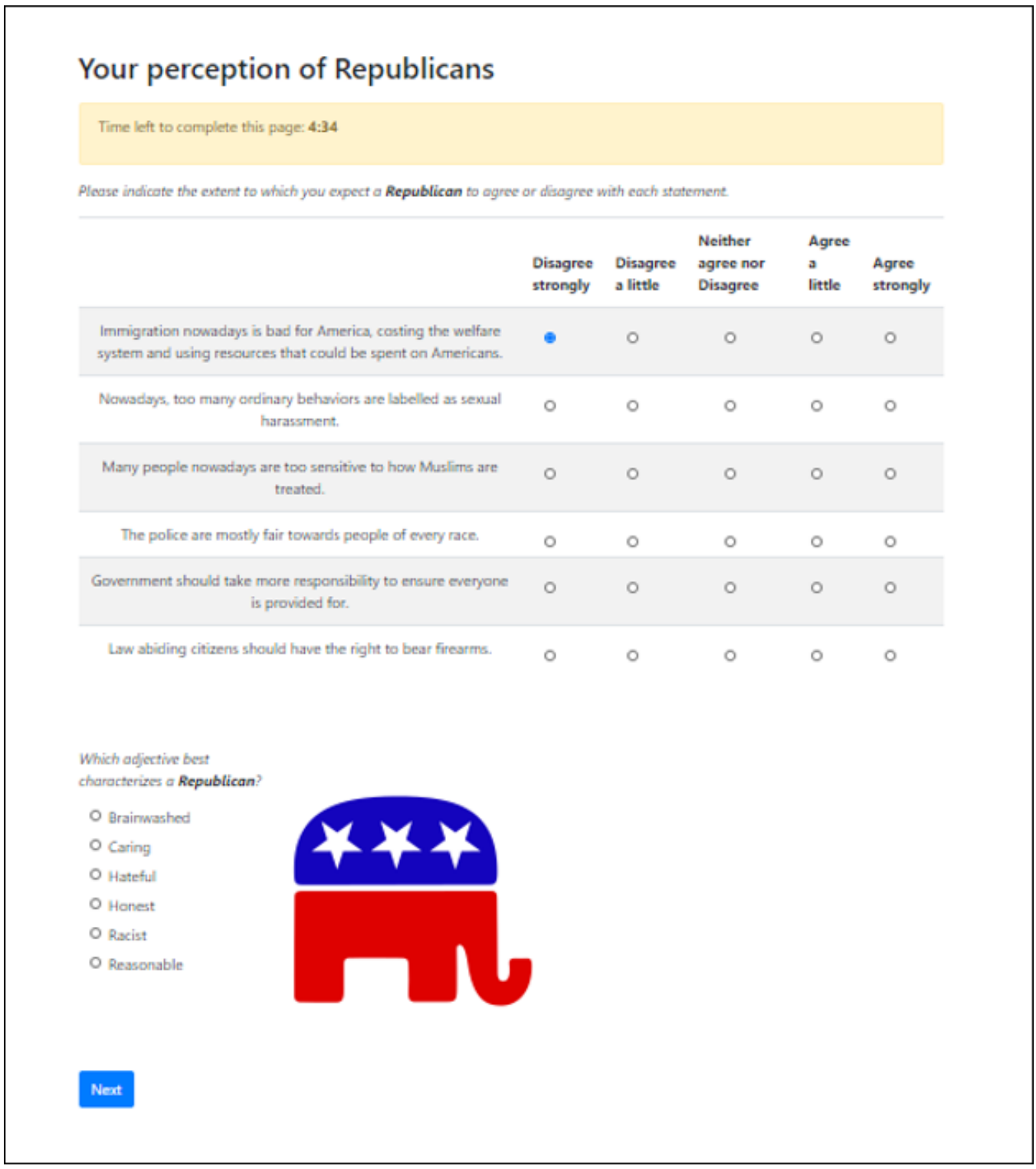


Figure 14: Identity salience pages - participants' perception of their opponent, when, alternatively the latter is a Democrat or an American

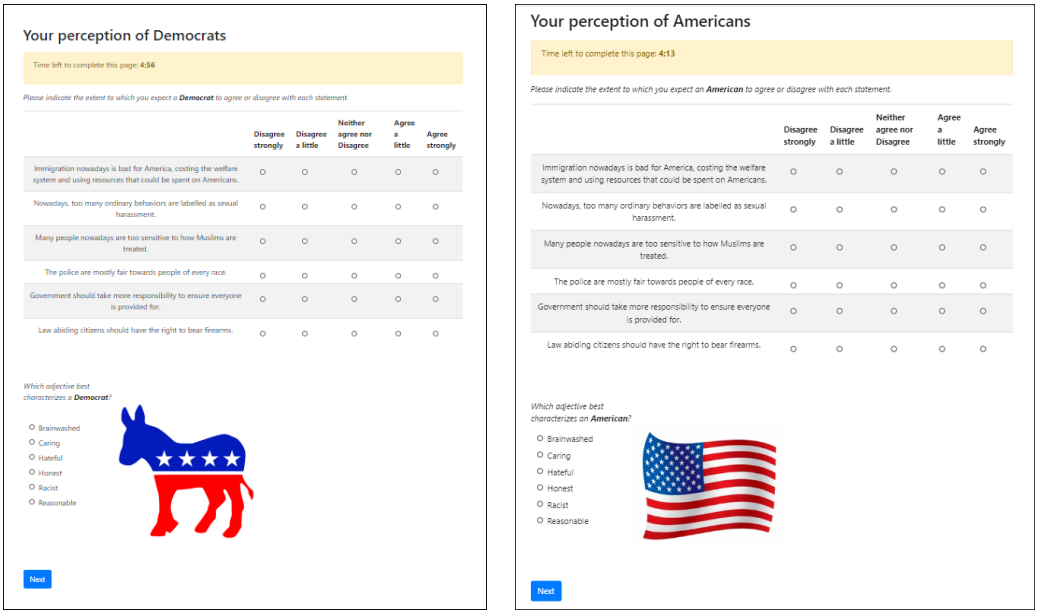


Figure 15: Identity salience pages - engagement in US politics

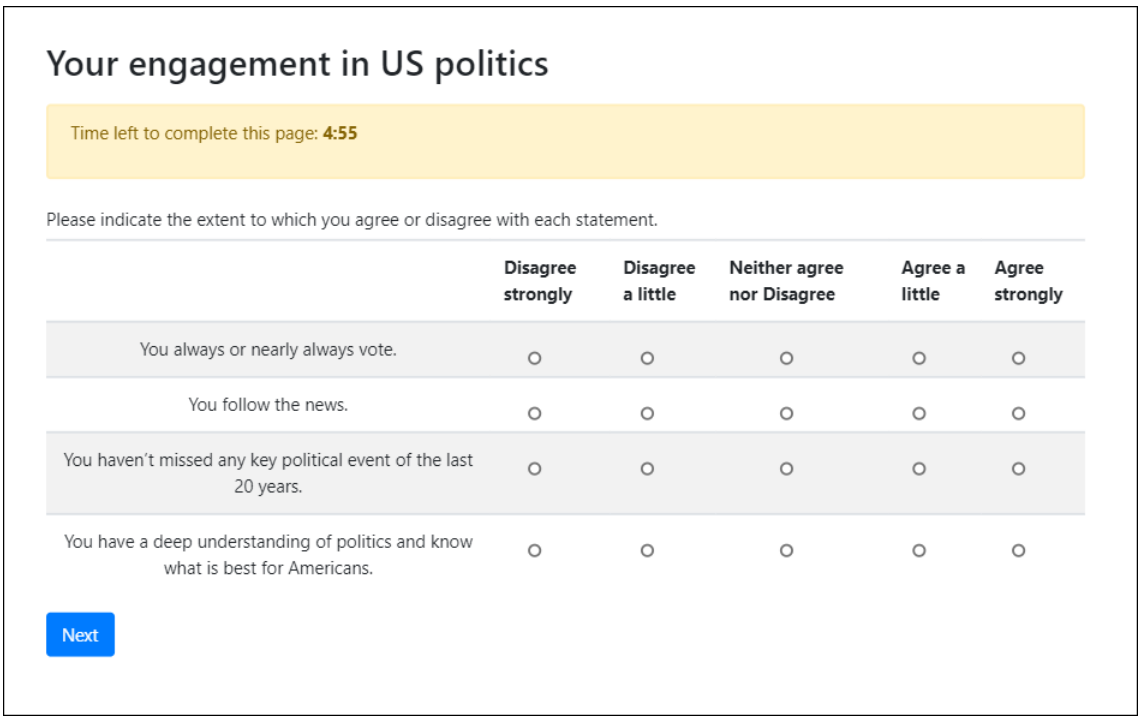


Figure 16: Initial elicitation of beliefs

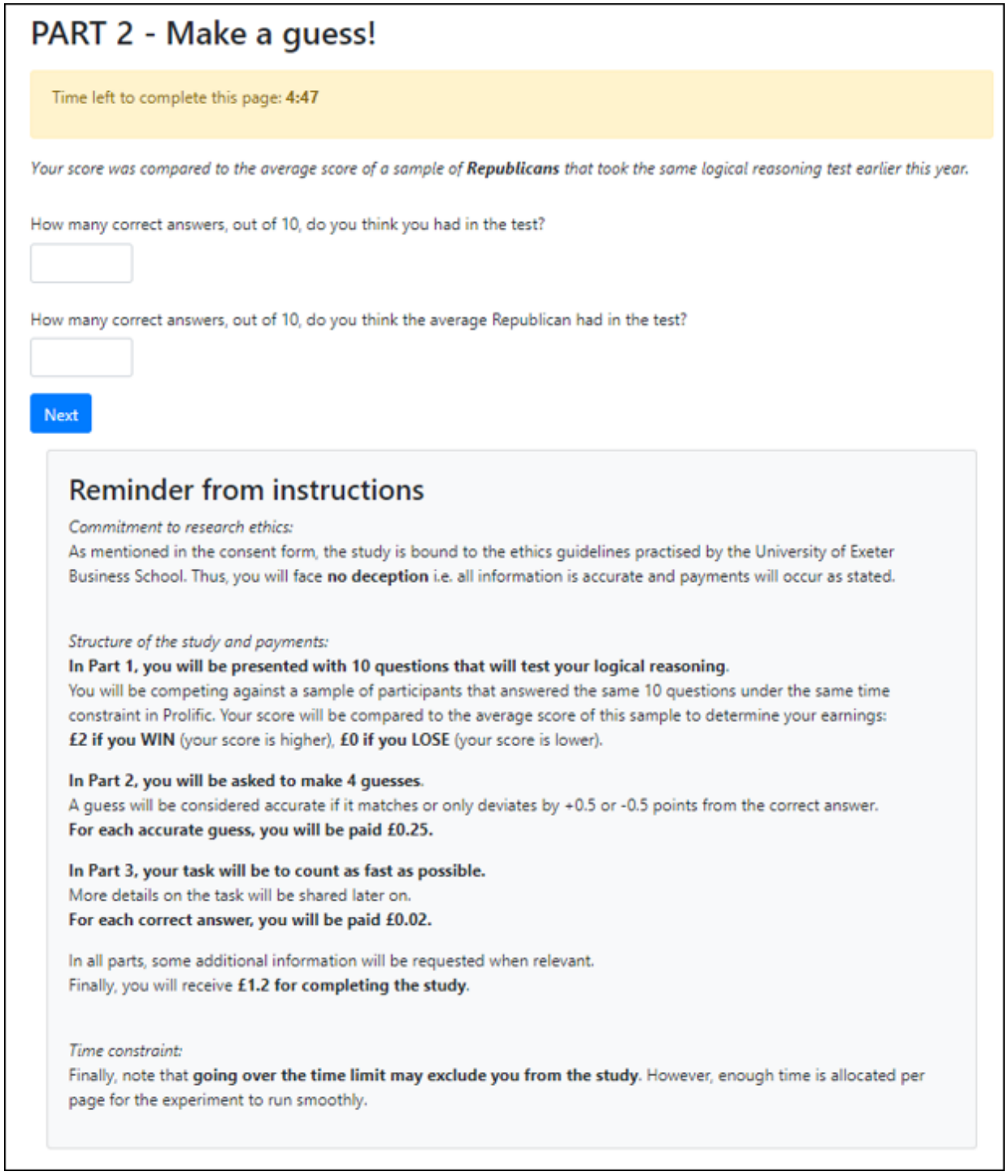


Figure 17: Revision of beliefs, when participants had a lower score than their opponent

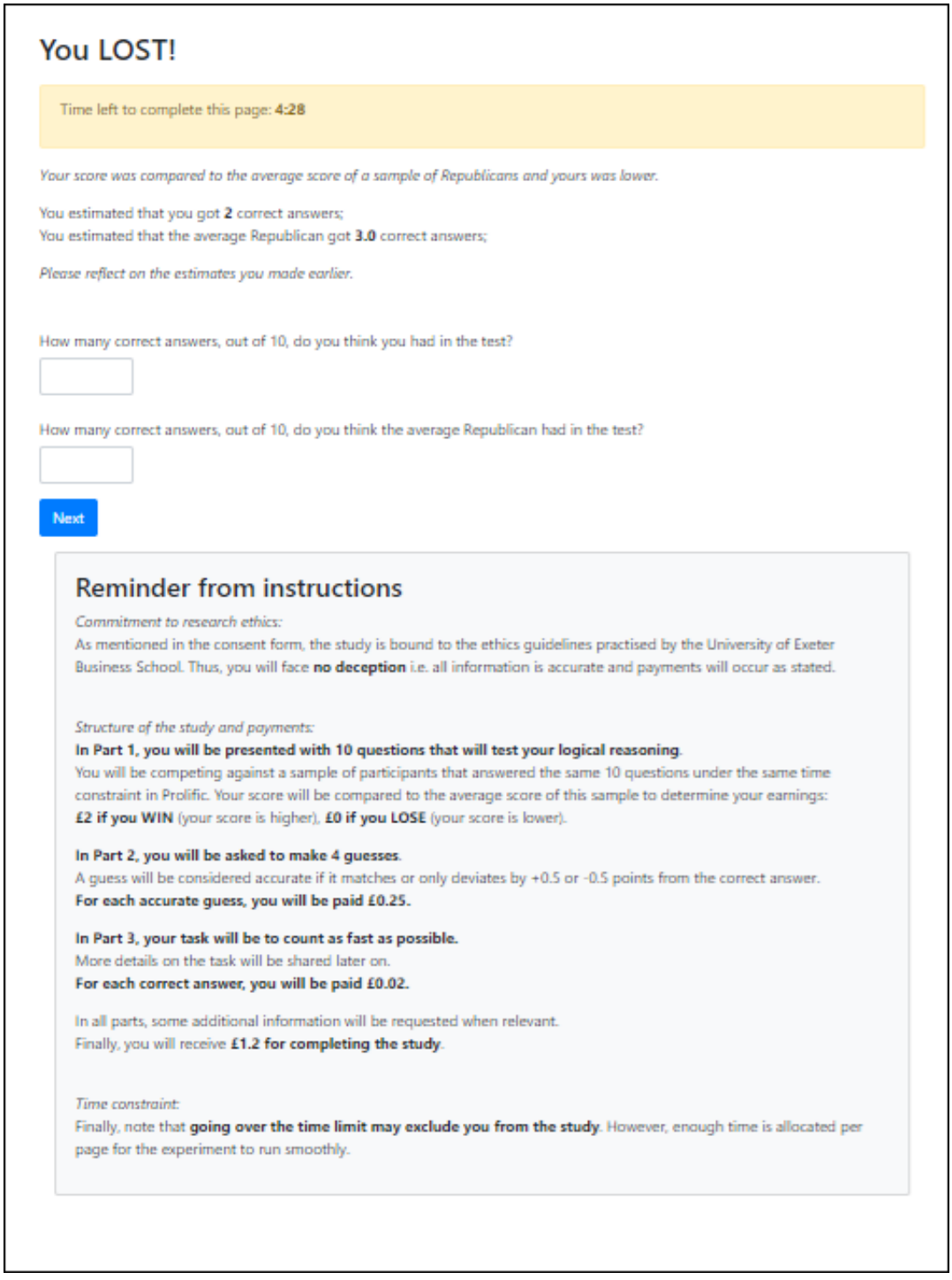


Figure 18: Revision of beliefs, when participants had a higher score than their opponent

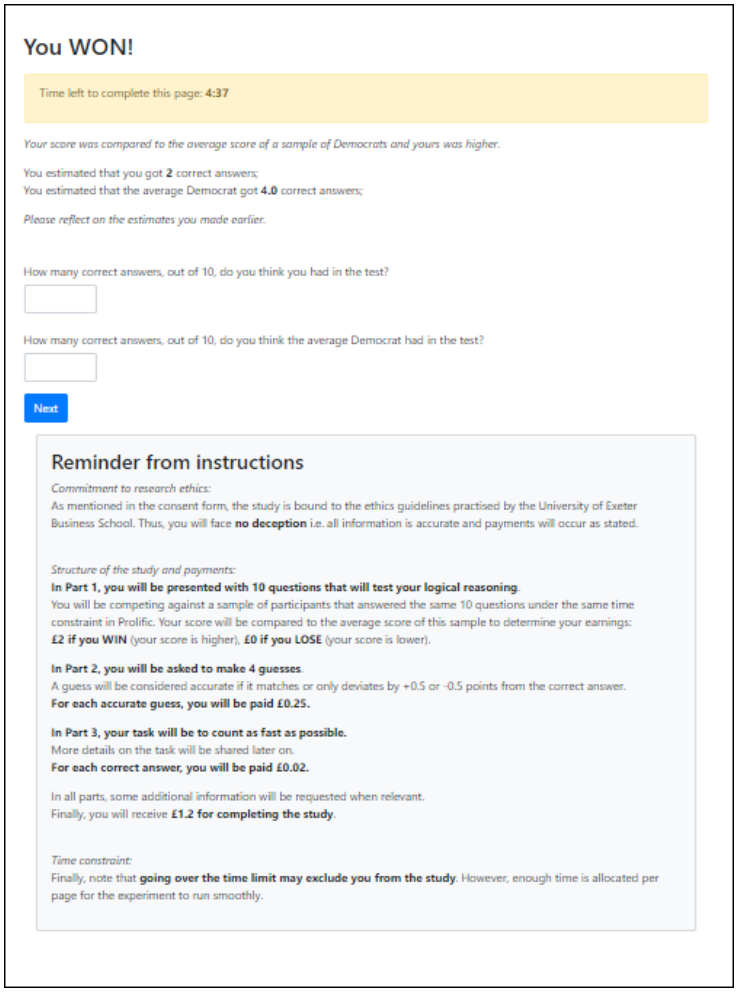


Figure 19: Page with option to reveal or avoid knowing one's score

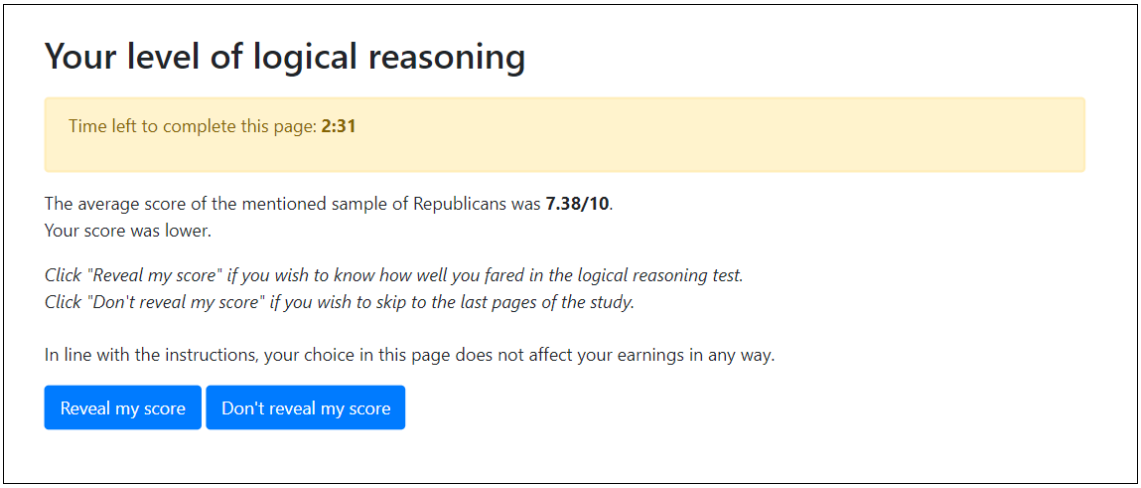


Figure 20: Revealed score page, only shown when participants chose to reveal in the former page

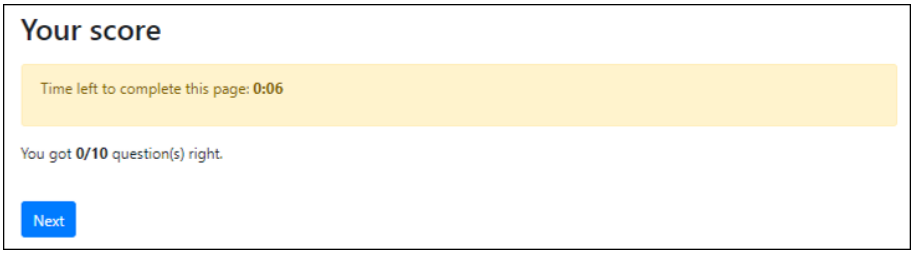


Figure 21: Big Five personality test

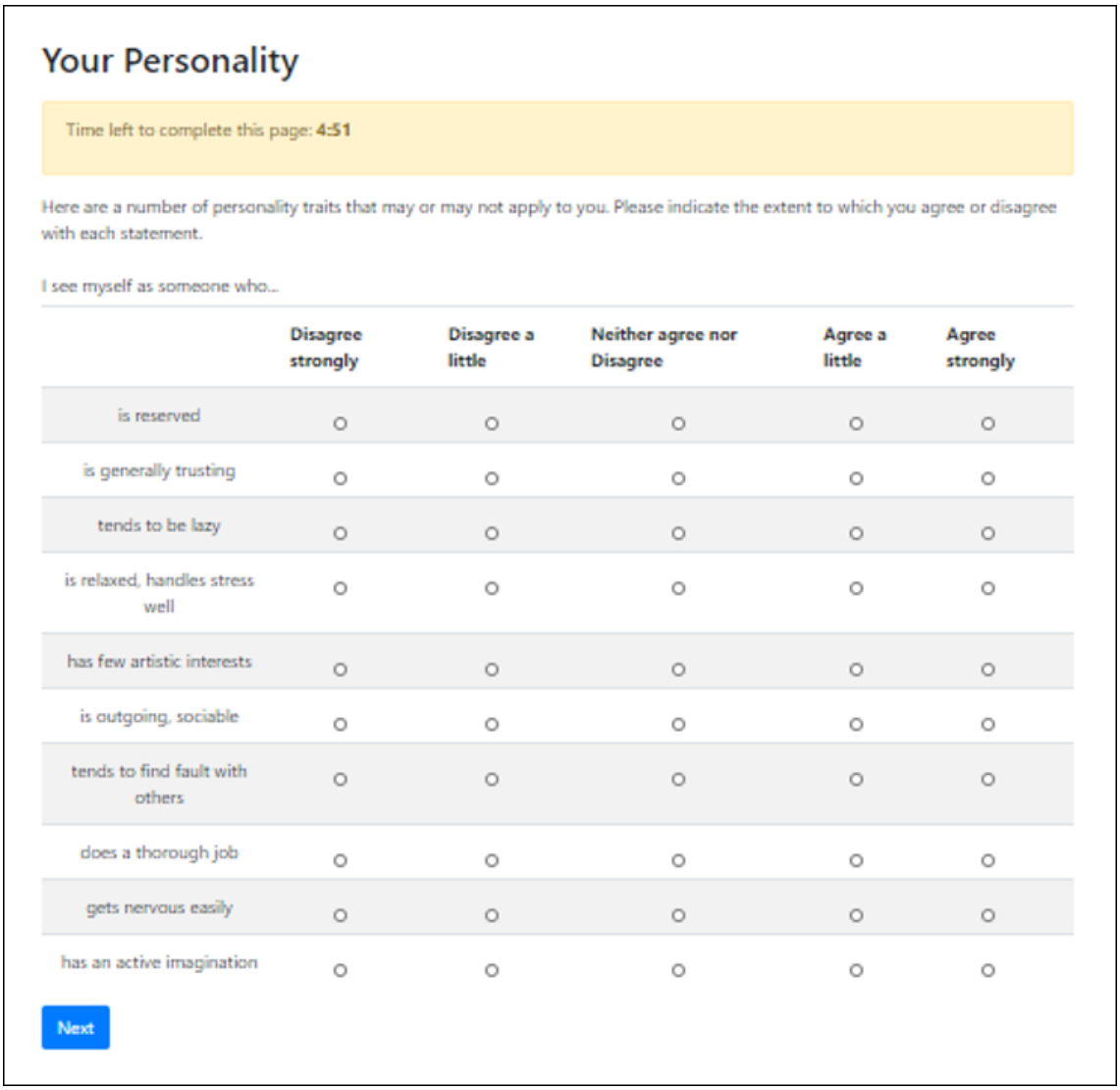


Figure 22: Vowel-counting effort task - instructions

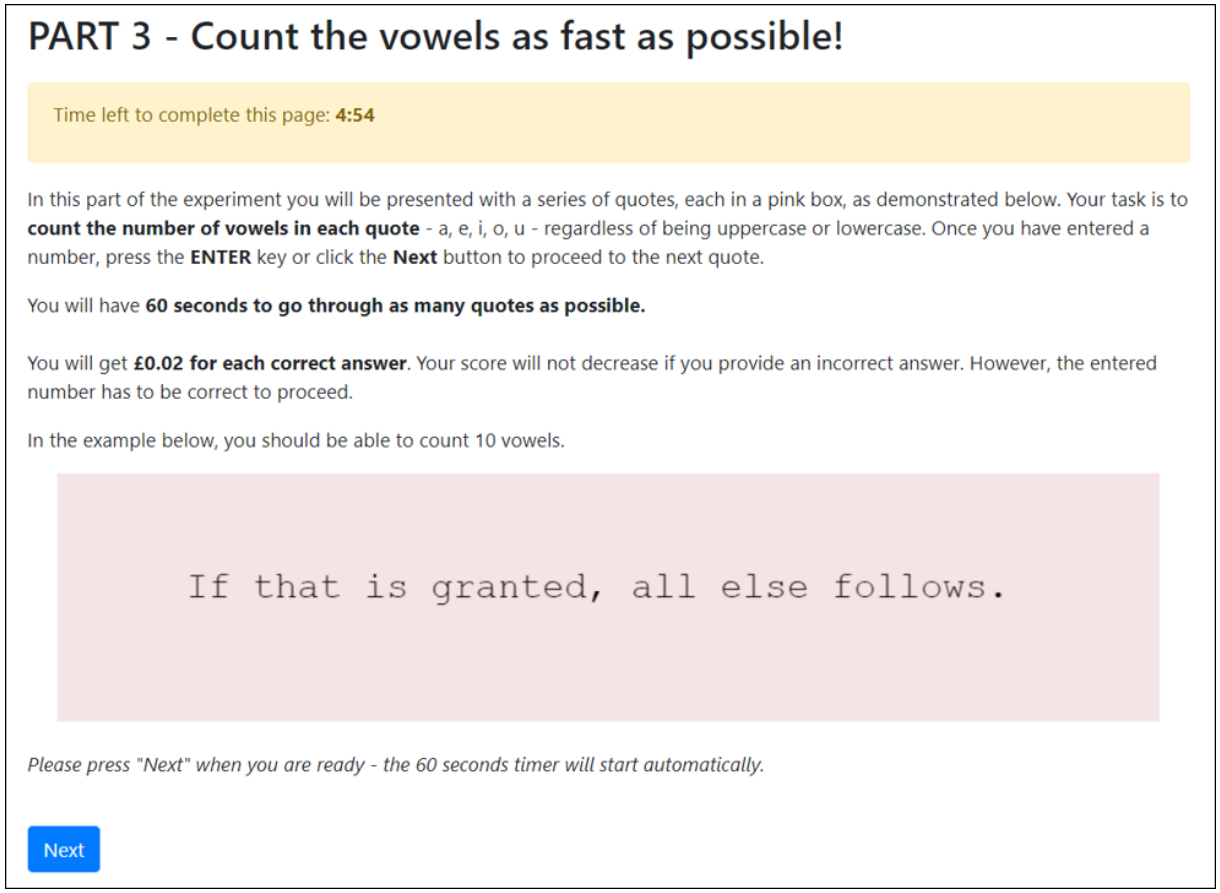


Figure 23: Vowel-counting effort task - first quote

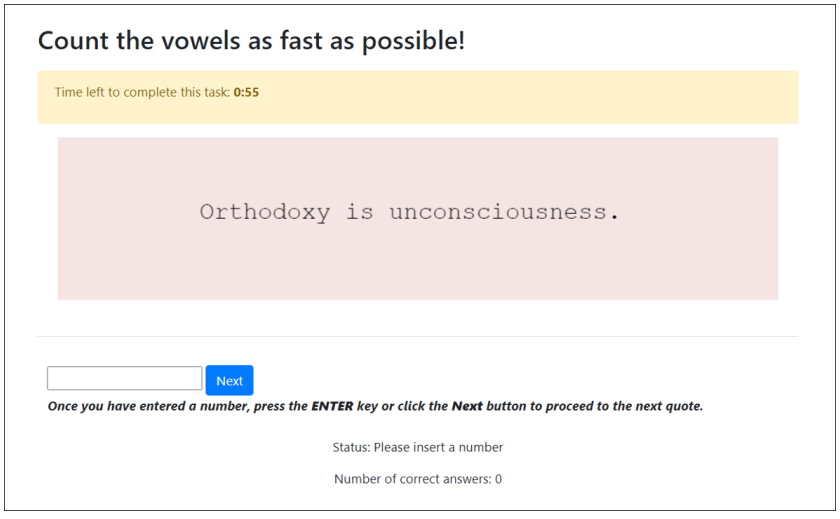


Figure 24: Vowel-counting effort task - display for invalid answer to first quote (participants could only move on to the next quote once answering correctly to the present quote)

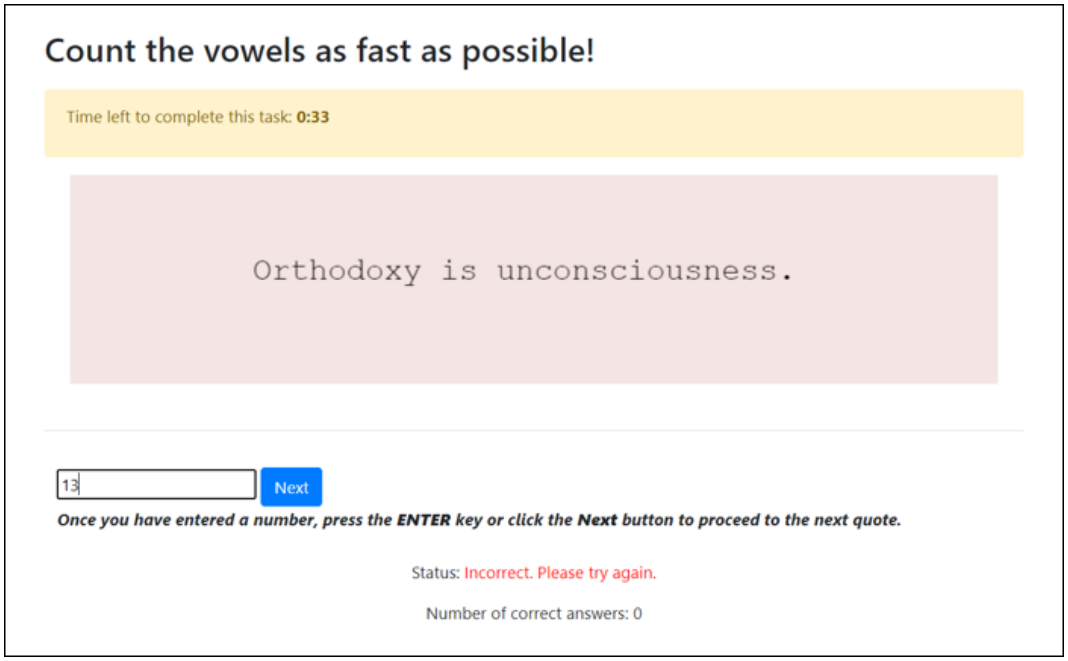


Figure 25: Vowel-counting effort task - second quote confirming number of correct answers

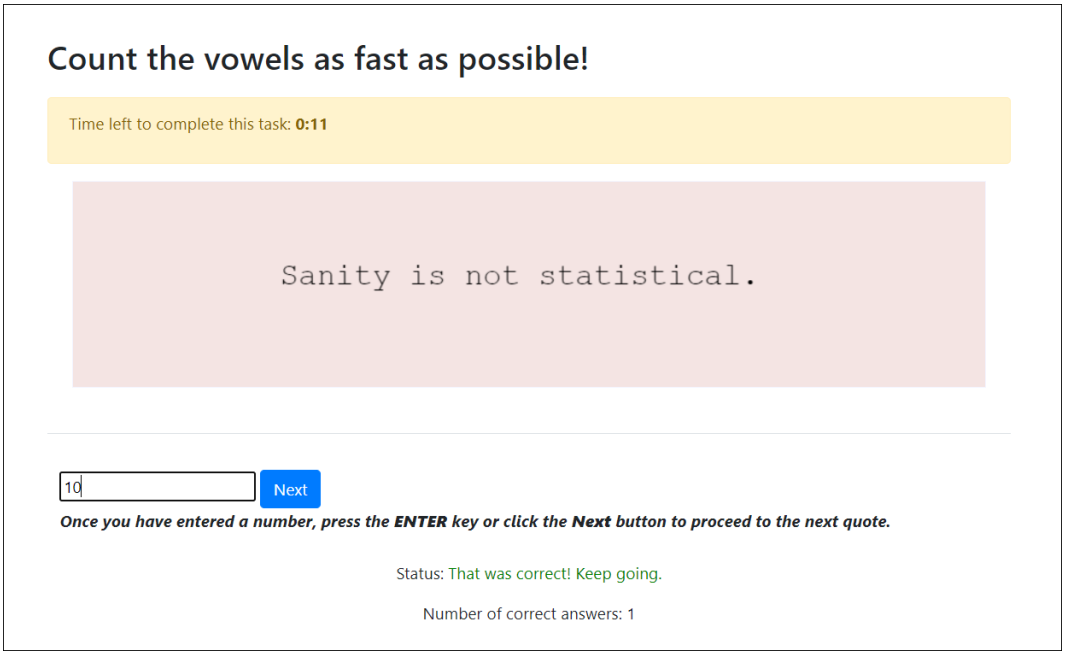


Figure 26: Participants' comments

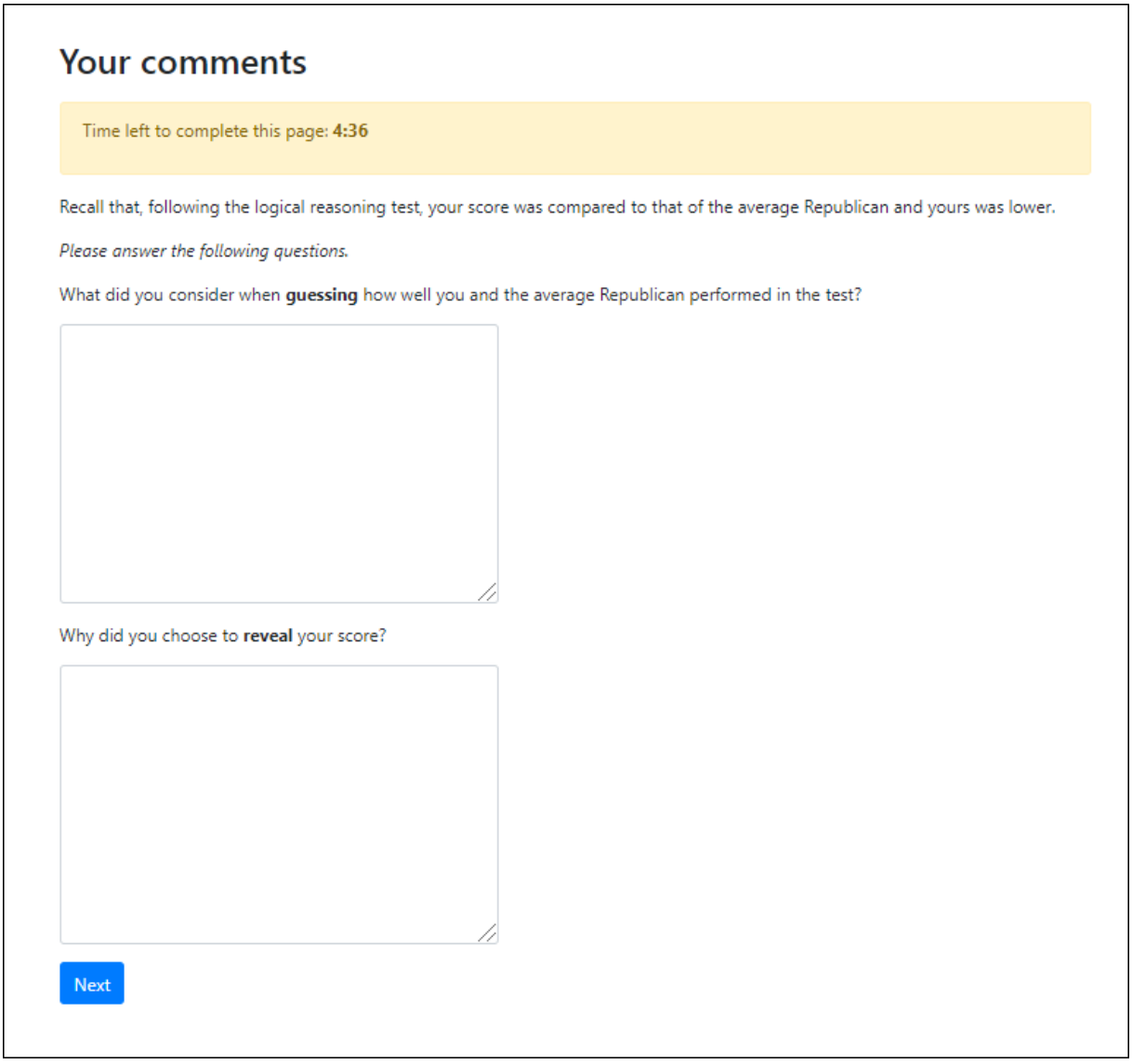


Figure 27: Payoffs (automatically computed)

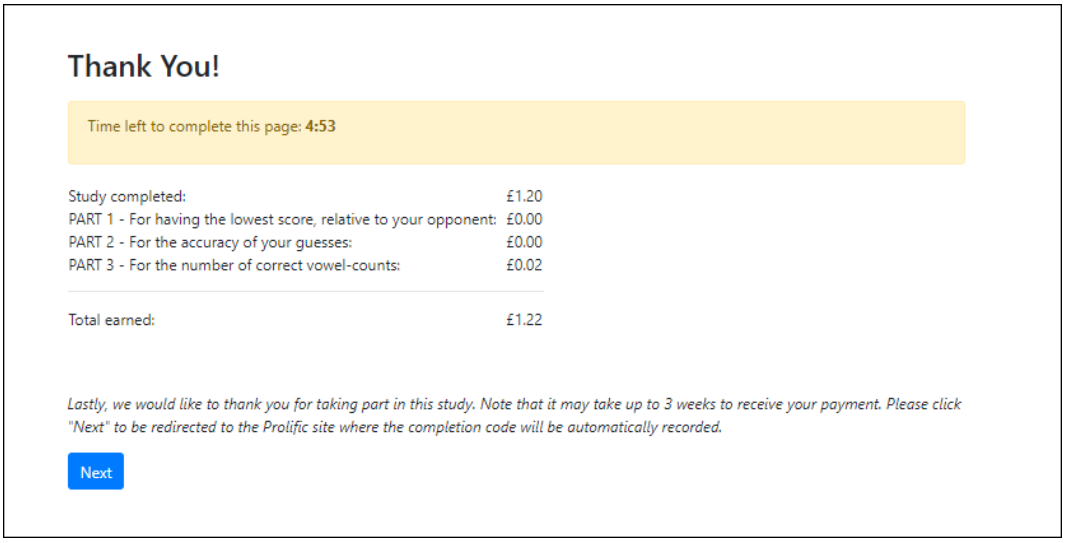


Figure 28: Prolific's confirmation page - clicking on the "Next" button in the former page would take participants back to Prolific webpage confirming completion of the study