**Fake news, fast and slow:**

**Deliberation reduces belief in false (but not true) news headlines**

**Supplementary Material**

Bence Bago1\*, David G. Rand2,3, & Gordon Pennycook4

1Institute for Advanced Study in Toulouse, University of Toulouse Capitole

2Sloan School, Massachusetts Institute of Technology

3Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

4Hill/Levene Schools of Business, University of Regina

\*Corresponding author: Bence Bago, Institute for Advanced Study in Toulouse, University of Toulouse Capitole, 21 Allée de Brienne, 31015 Toulouse, France, e-mail: bencebagok@gmail.com

1. **Supplementary Methods**

*Participants.* The preregistered sample for this Study was 1000 participants; 600 for the two response, and 400 for the one response baseline condition. In total, 1012 participants took part in this experiment, who were recruited through MTurk, and completed the experiment online. Out of the 1012, 359 participants took part in the one-response study (163 females, 192 males and 4 other, Mean age = 36.4 years, SD = 11.5 years) and 653 participants took part in the two-response condition (340 females, 307 males and 6 others, Mean age = 37.2 years, SD = 12.1 years). In the one response study 145 were Republican and 209 were Democrat leaning, while 5 did not respond. In the two response study, 228 were Republican and 416 were Democrat leaning, while 9 did not respond.

*News headlines.* We used a pool of 24 items, taken from Pennycook and Rand (2018), half of which were real and the other fake. Moreover, half of the items were Republican-consistent and the other half were Democrat-consistent items, based on a pretest. Participants were presented with 16 randomly selected headlines altogether; 4 from each category (i.e., Republican-consistent fake, Republican-consistent real, Democrat-consistent fake, Democrat-consistent real). Headlines were presented in a randomised order. All the items were included in a larger pretest (discussed in Pennycook & Rand, 2019) and were all rated by pretest participants to be either Pro-Democratic or Pro-Republican. For each headline, participants were asked the following question: “Do you think this headline describes an event that actually happened in an accurate way?” with the response options “Yes/No” (the order of Yes/No v. No/Yes was counterbalanced across participants).

*Load matrix.* As we wanted to minimize the impact of System 2 deliberation in the initial response stage of the two response experiment, participants were presented with a cognitive load. The rationale here is simple: Insofar as System 2 deliberation depends upon working memory to operate (Evans & Stanovich, 2013), restricting working memory capacity should increase reliance on System 1 intuitions (which do not depend upon working memory). As in other two response paradigm experiments (e.g., Bago & De Neys, 2018) we used a dot matrix task (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001) which has been shown to decrease analytical engagement in many tasks including probabilistic reasoning (De Neys, 2006) moral reasoning (Trémolière, De Neys, & Bonnefon, 2012) and the ultimatum game (De Neys, Novitskiy, Geeraerts, Ramautar, & Wagemans, 2011). In this task, before the headline is presented, participants are presented with a 4X4 matrix, with 5 dots in it, and they are instructed to memorise the dot pattern. After the initial response stage, participants are presented with a set of 4 matrices and they were asked to select the matrix that was presented to them in the beginning. After they made a decision, they were given a feedback if they selected the correct one or not. In cases where the participant failed to select the correct option, they were asked to pay more attention on the subsequent trials. Load was not applied during the final response stage.

*Response deadline.* To further assure the intuitive nature of the initial response, participants had to indicate their answer under a strict response deadline. The rationale here is that System 2 is argued to be relatively slower to produce a response than System 1 processing. Hence, by presenting participants with a deadline, we decrease the probability of System 2 engagement. To help us define the deadline we run a reading pre-test with N = 104 participants (41 females, 63 males, Mean age = 37.6 years, SD = 13.3 years) on MTurk. Participants were presented with the same headlines as in the actual experiment. They were instructed to simply read the material and then click on the “Next” button. We then logarithmically transformed RT data and back-transformed the data, thereby calculating the geometric mean. We found that, on average, people took 7.02 seconds (SD = 2) to read the headlines. Hence, we decided to set the response deadline to the closest integer of the average reading time: 7 seconds.

*Two response paradigm detailed procedure.* At the beginning of the experiment participants were told that they are going to be presented with each headline twice. First they should give an initial, intuitive response under response deadline, after which they will be presented with the same deadline again and can think about it as long as they want, without constraints. The literal instructions were:

“Welcome to the experiment!  
​Please read these instructions carefully!

This experiment is composed of 16 questions and a couple of practice questions. It will take about 25 minutes to complete and it demands your full attention. You can only do this experiment once.

In this task, we'll present you with a set of news headlines from 2017 / 2018. We are interested in how accurate you think the headline is. That is, do you think the headline is accurately describing something that actually happened? We want to know what your initial, intuitive decision is and how you respond after you have thought about the problem for some more time. Hence, as soon as the problem is presented, we will ask you to enter your initial response by clicking on it. We want you to respond with the very first answer that comes to mind. You don't need to think about it. Just give the first answer that intuitively comes to mind as quickly as possible. Next, the problem will be presented again and you can take all the time you want to actively reflect on it. Once you have made up your mind you can enter your final response. You will have as much time as you need to indicate your second response.  
To assure that the initial response is really intuitive, you will have 7 seconds to give a response. 1 second before the deadline passes, the background will turn yellow to warn you.  
In sum, keep in mind that it is really crucial that you give your first, initial response as fast as possible. Afterwards, you can take as much time as you want to reflect on the problem and select your final response.   
After you have entered your first and final answer we will also ask you to indicate how confident you are in the correctness of your response.  
Please confirm below that you read these instructions carefully and then press the "Next" button. We will start with a couple of practice problems.”

After the instructions, participants were presented with two practice problems. After this, they were presented with two dot matrix practice problems in which they were only asked to memorize the dot pattern and try to select the correct one out of the four matrices presented afterwards. Then, they were presented with two new practice problems, in which they had to give an initial response under load. Each trial started with a presentation of a fixation cross which stayed on screen for 1000 ms. After it disappeared, the dot matrix was presented for 2000 ms. Then, the headline was presented and participants had 7 seconds to give a response; 1 second before the deadline, the background turned yellow, to warn participants of the approaching deadline. If they managed to give a response, they were then presented with a question asking for their confidence in the correctness of their response in a scale from 0 to 100. In case they did not manage to give a response before the deadline passed, they received a message saying “*You did not enter your response before the deadline. Try to respond within the deadline on the next trials. No big deal if you're not totally sure. Just enter your very first intuitive answer. You get more time to reflect on your answer afterwards.*” After the initial confidence or the message, they were presented with the dot matrix question and had to select the pattern they were presented with. They received feedback on whether or not they selected the correct pattern. In case they did not, they were warned to try to focus on recalling the correct pattern in subsequent rounds. After the feedback, they were presented with the same headline again and were asked to give a final response. As a last step, they were asked to indicate their final confidence in the correctness of their response.

After the actual experiment, participants were presented with the headlines again and were asked if they have seen or heard about that story before (that is, prior to the experiment). Next, participants were asked to complete a seven item version of the Cognitive Reflection Test: first, they were presented with content-modified versions (Pennycook & Rand, 2018) of the original three CRT items (Frederick, 2005) after which they were presented with a newer four item non-numerical CRT [(Thomson & Oppenheimer, 2016)](https://www.zotero.org/google-docs/?MJT73y).

At the end, among various demographic questions, we asked a question to measure participant’s political orientation: “Which of the following best describes your political preference?” (1 - Strongly Democratic, 6 - Strongly Republican). Based on their response on this scale, participants were categorised as “Republicans” or “Democrats”. We coded items as “politically concordant” if the item was consistent with the partisan identity (Republican or Democrat) of the individual, and “politically discordant” if it was not (i.e., Democrat-consistent items were categorised as politically concordant for Republicans and politically discordant for Democrats, and vice versa).

To avoid familiarity effects, participants who took part in the neutral headlines pre-test or participated in previous studies in which the same set of headlines were used, could not participate in this experiment.

*One response baseline condition.* We also included a one-response baseline condition in which participants were presented with each headline only once. In each trial, after a brief fixation cross period (1 sec) the headline appeared and they were asked to give a response without any (load or deadline) constraints. As a final step, participants were also asked about their confidence level. In short, participants in this condition were simply given the “final response” portion of the 2-response paradigm.

*Calculation of Confidence intervals.* CIs of perceived accuracies were computed using the DescTools package in R (Signorell, 2016); we used the default “Wilson” approach to calculate Cis (Brown, Cai, & DasGupta, 2001).

*Analysis strategy*. In all of our analysis, we used mixed-effect logistic regression models. We always included items and Subject IDs as random intercepts in our models. We used lme4 package [(Bates, Sarkar, Bates, & Matrix, 2007)](https://www.zotero.org/google-docs/?W6h1dO) in R [(R Core Team, 2018)](https://www.zotero.org/google-docs/?cUZ6J1). We preregistered our primary analysis - any analysis that was not preregistered is labeled as *post hoc*. We reported standardized beta coefficients and their 95% confidence intervals.

1. **Neutral headlines pre-test**

**Methods & Procedure**

*Participants.* The preregistered sample for this Study was 600 participants; 400 for the two response, and 200 for the one response experiment. In total, 623 participants took part in this study, who were recruited through Mechanical Turk (MTurk), and completed the experiment online. Out of the 623, 202 participants were randomly assigned to the one-response condition (100 females and 101 males and 1 other, Mean age = 36.9 years, SD = 12.1 years) and 421 participants were randomly assigned to the two-response experiment (164 females and 256 males, Mean age = 35.9 years, SD = 10.4 years). In the one response study 41 were Republican and 93 were Democrat leaning, while 67 were independent or in the other category, and 1 did not respond. In the two response study, 98 were Republican and 194 were Democrat leaning, while 129 reported to be independent or other.

*News headlines.* We used politically neutral headlines that actually appeared on social media, for example: “300.000 pounds of rat meat sold as chicken wings across America”. 10 politically neutral headlines were presented, half of which were real (true and from a mainstream source) and half of which were fake (false and from an illegitimate source). All the items were included in a larger pretest (discussed in Pennycook & Rand, 2019) and were all rated by pretest participants to be neither Pro-Democratic nor Pro-Republican. The headlines were presented in a format as they normally appear on social media; a picture with a headline, byline and source. For each headline, participants were asked the following question: “Do you think this headline describes an event that actually happened in an accurate way?” with the response options “Yes/No” (the order of Yes/No v. No/Yes was counterbalanced across participants). Headlines were presented in a random order for each participants. The two and one response procedures were the same as in the main experiment.

*Exclusion.* We preregistered that we would exclude all trials with incorrect responses on the dot matrix question and the ones in which participants did not manage to give a correct response. Unexpectedly, however, we found a positive correlation between score on the CRT and accuracy of responding on the load question, r = 0.11, p < 0.0001. Hence, to minimise the possibility of a selection bias, we included all trials with in incorrect response on the load question in our analysis. However, we still had to exclude the missed deadline trials, as their initial response is N/A, and thus, cannot be compared to the final response. In total, we excluded (4.2%) of trials, and 4032 were further analysed (out of the 4210). This does not meaningfully change the analysis. Note that, as preregistered, we did not exclude any trials for the analyses involving a comparison of the one vs two response experiments. Due to experimenter error, one of the load items was not correctly displayed: during the response selection, the correct solution was not presented. This item was not taken out of the analysis.

**Results**

**Correlational evidence.** First, we tested the predicted interaction between CRT performance, response type (intuitive or final), and headline veracity (fake or real). The classical reasoning account claim that System 2 processing is important for overriding (and correcting) inaccurate intuitive responses implies that we should find a significant interaction between response type and headline veracity (i.e., people should be worse at discerning between fake and real news if forced to give an intuitive response). Furthermore, we hypothesized that if more analytic people are more likely to generate an accurate intuitive response, we should find a significant interaction between CRT performance and headline veracity. That is, highly reflective people should judge fake headlines to be less accurate and real headlines as more accurate and this should occur even for intuitive responses. Visual inspection of Table S2 supports both of these hypothesis: Perceived accuracy increased for real and decreased for fake news for final responses relative to intuitive responses and more analytic individuals were better at discerning between fake and real news than more intuitive individuals for both intuitive and final answers. Specifically, we found a significant interaction between CRT performance and veracity of headlines, *b* = 1.34, 95% CI = [0.99, 1.7], *p* < 0.0001 which suggests that high CRT people were more likely to produce a correct response than low CRT people. We also found a significant main effects of CRT score, *b* = -1.1, 95% CI = [-1.37, -0.81], *p* < 0.0001, and response type, *b* = -0.32, 95% CI = [-0.64, -0.003], *p* =0.048, but not headline’s veracity, *b* = 0.3, 95% CI = [-0.29, 0.89], *p* = 0.31. We did not find a significant interaction between CRT score and response type, *b* = -0.08, 95% CI = [-0.44, 0.27], *p* = 0.64, or a three way interaction between CRT performance, response type, and headline’s veracity; *b* = 0.32, 95% CI = [-0.1, 0.73], *p* = 0.13. More directly, we also did not find an interaction between response type and headlines veracity, even though the trend was as would be expected if people’s judgments improved for the final answer relative to the intuitive response, *b* = 0.21, 95% CI = [-0.19, 0.61], *p* = 0.3. This could be for many reasons, but one of them is that CRT performance explains a bigger chunk of variance and we do not have the power to detect an effect over and above that. Hence, we ran a post-hoc analysis that excluded CRT performance in the model. In this case, we found a very robust significant interaction between response type and headline’s veracity, *b* = 0.47, 95% CI = [0.29, 0.65], *p* < 0.0001 (along with a significant main effect of response type, *b* = -0.38, 95% CI = [-0.52, -0.23], *p* < 0.0001 and a significant effect of veracity *b* =1.4, 95% CI = [0.9, 1.91], *p* < 0.0001). Tables S1 shows the results.

It is important to note that fake and true headlines were seemingly not affected by increased deliberation in the same way. For true headlines, CRT performance is not very predictive at the initial response stage, unlike for fake items where it is clearly associated with better judgments. For both true and fake headlines, more reflective people are more likely to correct an erroneous intuitive response at the final response stage. Hence, being more reflective means that in some cases you have better intuitive responses, but also means that you are more likely to engage in deliberative correction when needed.

**Final response versus one response baseline.** We also compared perceived accuracy for fake and real news for the final response stage of the two response paradigm condition with the single (“final”) response from the one response paradigm condition (we will call these conditions as “no prior intuitive response”, “prior intuitive response”, respectively). Table S3 shows the results. This allows us to test whether being forced to give an initial response changes people’s final answers. That is, the mere act of providing an intuitive response may impact how people respond when giving a final (more reflective) answer. As is evident in Table S3, we found a significant interaction between the headline’s veracity and condition (prior intuitive response vs no prior intuitive response), *b* = 0.23, 95% CI = [0.04; 0.41], *p* = 0.018, as well as a significant main effect of headline’s veracity, *b* = 1.88, 95% CI = [1.42, 2.34], *p* < 0.0001, while the main effect of condition was not significant, *b* = -0.08, 95% CI = [-0.3, 0.13], *p* = 0.44. People’s final answers were less accurate (more likely to believe fake news and less likely to believe real news) if they first gave an intuitive answer. However, the effect was also not the same size for real and fake headlines: There was a 4.4% decrease in perceived accuracy for real news and 2.1% increase in perceived accuracy for fake news. This is interesting because, if this difference emerged as a result of people simply wanting to seem more consistent in the prior intuitive response condition, one would expect these differences to be roughly the same. A more parsimonious explanation is that forcing people to provide an intuitive response increases the salience of their potentially biased or incorrect prior beliefs and therefore undermines the effectiveness of later analytic reasoning.

**Initial response versus one response baseline.** As an additional manipulation check (not pre-registered), we compared initial response perceived accuracies on the two response condition with the perceived accuracies on the one response condition. If our intuition manipulation is working, we should see that when people are forced to response intuitively, the perceived accuracy for fake headlines is increasing while the perceived accuracy for real headlines is decreasing. Results are summarized in Table S3. We found a significant main effect of veracity *b* =1.31, 95% CI = [0.82, 1.8], *p* < 0.0001 and a significant main effect of experimental condition, *b* = -0.41, 95% CI = [-0.6, -0.22], *p* < 0.0001. Most importantly, we found a significant interaction between headline’s veracity and condition, *b* = 0.61, 95% CI = [0.42, 0.79], *p* < 0.0001. This, together with results on Table S3 confirm that our intuition manipulation increases reliance on intuition and decreases the possibility for analytic engagement.

1. **Additional Analysis**
2. **Detailed analysis of political headlines**

**Correlational evidence.** Table S4 illustrates the results. When CRT performance was included in the model (along with response type, veracity and concordance), we still found a significant interaction of veracity and response type, *b* = 0.53, 95% CI = [0.21, 0.86], *p* = 0.001. We also found a significant interaction between veracity of the headline and CRT performance; *b* = 1.02, 95% CI = [0.72, 1.32], *p* < 0.0001. This latter result suggests that CRT performance has bigger effects for fake news than for real news (i.e., highly reflective people are more likely to judge a fake headline inaccurate than less reflective people; for real headlines CRT has no effect, or even is nominally associated with increased perceived accuracy). Interestingly, this was the case both at the initial and final response stages as well; meaning that higher CRT people are also more likely to come up with the correct solution when forced to rely on intuitive processing. Furthermore, we found a significant four-way interaction, *b* = 0.39, 95% CI = [0.03, 0.76], *p* = 0.033. In short, this interaction indicates that: a) high CRT people are more likely to perceive a fake headline as inaccurate (which does not apply to real headlines), and b) low CRT people are more affected by political consistency than high CRT people. Put differently, low CRT people, but not high CRT people, are more likely to correct their response on fake items when the headline is consistent with their beliefs, but this difference is not evident for politically discordant fake news regardless of CRT level. Even though this was not predicted, it is interesting insofar as it is flatly inconsistent with the motivated reasoning account, which would predict that high CRT people should be *more* (not less) sensitive to the political concordance of the headlines. We also found a significant main effect of headline’s veracity, *b* = 0.8, 95% CI = [0.31, 1.3], *p* = 0.001, response type, *b* = -0.51, 95% CI = [-0.77, -0.25], *p* = 0.0001, and CRT performance, *b* = -1.12, 95% CI = [-1.38, -0.86], *p* < 0.0001. None of the other main effects or interactions were significant.

**Final response versus one-response baseline.** Table S3 shows the results. We predicted based on our neutral headline pre-test that we should find a significant interaction between headlines veracity and experimental condition (i.e., comparing the final answer in the two-response condition with the single answer from the one-response baseline). Note that to avoid selection bias, for this analysis, we did not exclude any trials. However, this interaction was not significant *b* = 0.03, 95% CI = [-0.14, 0.2] *p* = 0.74. Interestingly, however, we did find a three-way interaction between condition, political concordance, and headline veracity, *b* = 0.19, 95% CI = [0.01, 0.37], *p* = 0.037. As is evident in Table S3, the predicted anchoring effect is evident for politically discordant items; here, the perceived accuracy of fake headlines was lower and perceived accuracy for real headlines was higher for the no prior intuitive response condition relative to the prior intuitive response condition. For politically concordant items, however, there was no such biasing effect of intuitive responding. Together with the significant anchoring effect among politically neutral headlines in our pre-test, the present results suggest that there is something unique about politically concordant items that do not produce the carry-over anchoring effect. We also found a significant main effect of headline’s veracity, *b* = 1.95, 95% CI = [1.52, 2.38], *p* < 0.0001,concordance, *b* = -0.18, 95% CI = [-0.32, -0.04], *p* = 0.01, and interaction between veracity and concordance, *b* = -0.28, 95% CI = [-0.44, -0.11], *p* = 0.0009. We found no interaction between concordance and condition, *b* = -0.15, 95% CI = [-0.33, 0.02], *p* = 0.085, and no main effect of condition, *b* = 0.01, 95% CI = [-0.2, 0.2], *p* = 0.94.

**Initial response vs one-response baseline.** For completeness, we conducted an analysis in which we compare responses given at the initial response stage (the one given under load and response deadline) with one response results. We found a significant main effect of condition (response under load and pressure vs response on one response paradigm), *b* = -0.35, 95% CI = [-0.54, -0.16], *p* = 0.0004, a main effect of headline’s veracity, *b* = 1.54, 95% CI = [1.12, 1.95], *p* < 0.0001, and a main effect of concordance, *b* = -0.19, 95% CI = [-0.33, -0.06], *p* = 0.005. Besides, we found a significant interaction of veracity and condition; *b* = 0.34, 95% CI = [0.17, 0.51], *p* < 0.0001, concordance, veracity and condition, *b* = 0.21, 95% CI = [0.03, 0.39], *p* = 0.02, and concordance and veracity, *b* = -0.31, 95% CI = [-0.48, -0.15], *p* = 0.0002. The interaction of concordance and condition, *b* = -0.14, 95% CI = [-0.31, 0.04], *p* = 0.12, was not significant. Figure 1 and S5 summarize the average perceived accuracies. This suggests that load and time pressure increased perceived accuracy for fake headlines; both for discordant and concordant; but not for real ones. On real headlines, load and response deadline caused a difference of 5.2% on discordant headlines, but only a 1.2% difference for concordant headlines.

1. **Load exclusion analysis with neutral and politically relevant headlines**

**Neutral headlines pre-test**

We added Tables S5-S7 with excluding all participants who gave an incorrect response on the load questions. In the neutral headlines pre-test, we excluded 32.9% additional trials (the missed initial deadline trials remained excluded) and analysed 2706. Averages and confidence intervals show no notable deviations from the ones in which we included all incorrect load data. Note that the load question (in which participants had to choose one of the correct dot pattern out of 4 options) on one of the real headline was incorrectly presented; the correct solution was not presented there. Hence, this item was not included in this analysis (item nr. 4).

**Main study - Political headlines**We also present additional tables and averages for politically relevant in which we did exclude trials with incorrect load response. In total, we excluded 17.8% of trials from this analysis (the missed initial deadline trials remained excluded) and 8233 were further analysed. There are no notable differences between these results and the results when we applied only the missing initial response criteria.

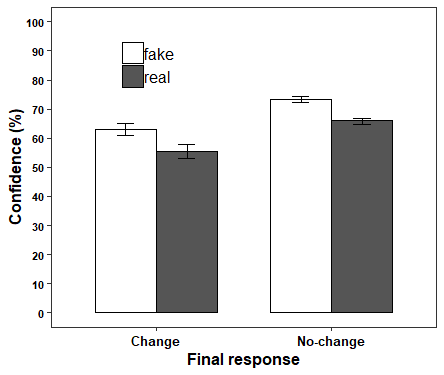
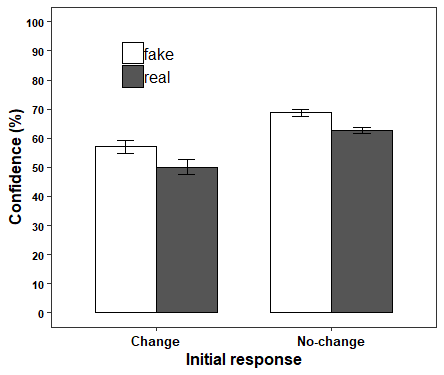
1. **Effect of knowledge**

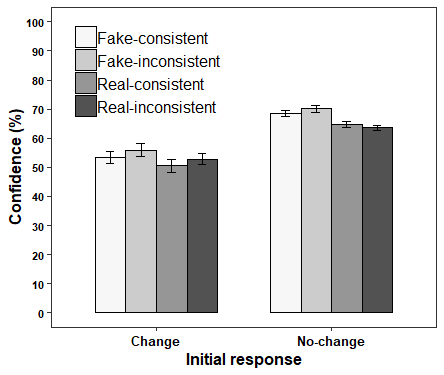
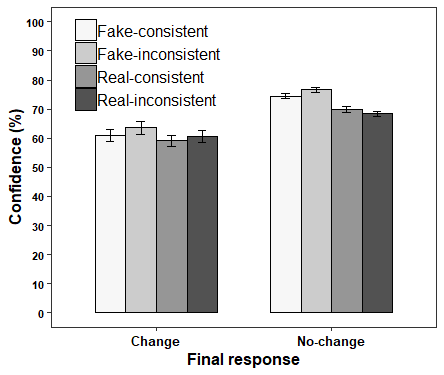
As explained in the main text, we used headlines that appeared online in real life. These headlines were also used in previous experiments as well. Hence, this might give rise to the suspicion that some of our participants might have known these headlines and that their knowledge altered the way they responded. Thus, people who took part in previous experiments in which the same set of headlines were applied could not participate in our main experiment. Furthermore, at the end of the experiment, we showed all the headlines people were presented with, and asked if they have seen or heard about this story prior to the experiment. In our neutral headline pre-test, familiarity with these problems were quite low (11.3%), as well as in the main experiment (19.1%). We also computed averages and 95% confidence intervals with excluding these responses. Even though perceived accuracy decreased on fake (~5% difference) and real headlines, it did not affect any of the differences we found. Table S8 and S9 shows the results.

1. **Confidence analysis**

Research on metacognition in reasoning showed that when people decide to change their intuitive, initial response, that response is associated with decreased level of confidence compared to those trials on which people do not change (Bago & De Neys, 2017; Thompson, Turner, & Pennycook, 2011). This led some theorists to believe that decreased level of initial confidence make people to engage in analytic thinking. We aimed to test whether this finding replicates with a different set of stimuli, namely, news headlines. Do people who are less confident at the initial response stage were more likely to change their response? Figure S1 and S2 represents the findings of neutral and politically relevant headlines, respectively, both at the initial and final response stage.

For neutral headlines, we indeed found that people were less confident when they ended up changing their initial response compared to when they ended up not changing it. change and no-change categories (by using mixed effect regression models, in which we added the random effect of items and Subject IDs), *b* = 0.18, 95% CI = [0.15, 0.2], *p* < 0.0001. We found the same difference regarding politically relevant headlines, *b* = 0.17, 95% CI =[0.15, 0.18], *p* < 0.0001. Hence, participants feeling less confident in their initial response will be more likely to change it, and this effect was present both for politically relevant and neutral headlines.

**Figure S1.** Figure illustrates confidence levels for change and no-change categories for neutral headlines in Study 1, separately for fake and real and for initial and final responses. Error bars are 95% confidence intervals.

**Figure S2.** Figure illustrates confidence levels for change and no-change categories for politically relevant items as a function of response type and political concordance. Error bars are 95% confidence intervals.

1. **Ideology analysis**

We tested whether political ideology has an effect in perceived accuracy of fake and real news headlines. We tested neutral and politically relevant headlines separately. For neutral headlines, we averaged over two measures of political ideology for each participants; 5 level scale measures of conservatism-liberalism on economic issues and on social issues (measured on different scales).

For politically relevant headlines we used a continuous, 6 level measure of political ideology, which was used to classify participants as Republicans/Democrats (this question was not asked in the neutral headlines experiment).

We added these continuous measures of ideology as a predictor to the model along with veracity of the headlines and response type for neutral headlines. For political headlines, we also added consistency to the model.

On neutral headlines, we found a main effect of ideology, *b* = 0.36, 95% CI = [0.08, 0.63], *p* = 0.01, main effect of headline’s veracity, *b* = 1.84, 95% CI = [1.25, 2.44], *p* < 0.0001 and a significant interaction between veracity and response type, *b* = 0.55, 95% CI = [0.13, 0.98], *p* = 0.01. Interestingly, we found a significant interaction between ideology and veracity, *b* = -0.52, 95% CI = [-0.89, -0.15], *p* = 0.006. Table S10 shows the results. Conservatives seem to be more susceptible to perceive fake headlines as real ones than liberals. However, there is no clear effect of ideology on the perceived accuracy of real headlines.

For political headlines, we found a significant main effect of veracity, *b* = 1.93, 95% CI = [1.36, 2.49], *p* < 0.0001, response type, *b* = -0.52, 95% CI = [-0.84, -0.2], *p* = 0.001, concordance, *b* = -0.73, 95% CI = [-1.27, -0.2], *p* = 0.001. We also replicated the interaction between veracity and response type *b* = 0.68, 95% CI = [0.29, 1.07], *p* = 0.0006. More interestingly, ideology significantly interacted with veracity of headlines, *b* = -0.44, 95% CI = [-0.87, -0.02], *p* =0.039, and concordance, *b* = 0.56, 95% CI = [0.06, 1.06], *p* = 0.027. Table S10 represents the findings. It suggests that ideology have a clear effect on perception of fake news, but less effect on real headlines; that is, more conservative participants are more likely to perceive fake as accurate, which does not seem to be true for real headlines. Conservative participants seem to be more susceptible to politically discordant headlines, than their liberal counterparts; while this ideology difference is smaller for politically concordant headlines.

No other interaction or main effect reached significance.

Note, however, that as this sample is not nationally representative, the results of the ideology analysis should be interpreted with great caution.

**Table S1.** Table shows perceived accuracy averages for initial and final responses and both on neutral and politically relevant real and fake headlines. 95% confidence intervals are in brackets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Fake headlines | | Real headlines | |
|  | Initial response | Final response | Initial response | Final response |
| Neutral | 39.4% [37.3; 41.6] | 32.4% [30.4; 34.5] | 67.4% [65.4; 69.4] | 70.4% [68.4; 72.4] |
| Concordant | 37.6% [35.7, 39.5] | 30.9% [29.1, 32.7] | 67.5% [65.6, 69.3] | 68.1% [66.2, 69.9] |
| Discordant | 33.3% [31.4, 35.1] | 27.1% [25.4, 28.9] | 56.7% [54.7, 58.6] | 58.4% [56.5, 60.3] |

**Table S2.** Table shows mean perceived accuracy for initial and final responses, both on fake and real headlines, separately for “intuitive” reasoners (below median CRT score) and “reflective” reasoners (median CRT score or above) on neutral headlines. 95% confidence intervals are in brackets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Intuitive | | Reflective | |
|  | Initial response | Final response | Initial response | Final response |
| Fake | 48.6% [45.2; 51.9] | 41.9% [38.6; 45.2] | 32.6% [30; 35.4] | 25.3% [22.9; 27.9] |
| Real | 66.2% [63; 69.2] | 66.9% [63.7; 69.9] | 68.4% [65.6; 71] | 73.1% [70.5; 75.6] |

**Table S3.** Table shows perceived accuracy averages for one response pre-tests (no prior intuitive response), initial (response under pressure and load) and final responses (prior intuitive response). As in the final response vs one response baseline test, trials with missed initial responses were not excluded from the “prior intuitive response” condition. 95% confidence intervals are in brackets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | No prior intuitive response | Prior intuitive response | Response under pressure and load |
| Neutral | Fake | 30.7% [27.9; 33.6] | 32.8% [30.9; 34.9] | 39.4% [37.3; 41.6] |
| Real | 74.6% [71.8; 77.1] | 70.2% [68.2; 72.1] | 67.4% [65.4; 69.4] |
| Concordant | Fake | 30.9% [28.5; 33.3] | 31% [29.2; 32.8] | 37.6% [35.7, 39.5] |
| Real | 68.7% [66.3; 71.1] | 67.9% [66; 69.6] | 67.5% [65.6, 69.3] |
| Discordant | Fake | 24.7% [22.5; 27] | 27.6% [26; 29.4] | 33.3% [31.4, 35.1] |
| Real | 61.9% [59.4; 64.4] | 58.4% [56.4; 60.2] | 56.7% [54.7, 58.6] |

**Table S4.** Mean perceived accuracy for initial and final responses, both on fake and real, politically concordant, and discordant headlines, separately for “intuitive” reasoners (below median CRT performance) and “reflective” reasoners (median CRT performance or above). 95% confidence intervals are in brackets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Intuitive | | | | |
|  | Politically Concordant | | Politically Discordant | |
|  | Initial response | Final response | Initial response | Final response |
| Fake | 46% [43.1; 48.8] | 37.2% [34.4; 40] | 41.3% [38.5; 44.2] | 36.1% [33.4; 38.9] |
| Real | 67.2% [64.4; 69.8] | 68.5% [65.8; 71.1] | 59% [56.2; 61.8] | 58.9% [56.1; 61.7] |
| Reflective | | | | |
|  | Politically Concordant | | Politically Discordant | |
|  | Initial response | Final response | Initial response | Final response |
| Fake | 30.1% [27.7; 32.6] | 25.2% [23; 27.7] | 26% [23.7; 28.5] | 19.1% [17; 21.3] |
| Real | 67.8% [65.2; 70.3] | 67.7% [65.1; 70.2] | 54.6% [51.9; 57.3] | 57.9% [55.2; 60.6] |

**Table S5.** Table shows perceived accuracy averages for initial and final responses when excluding trials with incorrect load response on neutral headlines. 95% confidence intervals are in brackets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Fake headlines | | Real headlines | |
|  | Initial response | Final response | Initial response | Final response |
| Neutral | 39.7% [37.2; 42.3] | 31.8% [29.4; 34.2] | 66.5% [63.9; 69] | 69.8% [67.2; 72.2] |
| Concordant | 37.6% [35.5; 39.8] | 31.2% [29.2; 33.2] | 67.6% [65.6; 69.6] | 69% [67; 71] |
| Discordant | 31.9% [29.9; 34] | 26.3% [24.4; 28.3] | 56.4% [54.3; 58.5] | 58.7% [56.5; 60.7] |

**Table S6.** Table shows perceived accuracy averages on neutral headlines, for initial and final responses, both on fake and real headlines, separately for “intuitive” reasoners (below median CRT score) and “reflective” reasoners (median CRT score or above), when excluding trials with incorrect load response. 95% confidence intervals are in brackets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Intuitive | | Reflective | |
|  | Initial response | Final response | Initial response | Final response |
| Fake | 42.3% [43.2; 51.4] | 39.7% [35.7; 43.8] | 34.7% [31.6; 38] | 26.5% [23.7; 29.6] |
| Real | 66.7% [62.8; 70.5] | 66.9% [62.7; 70.7] | 67.3% [64; 70.4] | 71.8% [68.5; 74.9] |

**Table S7.** Table shows perceived accuracy averages for initial and final responses, both on fake and real, politically concordant and discordant headlines, separately for “intuitive” reasoners (below median CRT performance) and “reflective” reasoners (median CRT performance or above), when excluding trials with incorrect load response. 95% confidence intervals are in brackets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Intuitive | | | | |
|  | Concordant | | Discordant | |
|  | Initial response | Final response | Initial response | Final response |
| Fake | 46.5% [43.3; 49.8] | 38.1% [35; 41.3] | 39.7% [36.6; 43] | 34.8% [31.7; 38] |
| Real | 67.9% [64.9; 70.8] | 69.5% [66.5; 72.4] | 58.1% [55; 61.2] | 59% [55.8; 62] |
| Reflective | | | | |
|  | Concordant | | Discordant | |
|  | Initial response | Final response | Initial response | Final response |
| Fake | 30.2% [27.5; 33] | 25.3% [22.8; 28] | 25.4% [22.9; 28.1] | 19.3% [17.1; 21.8] |
| Real | 67.5% [64.7; 70.1] | 68.6% [65.8; 71.2] | 54.9% [52; 57.8] | 58.4% [55.5; 61.2] |

**Table S8.** Table shows perceived accuracy averages for initial and final responses and both on neutral and politically relevant headlines when only including trials with which the participant was not familiar. 95% confidence intervals are in brackets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Fake headlines | | Real headlines | |
|  | Initial response | Final response | Initial response | Final response |
| Neutral | 35.8% [33.6, 38.1] | 27.5% [25.5, 29.6] | 65.5% [63.3, 67.6] | 68.5% [66.3, 70.6] |
| Concordant | 32.8% [30.8; 34.8] | 25.1% [23.3; 27] | 63.6% [61.4; 65.8] | 63.3% [61; 65.4] |
| Discordant | 28.1% [26.2; 30] | 21.6% [19.9; 23.4] | 52.4% [51.6; 56] | 53.8% [51.6; 56] |

**Table S9.** Table shows perceived accuracy averages for one response pre-tests (no prior intuitive response), initial (response under pressure and load) and final responses (prior intuitive response) when only including trials with which the participant was not familiar. As in the final response vs one response baseline test, trials with missed initial responses were not excluded from the “prior intuitive response” condition. 95% confidence intervals are in brackets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | No prior intuitive response | Prior intuitive response | Response under pressure and load |
| Neutral | Fake | 25.9% [23.2; 28.9] | 27.8% [25.8; 29.9] | 35.8% [33.6, 38.1] |
| Real | 68.1% [66; 70.2] | 68.2% [70.5; 76.3] | 65.5% [63.3, 67.6] |
| Concordant | Fake | 24.5% [22.2; 27] | 25.2% [23.4; 27.1] | 32.8% [30.8; 34.8] |
| Real | 64.2% [61.3; 67] | 63.1% [60.9; 65.2] | 63.6% [61.4; 65.8] |
| Discordant | Fake | 18.2% [16.1; 20.5] | 22.1% [20.4; 23.9] | 28.1% [26.2; 30] |
| Real | 56.3% [53.4; 59.2] | 53.8% [51.6; 56] | 52.4% [51.6; 56] |

**Table S10.** Table shows perceived accuracy averages as a function of response type, headline’s veracity and political ideology. 95% confidence intervals are in brackets. Positive correlation for fake headlines and negative correlation for real headlines mean that more conservative/Republican participants are more likely to judge the veracity of the headline accurately.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Fake headlines | | Real headlines | |
|  | Initial response | Final response | Initial response | Final response |
| Neutral | 0.13\*\* [0.04; 0.23] | 0.11\* [0.01; 0.2] | -0.02 [-0.12; 0.08] | -0.06 [-0.16; 0.03] |
| Concordant | 0.085\* [0.01; 0.16] | 0.11\*\* [0.03; 0.18] | -0.03 [-0.1; 0.05] | -0.09\* [-0.17; -0.01] |
| Discordant | 0.18\*\*\* [0.1; 0.25] | 0.16\*\*\* [0.09; 0.24] | -0.02 [-0.09; 0.06] | 0.004 [-0.07; 0.08] |

Note. \* p < 0.05, \*\*p < 0.01, \*\*\*p < 0.0001

**References**

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90–109.

Bago, B., & De Neys, W. (2018). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 101–117.

De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology*, *59*(6), 1070–1100.

De Neys, W., Novitskiy, N., Geeraerts, L., Ramautar, J., & Wagemans, J. (2011). Cognitive control and individual differences in economic ultimatum decision-making. *PloS One*, *6*(11), e27107.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621–640.

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50.

Signorell, A. (2016). DescTools: Tools for descriptive statistics. *R Package Version 0.99*, *17*.

Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140.

Trémolière, B., De Neys, W., & Bonnefon, J.-F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*(3), 379–384.