WORKING PAPERS

N° 126

Reasoning about climate change

Bence Bago, David Rand and Gordon Pennycook

IAST
INSTITUTE for ADVANCED STUDY in TOULOUSE

Reasoning about climate change

Bence Bago[1,2]*, David G. Rand[3,4], & Gordon Pennycook[5]

[1]Institute for Advanced Study in Toulouse, University of Toulouse 1 - Capitole
[2]Artificial and Natural Intelligence Toulouse Institute
[3]Sloan School, Massachusetts Institute of Technology
[4]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
[5]Hill/Levene Schools of Business, University of Regina

*Corresponding author: Bence Bago, Institute for Advanced Study in Toulouse, University of Toulouse 1 - Capitole, 1 esplanade de l'Université, 31080 Toulouse, France, e-mail: bencebagok@gmail.com

Abstract

Why is disbelief in anthropogenic climate change common despite broad scientific consensus to the contrary? A widely-held explanation involves politically motivated ("System 2") reasoning: Rather than helping uncover truth, people use their reasoning abilities to protect their partisan identities and reject beliefs that threaten those identities. Despite the popularity of this account, the evidence supporting it (i) does not account for the fact that partisanship is confounded with prior beliefs about the world, and (ii) is entirely correlational with respect to the effect of reasoning. Here, we address these shortcomings by (i) measuring prior beliefs and (ii) experimentally manipulating participants' extent of reasoning using cognitive load and time pressure while they evaluate arguments for or against anthropogenic global warming. The results provide no support for the politically motivated system 2 reasoning account: Engaging in more reasoning led people to have greater coherence between judgments and their prior beliefs about climate change - a process that can be consistent with rational (unbiased) Bayesian reasoning - and did not exacerbate the impact of partisanship once prior beliefs are accounted for. Thus, we challenge the dominant cognitive account of climate disbelief, and suggest that interventions aimed at providing accurate information about climate change may be effective in the long run.

Skepticism about climate change and its human origins represents a major impediment to the adoption of climate change mitigation policies [1–3]. One of the most commonly cited reasons for climate change denial is political partisanship or ideologies [4]. In the US, for example, people on the political right are more likely to believe that climate change is a hoax or that it is not caused by human activities [2,5–8]. What is more, people with greater numerical ability and cognitive sophistication show *increased* partisan differences in climate change beliefs, rather than greater agreement with the scientific consensus [9–13]. That is, having stronger cognitive ability appears to not protect against climate misperceptions; rather it may serve to bolster views that align with one's political identities.

The most popular explanation of this result is provided by the motivated system 2 reasoning framework[11,14–16]. This framework can be interpreted from the point of view of the dual process perspective [17–19], which distinguishes between two types of reasoning process: intuition ("System 1") and deliberation ("System 2"). While intuition is considered a low-effort, quick, automatic response to stimuli, deliberation is a more effortful, time-consuming process. The motivated system 2 reasoning framework asserts that cognitive abilities are linked to greater polarization because deliberation facilitates motivated reasoning: When faced with new evidence, engaging in deliberation better allows one to discredit the evidence if it is not congenial to one's identity and partisan commitments. As a result, there are large partisan differences in what evidence is deemed credible, eventually leading to substantial polarization in beliefs. In the language of dual-process theory, deliberative reasoning processes are triggered to rationalize or justify identity-consistent intuitive impulses. In the context of climate change, this would mean that deliberation leads Republicans to reject evidence in favor of climate change (to protect their partisan identity), while deliberation leads Democrats to reject evidence questioning climate change [10,11,20–23]. If more cognitively sophisticated people engage in more deliberation, they will be better at aligning their judgments of evidence about climate change with their respective political identities.

This theory has enormous practical importance because, if it is true, common strategies such as educating people, or making them more reflective, will not be effective against climate change denial. In fact, such strategies will only serve to *increase* partisan differences [10,23,24] (although there is evidence questioning this assumption; [25–27]. Furthermore, from a theoretical perspective, this "motivated system 2 reasoning" account stands in stark contrast to a common dual-process perspective - the "classical reasoning" view - whereby System 2 reasoning is thought to typically facilitate accuracy in a variety of decision-making tasks [18,28]. Put differently, the classical reasoning account posits that when people engage in deliberation, they tend to form more accurate beliefs, regardless of the partisan- or identity-alignment of the propositions they are deliberating about [29].

However, there are two serious limitations of the prior empirical research in this area. First, this work does not actually allow for the identification of politically motivated reasoning when people are evaluating evidence without making strong assumptions. In particular,

political identity is confounded with - but meaningfully separable from - people's prior beliefs about climate change [30]. In particular, Democrats are much more likely to believe that climate change is caused by human activity than Republicans, which means that partisanship and prior beliefs are correlated. Yet, many Republicans do believe in anthropogenic climate change, and some Democrats do not, meaning that partisanship and priors are meaningfully distinct constructs. For example, a recent Pew survey found that 53% of conservative Republicans believe that human activity contributes to global warming to at least some degree, while 8% of moderate Democrats think it does not [5]. Yet most studies claiming to provide evidence of politically motivated reasoning have not measured these prior beliefs. This is highly problematic [30–32], as convincing evidence of politically motivated reasoning must show that partisanship *per se* influences information evaluation, above and beyond the influence of prior beliefs about the specific topic. Although partisanship might influence prior beliefs, many other factors also contribute to beliefs, such as who people judge trustworthy as well as family environment or life experiences [12]; and prior beliefs may also influence partisanship. Thus effects driven by prior beliefs do not provide direct or compelling evidence for politically motivated reasoning.

Indeed, recent work finds that controlling for prior beliefs related to climate change nullifies the correlation between cognitive sophistication and partisan bias; instead, increased cognitive reflection was associated with placing greater emphasis on prior beliefs when evaluating new information.[30] While evaluating new evidence in light of prior beliefs is sometimes called "confirmation bias", in fact such evaluation can be entirely rational and unbiased from a Bayesian perspective[1] when there is uncertainty about the reliability of sources [33–37]: When considering evidence that is inconsistent with your prior beliefs, it can be rational to conclude that it is more likely that the information source is unreliable than that the accumulation of all your prior knowledge is wrong. For example, if a stranger tells you that he was abducted by aliens, is it not irrational to conclude that that information is probably unreliable. Thus, it is essential to account for prior beliefs when attempting to test for politically motivated reasoning.

Second, past research on motivated system 2 reasoning has relied upon correlating individual differences in cognitive sophistication (e.g., cognitive reflection, numeracy, education) with the extent of partisan differences on politicized issues [9,11,38]. Although it is generally thought that people scoring higher on cognitive sophistication scales are better at deliberation than people scoring lower on these scales, they also tend to differ in many other aspects. For example, they tend to generate different intuitions on many reasoning tasks (i.e., people who are more cognitively sophisticated also have different prior beliefs and knowledge than those

[1] Rationality here and throughout this manuscript refers to the normative nature of the thought process. The outcome of such a process can nevertheless be inaccurate. Imagine a 4 year old who is repeatedly told by his parents and grandparents that Santa Claus is real. If another kid, who is plainly less trustworthy than known adults, states that Santa is not real, it would not be irrational (absent additional exculpatory information) to maintain belief in Santa Claus. Hence, given strong existing prior beliefs, rational processes can arrive at the wrong conclusion.

who score lower; [39,40]. Thus, because this approach is correlational, it does not allow for the direct identification of causal effects of deliberation on polarization versus accuracy.

Current research

In the current research, we address both of these limitations. First, we provide a causal test of the role of intuition and deliberation on how people evaluate pro- and contra- arguments regarding climate change by forcing some participants to make judgments under cognitive load and time pressure. Second, we measure prior beliefs about climate change by asking how serious a risk participants believe climate change to be, and how much they agree that human activity causes climate change.

This paradigm allows us to shed new light on competing accounts of the role of deliberation in argument evaluation surrounding climate change: Does deliberation magnify partisan bias, consistent with the motivated system 2 reasoning framework [11]? Or does it facilitate accurate assessments or normatively rational evidence evaluation, consistent with a more classical perspective on reasoning [29,30,36]? Furthermore, we specify the classical reasoning account more precisely than in prior work. Previous research (e.g. studying blatantly false political news posts[29]) has argued that the classical reasoning approach simply predicts that more deliberation will lead to increased objective accuracy, defined here as holding a position more consistent with the scientific consensus on climate change. However, most people do not actually have direct access to the information needed to know the objectively accurate answer, particularly in the context of extremely complicated technical issues like climate change. Thus, the classical reasoning account would not necessarily predict that deliberation leads to more objectively accurate views. Instead, accuracy-motivated deliberation may lead to improved coherence between one's existing directly relevant beliefs and the stimuli being presented. That is, deliberation may increase the extent to which one evaluates whether new information makes sense in light of the relevant beliefs/knowledge that one has developed based on previous information one has encountered (a process that, as discussed above, can be consistent with unbiased, rational Bayesian updating [33–36]). In this case, deliberation should magnify differences based on prior beliefs - but *not* differences based on political identities *per se*. This is an essential distinction, because although partisanship and prior beliefs are correlated, they are still quite distinct. As reported in more detail below, for example, in our dataset a majority of Republicans do believe in anthropogenic climate change, and some Democrats do not. Thus, it is essential to disentangle the impact of prior beliefs from partisanship [32].

In our experiments, we asked participants to indicate how much they agreed with politically neutral arguments about climate change (meaning that there were no references in them to specific policies or to politics in any way). Arguments were content counter-balanced, such that for each statement we created a pro and contra version, one of which was randomly assigned to a given participant; participants never saw both the pro and contra version of the

same argument. Table 1 shows a pro and contra version of an example item from our experiment (for a complete set of statements, see Supplementary Materials Section A).

Table 1. Table shows an example of the argument items we used in this experiment.

Pro climate change	Contrary to climate change
Average temperatures on earth have increased at a rate far faster than can be explained by natural climate changes. A 2008 study compared data from tree rings, ice cores, and corals over the past millennium with recent temperature records. The study created the famous "hockey stick" graph, showing that the rise in earth's temperature over the preceding decade had occurred at a rate faster than any warming period over the last 1,700 years. In 2012 the Berkeley scientists found that the average temperature of the earth's land increased 2.5°F over 250 years (1750-2000), with 1.5°F of that increase in the last 50 years. Lead researcher Richard A. Muller, PhD, said "it appears likely that essentially all of this increase [in temperature] results from the human emission of greenhouse gases." In 2013, a surface temperature study published in Science found that global warming over the past 100 years has proceeded at a rate faster than at any time in the past 11,300 years. According to the IPCC's 2014 Synthesis Report, human actions are "extremely likely" (95-100% confidence) to have been the main cause of 20th century global warming, and the surface temperature warming since the 1950s is "unprecedented over decades to millennia."	Earth's climate has always warmed and cooled, and the 20th century rise in global temperature is within the bounds of natural temperature fluctuations over the past 3,000 years. Although the planet has warmed 1-1.4°F over the 20th century, it is within the +/- 5°F range of the past 3,000 years. A 2003 study by researchers at the Harvard-Smithsonian Center for Astrophysics found that "many records reveal that the 20th century is probably not the warmest nor a uniquely extreme climatic period of the last millennium." A 2005 study published in Nature found that "high temperatures - similar to those observed in the twentieth century before 1990 - occurred around AD 1000 to 1100" in the Northern Hemisphere. A 2013 study published in Boreas found that summer temperatures during the Roman Empire and Medieval periods were "consistently higher" than temperatures during the 20th century. According to a 2010 study in the Chinese Science Bulletin, the recent global warming period of the 20th century is the result of a natural 21-year temperature oscillation, and will give way to a "new cool period in the 2030s."

Since this is a novel way to measure reasoning about climate change, we started with a pilot study that conceptually replicated prior correlational results regarding cognitive sophistication (the extent to which people are able and willing to engage in deliberation, as measured by the Cognitive Reflection Test) and belief in climate change (see SM Section B). Consistent with prior work [30], we found that (i) cognitive sophistication is associated with politically polarized evaluations of our climate change arguments, but (ii) accounting for differences in prior beliefs about climate change nullifies this interaction between cognitive sophistication and political partisanship, and instead more cognitive sophisticated participants actually condition their judgments on coherence with prior beliefs rather than partisanship per se.

In our main experiments, we then recruited American participants from Lucid (quota matched to the national distribution on age, gender, ethnicity and geographic region) and experimentally induced intuitive responding for randomly selected participants to provide a *causal* test of the role of intuition versus deliberation on climate argument evaluation. Deliberation is highly dependent on available working memory capacities, and tends to take more time (compared to intuitive responses;[17,41]. Hence, to minimize the extent of deliberation, we applied a working memory load and a strict response deadline (28 seconds)[2]

[2] To define the response deadline, we conducted a reading pre-test experiment, in which we simply asked participants to read the material. The response deadline is the average of the reading times in this pre-test.

to participants in the "intuitive response" condition. Conversely, participants in the "deliberative response" condition were presented with the arguments without any constraint (and hence were free to deliberate). For details, see Methods (including discussion of a second response that was collected in the intuitive condition where participants could make a subsequent deliberative choice).

According to the motivated system 2 reasoning account, when people are able to engage in more reasoning in the deliberative response condition, they should show increased agreement with arguments that are concordant to their partisanship. This yields four specific predictions:

- MS2R 1) Deliberation will increase agreement with con arguments for Republicans, even when controlling for prior beliefs
- MS2R 2) Deliberation will decrease agreement with pro arguments for Republicans, even when controlling for prior beliefs
- MS2R 3) Deliberation will decrease agreement with con arguments for Democrats, even when controlling for prior beliefs
- MS2R 4) Deliberation will increase agreement with pro arguments for Democrats, even when controlling for prior beliefs

If deliberation simply facilitates accurate beliefs (defined in this case as beliefs that are more consistent with the scientific consensus), people will increase agreement with pro arguments and decrease agreement with contra arguments after deliberation, regardless of their partisanship or prior beliefs.

Finally, if deliberation facilitates the coherence between one's prior beliefs and evaluation of new information, we should find that deliberation increases agreement with arguments that are consistent with the participant's prior beliefs about climate change while decreasing agreement with arguments that are inconsistent with the participant's prior beliefs (as discussed in Footnote 1, bear in mind that a rational, Bayesian process does not necessarily improve objective accuracy). This yields four specific predictions:

- BC 1) Deliberation will increase agreement with con arguments for climate deniers, even when controlling for partisanship
- BC 2) Deliberation will decrease agreement with pro arguments for climate deniers, even when controlling for partisanship
- BC 3) Deliberation will decrease agreement with con arguments for climate believers, even when controlling for partisanship
- BC 4) Deliberation will increase agreement with pro arguments for climate believers, even when controlling for partisanship

To test these various predictions we used linear mixed effect models with crossed random effects for participant and item, including the most complex random effect terms that converged. For the specification of random effects used in the various models, see Methods

Section Table 2. Our models take agreement ratings as the dependent variable, and as independent variables we include prior belief in climate change, partisanship, argument type (pro or contra) and response type (intuitive or deliberative response condition). The motivated system 2 reasoning account predicts a three-way interaction between partisanship, response type and argument type, while the specific prior beliefs account predicts a three-way interaction between specific prior beliefs, argument type and response type. If deliberation facilitates accurate beliefs per se, we should find a two-way interaction between response type and argument type (in a model without partisanship or prior beliefs).

We conducted two studies that allow us to discern the causal effect of deliberation on climate change beliefs. These studies were equivalent except in two aspects. In Study 1, participants received both partisanship and prior belief questions at the end, after all the climate arguments items, while in Study 2, they received these questions prior to experimental manipulation. Moreover, in Study 1 prior belief was assessed using a single question assessing belief that human activity causes climate change, while in Study 2 prior belief was defined using an average of that question and a second question assessing the risk that climate change poses to humanity.

Results

Separability of partisanship and prior beliefs about climate change

We begin by documenting a motivation for our work - the disconnect between partisanship and prior beliefs. Although there is a moderate correlation between prior beliefs about climate change and partisanship, $r = 0.34$, $p < 0.0001$ (Study 1), $r = 0.44$, $p < 0.0001$ (Study 2), it is clear that these measures are meaningfully separable. As shown in Figure 1, a majority of participants from both parties believed in human-caused climate change. Thus, being Republican does not necessarily mean that one is a climate denier.

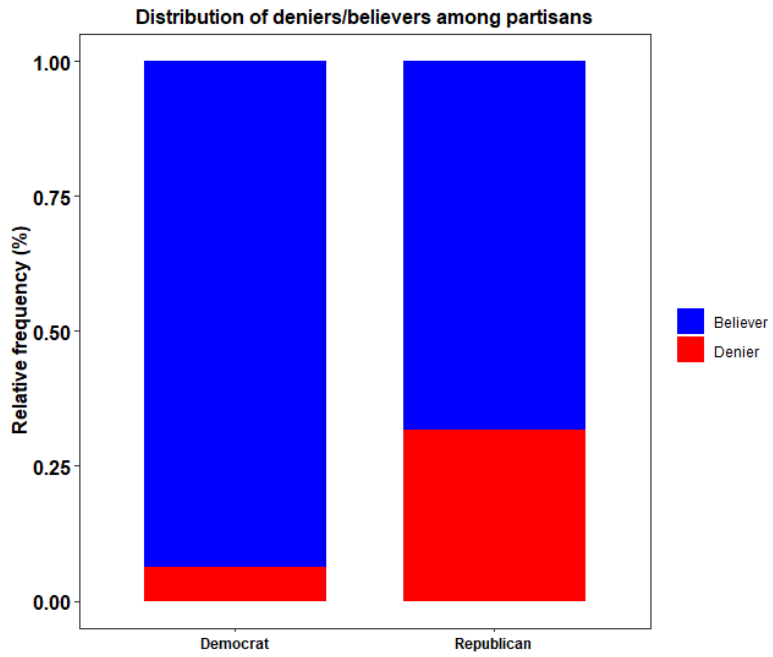

Figure 1. *Distribution of specific prior belief (i.e. people who believe vs who deny climate change and its consequences) plotted as a function of partisanship in our pooled studies. Although the two measures are moderately correlated, they are meaningfully distinct. While all statistical analysis are conducted using a continuous belief scale, for visualization we dichotomize this measure: those who scored 0 or below were assigned to the "Denier" category and those above 0 were assigned to the "Believer" group.*

Main experiments

We now turn to our key question of interest, and examine the difference in agreement with each statement between participants in the intuitive response condition and the deliberative response condition. This difference in agreement (i.e. the effect of deliberation) for each argument type is shown in Figures 2-3, split by partisanship and prior belief.

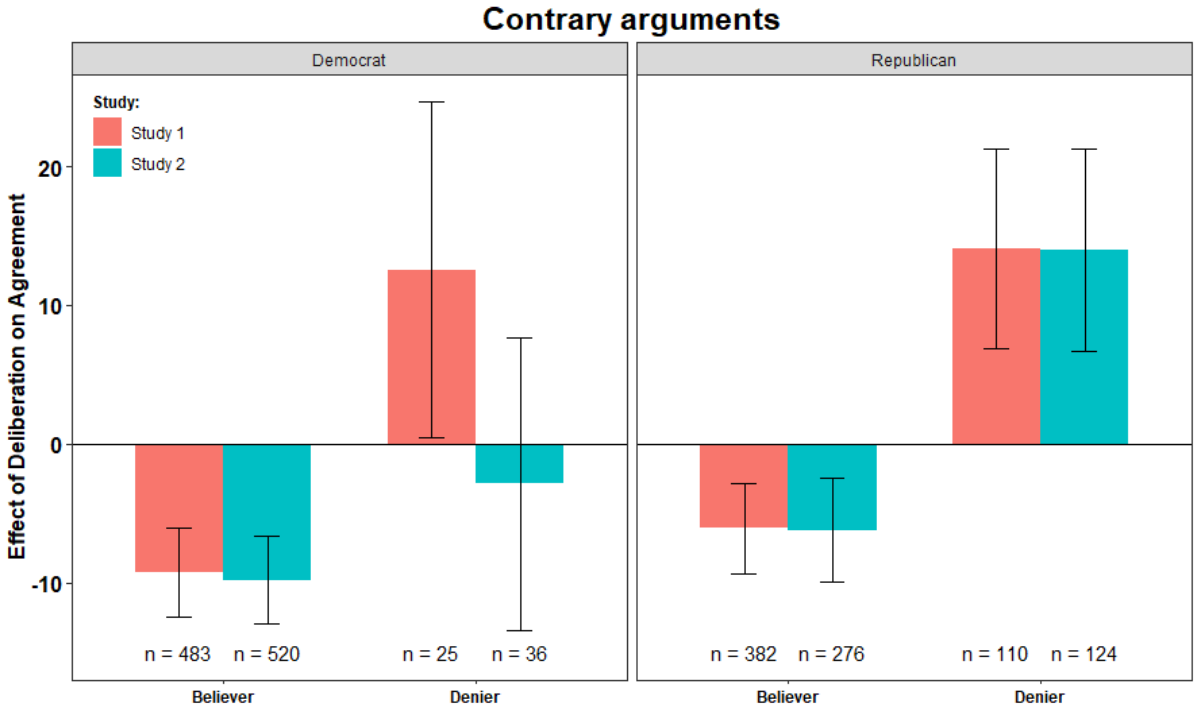

Figure 2. *Difference in agreement between intuitive and deliberative ratings as a function of prior belief (believer/denier) and politics (Democrats/Republicans) on CONTRARY arguments. Bars are mean differences between the deliberative and the intuitive response. Error bars are 95% confidence intervals. Negative values indicate a decrease, and positive values indicate an increase in agreement scores after deliberation.*

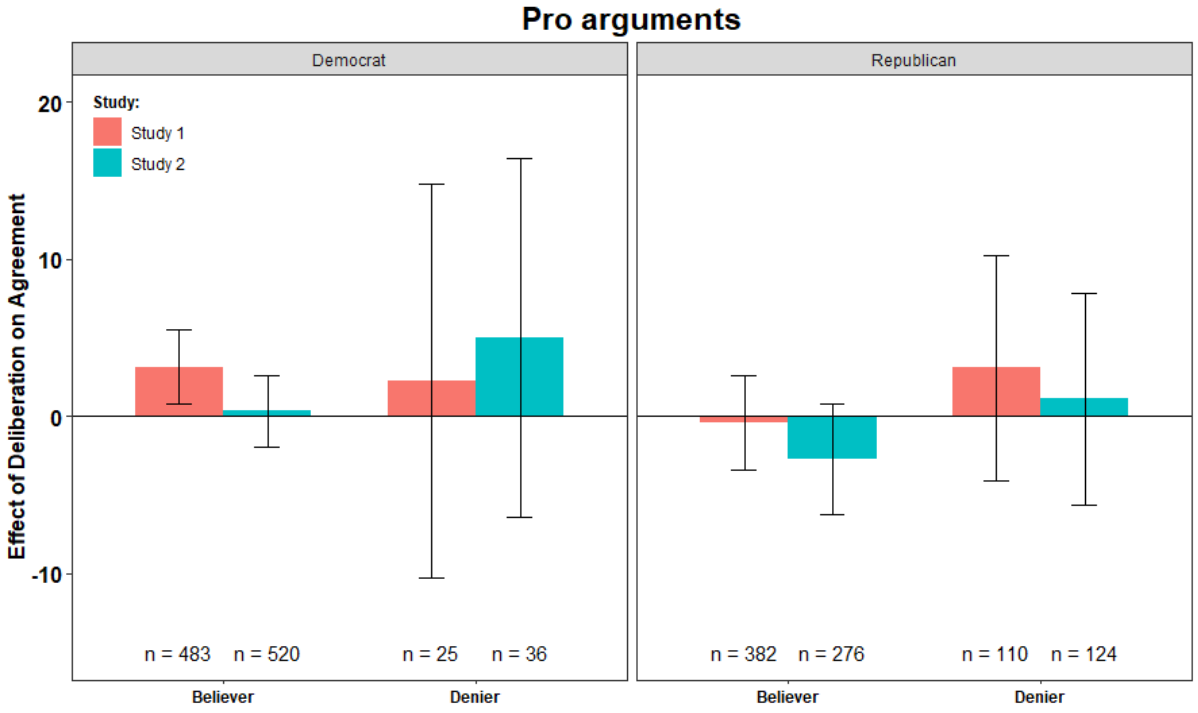

Figure 3. *Difference in agreement between intuitive and deliberative ratings as a function of prior belief (believer/denier) and politics (Democrats/Republicans) on PRO arguments. Bars are mean differences between the deliberative and the intuitive response. Error bars are 95% confidence intervals. Negative values indicate a decrease, and positive values indicate an increase in agreement scores after deliberation.*

We see similar results across both studies.

First, we ran a model including argument type, response condition, and their interaction (without including partisanship and prior belief in the model). Consistent with the proposal that deliberation simply facilitates accurate beliefs, we found a significant interaction between argument type and condition, such that that for the average participant deliberation increases the accuracy (Study 1: b = -5.63, p = 0.001, Study 2: b = -5.4, p = 0.004)[3].

We then turn to the full model including partisanship, prior beliefs, and relevant interactions. Contrary to the motivated system 2 reasoning account, there was no significant three-way interaction between partisanship, argument type and condition (Study 1: $b = 1.09$, $p = 0.77$; Study 2: $b = -1.27$, $p = 0.79$). Consistent with the belief-coherence deliberation account, however - and qualifying the two-way interaction reported above in the model without prior beliefs - we find a significant three-way interaction between prior beliefs, argument type and condition (Study 1: $b = -16.23$, $p < 0.0001$; Study 2: $b = -14.12$, $p = 0.0007$).

To further understand the key three-way interaction between prior belief, condition, and argument type, we compared intuitive and deliberative responses separately for believers versus deniers (above versus below 0 on prior beliefs) on each argument type. We found that on the contrary arguments, “believers” were more likely to decrease agreement after deliberation, (Study 1: b = -8.1, p < 0.0001; Study 2: b = -8.52, p < 0.0001), while “deniers” were more likely to increase agreement after deliberation, (Study 1: b = 13.23, p = 0.008; Study 2: b = 10.1, p = 0.03). On the pro arguments, neither believers (Study 1: b = 1.52, p = 0.275; Study 2: b = 0.07, p = 0.96) nor deniers (Study 1: b =3.14, p = 0.53; b = 0.32, p = 0.95) significantly changed their agreement after deliberation. Further subsetting on Republicans, we found that deliberation significantly increased agreement with contra arguments among Republicans who are deniers (Study 1: b = 14.34, p = 0.014, Study 2: b = 13.84, p = 0.009), but significantly decreased agreement with contra arguments among Republicans who are believers (Study 1: b = -6.18, p = 0.009, Study 2: b = -5.9, p = 0.03). This clearly demonstrates that deliberation does not increase agreement for Republicans per se, and does not increase agreement with the scientific consensus per se, but rather that the deliberation effect increases coherence with prior beliefs.

Finally, to avoid any potential selection bias arising from non-compliance with the time pressure manipulation [42], we replicated our main analyses while replacing any intuitive responses where the participant did not respond within the time limit (such that no response

[3] Standardized effects can be found in the Supplementary Materials

was recorded, and thus were not included in the main analyses above) with their corresponding final response - that is, assuming that missed responses correspond to no effect of deliberation. The results were qualitatively equivalent (significant two-way interaction between argument type and condition, Study 1: $b = -5.59$, $p = 0.001$, Study 2: $b = -4.99$, $p = 0.007$; after adding partisanship and beliefs in the models, we found non-significant three-way interaction between partisanship, argument type and condition, Study 1: $b = 0.64$, $p = 0.86$, Study 2: $b = -0.9$, $p = 0.84$; significant three-way interaction between prior belief, argument type and condition, Study 1: $b = -15.72$, $p < 0.0001$, Study 2: $b = -14.49$, $p = 0.0003$).

We thus found no evidence for motivated reasoning (none of the predictions of the MS2R account were supported). We did find support for the belief coherence account's predictions regarding the contra arguments, but - inconsistent with any of the accounts - we found no significant effects at all of deliberation on pro arguments. One potential explanation of this difference between argument types is that most participants may have already been exposed to the pro-climate change arguments - and, therefore, have stronger, and more quickly available, prior beliefs about those arguments, such that they would not need additional time to deliberate when evaluating them. Conversely, the contra-climate arguments were likely to have been more novel, such that evaluating them required more deliberation. In other words, prior exposure to pro arguments might have 1) prevented deliberation to occur because people had their intuitive judgements aligned with their prior beliefs (so they had no reason to deliberate) or 2) caused a floor/ceiling effect in the intuitive condition which prevented any detectable deliberation effect to arise. This is well illustrated by the fact that while for contra arguments, agreement scores are higher for believers than for deniers in the intuitive response condition (i.e., there is a misalignment between intuitive judgements and prior beliefs), for pro arguments, agreement scores are generally much lower for deniers (i.e., there is no misalignment between intuitive judgments and prior beliefs; see SM Figures S4-S7 for raw averages and response distributions). To further evaluate the possibility of prior exposure effect on argument evaluation, we conducted an additional experiment (Lucid sample, $N = 4733$) in which we measured people's familiarity with each argument (on a scale from 0 - absolutely not familiar, to 100 - very familiar). Consistent with the above familiarity account, people were more familiar with pro (Mean familiarity = 53.4) than contra items (Mean familiarity = 41.9), $t(4720.4) = -12.5$, $p < 0.0001$. Even more importantly, we found a strong argument-level correlation between an argument's average familiarity and the coefficient on the interaction between condition and prior belief for that specific item ($r = -0.64$, $p = 0.002$; Figure 4) . This shows that deliberation magnified the effect of prior beliefs to a greater extent for more unfamiliar arguments - and thus that differences in familiarity can explain the asymmetry we observed in our main studies between pro-climate change and contra-climate change arguments.

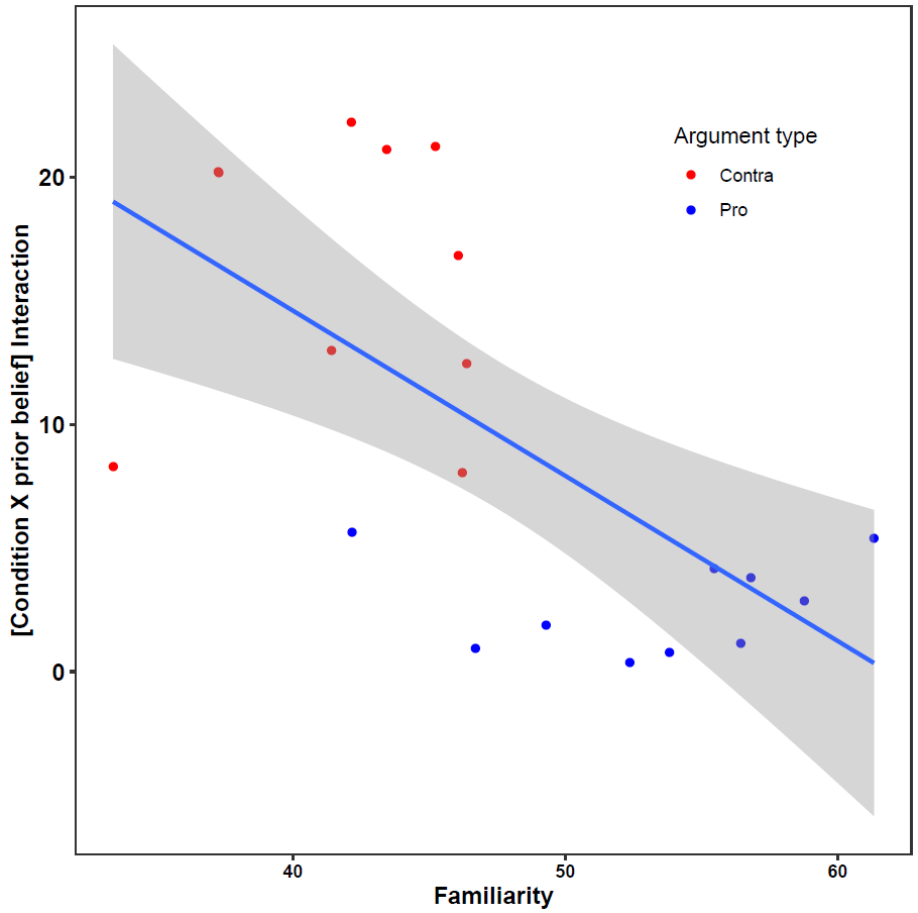

Figure 4. Deliberation magnifies the effect of prior beliefs to a greater extent for less familiar arguments. Shown is one dot per argument, with average out-of-sample familiarity rating on the x axis, and the coefficient on the interaction between condition and prior belief (indicating the extent to which deliberation magnifies the effect of prior beliefs) on the y axis. Contra-climate change arguments are shown in red, pro-climate change arguments are shown in blue.

In sum, our results suggest that deliberation moves agreement with anti-climate change arguments toward participants' pre-existing prior beliefs about climate change[4], supporting the belief-coherence account of reasoning, while we find no direct evidence for the classical reasoning or the motivated system 2 reasoning accounts.

Discussion

[4] Graphs on the means and distribution of responses for each response condition, party affiliation, prior belief and argument type can be found in Supplementary Materials, section D.

Here we investigated the role of reasoning in climate change (dis)belief, addressing two major limitations of prior work: We accounted for prior beliefs (which we demonstrate are correlated, but dissociable, from partisanship per se), and we experimentally manipulated extent of reasoning rather than relying on correlations with individual differences in reasoning. Our results showed that once prior beliefs were accounted for, an apparent causal effect of deliberation increasing reliance on partisan alignment disappeared. Instead, deliberation increased coherence between evaluation of novel climate arguments and one's pre-existing beliefs about climate change (regardless of one's partisanship).

These results show that patterns typically taken as evidence for the MS2R account - e.g. deliberation decreasing anti-climate agreement among Democrats but not Republicans - do not actually provide any positive evidence in favor of MS2R per se. Such patterns could be entirely explained by (even incidental) variation in people's prior beliefs. Hence, our results constitute a major challenge for the MS2R account. It is important to keep in mind, however, that partisanship is correlated with prior beliefs about climate - and furthermore it seems likely that different Republicans have different identity commitments (e.g., only climate deniers may consider climate denial to be crucial to Republican identity). Thus, our results do not unambiguously provide support for the belief coherence account over the MS2R account. However, we argue that in the absence of positive evidence for MS2R, it seems more parsimonious to conclude that reasoning about climate change is guided by coherence with prior beliefs, as prior work shows that coherence with priors guides reasoning in situations completely unrelated to partisan identities [35,43–46]. Future work should approach this issue in the same experimental fashion that we approached deliberation in the current work, namely by experimentally disentangling the effects of prior beliefs and partisan motivations.

Our experiment also does not allow us to identify what caused the differential priors that climate believers versus deniers brought into our study. It seems likely that exposure to differential information streams (e.g. right- versus left-leaning news channels) is a major contributor. It is also possible that political motivations influenced belief formation or updating at an earlier stage (that is, that prior beliefs mediate the effect of partisanship) - although the fact that a large fraction of Republicans are not climate deniers poses a challenge for this account. Note that besides differences in knowledge, there are other factors that could influence individual differences in prior beliefs, such as elite cues, extreme weather events or movements/countermovements [47]. It is also possible that potential cognitive asymmetries (e.g., differences in epistemic motivations) behind ideologies also play a role in prior belief formation [48]. Shedding light on how partisan (or ideological) differences in priors emerge is an extremely important direction for future work - but regardless of how they may have arisen, our results show that engaging in reasoning does not in and of itself lead to an increased direct effect of partisan alignment when people evaluate arguments relating to climate change.

While the behavior we observe may be rational, our data also illustrate that being "rational" (Bayesian) does not mean that deliberation will necessarily lead to greater belief in climate

change. Indeed, deliberation only leads to decreased belief in (incorrect) contra-climate change arguments among people who already believe that climate change is real - whereas deliberation drives deniers further away from accuracy. Our results therefore suggest that simply getting people to think more carefully will not, in and of itself, make people more likely to believe in the threat of climate change. But our results also suggest that things are not as bleak as suggested by the motivated system 2 reasoning account. If people are engaging in a good-faith effort to assess new information accurately, and relying on their prior beliefs to guide such judgments, then educational interventions could - in the long run - move people's priors and increase agreement with the scientific consensus. Accordingly, informing people about climate change has been shown to be an effective technique [25,26,37,49,50] to increase knowledge about issues surrounding climate change - that is, to shift people's prior beliefs regardless of how they judge a particular argument or piece of evidence at any particular time. Our results suggest that once this knowledge is obtained, inducing deliberation could help people resist climate misinformation and reduce their likelihood of engaging in science denial.

There are other limitations of our work that it is important to acknowledge. First, although Lucid uses quota matching on age, gender, ethnicity, and geographic region, our sample should not be considered fully representative of the US population. It should be noted, however, that our sample *is* ideologically balanced - there was an almost identical number of Republicans and Democrats, and our estimate of prior belief in anthropogenic climate change is reasonably close to the national average: According to US-wide nationally representative surveys [5], 20% of americans believe that human activity contributes no or very little to global warming (i.e., the people who would constitute as "deniers" in our sample), while we found 15.1%. Relatedly, we only recruited participants from the US, while climate change is obviously a global issue. Thus, it is important for future work to examine how our results generalize to other countries with differing political cultures. Another limitation involves the abstracted context (a survey experiment) in which we presented the pro- and contra-climate arguments. It would be fruitful for future work to examine how richer social contexts influence reasoning, such as the messenger who delivers the argument. Finally, we only examined prior beliefs about climate change. Future work should also examine other relevant beliefs, such as beliefs about the reliability of science and the trustworthiness of scientists.

In sum, the data presented here challenge the dominant motivated system 2 reasoning account of climate misbelief. Rather than being blinded by partisanship, our results suggest that climate deniers may in fact be making a good-faith effort to form accurate beliefs. The blame for climate denial would then rest on the producers and distributors of climate misinformation. This emphasizes the key role of effective climate communication - disseminating accurate information about climate change in an engaging manner is essential for winning the battle against climate denial.

Methods

Participants

Study 1. In total, 1007 participants were recruited from Lucid. In the two response paradigm, 629 participants (315 females, 308 males and 6 others, Mean age = 45.34 years, SD = 16.9 years) took part, while in the one response study 378 people (197 females, 180 males and 1 others, Mean age = 45.2 years, SD = 16.6 years) participated. In the two response condition, 314 people were Democrats, and 314 were Republicans, and in the one response test, 194 were Democrats and 179 were Republicans.

Study 2. In total, 1266 participants were recruited from Lucid who passed our first screening question (from a total of 1489 participants who clicked on the experiment link, 12 did not consent, and an additional 211 did not pass the first screening question and were not allowed to participate; these participants produced no data whatsoever). In the two response paradigm, 758 participants (387 females, 341 males and 8 others, Mean age = 45 years, SD = 17.4 years) took part, while in the one response study 508 people (251 females, 249 males and 1 others, Mean age = 45.7 years, SD = 17.2 years) participated. In the two response condition, 407 people were Democrats, and 318 were Republicans, and in the one response test, 306 were Democrats and 192 were Republicans.

This research was approved by The MIT Committee on the Use of Human Experimental Subjects. Before the experiment, we obtained informed consent from all participants.

Pre-registration for Study 2 can be found at: https://osf.io/4axr6/

Materials and procedures

Climate arguments. Altogether, participants were presented with 6 arguments in a randomised order. All the arguments we used in this study were taken from *procon.org* and can be seen in the Supplementary Material section A. Half of these arguments were supporting arguments for anthropogenic climate change. These were scientific arguments, mostly explaining how climate change affects our environment, or how human activity causes climate change, all related to climate change, while the other half of the arguments argued against that humans causing climate change or that it would have bad consequences. All of these arguments were presented in a non-partisan manner, there were no party cues, or politicians mentioned whatsoever. Arguments were content-counterbalanced, and participants never saw the pro and contra version of a given content. Participants were presented with only one of the two versions of one argument, but were always presented with 3 pro and 3 contra arguments. After each argument, we asked people "How much do you agree with this argument? (0=completely disagree, 100=completely agree)". Participants could give a response on a slider. To make sure that the position of the handler on the slide does not bias participants, the handler was not shown until participants first clicked on the slide.

Dot matrix task. 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 [17], 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., [51] we used a dot matrix task [52] which has been shown to decrease analytical engagement in many tasks including probabilistic reasoning [53] and moral reasoning [54]. In this task, before the argument 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 feedback as to whether 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 and reading pre-test. To further assure the intuitive nature of the initial response, participants had to indicate their answer under a strict response deadline [51,55,56]. 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 = 101 participants (35 females, 66 males, Mean age = 35.24 years, SD = 12 years) on MTurk. Participants were presented with the same arguments 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 27.3 seconds (SD = 34.02) to read the headlines [after excluding all trials three times the standard deviation above or below the mean]. Hence, we decided to set the response deadline to the closest integer of the average reading time: 28 seconds.

Partisanship and prior beliefs. We measured participants' political partisanship with a single question: "Which of the following best describes your political preference?" (1 - Strongly Democratic, 6 - Strongly Republican). To measure specific prior beliefs about climate change, in Study 1, we asked the following question: "How much do you agree with the idea that human activity causes global changes in climate? (Completely agree, Somewhat agree, Somewhat disagree, Completely disagree)". In Study 2, additionally, we also asked participants: "How significant is the risk climate change poses on humanity?[5] (Very significant, Somewhat significant, somewhat insignificant, Very insignificant)". In Study 2, we calculated the average of these two measures for every participant to create a "prior belief" score which we will later apply in the analysis. These questions reflect the position our climate arguments argue for or against the best, hence they constitute as the best way to measure specific prior belief in this specific context. We added the "risk" question in Study 2 precisely to make our belief scale more precise and better reflect the positions of the

[5] In the actual experiment, there was a grammatical error in this question: "How significant is the risk climate change pose on humanity?"

arguments. In Study 1, these questions were asked after climate change arguments, at the end of the experiment, while in Study 2, right at the beginning of the experiment, right before the instructions.

Two response paradigm procedure. Before the instructions, participants in Study 2 were presented two screening questions to screen careless respondents (Study 1 did not include screeners). Participants in the treatment were then told that they are going to be presented with arguments twice. The literal instructions were:

"Welcome to the experiment!

Please read these instructions carefully!

This experiment is composed of 6 questions and a couple of practice questions. It will take about 15 minutes to complete and it demands your full attention. You can only do this experiment once.

In this experiment, you will be presented with different arguments regarding climate change and its potential effects. You will be asked you to indicate how much you agree with the argument on a scale from 0 = completely disagree to 100 = completely agree. 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. 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 argument 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 28 seconds to give a response. 3 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.

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 a practice problems. After this, they were presented with a 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 a new practice problem, 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 argument was presented and participants had 28 seconds to give a response; 3 seconds before the deadline, the background turned yellow, to warn participants of the approaching deadline. 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 response 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 argument again and were asked to give a final response.

Next, participants were asked to complete a six item version of the Cognitive Reflection Test: first, they were presented with content-modified versions [29] of the original three CRT items [57] after which they were presented with a newer three-item non-numerical CRT [58].

In the two response paradigm, there is the potential for the second response to be anchored on the first response [55]. That is, the effect of deliberation may be under-estimated due to having explicitly stated one's intuitive response initially (and then being anchored on that response). Therefore, we focus our main text discussion on comparing the first (intuitive) response with the one response control condition (described below). We report results for the second response in SI Section C, where indeed we do find evidence of anchoring (comparing second response to control), but nonetheless generally replicate our main results when testing the effect of deliberation by comparing the initial response to the final response.

One response paradigm procedure. In this (control) condition, participants were presented with each argument only once. These are the literal instructions participants received at the beginning of this experiment:

"**Welcome to this experiment!**

In this experiment, you will be presented with different arguments regarding climate change and its potential effects. You will be asked you to indicate how much you agree with the argument on a scale from 0 = completely disagree to 100 = completely agree. Then, you will be presented with the same problem again and we will ask you three other questions. Please carefully read through the arguments before responding.

You will be presented with 6 arguments. The experiment will take about 15 minutes to complete, and will demand your full attention.

Press 'Next' to continue!"

In each trial, after a brief fixation cross period (1 sec) the argument appeared and they were asked to give a response without any (load or deadline) constraints. In short, participants in this condition were simply given the "final response" portion of the 2-response paradigm. At the end of this experiment, participants were presented with the same CRT, science knowledge and demographic questionnaires as in the two response version.

Exclusion. For Study 1, we only excluded trials in which the initial trial was missed, and hence, 6.1% of trials were excluded. We first excluded participants who did not pass our second screening question (note that people who did not pass the first screening question, 211

participants altogether, were not allowed to participate in the experiment); from the two response experiment we excluded 164 people and from the one response version we excluded 125 participants. As stated in the pre-registration, we excluded all trials in which participants did not manage to give an initial response in time (hence, there was no data recorded). Only exception is when we compared baseline and final responses in which analysis we included all the data. Some participants also did not finish the experiment, and quit after giving an initial response. These data were also excluded from the initial vs final analysis but not from the rest. Altogether, 11.6% of trials were excluded from analysis for missing data.

Statistical analysis. The maximal model converged in none of the cases. In all models, we included argument condition (0.5 - contra, -0.5 - pro) and condition (initial, final, baseline). As stated in the pre-registration, belief was coded as a numeric variable from -1 (strong disbelief) to 1 (strong belief) and partisanship similarly as a numeric, from -1 (strongly Democratic) to 1 (strongly Republican). We compared conditions to each other, always including only two of them in the models (e.g., initial vs final, etc). Table 2 shows all the fixed and random effects that were included in the separate models. In all models, the intercept was allowed to vary over subjects and item contents.

Table 2. Table shows all the fixed and random effects that were included in the models in both studies. In all models, the intercept was allowed to vary over subjects and item contents.

	Fixed	Random slope over Subjects	Random slope over item contents
Study 1	Argument type, Condition (0 - intuitive, 1 - deliberative), Partisanship, Belief	Argument type,	Belief
	Argument type, Condition	Argument type	-
Study 2	Argument type, Condition (1 - intuitive, 1 - deliberative), Partisanship, Belief	Argument type	Argument type
	Argument type, Condition	Argument type	Argument type

Acknowledgements

All data and pre-registration are available at the OSF page of the project: osf.io/6fcre. BB was supported by ANR grant ANR-17-EURE-0010 (Investissementts d'Avenir program), and ANR Labex IAST. DGR and GP gratefully acknowledge funding from the William and Flora Hewlett Foundation, the John Templeton Foundation, and the Reset project of Omidyar Group's Luminate Project Limited. We are also deeply endebted to Antonio Arechar for assistance in running our online experiments.

Competing Interests Statement

We declare no competing interests.

Data Availability Statement

Data is made open and available at: https://osf.io/6fcre/

References

1. Ding, D., Maibach, E. W., Zhao, X., Roser-Renouf, C. & Leiserowitz, A. Support for climate policy and societal action are linked to perceptions about scientific agreement. *Nat. Clim. Change* **1**, 462–466 (2011).
2. McCright, A. M., Dunlap, R. E. & Xiao, C. Perceived scientific agreement and support for government action on climate change in the USA. *Clim. Change* **119**, 511–518 (2013).
3. Pidgeon, N. & Fischhoff, B. The role of social and decision sciences in communicating uncertain climate risks. *Nat. Clim. Change* **1**, 35–41 (2011).
4. Hornsey, M. J., Harris, E. A., Bain, P. G. & Fielding, K. S. Meta-analyses of the determinants and outcomes of belief in climate change. *Nat. Clim. Change* **6**, 622–626 (2016).
5. Funk, C. & Hefferon, M. *US Public Views on Climate and Energy: Democrats Mostly Agree the Federal Government Should Do More on Climate, While Republicans Differ by Ideology, Age and Gender*. (Pew Research Center, 2019).
6. Funk, C. & Kennedy, B. How Americans see climate change and the environment in 7 charts. *Pew Res. Cent. Fact Tank* (2020).
7. McCright, A. M. & Dunlap, R. E. Cool dudes: The denial of climate change among conservative white males in the United States. *Glob. Environ. Change* **21**, 1163–1172 (2011).
8. McCright, A. M. & Dunlap, R. E. The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. *Sociol. Q.* **52**, 155–194 (2011).
9. Drummond, C. & Fischhoff, B. Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proc. Natl. Acad. Sci.* **114**, 9587–9592 (2017).
10. Hamilton, L. C. Education, politics and opinions about climate change evidence for

interaction effects. *Clim. Change* **104**, 231–242 (2011).

11. Kahan, D. M. *et al.* The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nat. Clim. Change* **2**, 732–735 (2012).
12. Malka, A., Krosnick, J. A. & Langer, G. The association of knowledge with concern about global warming: Trusted information sources shape public thinking. *Risk Anal. Int. J.* **29**, 633–647 (2009).
13. McPhetres, J., Bago, B. & Pennycook, G. Science beliefs, political ideology, and cognitive sophistication. (2020).
14. Kahan, D. M. Ideology, motivated reasoning, and cognitive reflection. *Judgm. Decis. Mak.* **8**, 407–424 (2013).
15. Kahan, D. M. 'Ordinary science intelligence': A science-comprehension measure for study of risk and science communication, with notes on evolution and climate change. *J. Risk Res.* **20**, 995–1016 (2017).
16. Kahan, D. M. & Corbin, J. C. A note on the perverse effects of actively open-minded thinking on climate-change polarization. *Res. Polit.* **3**, 2053168016676705 (2016).
17. Evans, J. S. B. & Stanovich, K. E. Dual-process theories of higher cognition advancing the debate. *Perspect. Psychol. Sci.* **8**, 223–241 (2013).
18. Kahneman, D. *Thinking, fast and slow*. (Farrar, Straus and Giroux, 2011).
19. Pennycook, G., Fugelsang, J. A. & Koehler, D. J. What makes us think? A three-stage dual-process model of analytic engagement. *Cognit. Psychol.* **80**, 34–72 (2015).
20. Bolsen, T., Druckman, J. N. & Cook, F. L. Citizens', scientists', and policy advisors' beliefs about global warming. *Ann. Am. Acad. Pol. Soc. Sci.* **658**, 271–295 (2015).
21. Hamilton, L. C., Cutler, M. J. & Schaefer, A. Public knowledge and concern about polar-region warming. *Polar Geogr.* **35**, 155–168 (2012).
22. Kahan, D. M. Climate‐science communication and the measurement problem. *Polit. Psychol.* **36**, 1–43 (2015).

23. Kahan, D. M., Landrum, A., Carpenter, K., Helft, L. & Hall Jamieson, K. Science curiosity and political information processing. *Polit. Psychol.* **38**, 179–199 (2017).
24. Kahan, D. M. Why smart people are vulnerable to putting tribe before truth. *Sci. Am.* (2018).
25. Goldberg, M. H., van der Linden, S., Maibach, E. & Leiserowitz, A. Discussing global warming leads to greater acceptance of climate science. *Proc. Natl. Acad. Sci.* **116**, 14804–14805 (2019).
26. van der Linden, S., Leiserowitz, A. & Maibach, E. Scientific agreement can neutralize politicization of facts. *Nat. Hum. Behav.* **2**, 2–3 (2018).
27. Ranney, M. A. & Clark, D. Climate change conceptual change: Scientific information can transform attitudes. *Top. Cogn. Sci.* **8**, 49–75 (2016).
28. Evans, J. S. B. T. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu Rev Psychol* **59**, 255–278 (2008).
29. Pennycook, G. & Rand, D. G. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50 (2019).
30. Tappin, B. M., Pennycook, G. & Rand, D. Rethinking the link between cognitive sophistication and politically motivated reasoning. *J. Exp. Psychol. Gen.* (2020) doi:https://doi.org/10.1037/xge0000974.
31. Druckman, J. N. & McGrath, M. C. The evidence for motivated reasoning in climate change preference formation. *Nat. Clim. Change* **9**, 111–119 (2019).
32. Tappin, B. M., Pennycook, G. & Rand, D. G. Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often undermine causal inference. *Curr. Opin. Behav. Sci.* **34**, 81–87 (2020).
33. Gerber, A. & Green, D. Misperceptions about perceptual bias. *Annu. Rev. Polit. Sci.* **2**, 189–210 (1999).

34. Hahn, U. & Harris, A. J. What does it mean to be biased: Motivated reasoning and rationality. in *Psychology of learning and motivation* vol. 61 41–102 (Elsevier, 2014).
35. Koehler, J. J. The influence of prior beliefs on scientific judgments of evidence quality. *Organ. Behav. Hum. Decis. Process.* **56**, 28–55 (1993).
36. Tappin, B. M., Pennycook, G. & Rand, D. G. Bayesian or biased? Analytic thinking and political belief updating. *Cognition* **204**, 104375 (2020).
37. Cook, J. & Lewandowsky, S. Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Top. Cogn. Sci.* **8**, 160–179 (2016).
38. Nurse, M. S. & Grant, W. J. I'll See It When I Believe It: Motivated Numeracy in Perceptions of Climate Change Risk. *Environ. Commun.* **14**, 184–201 (2020).
39. Raoelison, M. T., Thompson, V. A. & De Neys, W. The smart intuitor: Cognitive capacity predicts intuitive rather than deliberate thinking. *Cognition* **204**, 104381 (2020).
40. Thompson, V. A., Pennycook, G., Trippas, D. & Evans, J. S. B. Do smart people have better intuitions? *J. Exp. Psychol. Gen.* **147**, 945–961 (2018).
41. Evans, J. S. B. T. Dual Process Theories: Perspectives and problems. in *Dual Process Theory 2.0* (ed. De Neys, W.) 137–155 (Routledge, 2018).
42. Rand, D. G. Reflections on the time-pressure cooperation registered replication report. *Perspect. Psychol. Sci.* **12**, 543–547 (2017).
43. Evans, J. & Feeney, A. The role of prior belief in reasoning. *Nat. Reason.* 78–102 (2004).
44. Kardash, C. M. & Scholes, R. J. Effects of preexisiting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *J. Educ. Psychol.* **88**, 260 (1996).
45. Klauer, K. C., Musch, J. & Naumer, B. On belief bias in syllogistic reasoning. *Psychol. Rev.* **107**, 852 (2000).
46. Schommer, M. Effects of beliefs about the nature of knowledge on comprehension. *J.*

Educ. Psychol. **82**, 498 (1990).

47. Brulle, R. J., Carmichael, J. & Jenkins, J. C. Shifting public opinion on climate change: an empirical assessment of factors influencing concern over climate change in the US, 2002–2010. *Clim. Change* **114**, 169–188 (2012).
48. Baron, J. & Jost, J. T. False equivalence: Are liberals and conservatives in the United States equally biased? *Perspect. Psychol. Sci.* **14**, 292–303 (2019).
49. Wu, J. S. & Lee, J. J. Climate change games as tools for education and engagement. *Nat. Clim. Change* **5**, 413–418 (2015).
50. Goldberg, M. H., Gustafson, A., Rosenthal, S. A. & Leiserowitz, A. Shifting Republican views on climate change through targeted advertising. *Nat. Clim. Change* 1–5 (2021).
51. Bago, B. & De Neys, W. The intuitive greater good: Testing the corrective dual process model of moral cognition. *J. Exp. Psychol. Gen.* **148**, 1782–1801 (2019).
52. Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P. & Hegarty, M. How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *J. Exp. Psychol. Gen.* **130**, 621–640 (2001).
53. De Neys, W. Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *Q. J. Exp. Psychol.* **59**, 1070–1100 (2006).
54. Trémolière, B., De Neys, W. & Bonnefon, J.-F. Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition* **124**, 379–384 (2012).
55. Bago, B., Rand, D. G. & Pennycook, G. Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *J. Exp. Psychol. Gen.* (2020).
56. Bago, B. & De Neys, W. Fast logic?: Examining the time course assumption of dual process theory. *Cognition* **158**, 90–109 (2017).
57. Frederick, S. Cognitive reflection and decision making. *J. Econ. Perspect.* **19**, 25–42 (2005).
58. Thomson, K. S. & Oppenheimer, D. M. Investigating an alternate form of the cognitive

reflection test. *Judgm. Decis. Mak.* **11**, 99–113 (2016).

Supplementary Materials for "Reasoning about Climate Change"

Bence Bago, David G. Rand, Gordon Pennycook

A. Climate change arguments

Supplementary Table 1. Table shows all the arguments that were used in the main experiment

Contrary to climate change	Pro climate change
Earth's climate has always warmed and cooled, and the 20th century rise in global temperature is within the bounds of natural temperature fluctuations over the past 3,000 years. Although the planet has warmed 1-1.4°F over the 20th century, it is within the +/- 5°F range of the past 3,000 years. A 2003 study by researchers at the Harvard-Smithsonian Center for Astrophysics found that "many records reveal that the 20th century is probably not the warmest nor a uniquely extreme climatic period of the last millennium." A 2005 study published in Nature found that "high temperatures - similar to those observed in the twentieth century before 1990 - occurred around AD 1000 to 1100" in the Northern Hemisphere. A 2013 study published in Boreas found that summer temperatures during the Roman Empire and Medieval periods were "consistently higher" than temperatures during the 20th century. According to a 2010 study in the Chinese Science Bulletin, the recent global warming period of the 20th century is the result of a natural 21-year temperature oscillation, and will give way to a "new cool period in the 2030s."	Average temperatures on earth have increased at a rate far faster than can be explained by natural climate changes. A 2008 study compared data from tree rings, ice cores, and corals over the past millennium with recent temperature records. The study created the famous "hockey stick" graph, showing that the rise in earth's temperature over the preceding decade had occurred at a rate faster than any warming period over the last 1,700 years. In 2012 the Berkeley scientists found that the average temperature of the earth's land increased 2.5°F over 250 years (1750-2000), with 1.5°F of that increase in the last 50 years. Lead researcher Richard A. Muller, PhD, said "it appears likely that essentially all of this increase [in temperature] results from the human emission of greenhouse gases." In 2013, a surface temperature study published in Science found that global warming over the past 100 years has proceeded at a rate faster than at any time in the past 11,300 years. According to the IPCC's 2014 Synthesis Report, human actions are "extremely likely" (95-100% confidence) to have been the main cause of 20th century global warming, and the surface temperature warming since the 1950s is "unprecedented over decades to millennia."
Even if humans are creating a slightly warmer climate, it's not necessarily a bad thing. The underlying assumption that virtually all climate alarmists operate under is that the warming Earth is experiencing now is harmful, destructive and dangerous, but there is much evidence to suggest that moderate warming benefits most plants, animals and humans. We know, for instance, that plants grow significantly better with higher carbon-dioxide concentrations, which is why many greenhouses pump additional CO2 into their buildings. It's also been confirmed by multiple studies that greening has increased in recent decades — and likely because of higher carbon-dioxide concentrations. According to a study by Martin Brandt et al., published in the journal Nature Ecology & Evolution in 2017, 36 percent of the continent of Africa became greener over the 20-year period from 1992 to 2011, while only 11 percent became "less green." Interestingly, the researchers found the increased greening was likely "driven" by higher carbon-dioxide levels and precipitation, and the decreased greening was largely a result of humans cutting down vegetation.	There are some clear disadvantages to a warmer climate: • Agriculture depends on a supply of steady water, and floods and droughts will disrupt that. • The polar regions may thaw, but the soil up there is poor and that portion of the planet receives little direct sunlight. • Fewer people will freeze to death, but more people will die from the heat. • Disease-carrying insects like mosquitoes will thrive in warmer temperatures and carry diseases to new places, an effect we are already seeing.
Global warming and cooling are primarily caused by fluctuations in the sun's heat (solar forcing), not by human activity. Over the past 10,000 years, solar minima (reduced sun spot activity) have been "accompanied by sharp climate changes." Between 1900 and 2000 solar irradiance increased 0.19%, and correlated with the rise in US surface temperatures over the 20th century. According to a 2007 study published in Energy & Environment, "variations in solar activity and not the burning of fossil	Natural changes in the sun's activity cannot explain 20th century global warming. According to a Dec. 2013 study in Nature Geoscience, the sun has had only a "minor effect" on the Northern Hemisphere climate over the past 1,000 years, and global warming from human-produced greenhouse gases has been the primary cause of climate change since 1900. Another 2013 study found that solar activity could not have contributed to more than 10% of the observed global warming over the 20th century.

fuels are the direct cause of the observed multiyear variations in climatic responses." In a 2012 study by Willie Soon, PhD, Physicist at the Harvard-Smithsonian Center for Astrophysics, a strong correlation between solar radiation and temperatures in the Arctic over the past 130 years was identified. According to a 2012 study published in the Journal of Atmospheric and Solar-Terrestrial Physics, "up to 70% of the observed post-1850 climate change and warming could be associated to multiple solar cycles."	Measurements in the upper atmosphere from 1979-2009 show the sun's energy has gone up and down in cycles, with no net increase. According to a 2013 IPCC report, there is "high confidence" (8 out of 10 chance) that changes in the sun's radiation could not have caused the increase in the earth's surface temperature from 1986-2008. Although warming is occurring in the lower atmosphere (troposphere), the upper atmosphere (stratosphere) is actually cooling. If the sun were driving global warming, there would be warming in the stratosphere also, not cooling.
Sea levels have been steadily rising for thousands of years, and the increase has nothing to do with humans. A 2014 report by the Global Warming Policy Foundation found that a slow global sea level rise has been ongoing for the last 10,000 years. When the earth began coming out of the Pleistocene Ice Age 18,000 years ago, sea levels were about 400 feet lower than they are today and have been steadily rising ever since. According to Professor of Earth and Atmospheric Sciences at the Georgia Institute of Technology, Judith Curry, PhD "it is clear that natural variability has dominated sea level rise during the 20th century, with changes in ocean heat content and changes in precipitation patterns." Freeman Dyson, Emeritus Professor of Mathematical Physics and Astrophysics at the Institute for Advanced Study at Princeton University, has stated that there is "no evidence" that rising sea levels are due to anthropogenic climate change.	Sea levels are rising at an unprecedented rate due to global warming. As human-produced greenhouse gases warm the planet, sea levels are rising due to thermal expansion of warming ocean waters as well as melt water from receding glaciers and the polar ice cap. According to the IPCC, there has been a "substantial" human contribution to the global mean sea-level rise since the 1970s, and there is "high confidence" (8 out of 10 chance) that the rate of sea-level rise over the last half century has accelerated faster than it has over the previous 2,000 years. A 2006 study found that "significant acceleration" of sea-level rise occurred from 1870 to 2004. Between 1961 and 2003 global sea levels rose 8 inches. An Oct. 2014 study published in the Proceedings of the National Academy of Sciences concluded that the rate of sea level rise over the past century is unprecedented over the last 6,000 years. A separate Oct. 2014 study said that the global sea level is likely to rise 31 inches by 2100, with a worst case scenario rise of 6 feet. Climate Central predicts that 147 to 216 million people live in areas that will be below sea level or regular flood areas by the end of the century if human-produced greenhouse gas emissions continue at their current rate.
Glaciers have been growing and receding for thousands of years due to natural causes, not human activity. The IPCC predicted that Himalayan glaciers would likely melt away by 2035, a prediction they disavowed in 2010. In 2014 a study of study of 2,181 Himalayan glaciers from 2000-2011 showed that 86.6% of the glaciers were not receding. According to a 2013 study of ice cores published in Nature Geoscience, the current melting of glaciers in Western Antarctica is due to "atmospheric circulation changes" that have "caused rapid warming over the West Antarctic Ice Sheet" and cannot be directly attributed to human caused climate change. According to one of the study authors, "[i]f we could look back at this region of Antarctica in the 1940s and 1830s, we would find that the regional climate would look a lot like it does today, and I think we also would find the glaciers retreating much as they are today." According to Christian Schlüchter, Professor of Geology at the University of Bern, 4,000 year old tree remains have been found beneath retreating glaciers in the Swiss Alps, indicating that they were previously glacier-free. According to Schlüchter, the current retreat of glaciers in the Alps began in the mid-19th century, before large amounts of human caused CO2 had entered the atmosphere.	Glaciers are melting at unprecedented rates due to global warming, causing additional climate changes. About a quarter of the globe's glacial loss from 1851-2010, and approximately two thirds of glacial loss between 1991-2010, is attributable directly to global warming caused by human-produced greenhouse gases. According to the National Snow and Ice Data Center, global warming from human-produced greenhouse gases is a primary cause of the "unprecedented" retreat of glaciers around the world since the early 20th century. Since 1980 glaciers worldwide have lost nearly 40 feet (12 meters) in average thickness. According to a 2013 IPCC report, "glaciers have continued to shrink almost worldwide" over the prior two decades, and there is "high confidence" (about an 8 out of 10 chance) that Northern Hemisphere spring snow continues to decrease. If the glaciers forming the Greenland ice sheet were to melt entirely, global sea levels could increase by up to 20 feet. Melting glaciers also change the climate of the surrounding region. With the loss of summer glacial melt water, the temperatures in rivers and lakes increase. According to the US Geological Service, this disruption can include the "extinction of temperature sensitive aquatic species."
More than one thousand scientists disagree that human activity is primarily responsible for global climate change. In 2010 Climate Depot released a report featuring more than 1,000 scientists, several of them former UN IPCC scientists, who disagreed that humans are primarily responsible for global climate change. The Cook review of 11,944 peer-reviewed studies found 66.4% of the studies had no stated position on anthropogenic global warming, and while 32.6% of the studies implied or stated that humans are contributing to climate change, only 65 papers (0.5%) explicitly stated "that humans are the primary cause of recent global warming." A 2012 Purdue University survey found that 47% of climatologists	Overwhelming scientific consensus says human activity is primarily responsible for global climate change. The 2010 Anderegg study found that 97-98% of climate researchers publishing most actively in their field agree that human activity is primarily responsible for global climate change. The study also found that the expertise of researchers unconvinced of human-caused climate change is "substantially below" that of researchers who agree that human activity is primarily responsible for climate change. The 2013 Cook review of 11,944 peer-reviewed studies on climate change found that only 78 studies (0.7%) explicitly rejected the position that humans are responsible for global warming. A separate review of 13,950 peer-reviewed studies on climate change found only 24 that rejected

challenge the idea that humans are primarily responsible for climate change and instead believe that climate change is caused by an equal combination of humans and the environment (37%), mostly by the environment (5%), or that there's not enough information to say (5%). In 2014 a group of 15 scientists dismissed the US National Climate Assessment as a "masterpiece of marketing," that was "grossly flawed," and called the NCA's assertion of human-caused climate change "NOT true."	human-caused global warming. A survey by German Scientists Bray and Von Storch found that 83.5% of climate scientists believe human activity is causing "most of recent" global climate change. A separate survey in 2011 also found that 84% of earth, space, atmospheric, oceanic, and hydrological scientists surveyed said that human-induced global warming is occurring.
Rising levels of atmospheric CO2 do not necessarily cause global warming, which contradicts the core thesis of human-caused climate change. Earth's climate record shows that warming has preceded, not followed, a rise in CO2. According to a 2003 study published in Science, measurements of ice core samples show that over the last four climactic cycles (past 240,000 years), periods of natural global warming preceded global increases in CO2. In 2010 the Proceedings of the National Academy of Sciences published a study of the earth's climate 460-445 million years ago which found that an intense period of glaciation, not warming, occurred when CO2 levels were 5 times higher than they are today. According to ecologist and former Director of Greenpeace International Patrick Moore, PhD, "there is some correlation, but little evidence, to support a direct causal relationship between CO2 and global temperature through the millennia."	The rise in atmospheric CO2 over the last century was clearly caused by human activity, as it occurred at a rate much faster than natural climate changes could produce. Over the past 650,000 years, atmospheric CO2 levels did not rise above 300 ppm until the mid-20th century. Atmospheric levels of CO2 have risen from about 317 ppm in 1958 to 415 ppm in 2019. CO2 levels are estimated to reach 450 ppm by the year 2040. According to the Scripps Institution of Oceanology, the "extreme speed at which carbon dioxide concentrations are increasing is unprecedented. An increase of 10 parts per million might have needed 1,000 years or more to come to pass during ancient climate change events." Some climate models predict that by the end of the 21st century an additional 5°F-10°F of warming will occur.
Increased hurricane activity and other extreme weather events are a result of natural weather patterns, not human-caused climate change. According to a 2013 report from the Tropical Meteorology Project at Colorado State University, the increase in human-produced CO2 over the past century has had "little or no significant effect" on global tropical cyclone activity. The report further states that specific hurricanes, including Sandy, Ivan, Katrina, Rita, Wilma, and Ike, were not a direct consequence of human-caused global warming. Between 1995-2015 increased hurricane activity (including Katrina) was recorded, however, according to the NOAA, it was not the result of human-induced climate change; it was the result of cyclical tropical cyclone patterns, driven primarily by natural ocean currents. Many types of recorded extreme weather events over the past half-century have actually become less frequent and less severe. Professor of Earth and Atmospheric Sciences at the Georgia Institute of Technology, Judith Curry, PhD, states that she is "unconvinced by any of the arguments that I have seen that attributes a single extreme weather event, a cluster of extreme weather events, or statistics of extreme weather events" to human-caused climate change. Richard Lindzen, PhD, Emeritus Professor of Meteorology at the Massachusetts Institute of Technology, also states that there is a lack of evidence connecting extreme weather events such as hurricanes, tornadoes, droughts, or floods, to human-caused global warming.	Dramatic changes in precipitation, such as heavier storms and less snow, are another sign that humans are causing global climate change. As human-produced greenhouse gases heat the planet, increased humidity (water vapor in the atmosphere) results. Water vapor is itself a greenhouse gas. In a process known as a positive feedback loop, more warming causes more humidity which causes even more warming. Higher humidity levels also cause changes in precipitation. According to a 2013 report published in the Proceedings of the National Academy of Sciences, the recorded changes in precipitation over land and oceans "are unlikely to arise purely due to natural climate variability." Higher temperatures from global warming are also causing some mountainous areas to receive rain rather than snow. According to researchers at the Scripps Institution of Oceanography, up to 60% of the changes in river flow, winter air temperature, and snow pack in the western United States (1950-1999) were human-induced. Since 1991, heavy precipitation events have been 30% above the 1901-1960 average in the Northeast, Midwest, and upper Great Plains regions. A 2015 study found that global warming caused by human actions has increased extreme precipitation events by 18% across the globe, and that if temperatures continue to rise an increase of 40% can be expected.
The acidity levels of the oceans are within past natural levels, and the current rise in acidity is a natural fluctuation, not the result of human caused climate change. The pH of average ocean surface water is 8.1 and has only decreased 0.1 since the beginning of the industrial revolution (neutral is pH 7, acid is below pH 7). In 2010 Science published a study of ocean acidity levels over the past 15 million years, finding that the "samples record surface seawater pH values that are within the range observed in the oceans today." Increased atmospheric CO2 absorbed by the oceans results in higher rates of photosynthesis and faster growth of ocean plants and phytoplankton, which increases pH levels keeping the water alkaline, not	Ocean acidity levels are increasing at an unprecedented rate that can only be explained by human activity. As excess human-produced CO2 in the atmosphere is absorbed by the oceans, the acidity level of the water increases. Acidity levels in the oceans are 25-30% higher than prior to human fossil fuel use. According to a 2014 US Government Accountability Office (GAO) report, oceans have absorbed about 30% of the CO2 emitted by humans over the past 200 years, and ocean acidity could rise approximately 100-200 percent above preindustrial levels by 2100. According to a 2013 report from the World Meteorological Organization, the current acceleration in the rate of ocean acidification "appears

acidic. According to a 2010 paper by the Science and Public Policy Institute, "our harmless emissions of trifling quantities of carbon dioxide cannot possibly acidify the oceans."	unprecedented" over the last 300 million years. High ocean acidity levels threaten marine species, and slows the growth of coral reefs. According to a 2014 report by the Convention on Biological Diversity, "it is now nearly inevitable" that within 50-100 years continued human produced CO2 emissions will increase ocean acidity to levels that "will have widespread impacts, mostly deleterious, on marine organisms and ecosystems."
CO2 is already saturated in earth's atmosphere, and more CO2, manmade or natural, will have little impact on climate. As CO2 levels in the atmosphere rise, the amount of additional warming caused by the increased concentration becomes less and less pronounced. According to Senate testimony by William Happer, PhD, Professor of Physics at Princeton University, "[a]dditional increments of CO2 will cause relatively less direct warming because we already have so much CO2 in the atmosphere that it has blocked most of the infrared radiation that it can. The technical jargon for this is that the CO2 absorption band is nearly 'saturated' at current CO2 levels." According to the Heartland Institute's 2013 Nongovernmental International Panel on Climate Change (NIPCC) report, "it is likely rising atmospheric CO2 concentrations will have little impact on future climate."	The specific type of CO2 that is increasing in earth's atmosphere can be directly connected to human activity. CO2 produced by burning fossil fuels such as oil and coal can be differentiated in the atmosphere from natural CO2 due to its specific isotopic ratio. According to the Intergovernmental Panel on Climate Change (IPCC), 20th century measurements of CO2 isotope ratios in the atmosphere confirm that rising CO2 levels are the result of human activity, not natural processes such as ocean outgassing, volcanic activity, or release from other "carbon sinks." US greenhouse gas emissions from human activities in 2012 totaled 6.5 million metric tons, which is equivalent to about 78.3 billion shipping containers filled with greenhouse gases.

B. Correlational evidence

Methods

Study 1a

Participants. In total, 300 participants (112 females, 186 males and 1 others, Mean age = 35.2 years, SD = 10.6 years) took part in this study, who were recruited through MTurk, and completed the experiment online. In total, 183 people were Democrats, and 116 were Republicans.

Materials & Procedure

Climate arguments. Altogether, participants were shown 12 arguments in a randomised order (6 contra and 6 pro). Note that two of these arguments were not included in the two response experiments.

Procedure. At the beginning of the experiment, participants were presented with the following instructions:

"Welcome to this experiment!

In this experiment, you will be presented with different arguments regarding climate change and its potential effects. You will be asked you to indicate how much you agree with the argument on a scale from 0 = completely disagree to 100 = completely agree. Then, you will be presented with the same problem again and we will ask you three other questions. Please carefully read through the arguments before responding.

You will be presented with 12 arguments. The experiment will take about 11 minutes to complete, and will demand your full attention.

Press 'Next' to continue!"

After they finished responding to the arguments, participants were presented with a 6 item version of the cognitive reflection test, that is widely used to measure reflective abilities. Next, they were presented with a series of question to measure political knowledge and at the end, a series of standard demographic questions, regarding gender, age, political preferences. Most notably, to measure partisan preferences, we used the question: "Which of the following best describes your political preference? (Strongly Democrat/Democrat/Lean Democrat/ Lean Republican/Republican/Strong Republican)", and to measure belief in anthropogenic climate change, we used the question: "How much do you agree with the idea that human activity causes global changes in climate? (Completely agree/Somewhat agree/Somewhat disagree/Completely disagree)". Additionally, we also measured ideology with two questions; one asking about economic and another asking about social ideology: "On social/economic issues, I am: Strongly Liberal/Somewhat Liberal/Moderate/Somewhat Conservative/Strongly Conservative).

Statistical analysis. We used mixed effect regression models. Political preference was coded by a numeric variable (-3, -2, -1, 1, 2, 3; from Democrat to Republican), and climate belief was coded by another variable (-2, -1, 1, 2; from believer to non-believer)., and argument type was also coded by a dummy (-0.5 - pro, 0.5 - contra). For ideology, we averaged the responses to the two questions above (from -2, Liberal, to +2 Conservative). All continuous variables were scaled before entered into the analysis. We allowed the intercept to vary over subjects and argument contents. Table S2 summarizes the random and fixed effect of each model.

Study 1b

Participants. In total, 605 participants (310 females, 295 males, Mean age = 45.5 years, SD = 16.8 years) took part in this study, who were recruited through MTurk, and completed the experiment online. In total, 357 people were Democrats, and 246 were Republicans.

Pre-registration of this study can be found at the OSF page of the project.

Materials and procedures are identical to Study 1a, with the exception of political knowledge scale which we did not administer to make the experiment shorter. The specifics of the statistical models are described in Table S2.

Supplementary Table 2. Table shows the specifications of each model we run in Study 1a and 1b.

	Fixed	Random slope over	Random slope over

		Subjects	item contents
Study 1a (MTurk)	Argument type, Partisanship	Argument type	Argument type
	Argument type, CRT score, Partisanship	Argument type	Argument type
	Argument type, CRT score, Partisanship, Belief	Argument type	Argument type
Study 2a (Lucid)	Argument type, Partisanship	Argument type	Argument type
	Argument type, CRT score, Partisanship	Argument type	CRT score
	Argument type, CRT score, Partisanship, Belief	Argument type	CRT score

Results

In a first experiment, we used two sources of American participants, MTurk (Study 1a; N=300) and Lucid (Study 1b; N = 605 quota-matched to the US on age, gender, ethnicity, and geographic region). The two samples were analysed separately. The pre-registration for Study 1b can be found at: https://osf.io/fqkne/

In Study 1a, we first tested whether political preference was associated with how people judge pro versus contra climate change arguments (Figure S1). As expected, when predicting agreement we found a significant interaction between political partisanship and whether arguments were pro versus contra (argument type), $b = 30.42$, $p < 0.0001$; Republicans were more likely to agree with contra arguments than Democrats, while it was the other way around with pro arguments. This shows that there are indeed partisan differences in the evaluation of arguments pro and contra climate change; hence, this item set can be used to test the competing hypotheses.

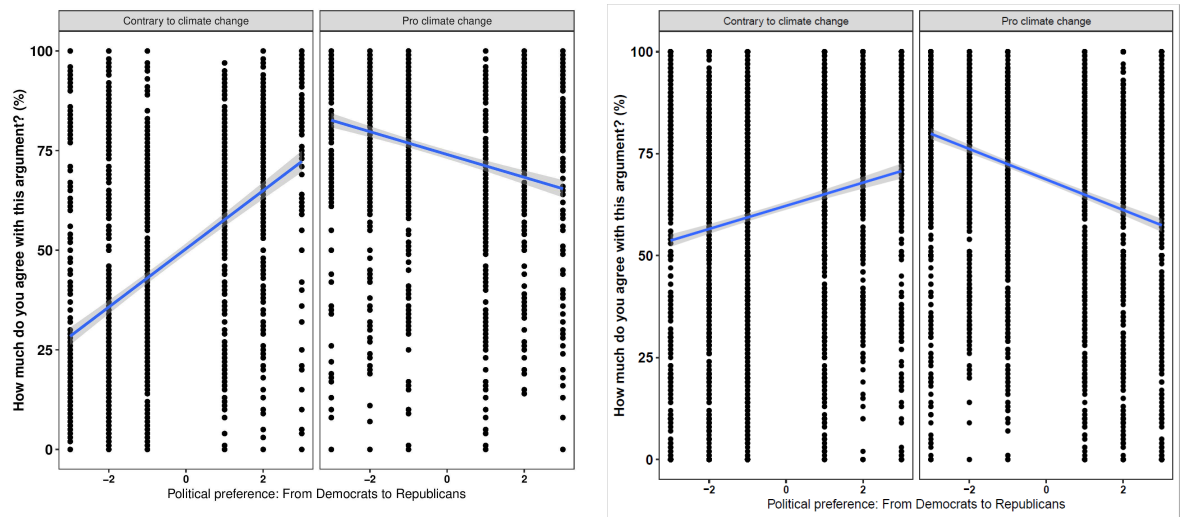

Figure S1. *Agreement scores as a function of political partisanship and argument type on Mturk (left) and Lucid sample (right). Points represent trials.*

When adding cognitive sophistication score to the model we found a significant 3-way interaction between cognitive sophistication, argument type, and political partisanship, b = 32.59, p < 0.0001 (Figure S2). That is, the already existing partisan differences in argument evaluation were magnified by cognitive sophistication, providing a conceptual replication of Kahan et al (2011). Most importantly, however, when we added prior beliefs to the model, the 3-way interaction between political partisanship, argument type and cognitive sophistication was no longer significant, b = -6.6, p = 0.39; whereas there was a significant 3-way interaction between prior beliefs, argument type, and cognitive sophistication, b = 51.68, p < 0.0001. This suggests that cognitive sophistication magnifies the effect of prior beliefs rather than partisanship per se (Figure S3). Importantly, the correlation between beliefs about climate change and political partisanship (using a continuous measure of Democrat-Republican partisanship strength) was positive, but modest in size (r = 0.29), demonstrating that they are clearly separate (albeit related) measures (see Tappin et al., 2020 for further discussion).

Study 1b replicated these results using a larger and more representative sample: Democrats were more likely to agree with pro than with contra arguments, and it was the other way around for Republicans (i.e., significant interaction between argument type and political partisanship, b = 19.63, p < 0.0001); and while cognitive sophistication increased partisan differences when modeled without priors (i.e., significant three-way interaction between partisanship, cognitive sophistication and argument type), b = 53.12, p < 0.0001 (Figure S2), including prior beliefs in the model eliminated this interaction, b = 3.4, p = 0.34 but themselves showed a strong interaction with argument type and crt score, b = 38.5, p < 0.0001 (Figure 3).

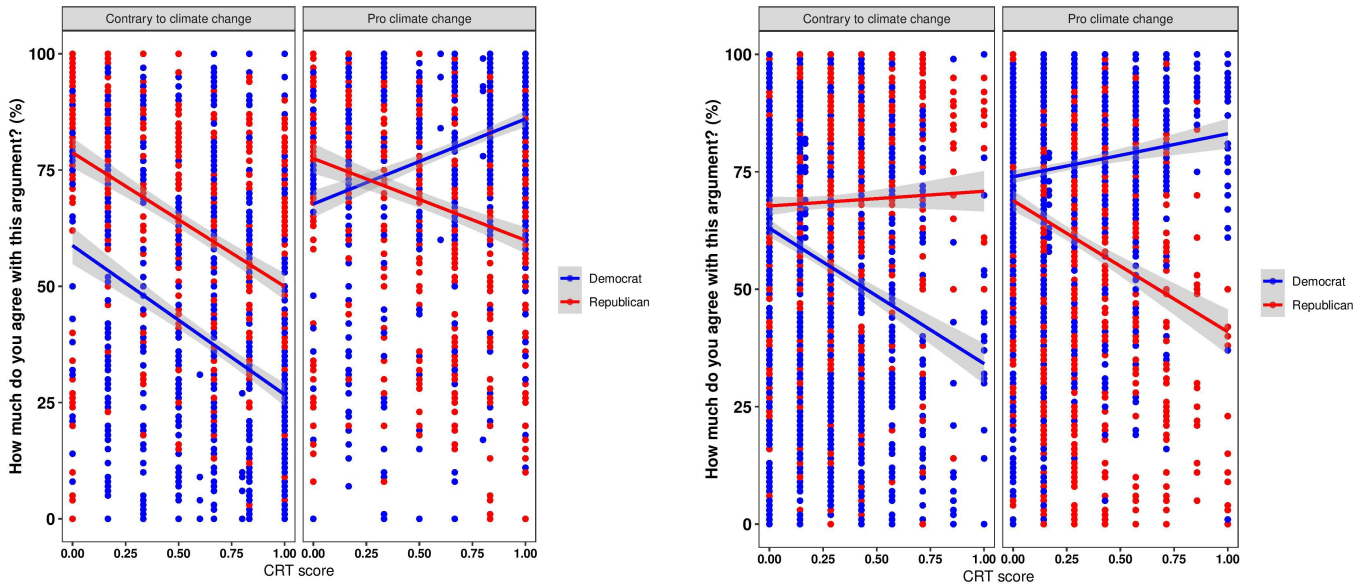

Figure S2. *Agreement scores as a function of political partisanship, CRT score and argument type on Mturk (left) and Lucid sample (right). Points represent trials.*

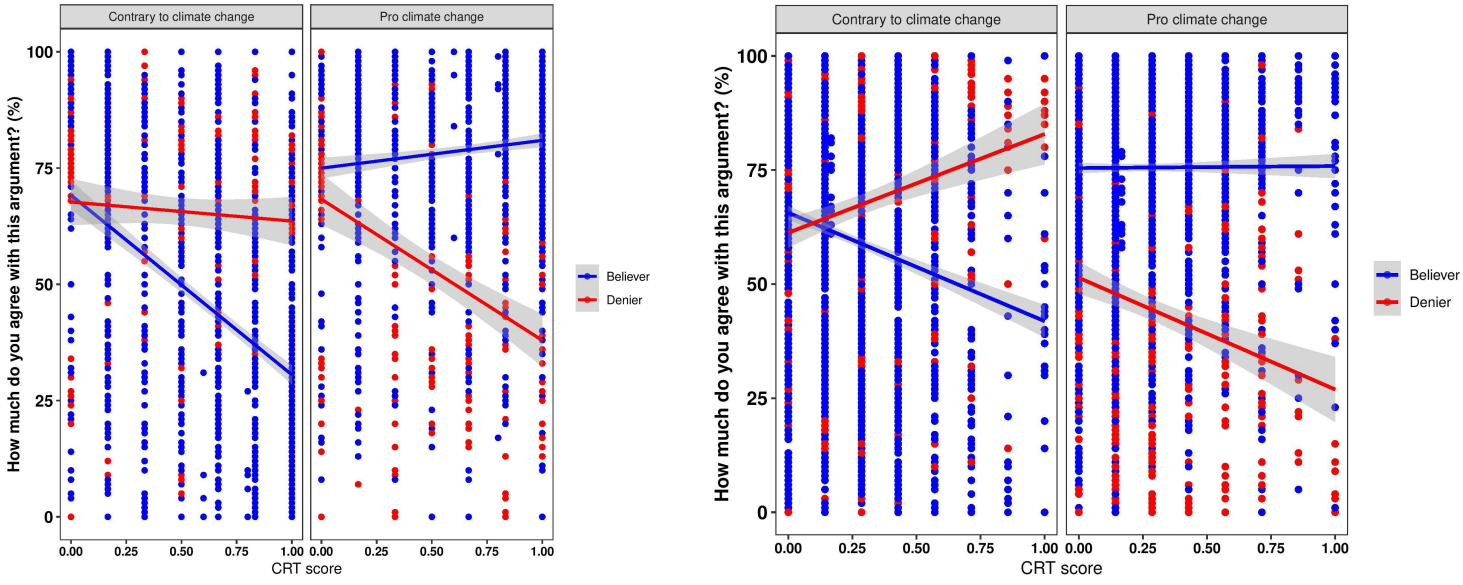

Figure S3. *Agreement scores as a function of prior beliefs, CRT score and argument type on Mturk (left) and Lucid (right) sample. Points represent trials.*

Additionally, we also conducted an analysis replacing partisanship with ideology (ideology was only recorded in Study 1a and not 1b). We found qualitatively equivalent results. There is a significant three way interaction between prior beliefs, argument type and CRT (measure of cognitive reflection abilities, see Supplementary Materials), $b = 53.79$, $p < 0.0001$, while there is no interaction between ideology, argument type and CRT, $b = 1.97$, $p = 0.82$. Hence,

this supports our conclusion that it is prior beliefs that drive reasoning directly and not ideology or partisanship.

C. Additional Results to the main, two response experiments.

Results

Supplementary Table 3. Table shows all the fixed and random effects that were included in the models in both studies. In all models, the intercept was allowed to vary over subjects and item contents.

	Fixed	Random slope over Subjects	Random slope over item contents
Study 1	Argument type, Response type (0 - final response, 1 - one response baseline), Partisanship, Belief	Argument type	Argument type, Belief
	Argument type, Response type (0 - intuitive response, 1 - final response), Partisanship, Belief	Argument type	Argument type, Belief
	Argument type, Response type	Argument type	Argument type
Study 2	Argument type, Response type (0 - final response, 1 - one response baseline), Partisanship, Belief	Argument type	Argument type, Belief
	Argument type, Response type (0 - intuitive response, 1 - final response), Partisanship, Belief	Argument type	Argument type, Belief

	Argument type, Response type	Argument type	-
Pooled	Argument type, Response type (0 - intuitive response, 1 - final response), Partisanship, Belief	Argument type	Argument type, Belief

Anchoring effect

We compared the final response of people in the two response paradigm to the responses of people in the baseline one response condition (between subject comparison) to see if the final response stage is affected by anchoring effect (that is, people anchored to their initial response in the two response paradigm, and gave a final response having that in mind, presumably, to appear consistent). We see largely similar results across both studies. First, we found a significant two way interaction between argument type and response type in Study 1, but not in Study 2 (no other fixed effects included in the model, Study 1: b = 4, p = 0.033, Study 2: b = 3.3, p = 0.11). We found a significant three way interaction between response type (final response vs one response baseline), argument type, and belief (Study 1: b = 10.9, p = 0.002, Study 2 = b = 11.1, p = 0.01), while no significant interaction between response type, argument type and partisanship (Study 1: b = -6.3, p = 0.083, Study 2: b = -3.4, p = 0.491). Most importantly, the significant three way interaction suggests that people in the two response paradigm were indeed anchored to their initial response, rendering the within-subject test of the effect of deliberation on agreement scores less sensitive.

Initial vs final response (within-subject deliberation effect)

Although we found evidence for anchoring effect in the final response stage, for completeness, we compared the initial and final response stage of the two response paradigm (within subject comparison). We found a significant interaction effect between argument type and response type in both studies (no other fixed effects included in the model, Study 1: b = -1.6, p = 0.016, Study 2: b = -1.8, p = 0.028). After adding partisanship and prior belief into the model, we found different results in the two studies. In Study 1, we found no significant interaction between response type (initial vs final response) and argument type, b = 1.2, p = 0.27, also no interaction between argument type, response type and partisanship, b = -0.93, p = 0.587, but we found a significant interaction between argument type, response type and prior belief, b = -4.6, p = 0.003. In Study 2, but we did find a significant interaction between argument type, response type and partisanship, b = -4.7, p = 0.029, and also between response type, argument type and prior belief, b =-3.9, p = 0.04.

Since the two analyses had conflicting results (in study 1 we found no significant interaction with partisanship, while we did in study 2), we pooled the two samples together, and conducted the same analysis again. In this pooled analysis, we found no significant interaction between argument type, response type and partisanship, b = -2.4, p = 0.07, but the interaction between response type, argument type and prior belief was significant, b = -4.4, p = 0.0003. This overall supports our main conclusion that deliberation is driven by belief-coherence effect.

D. Additional plots

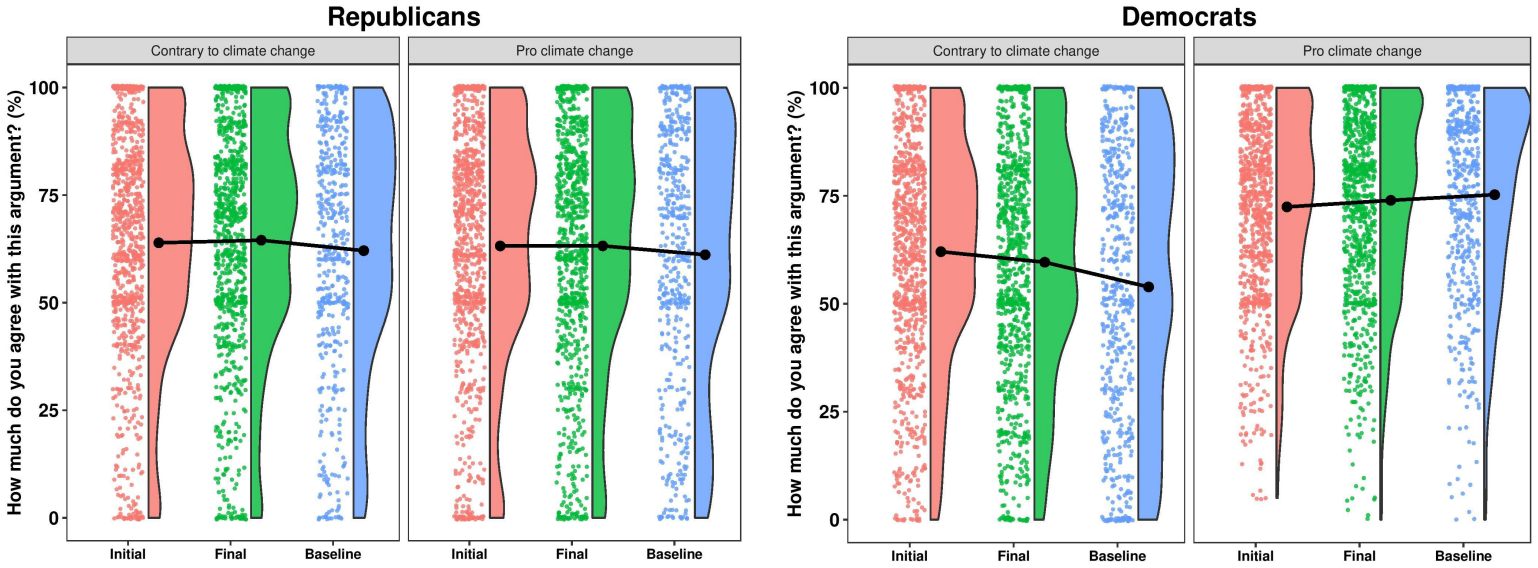

Figure S4. *Agreement scorers as a function of party affiliation, response condition (initial = intuitive response, final response, one-response baseline), and argument type (Study 1). Dots in the middle of the half-violin plots are averages. Positive trends mean increase in belief after deliberation, while negative trends mean a decrease in belief after deliberation.*

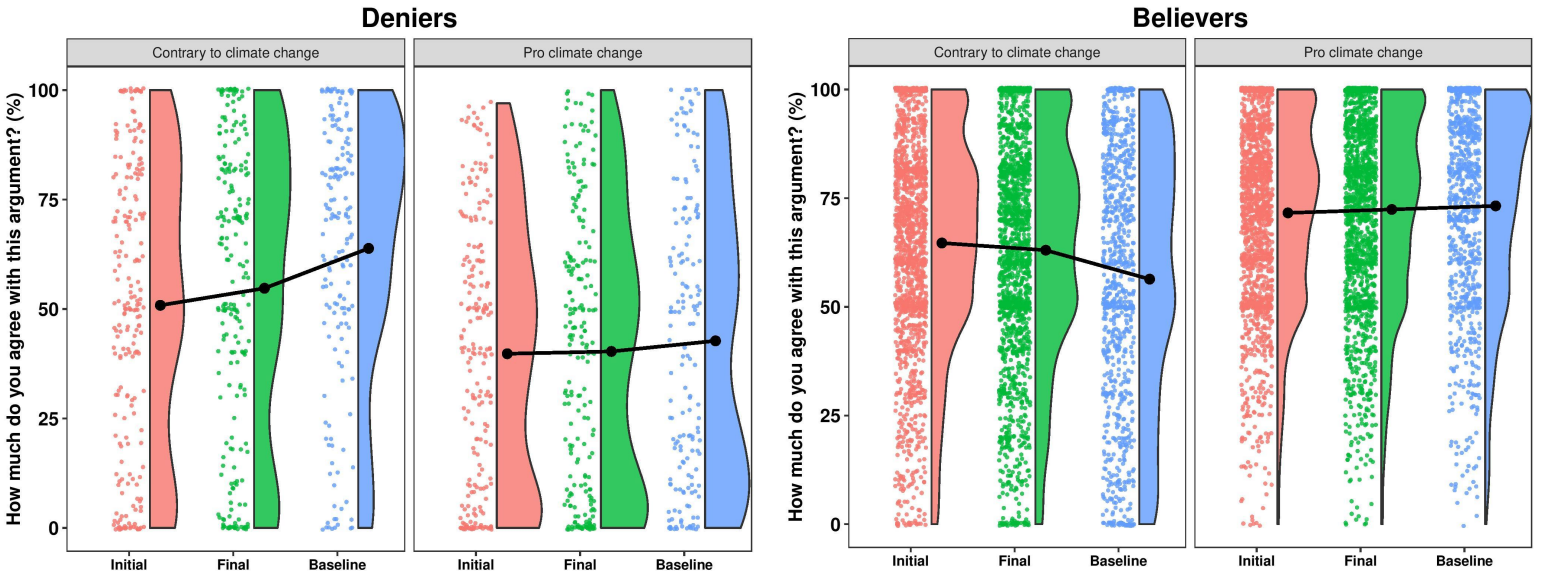

Figure S5. *Agreement scores as a function of prior beliefs, response condition (initial = intuitive response, final response, one-response baseline), and argument type (Study 1). Dots in the middle of the half-violin plots are averages. Positive trends mean increase in belief after deliberation, while negative trends mean a decrease in belief after deliberation.*

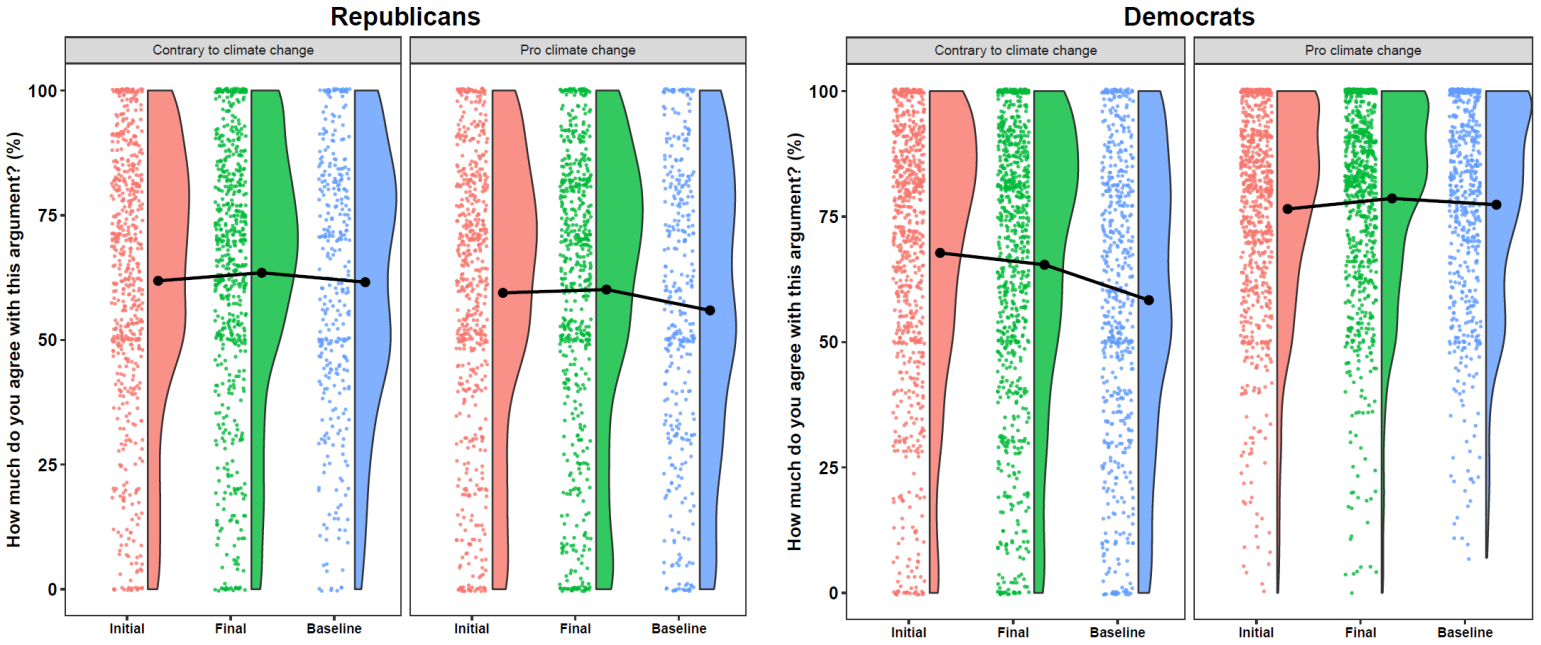

Figure S6. *Agreement scorers as a function of party affiliation, response condition (initial = intuitive response, final response, one-response baseline), and argument type (Study 2). Dots in the middle of the half-violin plots are averages. Positive trends mean increase in belief after deliberation, while negative trends mean a decrease in belief after deliberation.*

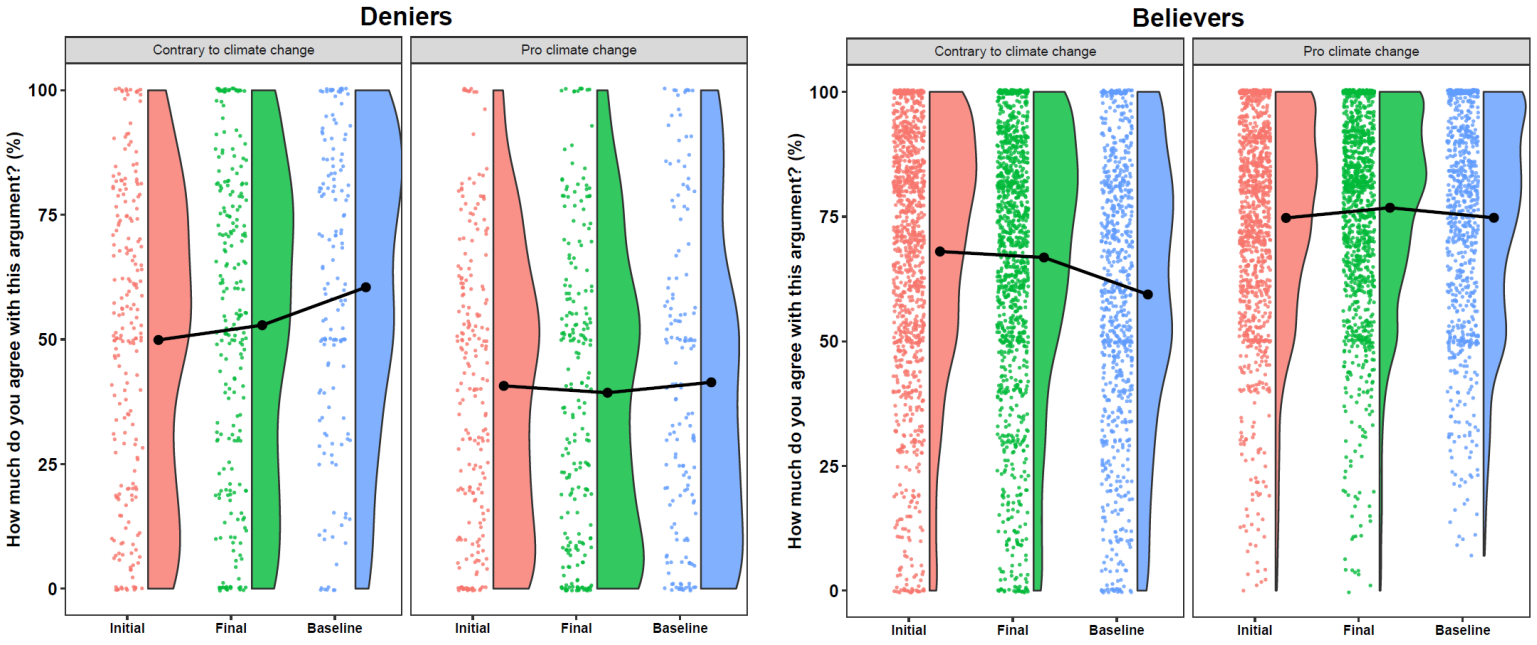

Figure S7. *Agreement scores as a function of prior beliefs, response condition (initial = intuitive response, final response, one-response baseline), and argument type (Study 2). Dots in the middle of the half-violin plots are averages. Positive trends mean increase in belief after deliberation, while negative trends mean a decrease in belief after deliberation.*

E. Standardized regression coefficients for Study 1 and 2, Intuitive vs Deliberative (initial response vs one response baseline) comparison.

	Study 1		Study 2	
	Missed trials ignored	Missed trials replaced	Missed trials ignored	Missed trials replaced
Argument type * response type	b = -0.21 p = 0.001	b = -0.21 p = 0.001	b = -0.19 p = 0.005	b = -0.18 p = 0.007
Politics * argument type * response type	b = -0.05 p = 0.43	b = -0.05 p = 0.39	b = -0.04 p = 0.61	b = -0.03 p = 0.66
Belief * argument type * response type	b = -0.34 p < 0.0001	b = -0.33 p < 0.0001	b = -0.28 p = 0.002	b = -0.29 p = 0.001
Deniers - Contra arguments: Response type effect	b = 0.41 p = 0.008	b = 0.41 p = 0.009	b = 0.33 p = 0.03	b = 0.38 p = 0.01

Believers - Contra arguments: Response type effect	b = -0.3 $p < 0.0001$	b = -0.3 $p < 0.0001$	b = -0.32 $p < 0.0001$	b = -0.33 $p < 0.0001$
Deniers - Pro arguments: Response type effect	b = 0.1 p = 0.53	b = 0.09 p = 0.56	b = 0.01 p = 0.95	b = 0.02 p = 0.9
Believers - Pro arguments: Response type effect	b = 0.07 p = 0.28	b = 0.07 p = 0.27	b = 0.003 p = 0.96	b = -0.03 p = 0.69
Republican Deniers - Pro arguments: response type effect	b = 0.43 p = 0.02	b = 0.42 p = 0.02	b = 0.44 p = 0.009	b = 0.47 p = 0.005
Republican Believers - Pro arguments: response type effect	b = -0.25 p = 0.009	b = -0.25 p = 0.01	b = -0.26 p = 0.03	b = -0.28 p = 0.02

F. Results using ideology rather than partisanship

We tested the robustness of our results to using ideology rather than partisanship in our models (as well as, of course, prior beliefs, condition [initial or baseline] and argument type). Although we did not measure ideology in our experiments, we managed to recover an ideology measure (7 point Likert scale from Extremely liberal to Extremely conservative) for 582 participants (pooling across S1 and S2) from records maintained by Lucid, the subject recruitment platform used for the study. Analyzing these the responses of these 582 participants, we found no significant three way interaction between ideology, condition and argument type, b = -5.03, p = 0.41, while finding a significant three way interaction between prior beliefs, condition and argument type, b = -13.7, p = 0.01. These results, along with the correlational evidence from Study 1a reported above in Supplementary Materials Section B, show that the same pattern observed for partisanship also holds for ideology: neither partisanship nor ideology are associated with the effect of deliberation once prior beliefs are accounted for.