Studying science denial with a complex problem-solving task

Justin Sulik (justin.sulik@gmail.com)
Cognition, Values & Behavior, Ludwig Maximilian University of Munich
Munich, Germany

Ryan McKay (ryan.mckay@rhul.ac.uk)
Department of Psychology, Royal Holloway University of London
Egham, UK

Abstract

Explanations of science denial rooted in individual cognition tend to focus on general trait-like factors such as cognitive style, conspiracist ideation or delusional ideation. However, we argue that this focus typically glosses over the concrete, mechanistic elements of belief formation, such as hypothesis generation, data gathering, or hypothesis evaluation. We show, empirically, that such elements predict variance in science denial not accounted for by cognitive style, even after accounting for social factors such as political ideology. We conclude that a cognitive account of science denial would benefit from the study of complex (i.e., open-ended, multi-stage) problem solving that incorporates these mechanistic elements.

Keywords: science denial; cognitive style, problem solving

Science denial — holding a belief that conflicts with established science, for instance on anthropogenic climate change or the general safety of vaccines — presents both a practical problem and a cognitive puzzle. The practical problem is that anti-scientific beliefs can exacerbate environmental, medical or other harms; the cognitive puzzle is how humans form these false beliefs in the first place. Solving the cognitive puzzle is key to reducing the practical problems.

Previous work on this issue has tended to focus on two main types of explanation: social factors and individual cognitive factors. **Social factors** include anything learnt from other people, including anecdotal accounts of others' experiences (Kubin, Puryear, Schein, & Gray, 2021); worldview-based justifications derived from religious or political allegiances (Kahan, 2013; Lewandowsky & Oberauer, 2016; Pasek, 2018; Rutjens, Sutton, & van der Lee, 2018); or information about the presence or absence of consensus (Ding, Maibach, Zhao, Roser-Renouf, & Leiserowitz, 2011; Lewandowsky, Gignac, & Vaughan, 2013). Research on **cognitive factors** frequently focuses on dispositional traits, such as cognitive style (e.g., whether people tend to process information reflectively or intuitively, Gervais, 2015; Pennycook, Fugelsang, & Koehler, 2015; Ståhl & van Prooijen, 2018); and delusional or conspiracist ideation (Jolley & Douglas, 2014; Lobato, Mendoza, Sims, & Chin, 2014; Prike, Arnold, & Williamson, 2018). These dispositions correlate with science denial (Bronstein, Pennycook, Bear, Rand, & Cannon, 2019; Gervais, 2015; Hartman, Dieckmann, Sprenger, Stastny, & DeMarree, 2017; McPhetres & Pennycook, 2019), and they are relatively stable (Peters, Joseph, Day, & Garety, 2004; Stagnaro, Pennycook, & Rand, 2018).

Though both social and cognitive factors likely contribute to science denial, we propose that a proper cognitive explanation will benefit from a shift in focus. We will argue that the commonly studied dispositional traits are not enough to explain science denial. Rather, as science denial involves false beliefs that are resistant to disconfirming evidence, a proper cognitive explanation must also incorporate the concrete mechanisms whereby information is gathered and processed to form beliefs. We then provide empirical evidence that this approach has the potential to throw new light on the cognitive underpinnings of science denial.

Traits vs. mechanisms in false belief formation

The core issue is that commonly studied dispositional traits are too far removed from actual beliefs to be full explanations of science denial. This distance between traits and beliefs is due to the fact that belief formation is a multi-stage process which includes hypothesis generation, data gathering, and hypothesis evaluation in light of the data. The false beliefs involved in science denial are just the final outcome of these various inferential steps.

A focus on traits leaves gaps in understanding how the false beliefs are formed. One gap is that a particular false belief could be due to inferential errors in various stages in the belief formation process: A false belief could arise due to someone being biased in the range of hypotheses they consider, or not collecting enough data, or underweighting evidence that disconfirms their hypotheses. Another gap is that cognitive traits — as commonly studied — do not diagnose just where in this belief formation process an error has occurred[1].

To illustrate this, compare a task measuring cognitive style with one that affords more fine-grained measurement of the specific *mechanisms* of belief formation — the inferential steps that explain *how* the belief was formed.

The Cognitive Reflection Test (CRT, Frederick, 2005) includes questions such as 'A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?' Commonly, respondents form the belief that the ball costs 10 cents, but this is incorrect: The ball costs 5 cents. The task indexes people's tendency to rely on an intu-

---

[1] We do not claim that traits cannot in principle predict specific missteps in belief formation. Just that, as things stand, the matter is underdetermined for dispositional traits like cognitive style or delusional ideation.

itive or non-reflective processing style, and the implication — if this is meant to help explain science denial — is that non-reflective processing is associated with a higher incidence of false beliefs[2].

There are presumably several inferential steps going on in the mind of someone tackling this problem but only the final answer is recorded: a guess about the number of cents. In contrast, these steps are made explicit in other tasks involving complex, open-ended problem solving. One such example is the 2-4-6 task (Wason, 1960). The aim in this task is to guess a specific rule that generates triplets of numbers. At the start, participants are told the rule generates the triplet 2-4-6. They are also told that they can propose additional triplets to test, and that they will be given feedback on whether or not those triplets match the generating rule. When they think they have done enough tests, they have one chance to announce their guess at the rule.

Commonly, participants start by proposing triplets such as 6-8-10. It turns out that 6-8-10 matches the rule, so they get positive feedback. However, most people ultimately fail at the task, because the generating rule is 'increasing numbers', whereas most people guess 'sequential even numbers' or 'multiples of 2'. Getting positive feedback on triplets such as 6-8-10 does not confirm the hypothesis 'even numbers', as positive feedback on such a test can be consistent with multiple hypotheses. Triplets such as 3-5-7 would get positive feedback, disconfirming the 'even numbers' hypothesis, yet few people propose such triplets. Triplets such as 6-4-2 would get negative feedback, narrowing the scope of the hypothesis, but even fewer people propose such triplets.

As with the CRT, participants may be tempted to go with their initial intuitive guess. However, in addition to the correctness of this guess — their belief about the rules governing the world of the problem — the 2-4-6 task affords measurement of multiple aspects of the problem-solving process, such as how many test triplets participants propose before announcing their guess. Further, as feedback to the test triplets can be positive or negative, we can measure not only how much evidence participants gather, but what kind. We can also measure their initial confidence (after learning that 2-4-6 matches the rule), as well as their confidence after each piece of feedback.

Finally, the 2-4-6 task is a full-information task: all the evidence for or against a specific hypothesis is available to anyone who wants it: They just have to think to ask for it. This is a useful feature, because one puzzling aspect of science denial is how people form false beliefs when there is ample empirical evidence supporting the scientific claims (Sulik, Efferson, & McKay, 2021).

---

[2]There is some debate on just what it is that this task indexes (McPhetres, 2018). Further, some people manage to form a correct answer intuitively (i.e., without having to engage reflective processing to suppress an incorrect intuition, Bago & De Neys, 2019). However, neither objection undermines our claims about the incompleteness of explanations of science denial based on cognitive traits.

Aims and overview of study

Our proposal is that, *in addition to* measuring traits such as cognitive style, researchers wanting to understand the cognitive aspects of science denial should also measure and analyze the specific mechanisms involved in belief updating: the inferential steps such as hypothesis generation, data gathering and hypothesis evaluation that shed light on *how* the false belief arises. We use the 2-4-6 task here as a proof of concept, but any multi-stage problem-solving task that has such steps might serve for future research. Using such a task will not only improve understanding of how dispositional traits are associated with the specific mechanisms of false belief formation, but also the specific ways in which these mechanisms are associated with false beliefs, even when disconfirmatory evidence is available.

In this study, we have participants complete the 2-4-6 task and a measure of cognitive style. They also respond to questions about science beliefs and political ideology, as this is a core social factor, and as our aim is not to rule out social factors, but rather to refine cognitive explanations.

We hypothesize, not only that overall failure to solve the 2-4-6 task will be associated with higher levels of science denial, but also that data gathering and hypothesis evaluating strategies measured during the 2-4-6 task will be associated with science denial. Finally, we hypothesize that they will predict unique variance in science denial, not accounted for by cognitive style or by political ideology.

Methods

Participants

We recruited participants in several stages as part of a larger project on science denial. Two of those stages are relevant to the tasks described here, and 415 participants did both those stages, though 51 of these failed attention checks, and 58 did not answer all questions, leaving 306 responses to be analyzed. Participants were recruited via Amazon's Mechanical Turk (MTurk) service, with participation managed by Turkprime (Litman, Robinson, & Abberbock, 2017). Participation was limited to those with IP addresses in the USA and over 95% approval rating on MTurk. Remuneration for both stages totaled $5.00. Participants provided informed consent at the beginning of each stage. The study passed the Psychology Department internal ethics procedure at Royal Holloway, University of London.

Procedure

In Stage 1, participants answered four multiple-choice questions (response options: true/false/don't know) on common topics of science denial: climate change, genetic modification of foods, vaccinations, and evolution. The science-denial score is the proportion of responses to these four questions where the participant's belief contradicted scientific consensus. To avoid revealing our interest in science denial, these were interspersed with general science knowledge questions (National Science Board, 2018; Shtulman & Valcarcel, 2012).

We also presented participants with a 7-item version of the CRT, combining a 3-item scale (Shenhav, Rand, & Greene, 2012) and a 4-item scale (Thomson & Oppenheimer, 2016). The analytic style score is the proportion of correct responses to these seven items.

In Stage 2, participants did the 2-4-6 task (Wason, 1960). They were given detailed instructions about the aim of the task (to guess a number-generating rule that we, the experimenters, were thinking of). Instructions included an example rule along with triplets that matched and didn't match the example rule (which was different from the 'increasing numbers' rule in the actual trial). Participants were also given an explanation of what they would do at each stage of the task: either decide if they were ready to guess — which they only had one chance at — or propose another test, in which case they would get feedback and rate how confident they were that they knew the rule. They also provided confidence ratings at the start, just after being given the initial 2-4-6 triplet. Confidence ratings were on a scale from 0 to 100%, with 100% representing absolute certainty. Variables derived from this task include:

1. '2-4-6 success': whether the participant ultimately guessed the rule correctly;

2. 'positive/negative tests': how many test triplets participants proposed, and whether these received positive or negative feedback[3];

3. 'initial confidence': participants' confidence (in whether they knew the rule) at the start of the task when they were told that triplet 2-4-6 matches the rule;

4. 'final confidence': participants' confidence at the point when they decided they had done enough tests, and were ready to guess the rule

As the 2nd–4th of these variables track the process of belief formation, we call them 'process variables' below.

Participants were also asked to rate their political ideology on a 7-point scale ranging from 'Very liberal' to 'Very conservative'.

Results

Descriptive overview & replication of previous findings

Table 1 summarizes the main variables on their original scales. Participants, on average, disagreed with scientific consensus on 31% of the science denial items. On average, they got 71% of the CRT questions correct, but only 7% of participants successfully solved the 2-4-6 problem, so the 2-4-6 is a more complex task. However, one of the virtues of the 2-4-6 task is that the low solution rate does not limit its usefulness

---

[3]This is not the same as whether participants were trying to confirm or disconfirm their hypotheses. That would require asking about their expectations, which we did not do here, though this is a useful avenue for further research.

in analysis, because it affords measurement of variables other than success. These variables explain *why* success was low, and ultimately, the aim is to see if any of them are also associated with science denial.

For instance, participants proposed just 1.97 tests on average before deciding they were ready to guess the rule. Most of these were positive tests (on average, participants proposed 1.66 positive tests; and just 0.31 negative tests). The low number of tests is consistent with findings that people have a general tendency to jump to conclusions (Furl & Averbeck, 2011; Ross, McKay, Coltheart, & Langdon, 2015), and the high proportion of positive tests is consistent with a general confirmation bias (Blanco, Barberia, & Matute, 2015; Jones & Sugden, 2001).

Consistent with previous findings that MTurk participants skew liberal (Berinsky, Huber, & Lenz, 2012), the average ideology rating was 2.27 (on a scale from 0-6, so this average lay between 2='slightly liberal' and 3='neutral').

Table 1: Descriptive overview of scores.

Variable	Mean	Bootstrapped 95% CIs
science denial	0.31	[0.28, 0.34]
analytic style	0.71	[0.67, 0.73]
2-4-6 success	0.07	[0.04, 0.09]
positive tests proposed	1.66	[1.53, 1.94]
negative tests proposed	0.31	[0.24, 0.42]
initial confidence	0.65	[0.61, 0.68]
final confidence	0.88	[0.86, 0.9]
ideology	2.26	[2.08, 2.45]

All numeric variables were standardized for the following analyses.

First, we checked that our data replicate previous findings that science denial is predicted both by social factors and by cognitive traits. We regressed denial on political ideology and analytic style. Both effects were significant (analytic style: $\beta = -0.11$, $t = -2.081$, $p = .038$; ideology: $\beta = 0.454$, $t = 8.549$, $p < .001$; model adjusted $R^2 = .22$). More analytic people displayed less science denial, more conservative people displayed more.

Is science denial predicted by mechanisms of belief formation during problem solving?

When 2-4-6 success was added to the above model, ideology remained a significant predictor ($\beta = 0.452$, $t = 8.518$, $p < .001$), analytic style became borderline ($\beta = -0.104$, $t = -1.964$, $p = .05$) and the effect of 2-4-6 success was not significant ($\beta = -0.058$, $t = -1.138$, $p = .256$; model adjusted $R^2 = .22$). However, the point of introducing the 2-4-6 task was that it affords measurement, not just of ultimate success vs. failure, but also of mechanistic aspects of the belief updating process.

Thus, we regressed science denial on ideology and analytic style, as previously, but added the four 'process variables' — variables that tap performance during the process of belief formation — which include the number of positive-feedback tests and negative-feedback tests proposed, as well as the participants' initial and final confidence. Science denial was still predicted by ideology ($\beta = 0.468$, $t = 8.835$, $p < 0.001$), but the effect of analytic style dropped out ($\beta = -0.099$, $t = -1.868$, $p = .062$). In addition, one process variable — the number of tests proposed that received positive feedback — was a significant predictor of science denial ($\beta = 0.142$, $t = 2.746$, $p = .006$). Proposing positive tests may seem as though it provides extra support for one's hypothesis, but as tests consistent with the commonest wrong hypotheses (e.g., 'increasing even numbers'; 'increasing multiples of 2') are also consistent with the true hypothesis ('increasing numbers'), this data-gathering strategy is misleading and likely to yield false beliefs. Indeed, the point is that this aspect of the belief formation process helps explain *why* people end up with false beliefs.

The other process variables were not significant predictors of science denial (negative-feedback tests $\beta = -0.119$, $t = -1.827$, $p = .069$; initial confidence: $\beta = 0.017$, $t = 0.286$, $p = .775$; final confidence $\beta = -0.08$, $t = -1.435$, $p = .152$; model adjusted $R^2 = .24$). This model with process variables ($AIC = 797.24$) represents an improvement on the initial model with just ideology and analytic style ($AIC = 800.21$, $\Delta_{AIC} = 2.97$).

Do such mechanisms go beyond dispositional traits in explaining science denial?

The CRT is quicker and easier to administer than the 2-4-6 task, so if the latter adds little to our understanding of science denial, it may still be worth focusing on the former. That might be the case if, for instance, analytic style predicts the same specific belief-updating mechanisms that in turn predict science denial. To see whether the process variables related to analytic style are also related to science denial, we conducted a path analysis comprising simultaneous regressions as shown in Fig. 1.

Given the large number of paths in Fig. 1, let us highlight the salient points. The only process variable predicted by analytic style is participants' final confidence ($\beta = 0.116$, $z = 2.047$, $p = 0.041$). However, final confidence does not significantly predict science denial ($\beta = -0.078$, $z = -1.581$, $p = 0.114$). Analytic style does not significantly predict the number of positive tests proposed ($\beta = 0.013$, $z = 0.220$, $p = 0.826$), though this is the process variable that significantly predicted science denial ($\beta = 0.162$, $z = 3.249$, $p = 0.001$). Thus, there is a disjunction between the process variables related to analytic style and the process variables related to science denial, so continuing to focus on instruments such as the CRT will not yield the same understanding as a more process-based or mechanistic approach.

Most of the process variables significantly predict 2-4-6 success (positive tests: $\beta = 0.183$, $z = 3.609$, $p < 0.001$; neg-

ative tests: $\beta = 0.411$, $z = 8.114$, $p < 0.001$; initial confidence: $\beta = -0.126$, $z = -2.499$, $p = 0.012$). This highlights how false beliefs can be explained by multiple mechanisms, whereas errors in single-response tasks like the CRT are underdetermined. The effect of ideology is consistent with the above models ($\beta = 0.45$, $z = 9.049$, $p < 0.001$).

Discussion

Two factors are important for understanding science denial: social information and individual cognition. Previous research on the latter aspect has tended to focus on dispositional traits, such as cognitive style or tendencies towards conspiracist or delusional ideation. We have argued that this leaves an explanatory gap: specific missteps in the belief-formation process are underdetermined by dispositions or traits as general as these, and a false belief may be due to various such missteps. For instance, the same false belief could be due to overconfidence, or due to biases in data gathering. It is useful to know which particular belief-formation mechanisms are associated with science denial, and this question is not adequately addressed by the current focus on dispositional traits.

We have shown that one such mechanism, a tendency to do more positive tests, significantly predicts science denial. However, a more important result for theory-development regarding science denial is that the role of this mechanism does not reduce to an effect of cognitive style. This is what we meant when we said that the specific mechanisms of false belief formation in the context of science denial are underdetermined by a focus on dispositional traits.

For the sake of brevity, given the scope of this paper, we have focused on a single dispositional variable (analytic cognitive style), a single example of complex problem solving with various measurable subparts (the 2-4-6 task), and a single social factor (political ideology). Thus, we consider these results to be more of a proof-of-concept, motivating the proposal that a promising direction for future work on cognition and science denial will be to study how people go about forming beliefs or solving problems, rather than just measuring the outcome belief or success at the problem.

It is true that some previous research has focused on specific belief-formation mechanisms (e.g., Prike et al., 2018; Sulik et al., 2021; Sulik, Ross, & McKay, 2020; Wagner-Egger, Delouvée, Gauvrit, & Dieguez, 2018), but a benefit of the 2-4-6 task, relative to these studies, is that it combines multiple such mechanisms in a single task that is complex and open-ended. It is, effectively, like doing science, in as far as it involves generating hypotheses, gathering data, and testing the hypotheses. This parallel between science and the 2-4-6 task has been described as dual search through both a hypothesis space and an experiment space (Klahr & Dunbar, 1988), and it is one reason we have focused on this task for our proof-of-concept.

One puzzling aspect of science denial is that we all inhabit the same objective reality, and yet still arrive at incompatible beliefs about that reality. In the 2-4-6 task, the relevant infor-
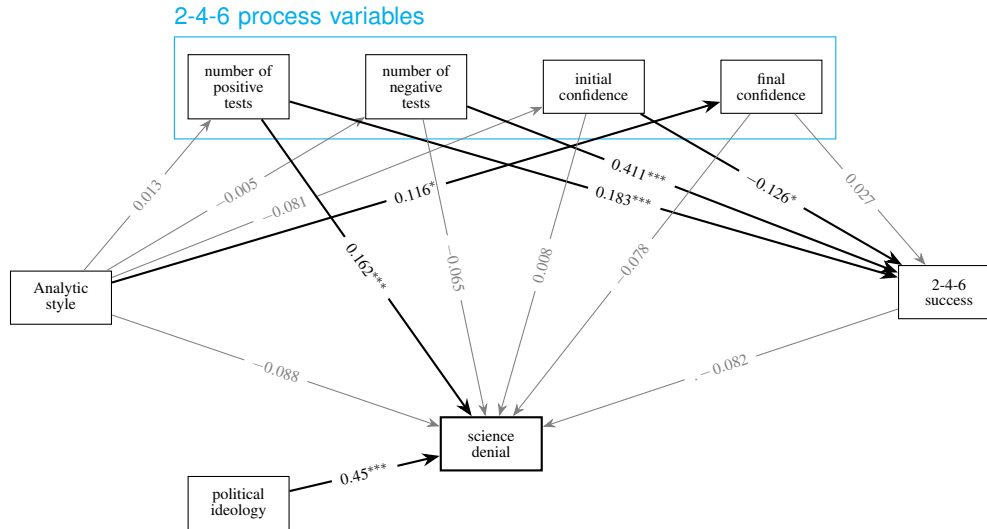
Figure 1: Path analysis with paths from analytic style to specific process variables in the 2-4-6 task, and from all variables to science denial. Path annotations are standardized regression coefficients, with significant paths in black (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

mation for solving the task is available to participants: They can propose whatever kind of (and as many) tests as they wish, and they always receive veridical feedback. And yet, because they gather and process information suboptimally, they can arrive at demonstrably false beliefs about the same underlying phenomenon (here, a number-generating rule).

We intend this as a complement to, rather than a replacement for, approaches focused on dispositional traits. For instance, it is quite plausible that a larger sample size would have revealed a significant effect of analytic style on science denial, but even then, this would not undermine the finding that analytic style does not predict the same process variables that predict science denial. We also acknowledge that socio-cognitive factors (such as social-dominance orientation, Jylhä, Cantal, Akrami, & Milfont, 2016) are important, but as such factors do not directly speak to the question of belief-formation mechanisms, they are somewhat orthogonal to our aim.

Future work would benefit from more nuanced quantification or categorization of data-gathering and hypothesis-testing strategies than we have described here. For instance, we have not discussed moves in experiment space (such that proposing 4-6-8 followed by 6-8-10 is a smaller move in experiment space than proposing 4-6-8 followed by 9-11-13). One thing that determines the rationality of such moves is the sparsity of the hypothesis space (Navarro & Perfors, 2011), and as we only included one such task, we did not manipulate this sparsity. We also did not ask participants for their leading hypothesis at each stage, or their prediction for whether the next piece of feedback would be positive or negative (cf. Gorman, Stafford, & Gorman, 1987). It would be useful to develop a task that has the useful features described above, but that tracks expectations about evidence better, varies in

difficulty, is less math-focused, or is more engaging. Finally, and more speculatively, it would be interesting to learn if manipulations that improve performance at the 2-4-6 task, such as setting up contrast classes (Gale & Ball, 2012) or conflicting perspectives (Lien & Lin, 2011), can be leveraged to improve public understanding of science.

References

Bago, B., & De Neys, W. (2019). The smart system 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, *25*(3), 257–299.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's mechanical turk. *Political analysis*, *20*(3), 351–368.

Blanco, F., Barberia, I., & Matute, H. (2015). Individuals who believe in the paranormal expose themselves to biased information and develop more causal illusions than nonbelievers in the laboratory. *PloS One*, *10*(7), e0131378.

Bronstein, M. V., Pennycook, G., Bear, A., Rand, D., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*, *8*, 108–117.

Ding, D., Maibach, E. W., Zhao, X., Roser-Renouf, C., & Leiserowitz, A. (2011). Support for climate policy and society action are linked to perceptions about scientific agreement. *Nature Climate Change*, *1*, 462–466.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25-42.

Furl, N., & Averbeck, B. B. (2011). Parietal cortex and insula relate to evidence seeking relevant to reward-related decisions. *Journal of Neuroscience*, *31*(48), 17572–17582.

Gale, M., & Ball, L. J. (2012). Contrast class cues and performance facilitation in a hypothesis-testing task: evidence for an iterative counterfactual model. *Memory & Cognition*, *40*, 408–419.

Gervais, W. M. (2015). Override the controversy: Analytic thinking predicts endorsement of evolution. *Cognition*, *142*, 312–321.

Gorman, M. E., Stafford, A., & Gorman, M. E. (1987). Disconfirmation and dual hypotheses on a more difficult version of wason's 2–4–6 task. *The Quarterly Journal of Experimental Psychology Section A*, *39*(1), 1–28.

Hartman, R. O., Dieckmann, N. F., Sprenger, A. M., Stastny, B. J., & DeMarree, K. G. (2017). Modeling attitudes toward science: development and validation of the credibility of science scale. *Basic and Applied Social Psychology*, *39*(6), 358–371.

Jolley, D., & Douglas, K. M. (2014). The social consequences of conspiracism: Exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprint. *British Journal of Psychology*, *105*(1), 35–56.

Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision*, *50*(1), 59–99.

Jylhä, K. M., Cantal, C., Akrami, N., & Milfont, T. L. (2016). Denial of anthropogenic climate change: Social dominance orientation helps explain the conservative male effect in brazil and sweden. *Personality and Individual Differences*, *98*, 184–187.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgement and Decision Making*, *8*, 407–424.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive science*, *12*(1), 1–48.

Kubin, E., Puryear, C., Schein, C., & Gray, K. (2021). Personal experiences bridge moral and political divides better than facts. *Proceedings of the National Academy of Sciences*, *118*(6), e2008389118.

Lewandowsky, S., Gignac, G. E., & Vaughan, S. (2013). The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change*, *3*(4), 399–404.

Lewandowsky, S., & Oberauer, K. (2016). Motivated rejection of science. *Current Directions in Psychological Science*, *25*(4), 217–222.

Lien, Y.-W., & Lin, W.-L. (2011). From falsification to generating an alternative hypothesis: Exploring the role of the new-perspective hypothesis in successful 2-4-6 task performance. *Thinking & Reasoning*, *17*(2), 105–136.

Litman, L., Robinson, J., & Abberbock, T. (2017). Turkprime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442.

Lobato, E., Mendoza, J., Sims, V., & Chin, M. (2014). Examining the relationship between conspiracy theories, paranormal beliefs, and pseudoscience acceptance among a university population. *Applied Cognitive Psychology*, *28*(5), 617-625.

McPhetres, J. (2018). What does the cognitive reflection test really measure: A process dissociation investigation.

McPhetres, J., & Pennycook, G. (2019, November 19). Science beliefs, political ideology, and cognitive sophistication. doi: https://doi.org/10.31219/osf.io/ad9v7

National Science Board. (2018). *Science & engineering indicators 2018.* Retrieved 3/6/2019, from https://www.nsf.gov/statistics/2018/nsb20181/report

Navarro, D., & Perfors, A. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychological Review*, *118*(1), 120-134.

Pasek, J. (2018). It's not my consensus: Motivated reasoning and the sources of scientific illiteracy. *Public Understanding of Science*, *27*(7), 786–806.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current Diretions in Psychological Science*, *24*(6), 425-432.

Peters, E., Joseph, S., Day, S., & Garety, P. (2004). Measuring delusional ideation: the 21-item peters et al. delusions inventory (pdi). *Schizophrenia Bulletin*, *30*(4), 1005.

Prike, T., Arnold, M. M., & Williamson, P. (2018). The relationship between anomalistic belief and biases of evidence integration and jumping to conclusions. *Acta Psychologica*, *190*, 217–227.

Ross, R. M., McKay, R., Coltheart, M., & Langdon, R. (2015). Jumping to conclusions about the beads task? A meta-analysis of delusional ideation and data-gathering. *Schizophrenia Bulletin*, *41*(5), 1183–1191.

Rutjens, B. T., Sutton, R. M., & van der Lee, R. (2018). Not all skepticism is equal: Exploring the ideological antecedents of science acceptance and rejection. *Personality and Social Psychology Bulletin*, *44*(3), 384–405.

Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in god. *Journal of Experimental Psychology: General*, *141*(3), 423–428.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*(2), 209–215.

Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). Performance on the cognitive reflection test is stable across time. *Judgment and Decision making*, *13*(3), 260.

Ståhl, T., & van Prooijen, J.-W. (2018). Epistemic rationality: Skepticism toward unfounded beliefs requires sufficient cognitive ability and motivation to be rational. *Personality and Individual Differences*, *122*, 155–163.

Sulik, J., Efferson, C., & McKay, R. (2021). Collectively

jumping to conclusions: Social information amplifies the tendency to gather insufficient data. *Journal of Experimental Psychology: General*.

Sulik, J., Ross, R., & McKay, R. (2020). The contingency illusion bias as a potential driver of science denial. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd annual conference of the cognitive science society* (pp. 829–835). Cognitive Science Society.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

Wagner-Egger, P., Delouvée, S., Gauvrit, N., & Dieguez, S. (2018). Creationism and conspiracism share a common teleological bias. *Current Biology*, *28*(16), R867–R868.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140.