

Factual Knowledge can reduce Attitude Polarization

Corresponding Author: Dr Michael Stagnaro

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This manuscript presents new evidence that speaks to an important debate about factual beliefs and motivated reasoning around contentious political topics. One of the big challenges in disentangling politically motivated reasoning (PMR) vs. rational updating is the possibly confounding role of factual beliefs (see Pennycook & Rand 2021, Box 2). This study cleverly aims to settle this debate by incentivizing learning about a variety of facts related to the selected topic (gun control) and then measuring effects on attitudes. The authors find that participants widely engage with facts and that there is a mostly depolarizing effect on attitudes. Interestingly, there does not seem to be any movement on affective polarization.

I think this is a good contribution and will help to move forward the debate on this topic. I have one comment on the presentation of the results. Given that the authors incentivize correct responses on the factual Qs (which should be emphasized in the main text), it is not surprising that there were strong learning effects across the board. I interpreted these findings (i.e., questions 1 and 2 in the text) as manipulation checks: the treatment was indeed powerful and worked as intended. I would caution against a conclusion here that people generally want to engage with diverse facts and are willing to learn counter-attitudinal information. Rather, the authors have created an important counterfactual by incentivizing people to (perhaps unnaturally) learn these facts, then observing the resulting changes in attitudes.

Analysis:

- Though overall attrition seems fairly low, I would like to see tests for differential attrition by condition.
- Given that the study is somewhat underpowered, I do not understand why the main reported estimates are from the unadjusted models.

A thought for the future: It would be interesting to build on this design by creating different learning incentive conditions: incentives only for pro-gun facts, only for anti-gun facts, etc. This would help integrate the findings reported here with some of the other cited literature.

Reference:

Pennycook, G. and Rand, D.G., 2021. The psychology of fake news. Trends in cognitive sciences, 25(5), pp.388-402.

Reviewer #2

(Remarks to the Author)

This manuscript explores the effect of political knowledge related to the topic of gun control on attitude polarization on the issue. It reports the results from a two-condition between-subjects experiment in which a representative sample of US partisans was randomly assigned to a treatment that included "a training session" that exposed participants to a high volume of competing issue-relevant information that included verifiable facts and in which those respondents were paid for reporting correct answers to increase relevant learning. The treatments included a mix of arguments for and against people's initial attitudes to assess what types of information people attend to, how much they "value" the information, how much they learn, and how that information affects their attitudes on the issue. The study included a second wave administered one month

following the initial survey to evaluate the longevity of the effects and retention of topic-relevant information.

The paper is framed as a test of the robustness of empirical findings that have been identified linking higher levels of knowledge to increased levels of attitude polarization via “partisan” (directional) motivated reasoning (e.g., climate change attitude polarization); more knowledge or higher levels of cognitive reflection or numeracy also have widely been shown to generate increased forms of “directional” motivated reasoning. The manuscript points out that prior correlational evidence that measures knowledge and that has identified interaction effects with “fact-based” treatments has often shown that higher levels of attitude polarization due to these individuals who are higher in knowledge/sophistication attending more closely to their partisan in-group’s position on any issue (notably, most studies do not directly manipulate the underlying “directional” motivation as explained by Druckman and McGrath 2019; and Kahan 2015/2016). The novel experimental design employed in this work was designed to exogenously vary factual knowledge about gun control relevant information to determine its independent causal effect on related belief polarization. The results demonstrate, counter to what has been demonstrated in prior work, that increases in factual knowledge have a depolarizing effect and that people often attend to and learn from counter-attitudinal arguments on this issue, and that these effects persist over time.

The results are described as “quite contrary to what PMR theory would predict” (line 235); however, a concern is that “people engaged with” or “attended to” and “valued” information that was counter to their general attitude because they were being paid for correct answers. By offering incentives for correct answers to the knowledge battery, this likely shifted the underlying (baseline?) “motivation” in how respondents engaged with the mix of arguments and verifiable facts on the issue – that is, couldn’t that induce an “accuracy motivation” if respondents learn in the training session prior to evaluating the mix of arguments that they will be incentivized to learn and answer questions correctly? If so, it seems a problematic confound in the two-condition design in which the treatment is not only manipulating exposure to a lengthy battery of “mixed” information, but also the underlying motivation related to processing any arguments/facts/ information on the issue. In addition, it has been demonstrated that exposure to mixes of strong competing arguments can lead to opinion moderation when people are motivated to attend to the information itself (as opposed to being motivated / induced to process information with the “directional” goal of defending one’s partisan identity) (see Druckman, 2012 for a review on motivated reasoning; and, Druckman 2014, for a broader discussion of “normative” criteria scholars have used to evaluate the quality of opinions, such as greater “factual knowledge”). Given that this study does not directly manipulate respondents’ underlying information processing motivation/goal, and that motivated reasoning theory does not assume that “partisan motivations” are the only motivations employed in political reasoning across all “contexts”, it does not provide information that is necessarily contrary to motivated reasoning theory more generally (e.g., are partisans “always” selectively resistant to counter-attitudinal facts (line 256) – and, what about the many moderators of PMR). The design and results could be seen as accentuating how generating an “accuracy motivation” (through monetary incentives) may exert a depolarizing impact on evaluating mixes of arguments with verifiable facts on a highly salient and controversial issue.

I’d like to see the exact timing of this study made more explicit in the main text. I was curious about when the waves were fielded (since preregistration was dated 8/21), especially given the many high-profile events surrounding this issue (tragic school shootings, etc.) and related legislation that was passed with bi-partisan support in 2022. Is it possible that the “context” (or any major tragic events reported in national news) surrounding this issue when the studies were fielded had any impact on the results?

In sum, the manuscript reports the results from a novel experiment to test pre-registered hypotheses about the depolarizing effect of increased factual knowledge on related attitudes about gun control. The manuscript could be re-framed to address some of my concerns above, but the experimental design is problematic due to the inclusion of incentives for attending to and learning from the treatments, and the possibility that this aspect of the design introduced an accuracy motivation among those in the control condition. Regrettably, I am not convinced from the results that “the impact of directional motivations on political cognitions may be more limited than previously thought” (concluding paragraph), nor that depolarization will be as “easy” as increasing the “political and scientific literacy” of the population on issues in highly polarized settings. The design does not allow for a “direct test of politically motivated reasoning”, and therefore, I do not recommend it for further consideration at this time.

Druckman, James N. "The politics of motivation." *Critical Review* 24.2 (2012): 199-216.

Druckman, James N. "Pathologies of studying public opinion, political communication, and democratic responsiveness." *Political Communication* 31, no. 3 (2014): 467-492.

Reviewer #3

(Remarks to the Author)

This is a polished, well-executed paper that taught me something new about how people use factual knowledge to inform their attitudes about public policy. I have some suggestions for how the theoretical framework could be further developed, which the authors hopefully find useful.

The paper is primarily interested in pushing back against the politically motivated reasoning (PMR) hypothesis that could explain the relationship between political knowledge and political polarization. Under the PMR hypothesis, political knowledge leaves individuals better-equipped to incorporate pro-attitudinal information into their belief systems and reject counter-attitudinal information. If the PMR hypothesis is right, then more information — even balanced information — should lead to more polarized attitudes. The experimental evidence the authors provide convincingly shows that this is not the case:

exposure to balanced information leads to less polarized attitudes.

So what, then, would account for this correlation? Put another way, how would the authors draw their DAG? I see a few potential alternatives because, as the authors lay out, there are many good reasons to be skeptical of the PMR hypothesis. The first is that it is not issue-specific knowledge at all. Rather, general political socialization/interest increases both political knowledge and attitude extremity. We see a correlation between political knowledge and attitude polarization because as someone becomes a more “political person” both of those things increase independent from one another. Another related possibility is that we observe this relationship due to selective exposure to pro-attitudinal information (especially when presented with counter-attitudinal information; Garrett 2009), which is less about how information is processed in the PMR account, and more about how politically knowledgeable people decide what information to consume in the first place. The authors are experimentally exposing people to balanced information, and the reason they find significant effects is precisely because people don’t typically choose to consume a balanced information diet. This is not a criticism of the experiment itself — quite the opposite! — but this is to say that I think the authors could do more to develop some alternatives regarding what *could be* accounting for this relationship rather than solely focusing on why the PMR hypothesis is wrong.

All this is to say, I don’t have any issues with the analysis as it was conducted or presented, I could have just used a bit more theoretical development beyond what is being taken down.

Reference

Garrett, R. Kelly. 2009. “Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate.” *Journal of Communication* 59: 676-699

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Thanks to the authors for addressing all of my comments. I'm satisfied with their changes and think this is a strong contribution.

Reviewer #2

(Remarks to the Author)

The manuscript has been significantly improved from the revisions. I appreciate the authors' responses to my earlier comments and the clarification that was provided in the revised manuscript to reflect what the experiment demonstrates. The results contribute to the existing literature by showing that incentivizing learning can lead to depolarized attitudes on a controversial topic. I thank the authors for clarifying that there was not a confound in the study because learning was incentivized for everyone and the facts provided did not consist of competing arguments. It was helpful to read the other comments and the responses from the authors. The shift in focus on the effects of learned information on attitudes is a major improvement. After reading the responses to the Reviewers and the paper several times over the last week, I am persuaded that the revised manuscript makes an important contribution to the existing corpus of literature on this topic.

Reviewer #3

(Remarks to the Author)

I found “Factual Knowledge and Attitude Polarization: An Experimental Test of Politically Motivated Reasoning” to be an interesting and well-executed paper on the first read, and had some suggestions for theoretical development. Those suggestions have been addressed in this revision, which I am even more favorable toward on a second read.

If I may offer one more suggestion for theoretical development that might help situate this manuscript in the broader literature, it would be to draw a sharper distinction between information *seeking* and information *processing* — or between *selective exposure* and *motivated reasoning*. This paper is precisely about the latter, showing that individuals attend to both pro- and counter-attitudinal information in similar ways when they are exposed to it. But some of the work the authors are putting themselves in conversation with (Peterson and Lyengar 2021, e.g.) is about the former (or even conflates the two), showing that individuals are — for potentially a variety of reasons — likelier to encounter pro-attitudinal information. I think it’s easy to conflate these two things, and for the authors to sharpen this theoretical contrast could arguably be a contribution in and of itself. Put another way, the authors helpfully show that even for individuals who we’d expect to prefer to *encounter* pro-attitudinal factual information, they *process* pro- and counter-attitudinal factual information the same way. This suggests a need for greater precision in what we mean when we say “motivated reasoning” — it isn’t a catch-all explanation for political differences, and when we precisely define it as politically biased information processing it’s less prevalent than the catch-all term implies.

In any case, this is a relatively minor point — certainly not worth holding things up over.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Replies to Reviewers

Reviewer #1:

This manuscript presents new evidence that speaks to an important debate about factual beliefs and motivated reasoning around contentious political topics. One of the big challenges in disentangling politically motivated reasoning (PMR) vs. rational updating is the possibly confounding role of factual beliefs (see Pennycook & Rand 2021, Box 2). This study cleverly aims to settle this debate by incentivizing learning about a variety of facts related to the selected topic (gun control) and then measuring effects on attitudes. The authors find that participants widely engage with facts and that there is a mostly depolarizing effect on attitudes. Interestingly, there does not seem to be any movement on affective polarization.

I think this is a good contribution and will help to move forward the debate on this topic. I have one comment on the presentation of the results. Given that the authors incentivize correct responses on the factual Qs (which should be emphasized in the main text), it is not surprising that there were strong learning effects across the board. I interpreted these findings (i.e., questions 1 and 2 in the text) as manipulation checks: the treatment was indeed powerful and worked as intended. I would caution against a conclusion here that people generally want to engage with diverse facts and are willing to learn counter-attitudinal information. Rather, the authors have created an important counterfactual by incentivizing people to (perhaps unnaturally) learn these facts, then observing the resulting changes in attitudes.

We have refocused our argument from how some people want to engage with diverse facts and learn counter-attitudinal information, to how individuals are simply *willing* to do so when given sufficient (yet minimal) incentives. The study is more focused now on the effects of the learned information on attitudes.

We also note, in the first paragraph of the Results section, that the learning analyses we present can be seen as a type of manipulation check, showing that the treatment was indeed powerful and worked as intended. As noted on page 4, we present these analyses in detail because we want to know not only whether people learned, but also how much and what type of facts they have retained (given the diverse set of facts we presented to them).

Analysis:

- Though overall attrition seems fairly low, I would like to see tests for differential attrition by condition.

SI section 2 presents analyses showing that all results hold when using instrumental variables to estimate the complier average treatment effects, which takes into account non-compliance (that is, those who attrited from the main study but for whom we obtained outcome scores right before leaving). On page 4, we refer readers to this supplemental analysis.

With regard to differential attrition, for the 8% of the overall sample who dropped out but from whom we were able to preserve their outcomes, we found no evidence of differential attrition ($p = .57$), with 40 individuals dropping out of the control group and 38 dropping out of the treatment group. Of the .5% of the overall sample who dropped out for whom we were *not* able to preserve their outcomes, one individual dropped out of the treatment group and 5 dropped out of the control group. These numbers are at such low rates (less than 1%) that there is little reason to think they could be driving any affect.

- Given that the study is somewhat underpowered, I do not understand why the main reported estimates are from the unadjusted models.

As reported in the pre-registered protocol (available here: https://aspredicted.org/WGT_BQ2), to determine the necessary sample size for our study, we have conducted a priori power analysis based on a pilot study. The result of that analysis indicated that to detect our anticipated treatment effects with 80% power and a conventional alpha of .05, a sample of 787 respondents would be needed. To ensure we have sufficient power and to also increase precision, we aimed for a final sample size of $N = 1,000$, and ended up with $N = 1,011$. We thus believe our study is sufficiently powered.

As for unadjusted models: if the reviewer refers to survey weights, they are unnecessary in this context because we recruited a quota matched, representative sample while maintaining very low levels of attrition. If, however, the reviewer refers to adjusting for covariates, SI section 1 presents analyses indicating that our results either do not change or become even stronger when including control variables in the model. If the reviewer is referring to some other quantity of interest, we apologize for the lack of understanding and would be happy to include additional models.

A thought for the future: It would be interesting to build on this design by creating different learning incentive conditions: incentives only for pro-gun facts, only for anti-gun facts, etc. This would help integrate the findings reported here with some of the other cited literature.

We are currently working on that exact study and are heartened that the reviewer shares the same intuition of such a study's value.

Reference: Pennycook, G. and Rand, D.G., 2021. The psychology of fake news. Trends in cognitive sciences, 25(5), pp.388-402.

Reviewer #2:

This manuscript explores the effect of political knowledge related to the topic of gun control on attitude polarization on the issue. It reports the results from a two-condition between-subjects experiment in which a representative sample of US partisans was randomly assigned to a treatment that included “a training session” that exposed participants to a high volume of competing issue-relevant information that included verifiable facts and in which those respondents were paid for reporting correct answers to increase relevant learning. The treatments included a mix of arguments for and against people’s initial attitudes to assess what types of information people attend to, how much they “value” the information, how much they learn, and how that information affects their attitudes on the issue. The study included a second wave administered one month following the initial survey to evaluate the longevity of the effects and retention of topic-relevant information.

The paper is framed as a test of the robustness of empirical findings that have been identified linking higher levels of knowledge to increased levels of attitude polarization via “partisan” (directional) motivated reasoning (e.g., climate change attitude polarization); more knowledge or higher levels of cognitive reflection or numeracy also have widely been shown to generate increased forms of “directional” motivated reasoning. The manuscript points out that prior correlational evidence that measures knowledge and that has identified interaction effects with “fact-based” treatments has often shown that higher levels of attitude polarization due to these individuals who are higher in knowledge/sophistication attending more closely to their partisan in-group’s position on any issue (notably, most studies do not directly manipulate the underlying “directional” motivation as explained by Druckman and McGrath 2019; and Kahan 2015/2016). The novel experimental design employed in this work was designed to exogenously vary factual knowledge about gun control relevant information to determine its independent causal effect on related belief polarization. The results demonstrate, counter to what has been demonstrated in prior work, that increases in factual knowledge have a depolarizing effect and that people often attend to and learn from counter-attitudinal arguments on this issue, and that these effects persist over time.

The results are described as “quite contrary to what PMR theory would predict” (line 235); however, a concern is that “people engaged with” or “attended to” and “valued” information that was counter to their general attitude because they were being paid for correct answers. By offering incentives for correct answers to the knowledge battery, this likely shifted the underlying (baseline?) “motivation” in how respondents engaged with the mix of arguments and verifiable

facts on the issue – that is, couldn't that induce an "accuracy motivation" if respondents learn in the training session prior to evaluating the mix of arguments that they will be incentivized to learn and answer questions correctly?

If so, it seems a problematic confound in the two-condition design in which the treatment is not only manipulating exposure to a lengthy battery of "mixed" information, but also the underlying motivation related to processing any arguments/facts/ information on the issue.

We completely agree that our use of incentives for correct responding induced an accuracy mindset. As we now discuss on page 10 of the revised manuscript, this was in fact a key feature of our design. We realize that we did not clearly explain the rationale or implementation in our original manuscript.

Most importantly, from a methodological perspective, we incentivized correct responding - and thus induced an accuracy mindset - in both the treatment AND the control conditions, and the only difference between the groups was the topic of information presented. Thus, the accuracy mindset does not present a confound. Furthermore, we emphasize that the incentives were just to correctly recall of the information, which is not one of our key outcomes. That is, the attitude measures, which *are* our focus, were not incentivized - and so, what we document is that getting people to pay attention to facts via a small incentive results in meaningful changes in subsequent (non-incentivized) attitudes.

From a theoretical perspective, we also appreciate your feedback which helped us realize that we were not precise in the framing of the prior version of the manuscript. We agree that our results do not mean that PMR never occurs under any condition. Rather than disproving PMR, we provide bounds on it: we show that a small incentive (just a few dollars) is sufficient to induce enough of an accuracy mindset to allow factual information to be encoded and subsequently have a lasting impact on attitudes for the extremely contentious topic of gun control. This suggests that to the extent that PMR does occur in other contexts, it is not insurmountable (see the revised Discussion section, especially page 11). That is, in contrast to the perspective often presented by many academics and popular press, our results suggest that attitudes on highly polarized topics can be changed with exposure to credible facts. We show that, when accompanied by a small incentive to pay attention to them, facts have power.

We have substantially revised the manuscript to make all of this clearer and to reflect this understanding throughout.

Given that this study does not directly manipulate respondents' underlying information processing motivation/goal, and that motivated reasoning theory does not assume that "partisan motivations" are the only motivations employed in political reasoning across all "contexts", it does not provide information that is necessarily contrary to motivated reasoning theory more generally (e.g., are

partisans “always” selectively resistant to counter-attitudinal facts (line 256) – and, what about the many moderators of PMR).

In addition, it has been demonstrated that exposure to mixes of strong competing arguments can lead to opinion moderation when people are motivated to attend to the information itself (as opposed to being motivated / induced to process information with the “directional” goal of defending one’s partisan identity) (see Druckman, 2012 for a review on motivated reasoning; and, Druckman 2014, for a broader discussion of “normative” criteria scholars have used to evaluate the quality of opinions, such as greater “factual knowledge”). The design and results could be seen as accentuating how generating an “accuracy motivation” (through monetary incentives) may exert a depolarizing impact on evaluating mixes of arguments with verifiable facts on a highly salient and controversial issue.

As noted in the manuscript (see page 4 as well as the entire Methods section), there were no arguments being deployed in this corpus of information; we presented factual content and allowed study participants to decide how to use this material to inform their perspective. We hope the reviewer can see how this is meaningfully different. While past work shows people can be swayed, indifferent, or overwhelmed by multiple arguments, simply presenting them with balanced factual content and letting them decide how to update their perspective on the topic (versus guide their updating via argumentation), we believe, is both more foundational to how people engage in basic day-to-day decision making (e.g. in an educational contexts where individuals learn to both expand their understanding, as well as obtain better grads), and is a stronger test for motivated reasoning. Given this information is presented on its own, with no directional goals for how one “should” interpret it, people have many more directions to take the motivated path compared to situations where they are exposed to a single directional argument.

I’d like to see the exact timing of this study made more explicit in the main text. I was curious about when the waves were fielded (since preregistration was dated 8/21), especially given the many high-profile events surrounding this issue (tragic school shootings, etc.) and related legislation that was passed with bi-partisan support in 2022. Is it possible that the “context” (or any major tragic events reported in national news) surrounding this issue when the studies were fielded had any impact on the results?

We now report, on page 11, that the study was conducted in September 2021, well before the 2022 legislation the reviewer mentions here. Given the variation in definition of mass shootings, it is not always easy to quantify the number of events in a given time frame. Across a number of data-bases, there were several mass shootings during 2021, with notable

variation in coverage. Still, we believe that there has been consistent and increasing concern regarding this topic and, as we show in SI section 11, this topic is quite representative of a highly polarized issue. Further, by testing two waves, we attempt to have our results generalizable beyond a single time frame.

Reviewer #3:

This is a polished, well-executed paper that taught me something new about how people use factual knowledge to inform their attitudes about public policy. I have some suggestions for how the theoretical framework could be further developed, which the authors hopefully find useful.

The paper is primarily interested in pushing back against the politically motivated reasoning (PMR) hypothesis that could explain the relationship between political knowledge and political polarization. Under the PMR hypothesis, political knowledge leaves individuals better-equipped to incorporate pro-attitudinal information into their belief systems and reject counter-attitudinal information. If the PMR hypothesis is right, then more information — even balanced information — should lead to more polarized attitudes. The experimental evidence the authors provide convincingly shows that this is not the case: exposure to balanced information leads to less polarized attitudes.

So what, then, would account for this correlation? Put another way, how would the authors draw their DAG? I see a few potential alternatives because, as the authors lay out, there are many good reasons to be skeptical of the PMR hypothesis. The first is that it is not issue-specific knowledge at all. Rather, general political socialization/interest increases both political knowledge and attitude extremity. We see a correlation between political knowledge and attitude polarization because as someone becomes a more “political person” both of those things increase independent from one another. Another related possibility is that we observe this relationship due to selective exposure to pro-attitudinal information (especially when presented with counter-attitudinal information; Garrett 2009), which is less about how information is processed in the PMR account, and more about how politically knowledgeable people decide what information to consume in the first place. The authors are experimentally exposing people to balanced information, and the reason they find significant effects is precisely because people don’t typically choose to consume a balanced information diet. This is not a criticism of the experiment itself — quite the opposite! — but this is to say that I think the authors could do more to develop some alternatives regarding what *could be* accounting for this relationship rather than solely focusing on why the PMR hypothesis is wrong.

We’ve added a paragraph to the Discussion section (see page 11) where we attempt to account for the discrepancy between the results of prior studies – which find a positive

correlation between domain-specific knowledge and attitude polarization – and our results. The text suggests that this might be due to a third variable affecting both knowledge and polarization (i.e., political interest) or because of greater partisan selective exposure among more knowledgeable individuals. Our experimental design attempts to address these two things.

Reviewer #1 (Remarks to the Author) Thanks to the authors for addressing all of my comments. I'm satisfied with their changes and think this is a strong contribution.

Reply: We thank the reviewer for their helpful suggestions and are happy they are satisfied with these changes.

Reviewer #2 (Remarks to the Author) The manuscript has been significantly improved from the revisions. I appreciate the authors' responses to my earlier comments and the clarification that was provided in the revised manuscript to reflect what the experiment demonstrates. The results contribute to the existing literature by showing that incentivizing learning can lead to depolarized attitudes on a controversial topic. I thank the authors for clarifying that there was not a confound in the study because learning was incentivized for everyone and the facts provided did not consist of competing arguments. It was helpful to read the other comments and the responses from the authors. The shift in focus on the effects of learned information on attitudes is a major improvement. After reading the responses to the Reviewers and the paper several times over the last week, I am persuaded that the revised manuscript makes an important contribution to the existing corpus of literature on this topic.

Reply: We thank the reviewer for their helpful suggestions and feel our manuscript is better for their comments. Specifically, they helped us see the need for greater clarification. We are very happy the reviewer sees this work as an important contribution.

Remaining comments by R3

Reviewer #3 (Remarks to the Author) I found “Factual Knowledge and Attitude Polarization: An Experimental Test of Politically Motivated Reasoning” to be an interesting and well-executed paper on the first read, and had some suggestions for theoretical development. Those suggestions have been addressed in this revision, which I am even more favourable toward on a second read.

If I may offer one more suggestion for theoretical development that might help situate this manuscript in the broader literature, it would be to draw a sharper distinction between information **seeking** and information **processing** — or between **selective exposure** and **motivated reasoning**. This paper is precisely about the latter, showing that individuals attend to both pro- and counter-attitudinal information in similar ways when they are exposed to it. But some of the work the authors are putting themselves in conversation with (Peterson and Iyengar 2021, e.g.) is about the former (or even conflates the two), showing that individuals are — for potentially a variety of reasons — likelier to encounter pro-attitudinal information. I think it's easy to conflate these two things, and for the authors to sharpen this theoretical contrast could arguably be a contribution in and of itself. Put another way, the

authors helpfully show that even for individuals who we'd expect to prefer to *encounter* pro-attitudinal factual information, they *process* pro- and counter-attitudinal factual information the same way. This suggests a need for greater precision in what we mean when we say "motivated reasoning" — it isn't a catch-all explanation for political differences, and when we precisely define it as politically biased information processing it's less prevalent than the catch-all term implies.

Reply: We thank the reviewer for this helpful suggestion and agree this is an important distinction. In response, we added the following text to the Discussion section (page 13): "Our results also suggest a need for greater clarity and precision in what we mean when using the term "politically motivated reasoning." In the existing literature, the term is often used to describe different things, including people's tendency to actively seek pro-attitudinal information and avoid counter-attitudinal information (i.e., selective exposure; Peterson & Iyengar, 2021) and the way people engage with pro- and counter-attitudinal information once they encounter it (i.e., information processing; Taber & Lodge, 2006). In this paper, we have focused on the latter. We show that once individuals have been exposed to both pro- and counter-attitudinal information, they attend to, internalize, and update on both types of facts. Hence, when PMR is defined as politically biased information processing, it is less prevalent than the catch-all term implies."