

Peer Review File

Manuscript Title: Historical narratives about the COVID-19 pandemic are motivationally biased

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

This paper reports three studies examining how recall of various pandemic-related perceptions and behaviors is influenced by current assessments and by identification with vaccination status. Using large sample sizes numbering in the thousands, the studies provide clear and consistent evidence of biased recall such that what people remembered about items including infection probability and severity, mask wearing behavior, and trust in government and science among others was strongly skewed by their current perceptions and their vaccination status. The effect was resistant to a cash incentive or providing information about biased recall. A third study involving participants from 10 countries revealed evidence of bias by comparing current recall of probability and severity of infection as well as government effectiveness to benchmark values obtained in a separate 2020 survey.

The studies reported here are both novel and timely; they should be of broad interest to researchers from various disciplines and also society more generally. The writing is clear, the abstract is informative, conclusions are appropriate, and some limitations of the study are acknowledged. The methods used to obtain the key data are solid, the evidence for recall bias is robust, and the statistical/analytic procedures used to demonstrate these effects are rigorous and appropriate. The depiction of the key results on biased recall in Figures 1 and 2 is economical, but not highly intuitive, at least to this reader – it took me some time to fully grasp exactly what was being depicted and to map the lines depicted to the individual conditions. Figuring out the meaning of “predicted recall at the scale midpoint” is what made this depiction somewhat opaque. I do not have a specific suggestion concerning an alternative way to display the key results, but I do encourage the authors to consider a more intuitive way to show the data.

Evidence on biased recall of the kind shown here, where current perceptions drive recall of past attitudes, has been reported in numerous previous studies in the literature, some of which are cited by the authors. From a purely theoretical perspective, the key effects demonstrated in this study are as would be predicted on the basis of much prior research in cognitive and social psychology; there isn't much new here in terms of support for a novel idea or a new theoretical construct. If there is a new idea that emerges from this work, the authors should elaborate on what it is. However, the study provides a compelling example of how the well-known phenomenon of recall bias is relevant to a societal event of great magnitude such as the Covid-19 pandemic, with implications for future pandemic preparation noted by the authors. Thus to my mind, the contribution is both substantial and important despite the lack of a truly novel theoretical contribution.

Nonetheless, I think that the authors need to provide at least some discussion of the relation of the present results to prior research on consistency biases in recall. This discussion could touch on the cited work by Wilson & Ross (2003; ref. 5), and even more importantly, recent studies that the authors don't cite concerning consistency bias as it relates to false memories for fake news involving a variety of public events, the most directly relevant concerning biased recall regarding Covid-19 vaccines (Greene et al., Applied Cognitive Psychology, 2022) as well as other public

events such as the Irish abortion referendum (Murphy et al., Psychological Science, 2018) and the US capitol riots (Calvillo et al., Memory, 2022). Broadening the discussion to include these highly relevant studies would strengthen the paper.

Referee #2 (Remarks to the Author):

Report on
"Mapping motivational bias in recalling the SARS-CoV-2 pandemic"

Summary

This paper studies how people recall their own assessments of risks and policy interventions during the Covid pandemic. It shows systematic differences between vaccinated and unvaccinated individuals, who display opposite biases in their recall. Two interventions (higher incentives, information about recall bias) fail to dent these biases. Across countries, participants show similar bias in predicting the past attitudes of others.

Comments

This is a timely and interesting study. The results speak to a literature in psychology on hindsight bias, biased memory and motivated cognition, as well as a literature on perception of public policies. The results are of interest to policy makers, although given the null results on the interventions, it is not exactly clear how they can address these issues. As far as I can tell, the results are novel, although there is another working paper on pandemic recall that, in some ways, goes beyond the methods in the current paper (see below). The paper is mostly well-written, although I found some aspect difficult to follow, as I detail below.

While I read the paper with interest, I have a number of concerns that make me hesitant to recommend publication:

1. The paper does not talk at all about the overall recall effect. Given the relatively small number of unvaccinated in the population (oversampled in the study), I presume people generally overestimated their past perceived infection risk and the appropriateness of interventions. So a different story from the one stressed by the authors, is that hindsight bias could lead to more general support for the interventions. To evaluate this better, there should be more information about the distribution of vaccinations, both in the sample and in the general population and the direction of the overall effects.
2. The paper relies exclusively on regression evidence. It would be good to provide raw evidence on the distribution of past ratings, the current ratings, and the recall bias, showing how distributions shift in different groups. In particular, while the main graphs are clever, they are also very dense and difficult to interpret (it took me quite a while to grasp them). More explicit explanation may help: For instance the comment in the methods that these lines represent all combinations of present and past attitudes that would lead to a recall at the scale midpoint would be helpful to state explicitly. Also, if I am not mistaken, the slope of the line is equal to the ratio of coefficients if interaction terms, while the position is determined by the coefficients of the baseline. Even so, I think a more "raw" representation of the data, if it can be produced in succinct format, may be more accessible to readers.
3. The correlations are not always easy to interpret. While the effect of current attitudes is clearly indicative of motivational biases, I am not sure how to interpret recall biases of the (un)vaccinated conditional on current attitudes, which, if I am not mistaken, drives the position of the lines in the main figures – perhaps the most striking visual effect. In the Methods section, we learn that recall correlates highly with life satisfaction. This seems important to report in more detail, as it means the survey answers may not be so much about recall, as about voicing some general feeling of

(dis)contentment. I also find the correlations with appropriateness of interventions hard to evaluate. Vaccinated and unvaccinated may differ on many dimensions that drive both their recall and their appropriateness (life satisfaction comes to mind again). The paper does not make explicit causal claims, but it could at least show that there exists a relation conditional on measured variables such as life satisfaction. Note also that the current manuscript is less ambitious than a recent working paper on pandemic recall by Herz et al. (2022), which experimentally manipulates recall rates and shows a causal impact of recall on policy preferences.

4. Related to the previous point, I am not sure I understand the rationale behind the regression specifications. The main regressions use recall as a DV and control for current attitudes, but not all interpretations are clear to me (see previous point). In a second set of regressions, recall biases are estimated by standardizing based on the "grand mean" of attitudes. I am not exactly sure what this means or why it is done. In the literature on hindsight/memory bias, people have used as a DV the "shift index" (individual recall error), or the "proximity index", which measures how close current preferences are to the current attitudes. It would be good to relate the analyses to these approaches (Pohl, 2007).

5. In Study 3, people predict the ratings of others in the past. This may have little to do with recall bias or even any temporal dimension, as there are many reasons why people may mispredict the beliefs of others. Indeed the results seem to be consistent with some form of projection bias of one's own attitude on others. Thus, I don't think this can be presented as a straightforward extension of the results, and I am not sure if this study should be included in the paper at all.

6. In Study 2, it is not clear to me how strong the treatment manipulations were, which makes it hard to evaluate the null effects. In the study by Zimmermann cited by the authors participants can win 50 euros each for correct recall. Here, it seems that one prize of 100 euros was available for all participants, which may not have been a very strong motivator. Was there other evidence people engaged with the incentives (e.g. longer decision times)? Similarly, regarding the information, was there some sort of manipulation check that people engaged with the information at all?

References

Herz, H., D. Kistler, C. Zehnder, C. Zihlmann. 2022. Hindsight Bias and Trust in Government: Evidence from the United States. CESifo working paper 9767.

Pohl, R. F. "Ways to assess hindsight bias." *Social Cognition* 25.1 (2007): 14-31.

Referee #3 (Remarks to the Author):

This manuscript reported three studies that examined biases in recalling COVID-related perceptions. In Studies 1 and 2, participants previously rated some COVID perceptions and they were asked to recall those ratings in the present study. In Study 3, previous participants had made judgments and new participants were asked to estimate those judgments. In Study 1, participants' recall was biased by their current perceptions and by their vaccination status and identification and the magnitude of this bias was related to greater identity with vaccination status. Study 2 examined two methods to reduce bias in recall: providing financial incentives for accuracy and providing information about recall biases. The results mirrored those in Study 1 and neither providing incentives nor information did not reduce bias. Study 3 examined the stability of biased judgments across countries. In Study 3, participants estimated other participants' predictions of COVID-19 risks. Participants from all countries showed some biased judgments, all overestimating the probability of infection and all except Japan underestimating the severity of COVID-19.

I enjoyed reading this manuscript and believe that it is important to understand memory for threat

perceptions. I thought that the studies described were well-conducted and the manuscript was clear and concise. I have just a few comments that could be addressed in a revision.

1. On p. 10, line 222, the authors wrote "biased recall..." but because the initial ratings and the subsequent estimates were made by different participants in Study 3, this isn't recall.
2. One additional limitation could have been the effectiveness of the incentive. Correct recall increased the chance of winning money. Perhaps a more certain incentive would have been more effective at reducing bias.
3. The memory distortion in the present study seemed similar to hindsight memory distortion. I think some connection to the hindsight literature would help situation the present findings in the literature. This study seems particularly relevant: <https://doi.org/10.1037/mac0000033>

Referee #4 (Remarks to the Author):

A. Summary of the key results

This interesting study demonstrates bias in recall of experiences during the COVID-19 pandemic are strongly biased by current evaluations and motivations.

B. Originality and significance: if not novel, please include reference

Very novel and significant, I think this is likely to be of interest to a wide range of readers.

C. Data & methodology: validity of approach, quality of data, quality of presentation

Very clear and valid approach and methodology. Lovely graphs that are well introduced and decoded for the reader.

D. Appropriate use of statistics and treatment of uncertainties

Appropriate analyses, well explained

E. Conclusions: robustness, validity, reliability

Robust and replicable studies reported here

F. Suggested improvements: experiments, data for possible revision

I would like to see a little more detail in the discussion about what this means in practice – particularly with some reference to effect sizes. This work is novel and sheds new light on how our reconstructions of major historical events can become distorted very shortly afterwards, but I would like to see more specific recommendations. The implications listed are rather vague, but maybe some more detail could be shared about how we evaluate perceptions of the past and how likely they are to be biased (plus what we should do about that).

I think a major novel contribution to the literature here is linking these findings of biased recall to an adaptive, functional account of memory. This comes up in the introduction but is not revisited in great detail in the analysis or discussion – do the authors have any data that might shed light on *how* these memory errors might be functional (for example what might be different about those who experience distorted memory vs those who did not?). And/or in the discussion, could this be unpacked in some more detail?

Finally, I was surprised not to see some reference to emotional recall literature (e.g. Levine et al., 2001 <https://doi.org/10.1080/02699930125955>). The methods and findings in that field are so similar to what has been done here and I think might be useful in guiding interpretation/implications, especially from a functional perspective.

G. References: appropriate credit to previous work?

Yes

H. Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions

Yes – with some more concrete recommendations needed in the discussion.

Author Rebuttals to Initial Comments:

Response Letter for Nature Manuscript 2023-05-08562B

Editor

Editor, Comment 1

Your manuscript, "Mapping Motivational Bias in Recalling the SARS-CoV-2 Pandemic", has now been seen by 4 referees, whose comments are attached below. While they find your work of potential interest, as do we, they have raised important concerns that in our view need to be addressed before we can consider publication in Nature.

We are grateful for the opportunity to revise our manuscript. Moreover, we would like to thank the reviewer team for their constructive comments and suggestions. In the revised manuscript, we have addressed all of them.

Below, you find a summary of our revisions in response to each of the reviewers' comments. Additionally, we made sure to adhere to Nature's formatting and reporting guidelines.

Editor, Comment 2

More specifically, you will see that while all the referees find your work timely and interesting, especially R2 is concerned that your findings may be explained by other factors, such as general hindsight bias (point 1), life-satisfaction (point 3), and thinks that study 3 may not have to do with recall bias. Should further experiments allow you to address these and the referees' other criticisms - thereby providing strong evidence for your claims - we would be happy to consider a revised manuscript (unless something similar has been accepted at Nature or appeared elsewhere in the meantime).

As detailed below, existing literature, additional analyses and new data can dispel these issues:

Hindsight bias is conceptually different from the recall bias that we study here. We discuss this in detail in response to Reviewer 2, Comments 4 and 5.

Moreover, we controlled our main analyses for life satisfaction and found the same patterns, suggesting that answers are indeed about recall and not general feelings of (dis)contentment. We discuss this in detail in response to Reviewer 2, Comment 4.

Additional analyses on the question of whether population-based estimates are a good proxy for recall (former Study 3, now Study 4) suggest that our measure does not capture a general form of projection bias but is indeed related to recall. We discuss this in detail in response to Reviewer 2, Comment 6; Editor, Comment 3 and Reviewer 3, Comment 2.

Finally, as suggested, we collected new experimental data to gain further evidence that the recall bias demonstrated in our research has a motivational foundation. While bias-awareness information or minor incentives were not found to correct recall biases in Study 2, providing larger monetary incentives for accurate recall in the new Study 3 improved recall. In our study design,

we also included a non-pandemic neutral recall item, in which inaccurate recall should not have a motivational basis. Indeed, we found that large incentives had no effect on recall accuracy for this neutral recall item. Taken together, this new evidence indicates that (i) biases are strong, (ii) can be corrected only with large incentives and (iii) have a motivational foundation. See our detailed response to Reviewer 2, Comment 7.

Editor, Comment 3

There are two other points that we would like to give specific guidance on:

R2 (point 5) is unsure as to whether study 3 should be included in the paper. We feel that it has the potential to strengthen the paper so ask that, rather than removing it, you consider ways of strengthening the findings (with additional analyses or experimental data) to provide a stronger demonstration of the effect of recall bias on peoples' memories of the pandemic.

Based on additional analyses of data from Study 1, we argue that biased estimates of past population-based perceptions and behaviours, as used in former Study 3 (now Study 4), are empirically related to and, therefore, the best available proxy for biased individual-level recall across several countries.

Specifically, in Study 1, we collected both individual-level recall and population-level estimates of the past. The latter was elicited in the same way as in Study 4. This allows us to assess the validity of the population-level estimates. We do so by regressing population-level estimates on individual-level recall while controlling for present and past ratings (see section S1.6 in the supplement). We find that for all four items, individual-level recall significantly predicts the population-level estimate. At the same time, in three of the four cases, the present rating does not predict the population-level estimate on top of the individual-level recall. This suggests that our measure does not capture a general form of projection bias but is indeed related to recall. Therefore, we believe that the individual-level estimates of population-based ratings can be seen as proxies for individual-level recall. Nevertheless, to be transparent about the differences in these measures, we now explicitly discuss this as a limitation in the Discussion section; see our detailed response to Reviewer 2, Comment 6.

Editor, Comment 4

R4 (point F) asks for specific recommendations based on these data. We do not believe that you have data that can generate recommendations (as this is an observational study) so ask that you refrain from speculating and providing recommendations that are not supported by your data.

As suggested, we refrained from engaging in speculative discussions regarding potential policy implications, recognizing the importance of grounding recommendations solely in the data and evidence presented.

In the Discussion section, we explore the potential implications of our findings (e.g. for the climate crisis). However, we want to assure you that we have approached this issue with caution, avoiding overgeneralization and any unwarranted extrapolations. Instead, our intention in hinting at potential implications is to highlight avenues for future research and stimulate interest in further investigations. By doing so, we hope to contribute to the development of a broader understanding of the implications of our findings beyond the scope of this particular study.

Reviewer 1

Reviewer 1, Comment 1

This paper reports three studies examining how recall of various pandemic-related perceptions and behaviors is influenced by current assessments and by identification with vaccination status. Using large sample sizes numbering in the thousands, the studies provide clear and consistent evidence of biased recall such that what people remembered about items including infection probability and severity, mask wearing behavior, and trust in government and science among others was strongly skewed by their current perceptions and their vaccination status. The effect was resistant to a cash incentive or providing information about biased recall. A third study involving participants from 10 countries revealed evidence of bias by comparing current recall of probability and severity of infection as well as government effectiveness to benchmark values obtained in a separate 2020 survey.

The studies reported here are both novel and timely; they should be of broad interest to researchers from various disciplines and also society more generally. The writing is clear, the abstract is informative, conclusions are appropriate, and some limitations of the study are acknowledged. The methods used to obtain the key data are solid, the evidence for recall bias is robust, and the statistical/analytic procedures used to demonstrate these effects are rigorous and appropriate. The depiction of the key results on biased recall in Figures 1 and 2 is economical, but not highly intuitive, at least to this reader – it took me some time to fully grasp exactly what was being depicted and to map the lines depicted to the individual conditions. Figuring out the meaning of “predicted recall at the scale midpoint” is what made this depiction somewhat opaque. I do not have a specific suggestion concerning an alternative way to display the key results, but I do encourage the authors to consider a more intuitive way to show the data.

We thank the reviewer for the overall positive evaluation. We have addressed all the thoughtful comments and suggestions (for details, see below). Please note that page numbers refer to the clean version of the revised manuscript (without tracked changes).

Regarding the figures, we agree that they are quite complex. We aim to provide a comprehensive description of the factors affecting recall, both in our analyses and in the figures. Therefore, we worked hard to better explain the figures by providing an example and revising Figure 1 to match the example. With this, we hope that we can facilitate understanding and grasp the intuition of the figures more easily. We prefer keeping the figures rather than providing only a simplified but potentially misleading snapshot of the complex nexus of the relevant factors.

Additionally, we created “raw” representations of the data (as suggested by Reviewer 2) showing the distribution of the recall bias and added them to the supplement (see S1.3.1). We added them to the supplement because these somewhat simpler visualizations of the results cannot capture how recall depends on both present and past ratings.

Reviewer 1, Comment 2

Evidence on biased recall of the kind shown here, where current perceptions drive recall of past attitudes, has been reported in numerous previous studies in the literature, some of which are cited by the authors. From a purely theoretical perspective, the key effects demonstrated in this study are as would be

predicted on the basis of much prior research in cognitive and social psychology; there isn't much new here in terms of support for a novel idea or a new theoretical construct. If there is a new idea that emerges from this work, the authors should elaborate on what it is. However, the study provides a compelling example of how the well-known phenomenon of recall bias is relevant to a societal event of great magnitude such as the Covid-19 pandemic, with implications for future pandemic preparation noted by the authors. Thus to my mind, the contribution is both substantial and important despite the lack of a truly novel theoretical contribution.

We thank the reviewer for this assessment. We also believe that the robustness of recall bias in the context of one of the largest health threats and global societal upheavals in contemporary history, together with unique and large datasets across several countries, provides a compelling investigation into this phenomenon. Therefore, we hope this research will be of interest to many people, even beyond those fields that typically study such phenomena, especially because the results may help to understand the tensions in many societies in the post-pandemic period.

Reviewer 1, Comment 3

Nonetheless, I think that the authors need to provide at least some discussion of the relation of the present results to prior research on consistency biases in recall. This discussion could touch on the cited work by Wilson & Ross (2003; ref. 5), and even more importantly, recent studies that the authors don't cite concerning consistency bias as it relates to false memories for fake news involving a variety of public events, the most directly relevant concerning biased recall regarding Covid-19 vaccines (Greene et al., Applied Cognitive Psychology, 2022) as well as other public events such as the Irish abortion referendum (Murphy et al., Psychological Science, 2018) and the US capitol riots (Calvillo et al., Memory, 2022). Broadening the discussion to include these highly relevant studies would strengthen the paper.

Following this suggestion, we expanded our discussion of and references to prior work on recall bias in the Introduction section. That is, when reviewing the previous literature, we provide specific examples of other domains in which biased recall has been observed, including the suggested references (p. 3). Moreover, we now discuss Wilson and Ross (2003) when we elaborate on the bidirectionality of effects in the Limitations section (p. 14). Finally, we refer to other research suggesting a functional role of biased recall in the Discussion section and call for future research on this important topic (p. 14).

Reviewer 2

Reviewer 2, Comment 1

Report on
"Mapping motivational bias in recalling the SARS-CoV-2 pandemic"

Summary

This paper studies how people recall their own assessments of risks and policy interventions during the Covid pandemic. It shows systematic differences between vaccinated and unvaccinated individuals, who display opposite biases in their recall. Two interventions (higher incentives, information about recall bias) fail to dent these biases. Across countries, participants show similar bias in predicting the past attitudes of others.

Comments

This is a timely and interesting study. The results speak to a literature in psychology on hindsight bias, biased memory and motivated cognition, as well as a literature on perception of public policies. The results are of interest to policy makers, although given the null results on the interventions, it is not exactly clear how they can address these issues. As far as I can tell, the results are novel, although there is another working paper on pandemic recall that, in some ways, goes beyond the methods in the current paper (see below). The paper is mostly well-written, although I found some aspect difficult to follow, as I detail below.

We thank the reviewer for the overall positive evaluation. We have addressed all the thoughtful comments and suggestions (for details, see below). Please note that page numbers refer to the clean version of the revised manuscript (without tracked changes).

Reviewer 2, Comment 2

While I read the paper with interest, I have a number of concerns that make me hesitant to recommend publication:

The paper does not talk at all about the overall recall effect. Given the relatively small number of unvaccinated in the population (oversampled in the study), I presume people generally overestimated their past perceived infection risk and the appropriateness of interventions. So a different story from the one stressed by the authors, is that hindsight bias could lead to more general support for the interventions. To evaluate this better, there should be more information about the distribution of vaccinations, both in the sample and in the general population and the direction of the overall effects.

Our results suggest that the overall recall bias strongly depends on a population's vaccine uptake as well as individuals' identification with their vaccination status. Both of these variables vary by geographical location (e.g. country or region) and time. Therefore, we believe it is much more accurate to show how recall depends on these factors. Accordingly, we would prefer to refrain from any statements that suggest something along the lines 'people dramatize the pandemic' or the opposite. We understand that some sentences in the previous version could be interpreted along these lines, so we have changed the respective wording. In addition, we describe the determinants of people's bias in recall and show that the directionality of biased recall is linked to opposing evaluations of the pandemic and the political actions taken. The understanding of the drivers and the magnitude of this polarization, we believe, is an important consequence of our findings. This is now also mentioned in the revised manuscript (p. 12): *"The results suggest that while a vaccinated majority has a more positive view of the measures taken during the pandemic, as warranted by respective perceptions of the past, a small segment of society has a strong desire to take revenge on those who spoke out or took responsibility during the pandemic."*

Reviewer 2, Comment 3

The paper relies exclusively on regression evidence. It would be good to provide raw evidence on the distribution of past ratings, the current ratings, and the recall bias, showing how distributions shift in different groups. In particular, while the main graphs are clever, they are also very dense and difficult to interpret (it took me quite a while to grasp them). More explicit explanation may help: For instance the comment in the methods that these lines represent all combinations of present and past attitudes that would lead to a recall at the scale midpoint would be helpful to state explicitly. Also, if I am not mistaken,

the slope of the line is equal to the ratio of coefficients if interaction terms, while the position is determined by the coefficients of the baseline. Even so, I think a more “raw” representation of the data, if it can be produced in succinct format, may be more accessible to readers.

We agree that the figures are quite complex. Thus, we worked hard to better explain the figures. Specifically, we now provide an example in the main text and revised Figure 1 to match the example. We hope that these changes foster a more intuitive understanding of these complex figures.

Additionally, we created the requested “raw” representation of the data in the form of figures that show the distribution of the recall bias and added it to the supplement (see S1.3.1). The reason we added it only to the supplement is that these somewhat simpler visualizations of the results cannot capture how recall depends on *both* present and past ratings.

Reviewer 2, Comment 4

The correlations are not always easy to interpret. While the effect of current attitudes is clearly indicative of motivational biases, I am not sure how to interpret recall biases of the (un)vaccinated conditional on current attitudes, which, if I am not mistaken, drives the position of the lines in the main figures – perhaps the most striking visual effect. In the Methods section, we learn that recall correlates highly with life satisfaction. This seems important to report in more detail, as it means the survey answers may not be so much about recall, as about voicing some general feeling of (dis)contentment. I also find the correlations with appropriateness of interventions hard to evaluate. Vaccinated and unvaccinated may differ on many dimensions that drive both their recall and their appropriateness (life satisfaction comes to mind again). The paper does not make explicit causal claims, but it could at least show that there exists a relation conditional on measured variables such as life satisfaction. Note also that the current manuscript is less ambitious than a recent working paper on pandemic recall by Herz et al. (2022), which experimentally manipulates recall rates and shows a causal impact of recall on policy preferences.

We have added analyses to the supplement, in which we show that our results remain unaffected by the inclusion of control variables that might drive both recall and current variables (see additional regression tables in S1.3.2). In particular, we control for demographic and socioeconomic variables (age gender, education and household income), past and present life satisfaction and cognitive ability (proxied by a three-item cognitive reflection task). Together with the results from our new study (discussed in response to this reviewer's comment 7), we believe these results provide evidence that the answers are indeed about recall and not general feelings of (dis)contentment.

Regarding the working paper by Herz et al. (2022), we want to note some important differences between their studies and ours:

Herz et al. focused on hindsight bias, which is conceptually different from the recall bias that we study here. Hindsight bias describes the tendency to perceive past events as more predictable than they were. Hence, hindsight bias serves as a good explanation if there is some kind of ‘objective truth’ to which individuals can anchor their recall.

While Herz et al. study general hindsight bias (as the objective truth is the same for all participants, e.g. whether a certain policy has been implemented or not), we investigate the motivational nature of biased individual recall (i.e. personal memory distortions) and relate it to

present evaluations that subjectively differ between different population groups (e.g. vaccinated and unvaccinated individuals). Specifically, we investigate how people's identities (vaccination status and vaccination status identification) affect individual recall. We calculated the effect of present estimations ('bias force' according to the Truth and Bias model; West and Kenny, 2011, Psychological Review) and past estimations ('truth force') on recall, as detailed below (Reviewer 2, Comment 5). Hence, the very nature of the recall situation in our studies differs from hindsight bias, as studied by Herz et al. (2022), and leads to very different results and also provides very different conclusions and implications. Although the two concepts are related, we refrained from discussing the hindsight bias literature to avoid confusion.

That said, the working paper by Herz et al. (2022) is certainly relevant to our studies, so we now acknowledge their work when motivating our research. We thank the Reviewer for pointing us to this important piece.

Reviewer 2, Comment 5

Related to the previous point, I am not sure I understand the rationale behind the regression specifications. The main regressions use recall as a DV and control for current attitudes, but not all interpretations are clear to me (see previous point). In a second set of regressions, recall biases are estimated by standardizing based on the "grand mean" of attitudes. I am not exactly sure what this means or why it is done. In the literature on hindsight/memory bias, people have used as a DV the "shift index" (individual recall error), or the "proximity index", which measures how close current preferences are to the current attitudes. It would be good to relate the analyses to these approaches (Pohl, 2007).

The main difference between the previous research on hindsight/memory/recall bias and our studies is that our data have three instead of just two pieces of statistical information: present estimates (e.g. how I perceive the infection risk today), past estimates (e.g. how I really perceived the infection risk at a previous time = previously collected data) and the recall of past estimates (e.g. what I think today: how have I perceived risk in the past?). Note that research on hindsight/memory/recall bias usually does not assess present estimates (because there is some objective value or truth, as explained above). In our case, however, individuals may differ in their current evaluations, which may spark different motivations to distort recall of the past in order to have a consistent self-view.

The availability of all three pieces of information allows for their combination by utilizing the Truth and Bias model, as proposed by West and Kenny (2011, Psychological Review). This statistical model distinguishes between the 'directional bias' (i.e. the extent to which responses tend towards scale endpoints), the 'bias force' (i.e. the effect of present estimates on recall) and the 'truth force' (i.e. the effect of past estimates on recall) as predictors of people's recall. To do so, it regresses recalled values on past and present values; the intercept denotes directional bias, the regression coefficient of present values denotes the bias force and the regression coefficient of the present values denotes the truth force. To assess the effects of individual characteristics (e.g. vaccination status and identification) or experimental manipulations, these variables are added as moderators to the model. For instance, in the new Study 3, recall is predicted by past and present ratings plus experimental manipulation plus its interactions with past and present ratings.

The regression results are shown in Figures 1 and 2. All models were estimated a second time with recall, with past and present ratings being centered by the grand mean of past ratings, as recommended by West and Kenny (2001). In this way, the directional bias (intercept) presented

in the supplemental regression tables can be interpreted as an over- or underestimation of recall compared to actual past ratings. Importantly, centering does not change the results; it only facilitates interpretation.

Although we do not have the space to explain all the details of the statistical procedure in the main text, we have worked hard to provide a basic intuition about the estimation and its interpretation in the revised main manuscript. Moreover, we refer to the relevant literature for interested readers. There is a more detailed explanation in the Methods section.

Taken together, because of the interplay between present perceptions, estimates about past perceptions and recall of past experiences, we cannot use simple two-dimensional indices as sometimes used in the literature. Yet, addressing this reviewer's concern, we now provide additional two-dimensional figures (ignoring the present estimates) in the supplement (see S1.3.1). For the statistical analyses, we still use a more accurate three-dimensional model whenever the necessary data are available.

Reviewer 2, Comment 6

In Study 3, people predict the ratings of others in the past. This may have little to do with recall bias or even any temporal dimension, as there are many reasons why people may mispredict the beliefs of others. Indeed the results seem to be consistent with some form of projection bias of one's own attitude on others. Thus, I don't think this can be presented as a straightforward extension of the results, and I am not sure if this study should be included in the paper at all.

We agree that the prediction of others' past (although population-representative) ratings comes with some limitations. The main reason we decided to use such data rather than individual-level data from the past and its respective recall is that such data with the exact same measures are not available across several countries.

However, we argue that such data provide at least a proxy for individual-level recall bias. In fact, identifying individual differences in the deviation from estimated population ratings speaks to an individual-level recall bias, particularly if recall is different for different groups, which can be assumed to have different motivations for recall bias. To support this argument, we use additional data from Study 1, in which we collected both individual-level recall and population-level estimates of the past. The latter was elicited in the same way as in Study 4 (previously labelled Study 3). This allows us to assess the validity of the population-level estimates. In particular, we can test whether there is indeed some form of projection bias. We do so by regressing population-level estimates on individual-level recall while controlling for present and past ratings (see section S1.6 in the supplement). We find that for all four items, individual-level recall significantly predicts the population-level estimate. At the same time, in three of the four cases, the present rating does not predict the population-level estimate on top of the individual-level recall. This suggests that our measure does not capture a general form of projection bias but is indeed related to recall.

Therefore, we believe that the individual-level estimates of population-based ratings (and individual differences thereof), as reported in Study 4, can be seen as a proxy for individual-level recall of (the very same) individual-based ratings (and individual differences thereof), as reported in Studies 1–3.

Nevertheless, as we agree that these methodological differences are important to address, we added this as a limitation to the Discussion section (p. 14): *"The studies reported here have some limitations. [...] the estimate of others' past ratings used in Study 4 can be considered a proxy for individual recall. However, the two concepts may differ, as estimates for the general population may also depend on other factors, such as feelings and thoughts about this population."*

Reviewer 2, Comment 7

In Study 2, it is not clear to me how strong the treatment manipulations were, which makes it hard to evaluate the null effects. In the study by Zimmermann cited by the authors participants can win 50 euros each for correct recall. Here, it seems that one prize of 100 euros was available for all participants, which may not have been a very strong motivator. Was there other evidence people engaged with the incentives (e.g. longer decision times)? Similarly, regarding the information, was there some sort of manipulation check that people engaged with the information at all?

We are grateful for these important suggestions. In Study 2, we checked how long people engaged with the respective intervention tasks. In the information condition, participants took a median time of 94 seconds; in the incentive condition, participants took a median time of 61 seconds and in the control condition, participants spent a median time of 45 seconds reading the instructions and recalling their answers. Thus, participants in the intervention conditions invested a considerably greater amount of time, suggesting that they engaged with the tasks.

However, we agree that the incentives were much smaller in Study 2 than in Zimmermann's study (2020). Zimmermann varied the size of incentives to improve recall and found that only large incentives are sufficient to counteract the motivational basis for recall bias.

We thus conducted a new experiment in which each participant in the incentive condition could win 25 euros for correct recall (see the new Study 3 in the revised manuscript). Hence, the expected value is about 200 times larger than in Study 2. We investigated the effect of the prospect of receiving an incentive for correct recall relative to a control group, which did not receive any incentive. We measured recall for pandemic-related items, as in our previous studies, and, importantly, for a neutral item (the time of day when participants took the survey in the past), for which there is no basis for motivated recall. Using this experimental design, we found evidence of reduced recall bias in the incentive condition (compared to the no-incentive control condition). However, this positive effect of incentives on recall accuracy occurred only for pandemic-related items—where bias should have a motivational basis—but not for the neutral item. We interpret this as clear evidence that recall bias is indeed motivated.

We would like to thank the reviewer again for pointing to this important limitation, which made us conduct an additional study. We hope the reviewer agrees that Studies 2 and 3 together provide important insights into the very nature of recall bias in the context of the COVID-19 pandemic.

References

Herz, H., D. Kistler, C. Zehnder, C. Zihlmann. 2022. Hindsight Bias and Trust in Government: Evidence from the United States. CESifo working paper 9767.

Pohl, R. F. "Ways to assess hindsight bias." *Social Cognition* 25.1 (2007): 14-31.

Reviewer 3

Reviewer 3, Comment 1

This manuscript reported three studies that examined biases in recalling COVID-related perceptions. In Studies 1 and 2, participants previously rated some COVID perceptions and they were asked to recall those ratings in the present study. In Study 3, previous participants had made judgments and new participants were asked to estimate those judgments. In Study 1, participants' recall was biased by their current perceptions and by their vaccination status and identification and the magnitude of this bias was related to greater identity with vaccination status. Study 2 examined two methods to reduce bias in recall: providing financial incentives for accuracy and providing information about recall biases. The results mirrored those in Study 1 and neither providing incentives nor information did not reduce bias. Study 3 examined the stability of biased judgments across countries. In Study 3, participants estimated other participants' predictions of COVID-19 risks. Participants from all countries showed some biased judgments, all overestimating the probability of infection and all except Japan underestimating the severity of COVID-19.

I enjoyed reading this manuscript and believe that it is important to understand memory for threat perceptions. I thought that the studies described were well-conducted and the manuscript was clear and concise. I have just a few comments that could be addressed in a revision.

We thank the reviewer for the overall positive evaluation. We have addressed the thoughtful comments and suggestions (for details, see below). Please note that page numbers refer to the clean version of the revised manuscript (without tracked changes).

Reviewer 3, Comment 2

On p. 10, line 222, the authors wrote “biased recall...” but because the initial ratings and the subsequent estimates were made by different participants in Study 3, this isn't recall.

We agree that the individual-level recall of population-based ratings (and individual differences thereof), as reported in (now) Study 4, is somewhat different from the individual-level recall of own ratings (and individual differences thereof), as reported in Studies 1–3.

Still, we argue that bias in these estimations is conceptually related, as it should be affected by the same motivational determinants as the individual-level recall bias assessed in the other studies. To support this argument, we use additional data from Study 1, where we collected both individual-level recall and population-level estimates of the past. The latter was elicited in the same way as in Study 4 (previously labelled Study 3). This allows us to assess the validity of the population-level estimates. We do so by regressing population-level estimates on individual-level recall while controlling for present and past ratings (see section S1.6 in the supplement). We find that for all four items, individual-level recall significantly predicts the population-level estimate. At the same time, in three of the four cases, the present rating does not predict the population-level estimate on top of the individual-level recall. This suggests that our measure does not capture a general form of projection bias but is indeed related to recall. Therefore, we believe that the individual-level estimates of population-based ratings (and individual differences thereof), as reported in Study 4, can be seen as proxies for individual-level recall of (the very same) individual-based ratings (and individual differences thereof), as reported in Studies 1–3.

That said, addressing this reviewer's concern and for more conceptual clarity, we have changed the wording from "recall" to "estimate" in Study 4.

Reviewer 3, Comment 3

One additional limitation could have been the effectiveness of the incentive. Correct recall increased the chance of winning money. Perhaps a more certain incentive would have been more effective at reducing bias.

We agree. Based on this and feedback from Reviewer 2, we conducted a new experiment in which every participant in the treatment group could win 25 euros for correct recall (see the new Study 3 in the revised manuscript). We investigated the effect of this incentive on recall relative to a control group, which did not receive any incentive. We measured recall both for pandemic-related items and for a neutral item (the time of day when participants took the survey in the past), for which there is no basis for motivated recall.

We found that the high incentive improved recall for pandemic-related items but not for the neutral item. We see this as evidence that recall indeed has a motivational foundation.

We would like to thank the reviewer for pointing to this important limitation, which made us conduct the additional study.

Reviewer 3, Comment 4

The memory distortion in the present study seemed similar to hindsight memory distortion. I think some connection to the hindsight literature would help situation the present findings in the literature. This study seems particularly relevant: <https://doi.org/10.1037/mac0000033>

Our studies focus on motivated recall of the pandemic (e.g. having perceived pandemic measures as exaggerated). This is conceptually different from hindsight bias, which describes the tendency to perceive past events as more predictable than they were. The two concepts may be related but we refrained from discussing the hindsight bias literature to avoid confusion. Furthermore, from our perspective, hindsight bias serves as a good explanation if there is some kind of objective truth to which individuals can anchor their recall. However, in our studies, we can show that recall strongly varies across individuals, as recall is motivated by vaccination status and vaccination status identification.

That said, given that this and related studies on hindsight bias are still relevant to motivating our research, we now refer to it when motivating our research in the Introduction section.

Reviewer 4

Reviewer 4, Comment 1

A. Summary of the key results

This interesting study demonstrates bias in recall of experiences during the COVID-19 pandemic are strongly biased by current evaluations and motivations.

B. Originality and significance: if not novel, please include reference

Very novel and significant, I think this is likely to be of interest to a wide range of readers.

C. Data & methodology: validity of approach, quality of data, quality of presentation

Very clear and valid approach and methodology. Lovely graphs that are well introduced and decoded for the reader.

D. Appropriate use of statistics and treatment of uncertainties

Appropriate analyses, well explained

E. Conclusions: robustness, validity, reliability

Robust and replicable studies reported here

We thank the reviewer for the overall positive evaluation of our results, their originality, data and methodology, as well as statistical methods used and conclusions drawn from them. We have addressed the thoughtful comments and suggestions below. Please note that page numbers refer to the clean version of the revised manuscript (without tracked changes).

Reviewer 4, Comment 2

F. Suggested improvements: experiments, data for possible revision

I would like to see a little more detail in the discussion about what this means in practice – particularly with some reference to effect sizes. This work is novel and sheds new light on how our reconstructions of major historical events can become distorted very shortly afterwards, but I would like to see more specific recommendations. The implications listed are rather vague, but maybe some more detail could be shared about how we evaluate perceptions of the past and how likely they are to be biased (plus what we should do about that).

The editor suggested refraining from speculative discussions regarding potential policy implications. We recognize the importance of grounding recommendations solely in the data and the evidence presented. To avoid overgeneralization and unwarranted extrapolations, we discuss potential implications of our findings (e.g. for the climate crisis) with great caution and highlight avenues for future research. By doing so, we hope to contribute to the development of a broader understanding of the implications of our findings beyond the scope of this particular study.

Reviewer 4, Comment 3

I think a major novel contribution to the literature here is linking these findings of biased recall to an adaptive, functional account of memory. This comes up in the introduction but is not revisited in great detail in the analysis or discussion – do the authors have any data that might shed light on *how* these memory errors might be functional (for example what might be different about those who experience distorted memory vs those who did not?). And/or in the discussion, could this be unpacked in some more detail?

This is a very interesting thought. Although our data do not provide insights into the functional value of biased recall, we at least point to this possibility and call for further research in this direction in the Discussion section (p. 14): *“The strong directional character of this bias and the influence of identity in this regard have major implications that warrant further exploration. For*

instance, memory distortions can be functional in coping with major life events.²⁵ Therefore, future research could investigate whether people with a larger (vs. smaller) tendency for biased recall of the COVID-19 pandemic may have had some advantages on other psychological dimensions, including psychological functioning during or after the pandemic."

Reviewer 4, Comment 4

Finally, I was surprised not to see some reference to emotional recall literature (e.g. Levine et al., 2001 <https://doi.org/10.1080/02699930125955>). The methods and findings in that field are so similar to what has been done here and I think might be useful in guiding interpretation/implications, especially from a functional perspective.

We now refer to the emotional recall literature when we relate to the potential functional value of biased recall (see our response to the reviewer's previous comment).

Reviewer 4, Comment 5

G. References: appropriate credit to previous work?

Yes

H. Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions

Yes – with some more concrete recommendations needed in the discussion.

We again would like to express our gratitude for the reviewer's helpful comments and suggestions. Regarding potential policy recommendations, we follow the editor's advice to refrain from giving advice/suggestions that are not tested in our research.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have done a very good job of responding to my comments and those of the other reviewers. The additional study with a stronger incentive manipulation strengthens the paper, as do additional analyses, references, and clarifications to the main text.

Importantly, the authors have improved their characterization of the main results, especially those in Figure 1, which are complex and non-intuitive. The added text helps, and having spent a fair amount of time grappling with what is depicted, the main message conveyed by those data is clearer to me now as a result of the changes made in the revised manuscript. But I still wonder whether it will be entirely clear to the naïve reader. To further illustrate concretely what those figures depict, the authors might want to specify labels for each of the scale endpoints in the figure caption or corresponding results text. Another possibility is to call attention to Figure 1H on Life Satisfaction, where the near-horizontal line would appear to indicate that recall of life satisfaction is almost entirely influenced by present rating and hardly at all by past rating, especially in the unvaccinated sample. Anything further that can be done to make it easier for the naïve reader to see how the data reflect recall bias would be helpful.

Finally, I note that in the time since the initial review of this manuscript, a new article has been published that contains a wide-ranging and up-to-date review of bias effects similar to those cited in references 5-9:

Schacter, D.L., Greene, C.M, & Murphy, G. (2023). Bias and constructive processes in a self-memory system. *Memory*. <https://doi.org/10.1080/09658211.2023.2232568>.

The authors might want to cite this article so that readers can be directed toward a broad current assessment of bias effects in memory that are relevant to their findings.

Referee #2 (Remarks to the Author):

The authors made valuable changes to their manuscript, which went a long way to address my concerns. I evaluate their response to my comments in turn.

Overall recall effect: The authors argue that vaccination rates and people's identification differ by country and region, and hence the overall effect is different between these regions. I see this point. Still, if it is true that a majority was vaccinated everywhere, the overall memory bias leads to an increase in support, and I think it's worth pointing that out. The current language already goes a long way in that direction, but could be more informative, for instance by indicating the average size (or range) of the minority. I am not sure the current interpretation about "revenge" is supported by the data.

Graphs and raw data: The description of the graphs is now clearer, although of course I know too much to assess the experience of a first-time reader. I find the histograms in the supplemental data a useful addition.

Correlations and causation: it is encouraging that the results hold up with the controls, even if there can still not be a claim to causation. I am still somewhat confused by the following statement in the method section: ``life satisfaction was excluded, as this hardly differed between participants, and recall was based almost entirely on current life satisfaction at population level." The last sentence sounds to me like recall was driven by life satisfaction, but maybe I misunderstand.

Relation to hindsight bias: the conceptual and empirical clarifications are appreciated.

Recall bias in study 4: I am still not entirely convinced about the interpretation. The authors show regressions that show a statistically significant relation between estimates of population variables and individual recall in Study 1, and argue that the former is a proxy for the latter (for which they lack data in Study 4). This does indeed go some way to address the point. However, the regressions also show that the correlation is modest at best (with an r -squared between 2 and 6 percent, depending on the variable). So if it is a proxy, it's a very noisy proxy, and most of the variation in population estimates has nothing to do with recall. If the authors insist on keeping this in the paper, I think there should be strong disclaimers upfront, saying the findings are suggestive at best.

Strength of manipulations: It is good to see the new data, and I agree that the effect of stronger incentives points towards a motivated bias. Given how hard it is to interpret null results for the first two interventions, I would encourage the authors to report response times differences reported in their rebuttal, to make clear subjects engaged with manipulations.

Referee #3 (Remarks to the Author):

The authors have adequately addressed my previous concerns in the manuscript and/or in their response to them. I believe this paper makes an important and interesting contribution to the literature on recall biases.

Referee #4 (Remarks to the Author):

My comments on the previous draft were relatively minor and the authors have addressed them effectively. This is a very interesting paper and I recommend publication in the current form.

Author Rebuttals to First Revision:

Response Letter for Nature Manuscript 2023-05-08562C

Referee #1

The authors have done a very good job of responding to my comments and those of the other reviewers. The additional study with a stronger incentive manipulation strengthens the paper, as do additional analyses, references, and clarifications to the main text.

We thank the reviewer for the positive evaluation. We have addressed the final comments and suggestions in the new revision and below.

Importantly, the authors have improved their characterization of the main results, especially those in Figure 1, which are complex and non-intuitive. The added text helps, and having spent a fair amount of time grappling with what is depicted, the main message conveyed by those data is clearer to me now as a result of the changes made in the revised manuscript. But I still wonder whether it will be entirely clear to the naïve reader. To further illustrate concretely what those figures depict, the authors might want to specify labels for each of the scale endpoints in the figure caption or corresponding results text. Another possibility is to call attention to Figure 1H on Life Satisfaction, where the near-horizontal line would appear to indicate that recall of life satisfaction is almost entirely influenced by present rating and hardly at all by past rating, especially in the unvaccinated sample. Anything further that can be done to make it easier for the naïve reader to see how the data reflect recall bias would be helpful.

The figure note has been reworked and extended with an example to improve clarity.

Finally, I note that in the time since the initial review of this manuscript, a new article has been published that contains a wide-ranging and up-to-date review of bias effects similar to those cited in references 5-9:

Schacter, D.L., Greene, C.M, & Murphy, G. (2023). Bias and constructive processes in a self-memory system. Memory. <https://doi.org/10.1080/09658211.2023.2232568>.

The authors might want to cite this article so that readers can be directed toward a broad current assessment of bias effects in memory that are relevant to their findings.

Many thanks for the suggestion. We agree that this new article provides additional background information about memory biases. Therefore, we decided to add it to the introduction (new reference 6).

Referee #2

The authors made valuable changes to their manuscript, which went a long way to address my concerns. I evaluate their response to my comments in turn.

We thank the reviewer for the positive evaluation. We have addressed the final comments and suggestions in the new revision and below. Page numbers refer to the clean version of the revised manuscript.

Overall recall effect: The authors argue that vaccination rates and people's identification differ by country and region, and hence the overall effect is different between these regions. I see this point. Still, if it is true that a majority was vaccinated everywhere, the overall memory bias leads to an increase in support, and I think it's worth pointing that out. The current language already goes a long way in that direction, but could be more informative, for instance by indicating the average size (or range) of the minority. I am not sure the current interpretation about "revenge" is supported by the data.

We thank the reviewer for this evaluation. We now indicate the range of people being vaccinated in Study 4 (p.11):

"The majority (88%) had received at least one dose of a COVID-19 vaccine (ranging from 72% in Japan to 96% in Spain)."

We also added that a majority intended to comply with future pandemic measures; we believe that this now avoids the impression that the effects, in sum, may lead to a decrease in support of future pandemic measures.

We also supply the ranges for the recall items of Study 4, so that they now indicate the overall memory bias, as requested (p. 12):

"Comparing estimates with the benchmark values,²³ we found the majority of participants in all countries overestimated perceived probability of infection (ranging from 65% in the United Kingdom to 92% in Italy), while the majority of participants in all countries except Japan (24%) and Mexico (42%) underestimated perceived severity of the illness in 2020 (ranging from 74% in Spain to 97% in the United Kingdom, also see Figure 3A and Extended Data Table 24). Bias regarding government effectiveness varied by country (the share of participants overestimating this variable ranged from 31% in Italy to 81% in Japan)."

Concerning the "revenge" interpretation we provide at the end of Study 4, we do believe that the notion of revenge can be applied here. We understand revenge in the context of our text as the willingness to engage in punishing behavior as a response to a (perceived) mistreatment. This notion is precisely captured by our punishment intention variables, which measure participant's agreement to (i) *"Politicians should be punished for how they handled the corona pandemic"* and (ii) *"Scientists who gave advice to the government should be punished for how they handled the corona pandemic"*. Both items elicit explicit punishment intentions as a response to past behavior, and thus can be interpreted as a desire for "revenge". The share of people is small but considerable, and we now provide this in the text.

Graphs and raw data: The description of the graphs is now clearer, although of course I know too much to assess the experience of a first-time reader. I find the histograms in the supplemental data a useful addition.

We thank the reviewer for this evaluation.

Correlations and causation: it is encouraging that the results hold up with the controls, even if there can still not be a claim to causation. I am still somewhat confused by the following statement in the method section: "life satisfaction was excluded, as this hardly differed between participants, and recall was based almost entirely on current life satisfaction at population level." The last sentence sounds to me like recall was driven by life satisfaction, but maybe I misunderstand.

We are sorry for the confusion. Life satisfaction differs from the other variables as its recall almost entirely depends on current evaluations and does merely differ between individuals with different vaccination status and identification (see Figure 1). Therefore, we excluded it from estimating individual recall bias. To avoid confusions, the text has been changed (p. 22):

"In calculating each participant's T&B model, only seven variables were included; life satisfaction was excluded, as this hardly differed between participants, and recall of life satisfaction was based almost entirely on current life satisfaction at population level (see Figure 1)."

Relation to hindsight bias: the conceptual and empirical clarifications are appreciated.

Thank you!

Recall bias in study 4: I am still not entirely convinced about the interpretation. The authors show regressions that show a statistically significant relation between estimates of population variables and individual recall in Study 1, and argue that the former is a proxy for the latter (for which they lack data in Study 4). This does indeed go some way to address the point. However, the regressions also show that the correlation is modest at best (with an r-squared between 2 and 6 percent, depending on the variable). So if it is a proxy, it's a very noisy proxy, and most of the variation in population estimates has nothing to do with recall. If the authors insist on keeping this in the paper, I think there should be strong disclaimers upfront, saying the findings are suggestive at best.

We agree that population estimates are a noisy proxy for individual-level recall and added a disclaimer to warrant careful interpretation of the findings (p. 12):

"A pretest with additional data from Study 1 showed that such population-level estimates relate to individual-level recall, as used in Studies 1–3 (for details, see Study 1 Analyses in the Methods and Extended Data Table 9). While population estimates are likely to be influenced by other factors (e.g. media exposure and education) and can only be seen as a noisy proxy for individual-level recall, they were considered a viable replacement given that no individual data was available from the past."

Strength of manipulations: It is good to see the new data, and I agree that the effect of stronger incentives points towards a motivated bias. Given how hard it is to interpret null results for the first two interventions, I would encourage the authors to report response times differences reported in their rebuttal, to make clear subjects engaged with manipulations.

We agree. The information has been added to the revised manuscript (p. 7):

"Analyses of the response times indicate that participants read the interventions; in the information (incentive, control) condition, participants took a median time of 94 (61, 45) seconds to read the instructions and recall their answers."

Referee #3

The authors have adequately addressed my previous concerns in the manuscript and/or in their response to them. I believe this paper makes an important and interesting contribution to the literature on recall biases.

We thank the reviewer for the positive evaluation.

Referee #4

My comments on the previous draft were relatively minor and the authors have addressed them effectively. This is a very interesting paper and I recommend publication in the current form.

We thank the reviewer for the positive evaluation.