nature research

Peer Review Information

Journal: Nature Human Behaviour
Manuscript Title: Memorability shapes perceived time (and vice versa)
Corresponding author name(s): Martin Wiener

Reviewer Comments & Decisions:

Decision Letter, initial version:

24th October 2023

Dear Dr Wiener,

Thank you once again for your manuscript, entitled "Memorability shapes perceived time (and vice versa)", and for your patience during the peer review process.

Your Article has now been evaluated by 3 referees. You will see from their comments copied below that, although they find your work of potential interest, they have raised quite substantial concerns. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version if you are willing and able to fully address reviewer and editorial concerns.

We hope you will find the referees' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the referees again in the absence of major revisions. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to (1) identifying key priorities that should be addressed in revision and (2) overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritized set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

We ask that you address all reviewers' concerns in full, and that you pay particular attention to the following issues:
1) Reviewer #1 expresses concern regarding the robustness of the results relative to the thresholding procedure and how this affects the outcome of the results. Please address this and show that your

results are robust to different threshold choices.
2) Reviewer #3 requests to provide additional analysis of the neural dynamics of different layers/time steps to give more insight into the underlying mechanism.
3) Reviewer #1 highlights some important presentation issues, requesting to provide, for every experiment, figures showing the full distribution of the data, including individual data points.
4) Reviewer #2 requests an additional heatmap figure to provide a more complete picture of the results.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

If you wish to submit a suitably revised manuscript, we would hope to receive it within 3 months. I would be grateful if you could contact us as soon as possible if you foresee difficulties with meeting this target resubmission date.

With your revision, please:

• Include a "Response to the editors and reviewers" document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. When formatting this document, please respond to each reviewer comment individually, including the full text of the reviewer comment verbatim followed by your response to the individual point. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

• Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

Thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Sincerely,

Giacomo Ariani
Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #1: Time perception, human vision, memorability

Reviewer #2: Time perception, artificial neural networks

Reviewer #3: Time perception, memorability, neural networks

REVIEWER COMMENTS:

Reviewer #1:
Remarks to the Author:
Ma and colleagues in this work provide four behavioral studies and computational evidence that the higher-order semantic content of an image affects its perceived temporal duration. Among the semantic factors that were systematically investigated there were the size of the scene depicted and the level of clutter (Exp. 1 and 2), as well as the level of memorability (Exp. 3 and 4).
Results of Exp. 1 and 2 converge in showing that that the scene size dilates perceived time while clutter compresses it. Results of Exp. 3 and 4 show that the memorability of an image dilates the perceived duration (memorable image last for longer), and also that the perceived duration of a scene increases its memorability.

I think the behavioral methodology employed is solid, the sample sizes are adequate, and the results seem consistent. The article is well written, and it has a clear and broad narrative. There is a high degree of novelty in the reported findings. Still, I have few concerns that I would like the Authors to address before I can fully address the suitability of this article for publication.

1) Why in Exp. 1 and 2 data were analyzed using a generalized linear mixed model approach, while on the contrary in Exp. 3 data were analyzed using an ANOVA performed on bisection points (BP)? All the three experiments were apparently performed using a similar temporal categorization task, which differed only in the type of semantic manipulation used to create images bins (i.e. size, clutter, memorability), so I do not find reasons for this choice.
2) Connected to the previous point, could you better clarify why the need for introducing a BP as a measure to analyzed for Exp. 3? Not really clear from the manuscript.
3) I would strongly encourage the authors to display for each of the 4 behavioral experiments additional plots showing measures of central tendency along with their distribution or individual data. It was quite surprising to see that the only data representation used for Exp. 1 and 2, were data from 2 exemplary participants (out of the ~50 participants who took part in the study). For example, the authors could show that average psychometric curve or threshold values (e.g. the real ones or the ones estimated from the mixed model).
4) A better description of the participants characteristics should be made in the Methods section. Were participants balanced for sex/gender in the 4 experiments? No information is reported in the method, and this is a variable with potential effects on perception.
5) Page 10: "Yet, in our findings, we also observed that the durations of more memorable images were perceived with greater precision.". I think the term precision would need to be better explain throughout the text, given the relevance it has for some of your findings. What is exactly the measure

you refer to as 'precision'? How is it extracted from your data? This is not clear from reading the manuscript.
6) The results of the application of the rCCN model provides interesting support for your behavioural data and interesting hints about the possible underlying mechanisms. Still, I would like to Authors to provide some more details about the thresholding procedure, how it was chosen and how changing threshold in your model might affect the outcome displayed in figure 4. The description on page 4, especially, stimulates my observation: "Specifically, we set an arbitrary threshold and categorized all images as "long" if the entropy value fell below that threshold and "short" if it fell above it".

Minor points:
• Why the title focuses only on the relationship between memorability and perceived time, leaving aside the other aspects investigated?
• The references list is missing the authors name, and some citation in the text are incomplete (e.g. the very first citation "(van, 2009)").
• I find the last sentence of the abstract a bit too vague: "These findings thus provide a new avenue in vision research towards the study of perceived image durations as means of explaining visual system responses.". Would there be the chance for the Authors to rephrase it so that it better reflects the conclusions of the paper?
• Page 6: "who were able to more effectively able to ignore...": please correct the syntax

Reviewer #2:
Remarks to the Author:
In this manuscript, the authors address the intriguing relationship between the perception of time as it relates to visual perception. Specifically, they explore how changes in high-level features attributed to ventral visual stream processing interact with the perception of time. This is explored via a series of psychophysical experiments that study effects related to spatial scale, clutter and memorability of complex visual scenes. I found this study interesting, particularly the discussion section, which explores competing psychological theories in concert with supporting neuroscientific evidence.

Major Comments:

1. The Introduction section reads a lot like a discussion, with significant, yet seemingly unnecessary focus on background topics. Rather I would suggest a suggest a clearer definition and targeted exposition of the problem that the work seeks to address, together with the central hypothesis – that differences in subjective timing are a cause not effect of visual processing – which is only mentioned towards the end of the discussion section.
2. The first two experiments study the effects of scene clutter on timing perception, where it is repeatedly stated that this cannot be a magnitude effect as an increase in clutter magnitude lead to an increase in the probability of a short duration classification. However, this can be restated as an decrease in "open space" magnitude leads to a higher probability of short duration classification. The discussion section itself explores this hypothesis in terms of a mentalized navigation through space. So an interpretation is possible, where a monotonic relationship of magnitude with time dilation remains valid, but the question is with respect to the magnitude of what feature. A more nuanced interpretation of the results is suggested in the results section itself.
3. Figure 2B (experiment 3) is never described in the text, while a temporal reproduction task and its

results are described in the text pertaining to experiment 4. Are these two related? If so, this needs to be clearly stated.
4. It is unclear why the authors conclude that perceived duration affects memorability from the results of experiment 4. Is it because "the inclusion of average reproduced duration from the previous day's session also improved model fit"? If so, isn't it also valid to interpret this as a contribution of inter-subject differences in memorability which affect the reproduction duration?
5. To explain the "Simpson's paradox", the authors state "It is possible, therefore, that subjects who were able to more effectively able to ignore the images, and so reproduce longer duration estimates …". But why would ignoring the image and attending more to the duration lead to longer reproduction durations rather than more precise ones?
6. What do the dots in Figures 3B and 3C represent? Also, these figures paint an incomplete picture of the results - Would it be possible to present a heatmap that shows reproduced duration as a function of P(recall) and actual duration?
7. I found the network modeling results pertaining to the relationship between memorability and time dilation as being interpreted via circular logic – rate of change in entropy is used both to measure memorability (roughly, higher rate of change=more memorable) and to measure time dilation (roughly, when entropy falls below a threshold it is categorized as long). But if the entropy drops more rapidly, for a fixed threshold, it is expected that there will be more images categorized as long. So the result that more memorable images are more likely to be categorized as long is expected, and not so "Remarkable". It seems to be a result of the seemingly arbitrary choice of using values below an entropy threshold to categorize images as being perceived for longer. It is possible, instead, that the text here is a bit terse and I had a hard time making sense of it. Either way, a clearer explanation of the metric choices and description of the results is suggested.
8. The discussion suggests that using navigation distance to explain time dilation effects related to scene size and clutter "would predict an interaction between scene size and clutter". Why would such an interaction be necessary?
9. In relation to their result relating time dilation to network processing speed, the authors state "This finding mirrors recent neural recordings in nonhuman primates, as well as modeling with with RNNs, demonstrating that perceived duration is the result of changes in the speed of neural trajectories through state space". It appears however that this explanation conflates the speed of neural trajectories with processing speed, and this seems misleading.

Minor comments:
1. On page 6, interaction between memory and reproduction duration is attributed to Figure 3B instead of 3C.
2. On page 6, "able to more effectively able to ignore" -> "able to more effectively ignore"
3. On page 10, "modeling with with RNNs" -> "modeling with RNNs"

Reviewer #3:
Remarks to the Author:
This work studied the relationship of perception time and features of images in a series of behavior experiments and machine learning model of visual processing. The first two experiments found that size and clutter have opposite effects, with the former dilating and the later contracting perceived time. Such effects cannot be explained by magnitude accounts. Then the authors studied memorability

on perceived time, finding bidirectional effects that more memorable images lead to time dilation, while longer perceived time also results in better memory recall. Using machine learning neural network model (rCNN) that has forward layers and lateral connections within layers, pre-trained for image categorization, the authors found that images with higher memorability will lead to faster processing (quicker convergence to the categorization solution).
This is nice piece of work with new findings of visual semantic features on perception time, very well written and presented. However, there are a few questions that need to be addressed before it can be accepted for publication.

1) While experiments 1 and 2 motivated the study of memorability on perception time in experiments 3 and 4 and in computational model, the two parts seemed not coherently synergized, as the authors did not go back to more explicitly offer explanation of the contraction effect of clutter (during the reading, one would expected that the clutter maybe related to lower memorability?).

2) While the rCNN model provided insight into processing speed and an association of the entropy in different time steps to the decision processing can fit to psychrometric functions, we still do not understand the "mechanism" that should be inherent in some features of the image itself. Is more memorable images has certain lower dimensions features which allow them to be better clustered in categorization, thus has deeper minima in the loss functions, so the convergence speed is faster. More analysis of the neural dynamics of different layers/time steps representing the images may also offer deeper understanding.

3) Question in 2) is important for the interpretation of the results and the understanding of the visual processing system. In Discussion and Conclusion, the authors link and interpret the results to priority coding framework. (e.g., "These results point to a framework in which time dilation is both the result and the cause of priority coding in the visual system, which is verified by computational modeling of the ventral visual stream". In human biological visual systems, such priority coding (e.g., threatening/emotional/rewarding/appetitive) could be indeed built in through environment learning, likely involving top-down processes. However, the rCNN pre-trained on image categorisation is not supposed to have such "purpose" components, but just memory structures representing the distances of the representation space of the images used for training?

Minor points:
1. While details were presented in Methods, some key information can be mentioned in the main text. For example, in Experiment 3, the presentation time of the images could be mentioned in the main text; No mention about the training of the neural network model in the main text.
2. On page 11, the claim that "Time dilation thus results from a faster speed of the network, rather than an increase in neural firing" seemed not been substantiated, as no information and analysis is available about the firing rate of the neural network model.
3. Fig. 3, the dots with different color scales seemed not explained.
4. The claims of linear associations in Fig. 2BC (BP, CV, noise) seemed to be rather crude, though nonlinear associations might be hard to hypothesize and explain.

Author Rebuttal to Initial comments

We thank the reviewers for their thoughtful critiques of our paper. We have updated our manuscript in accordance with their comments and provide below our point-by-point responses. We believe our findings are now strengthened by these additions and that our paper now provides an even more compelling argument for the link between image memorability and perceived time. We hope that you find our updated results suitable for publication. Please note that all changes in the manuscript appear as **bold** text.

Before we enumerate our responses to specific comments, we wish to provide to all reviewers a general overview of some of the changes made to our manuscript. In particular, we have updated our analysis of the RCNN results, which now includes a more in-depth exploration of the changes observed in entropy timecourses for the different memorability stimuli. To this end, we have fit a simple two-parameter model to the entropy curves, which captures well the increased "speed" with which the network reduces entropy for stimuli of higher memorability. This model provides predictions regarding the reaction time for stimuli as a function of their memorability, which we validate in two of our experiments. Further, the model provides predictions on how psychometric curves should change as a function of altering either the threshold for the response or the speed of the entropy decrease, which we now provide in supplementary materials. Finally, we apply the RCNN and model now to the entire corpus of the LaMem dataset (58,740 images) which exhibits the same relationship between model speed and memorability as observed in our sample. From this larger dataset, we can select images that are predicted to have an exclusive effect on memorability (but not perceived duration), perceived duration (but not memorability), or both. We now provide this dataset along with our other experimental data on OSF for others to explore for future experiments. However, in analyzing these data and preparing code for OSF, as required by Nature Human Behaviour, we noticed that our image analysis pipeline for the RCNN was converting all images to grayscale before running them through the model. We fixed this step, which we found did not change the results - in fact, the relationship between memorability and speed became stronger - but in the interest of full disclosure wanted to let the reviewers know of this difference.

Reviewer #1:

1) Why in Exp. 1 and 2 data were analyzed using a generalized linear mixed model approach, while on the contrary in Exp. 3 data were analyzed using an ANOVA performed on bisection points (BP)? All the three experiments were apparently performed using a similar temporal categorization task, which

differed only in the type of semantic manipulation used to create images bins (i.e. size, clutter, memorability), so I do not find reasons for this choice.

We thank the reviewer for raising this point, and apologize for neglecting to explain this difference adequately in the methods section. For the first two experiments, the number of stimulus combinations (6 levels of clutter, size, and duration) led to a high total number of trials, but a low number per individual condition for characterizing a psychometric function with common fitting routines. Indeed, with 1512 trials in a given session, this left 42 trials for each of the 36 possible image combinations, which left only 7 trials per duration, below the number needed to best characterize a psychometric function without significant bias (~10 trials per condition; cf. Treutwein, 1995; Wichmann & Hill 2001). This is why we adopted a generalized linear mixed model approach, which can demonstrate robust effects in the presence of noise (Moscatelli, et al. 2012).

For experiment 3, we had 7 levels of memorability (rather than 36 levels of size/clutter), which permitted us to increase the number of durations (7 durations), and allowed a total of 1372 trials, with 196 trials in each memorability bin and 28 trials per duration. This increase allowed us to fit the data with psychometric curves, which provided the extra benefit of being able to calculate both the bisection point and the coefficient of variation as measures of accuracy and precision (we note that, in the GLMM, one cannot easily infer these values from fitted parameters). For the reviewer, we do note that we re-analyzed the results of Experiment 3 using the GLMM approach and found similar results, with memorability scores leading to an increase in "long" responses [β=0.216, SE = 0.047, t=4.555]. We now provide better clarification on this point in the manuscript.

Treutwein, B. (1995). Adaptive psychophysical procedures. Vision research, 35(17), 2503-2522.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. Perception & psychophysics, 63(8), 1293-1313.

2) Connected to the previous point, could you better clarify why the need for introducing a BP as a measure to analyzed for Exp. 3? Not really clear from the manuscript.

Thank you, we're happy to provide clarification on this. The need for introducing the BP follows from the above points in that it provides a better measure of the relative direction of bias in perceived estimates. Further, we note that the BP and CV accompany one another in this analysis, which provides measures of both bias and precision that can be compared across levels. We now provide better clarification on this point in the manuscript.

3) I would strongly encourage the authors to display for each of the 4 behavioral experiments additional plots showing measures of central tendency along with their distribution or individual data. It was quite surprising to see that the only data representation used for Exp. 1 and 2, were data from 2 exemplary participants (out of the ~50 participants who took part in the study). For example, the authors could show that average psychometric curve or threshold values (e.g. the real ones or the ones estimated from the mixed model).

We agree with this point, and now provide additional plots. Specifically, Figure 1 now includes two panels with the average proportion of "long" responses for the size and clutter levels for each experiment. Notably, when visualizing the data in this way, we observed a tendency to say "long" more often in Experiment 1 than in Experiment 2; however, this effect was not significant when conducting an analysis of between-group differences, which we now report. For the full distributions, we provide these now as supplementary figures.

4) A better description of the participants characteristics should be made in the Methods section. Were participants balanced for sex/gender in the 4 experiments? No information is reported in the method, and this is a variable with potential effects on perception.

We now provide demographic information in the Methods section. Subjects were sampled randomly without regard for sex/gender.

5) Page 10: "Yet, in our findings, we also observed that the durations of more memorable images were perceived with greater precision.". I think the term precision would need to be better explain throughout the text, given the relevance it has for some of your findings. What is exactly the measure you refer to as 'precision'? How is it extracted from your data? This is not clear from reading the manuscript.

We apologize for any lack of description, and now provide more clarity on what we mean by "precision". For psychometric data this is calculated as the difference limen divided by the bisection point, which yields the coefficient of variation. The difference limen itself is calculated as the half-difference between the 0.75 and 0.25 points on the psychometric curve for classifying a stimulus as "long". The interpretation of "precision" in this case means that the stimulus classifications were more "sharp" around a given value, which means that the window around the categorical boundary for misclassifications (i.e. saying a short stimulus was "long" and vice versa) was smaller.

6) The results of the application of the rCCN model provides interesting support for your behavioural data and interesting hints about the possible underlying mechanisms. Still, I would like to Authors to provide some more details about the thresholding procedure, how it was chosen and how changing threshold in your model might affect the outcome displayed in figure 4. The description on page 4, especially, stimulates my observation: "Specifically, we set an arbitrary threshold and categorized all images as "long" if the entropy value fell below that threshold and "short" if it fell above it".

We are happy to provide more details about the thresholding procedure. We note that our method is similar to that conducted by the researchers who first built this rCNN (Spoerer, et al. 2020; Karapetian, et al. 2023). In those papers, the authors matched the rCNN output to reaction time by adjusting the threshold of the rCNN for different images until the highest possible correlation was observed between model "reaction time" (defined as the time when the entropy curve crossed the threshold) and human reaction time. In our case, we built psychometric functions for the model by calculating, for each time bin, the proportion of images that were classified as "long" if the entropy for that time bin fell below the threshold, thus building a psychometric curve. This process was repeated separately for each set of images within each memorability bin. Next, we fit these data with sigmoid curves using the same methods as for human data, yielding the BP and CV. To quantify the similarity of these data to human results, we next calculated the slope of a linear regression for both BP and CV against memorability bins; finally, the difference between each of these slopes and the ones observed for human data was calculated. We repeated this process for a range of entropy thresholds spanning the range of entropy values observed. Finally, we selected the entropy threshold with the lowest common difference for BP and CV effects observed in human data, thus resulting in the psychometric curves plotted in figure 4b. To the reviewers question about changing the threshold, we now provide two supplementary figures which provide details on how changing either the threshold or the "A" parameter of our simple model changes the shape of the psychometric function.

Minor points:

• Why the title focuses only on the relationship between memorability and perceived time, leaving aside the other aspects investigated?

We felt that the title should direct readers to the memorability finding, as these exhibited a bidirectional relationship between memory and time, along with the rCNN results. The other aspects are certainly a part of this, but respectfully we would like to keep the title as is.

• The references list is missing the authors name, and some citation in the text are incomplete (e.g. the very first citation "(van, 2009)").

Our apologies, this was due to our LaTeX compiler. We have ensured that all references are correct

• I find the last sentence of the abstract a bit too vague: "These findings thus provide a new avenue in vision research towards the study of perceived image durations as means of explaining visual system responses.".

Would there be the chance for the Authors to rephrase it so that it better reflects the conclusions of the paper?

We thank the reviewer for suggesting this and have modified the last sentence to more accurately reflect the conclusions.

• Page 6: "who were able to more effectively able to ignore...": please correct the syntax

Thank you. This has been fixed.

Reviewer #2:

1. The Introduction section reads a lot like a discussion, with significant, yet seemingly unnecessary focus on background topics. Rather I would suggest a suggest a clearer definition and targeted exposition of the problem that the work seeks to address, together with the central hypothesis – that differences in subjective timing are a cause not effect of visual processing – which is only mentioned towards the end of the discussion section.

We thank the reviewer for this suggestion, and have reworked our introduction to be more clear.

2. The first two experiments study the effects of scene clutter on timing perception, where it is repeatedly stated that this cannot be a magnitude effect as an increase in clutter magnitude lead to an increase in the probability of a short duration classification. However, this can be restated as an decrease in "open space" magnitude leads to a higher probability of short duration classification. The discussion section itself explores this hypothesis in terms of a mentalized navigation through space. So an interpretation is possible, where a monotonic relationship of magnitude with time dilation remains valid, but the question is with respect to the magnitude of what feature. A more nuanced interpretation of the results is suggested in the results section itself.

We appreciate this point, and recognize that the reviewer is correct in asserting that the size/clutter effects could be due to the same underlying magnitude. Indeed, this is what we had meant to suggest in the discussion about mentalized navigation, which we now phrase more clearly throughout the manuscript.

3. Figure 2B (experiment 3) is never described in the text, while a temporal reproduction task and its results are described in the text pertaining to experiment 4. Are these two related? If so, this needs to be clearly stated.

We apologize to the reviewer for any confusion regarding this figure and the temporal reproduction task. We now more clearly describe this and other aspects of the study, and have updated Figure 2 as well. Experiment 3 was a Temporal Categorization task, whereas Experiment 4 was a Temporal Reproduction task with separate groups of subjects. Note that Experiment 4 had two parts: an initial session consisting of Temporal Reproduction, and a 1-day follow-up session consisting of a Surprise Memory Test.

4. It is unclear why the authors conclude that perceived duration affects memorability from the results of experiment 4. Is it because "the inclusion of average reproduced duration from the previous day's session also improved model fit"? If so, isn't it also valid to interpret this as a contribution of inter-subject differences in memorability which affect the reproduction duration?

We thank the reviewer for this comment and apologize for the lack of clarity. The reviewer is suggesting that the effect of longer reproduced durations associated with recall is due to the fact that higher memorability is associated with longer perceived durations; ergo, a subject sees an image high(low) in memorability, and so reproduces a duration that is long(short), and so longer reproduced durations *should* be associated with greater recall because longer reproduced durations are *already* associated with higher memorability (which are recalled better). Further, some subjects are more susceptible to memorability effects, and so these subjects should exhibit a larger effect of reproduced duration on recall. We understand the interpretation, but first note that this finding is what one would expect if one *didn't* include memorability in the model, as memorability would not be controlled for; indeed, a model that only includes reproduced duration for predicting recall finds a larger effect of reproduced duration than in the model with both covariates [β=0.891, SE = 0.303, t=2.943], which is likely driven by the effect the reviewer suggests. In our case, a model with both reproduced duration and memorability outperformed a model with either one (we now report a full model comparison in the results to demonstrate this). Further, if the effect of recall were driven entirely by memorability, of which reproduced duration was a byproduct, then one would predict a high Variance Inflation Factor (VIF) in our model. In our case, the VIF value was 1.0006 (which we now report), lower than a cutoff of 5 commonly used to indicate multicollinearity.

To additionally address this, we present two things: 1) in our GLMM we now include memorability by subject as a random effect (which also provides a better fit) and find that the significant effect of reproduced duration persists; by varying the slope per subject by memorability we thus control for the intersubject memorability effects as described by the reviewer. 2) We conducted a second analysis on these data where we median split each subject's data by reproduced duration within each memorability bin. In this way, we examined the within-subject effect of reproducing a relatively short or long interval for each bin. We again find that longer reproduced durations are associated with better recall, with no interaction with memorability bin (that is, regardless of how memorable the images were, longer reproduced durations were associated with better recall). We present these data now in Figure 4 as an alternative analysis that supports the main findings.

However, we note that, in this case, there is likely nothing that can be done to *truly* control for memorability when examining the effect of reproduced duration on recall (because one could assume that inter-*trial* differences in memorability drive the differences in duration), and so one would need to conduct an experiment where reproduced duration was independently manipulated via some other means (say, via click trains, attentional orienting, or brain stimulation) and a resultant change in recall found. We suggest this now in our discussion as avenues for future work.

5. To explain the "Simpson's paradox", the authors state "It is possible, therefore, that subjects who were able to more effectively able to ignore the images, and so reproduce longer duration estimates …". But why would ignoring the image and attending more to the duration lead to longer reproduction durations rather than more precise ones?

We thank the reviewer for this question. In timing tasks, the effect of attentional manipulations is typically twofold: attending to a stimulus's duration increases both the length of that duration *and* the precision of that duration (Brown, 1997). The heuristic explanation for this effect relies on Internal Clock models where attention is thought to widen the "gate" of pacemaker pulses into an accumulator module, such that more attention leads to more pulses and a longer perceived duration. We admit that our explanation initially was a bit offhand, as we had not expected this effect. But, based on the reviewer's comment, we examined the data further and found a between-subject correlation between reproduced duration and precision, such that subjects who reproduced longer intervals also reproduced them more precisely (Spearman rho: -0.6234, p=0.0031), which follows the possibility that these subjects were attending more to duration, and less to the image qualities themselves. We now report this and expand on our explanation. Critically, we note that this effect will need to be followed up on to determine the validity of an attentional explanation.

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. Perception & psychophysics, 59, 1118-1140.

6. What do the dots in Figures 3B and 3C represent? Also, these figures paint an incomplete picture of the results - Would it be possible to present a heatmap that shows reproduced duration as a function of P(recall) and actual duration?

The dots in Figures 3B and 3C represent memorability scores (3B) and reproduced durations (3C) for individual trials/participants (the color represents the participant, whereas the position on the x-axis represents the value for that trial, whereas the position on the y-axis represents whether the subject successfully recalled that image or not, represented as a 0 or a 1. Please note there is some jitter that was introduced in the plot, which is why points on the y-axis are not exactly at 0 or 1). We are sympathetic to the reviewer's suggestion of a heatmap, but humbly suggest that our addition of the median split data provides a more complete picture of the results.

7. I found the network modeling results pertaining to the relationship between memorability and time dilation as being interpreted via circular logic – rate of change in entropy is used both to measure memorability (roughly, higher rate of change=more memorable) and to measure time dilation (roughly, when entropy falls below a threshold it is categorized as long). But if the entropy drops more rapidly, for a fixed threshold, it is expected that there will be more images categorized as long. So the result that more memorable images are more likely to be categorized as long is expected, and not so "Remarkable". It seems to be a result of the seemingly arbitrary choice of using values below an entropy threshold to categorize images as being perceived for longer. It is possible, instead, that the text here is a bit terse and I had a hard time making sense of it. Either way, a clearer explanation of the metric choices and description of the results is suggested.

We apologize for the lack of clarity here. Please note that we have updated our modeling results, including supplementary figures demonstrating how changing either the threshold or the rate of change changes the shape of the resulting psychometric functions. We note that the "remarkable" aspect of the data is not that the model recapitulates the time dilation effect of memorability (with more memorable stimuli being classified as "long" more often), but that it *also* recapitulates the precision effect of memorability (with more memorable stimuli being classified more precisely). However, to the reviewer's point, we no longer refer to this as remarkable, and now provide a clearer explanation in the text.

8. The discussion suggests that using navigation distance to explain time dilation effects related to scene size and clutter "would predict an interaction between scene size and clutter". Why would such an interaction be necessary?

We have given this some additional thought and agree that an interaction between both is not necessary. Following a similar point by Reviewer 1, we now provide a stronger case for navigation distance as a possible driver of the effects of Experiments 1 and 2.

9. In relation to their result relating time dilation to network processing speed, the authors state “This finding mirrors recent neural recordings in nonhuman primates, as well as modeling with with RNNs, demonstrating that perceived duration is the result of changes in the speed of neural trajectories through state space”. It appears however that this explanation conflates the speed of neural trajectories with processing speed, and this seems misleading.

We agree with the reviewer. Our intention was to note that these things are similar, but not that they are necessarily the same. We now provide more clarity on this point. Importantly, we note that neural data would be necessary to determine if the changes in speed observed in the RCNN are actually instantiated in biological neural networks, or merely reflect a feature of the RCNN model.

Minor comments:

1. On page 6, interaction between memory and reproduction duration is attributed to Figure 3B instead of 3C.

Fixed. Thank you

2. On page 6, “able to more effectively able to ignore” -> “able to more effectively ignore”

Fixed.

3. On page 10, “modeling with with RNNs” -> “modeling with RNNs”

Fixed.

Reviewer #3:

1) While experiments 1 and 2 motivated the study of memorability on perception time in experiments 3 and 4 and in computational model, the two parts seemed not coherently synergized, as the authors did not go back to more explicitly offer explanation of the contraction effect of clutter (during the reading, one would expected that the clutter maybe related to lower memorability?).

We thank the reviewer for this point and now provide a stronger connection between the first and second parts of our study. Indeed, clutter effects have been inversely related to memorability, which we now state. Yet, we do not think that the results of Experiments 1 and 2 are driven solely by memorability, although it is possible that memorability played a role. Rather, following points by Reviewers 1 and 2 we now provide a stronger link between the results of the first two experiments and navigational affordances.

2) While the rCNN model provided insight into processing speed and an association of the entropy in different time steps to the decision processing can fit to psychrometric functions, we still do not understand the "mechanism" that should be inherent in some features of the image itself. Is more memorable images has certain lower dimensions features which allow them to be better clustered in categorization, thus has deeper minima in the loss functions, so the convergence speed is faster. More analysis of the neural dynamics of different layers/time steps representing the images may also offer deeper understanding.

We thank the reviewer for this point, which we considered very carefully in crafting a revision. Indeed, the true "mechanism" here is still not clear; that is, why should some features drive a faster processing speed than others? Yet, this question speaks to the larger overall one of what drives memorability, which is driven by a wide number of image features (Isola, et al. 2014). Regarding the underlying process of the model, we do note that the original creators of BLnet have gone into detail regarding channel properties of the various layers with recurrent time steps (Spoerer, et al. 2020), and so we chose not to repeat their analysis here. Rather, to better understand the results, we examined further the entropy time courses as described at the start of this document, where we found that a two-parameter power function fit the entropy data extremely well (mean R^2 = 0.99). One aspect of this is that previous research has linked entropy time courses to reaction time data, and so a prediction that follows is that subjects should have faster RTs for more memorable stimuli. We subsequently examined RT data for Experiment 3 (categorization) and Experiment 4 (surprise memory test) and found the predicted relationship between RT and memorability. Using the "A" parameter from this model, which captures the "speed" component (how fast the entropy decays across time steps), we reasoned that one could

predict time dilation effects for stimuli outside of our sample, and so we fit the model to the entire corpus of the LaMem dataset (58,740 images). Here, we found the same relationship as in our sample of 196 images, with faster speeds for more memorable stimuli. Yet, while this relationship was significant, we noted substantial variability in entropy decay across the dataset, with some images exhibiting very high memorability despite very low speed. If one assumes that speed predicts time dilation effects, then one could select images predicted to have large time dilation effects (but no effect on memorability), large memorability effects (but no time dilation effect), or both. We present some of these sample images in Figure 5. Further, we make this dataset available on OSF, from which other researchers may sample images for use in other experiments or image analysis. However, one would need to test if the images predicted to dilate time truly do (or do not, in the case of low speed images).

3) Question in 2) is important for the interpretation of the results and the understanding of the visual processing system. In Discussion and Conclusion, the authors link and interpret the results to priority coding framework. (e.g., “These results point to a framework in which time dilation is both the result and the cause of priority coding in the visual system, which is verified by computational modeling of the ventral visual stream”. In human biological visual systems, such priority coding (e.g., threatening/emotional/rewarding/appetitive) could be indeed built in through environment learning, likely involving top-down processes. However, the rCNN pre-trained on image categorisation is not supposed to have such “purpose” components, but just memory structures representing the distances of the representation space of the images used for training?

We thank the reviewer for this point and note that our connection to priority coding was merely speculatory. We agree that the rCNN model does not assign any purpose to the images, and so higher memorability is likely driven by intrinsic features of the images; on the other hand, memorability effects have been recently linked to the typicality of objects contained within (Kramer, et al. 2023), and so one possibility is that faster speeds result from a smaller distance between objects in the presented image and those in the training set; in this way, the speed effects of the RCNN depend on the “diet” of the model (Conwell, et al. 2023). We humbly suggest that, by providing the model predictions to the wider community, a whole host of studies can be done to better understand how the RCNN results contribute to scene memorability.

Kramer, M. A., Hebart, M. N., Baker, C. I., & Bainbridge, W. A. (2023). The features underlying the memorability of objects. Science advances, 9(17), eadd2981.

Conwell, C., Prince, J. S., Kay, K. N., Alvarez, G. A., & Konkle, T. (2022). What can 1.8 billion regressions tell us about the pressures shaping high-level visual representation in brains and machines?. BioRxiv, 2022-03.

Minor points:

1. While details were presented in Methods, some key information can be mentioned in the main text. For example, in Experiment 3, the presentation time of the images could be mentioned in the main text; No mention about the training of the neural network model in the main text.

We thank the reviewer for this and now provide further details regarding presentation time. With regards to the RCNN, we used pre-trained weights for the model, which we now clearly state.

2. On page 11, the claim that "Time dilation thus results from a faster speed of the network, rather than an increase in neural firing" seemed not been substantiated, as no information and analysis is available about the firing rate of the neural network model.

We agree, and now temper this claim.

3. Fig. 3, the dots with different color scales seemed not explained.

We apologize for not stating this (Reviewer 2 made a similar comment) and now state what the dots represent (they are the individual trial values of memorability and reproduced duration with the color scale representing individual subjects).

4. The claims of linear associations in Fig. 2BC (BP, CV, noise) seemed to be rather crude, though nonlinear associations might be hard to hypothesize and explain.

We agree, and now note that the linear associations are not perfect. With regard to the BP, please note that the figure of BP results in our first draft was incorrect, as we discovered when regenerating the

figures that some subjects had been left out accidentally (the statistics and all other plots were correct). This is now fixed.

Decision Letter, first revision:

12th February 2024

Dear Dr. Wiener,

Thank you for your patience as we've prepared the guidelines for final submission of your Nature Human Behaviour manuscript, "Memorability shapes perceived time (and vice versa)" (NATHUMBEHAV-23092982A). Please carefully follow the step-by-step instructions provided in the attached file, and add a response in each row of the table to indicate the changes that you have made. Please also address the additional marked-up edits we have proposed within the reporting summary. Ensuring that each point is addressed will help to ensure that your revised manuscript can be swiftly handed over to our production team.

We would hope to receive your revised paper, with all of the requested files and forms within two-three weeks. Please get in contact with us if you anticipate delays.

When you upload your final materials, please include a point-by-point response to any remaining reviewer comments.

If you have not done so already, please alert us to any related manuscripts from your group that are under consideration or in press at other journals, or are being written up for submission to other journals (see: https://www.nature.com/nature-research/editorial-policies/plagiarism#policy-on-duplicate-publication for details).

Nature Human Behaviour offers a Transparent Peer Review option for new original research manuscripts submitted after December 1st, 2019. As part of this initiative, we encourage our authors to support increased transparency into the peer review process by agreeing to have the reviewer comments, author rebuttal letters, and editorial decision letters published as a Supplementary item. When you submit your final files please clearly state in your cover letter whether or not you would like to participate in this initiative. Please note that failure to state your preference will result in delays in accepting your manuscript for publication.

In recognition of the time and expertise our reviewers provide to Nature Human Behaviour's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Memorability shapes perceived time (and vice versa)". For those reviewers who give their assent, we will be publishing their names alongside the published article.

Cover suggestions

We welcome submissions of artwork for consideration for our cover. For more information, please see our guide for cover artwork.

If your image is selected, we may also use it on the journal website as a banner image, and may need to make artistic alterations to fit our journal style.

Please submit your suggestions, clearly labeled, along with your final files. We'll be in touch if more information is needed.

ORCID

Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance:
https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

Nature Human Behaviour has now transitioned to a unified Rights Collection system which will allow our Author Services team to quickly and easily collect the rights and permissions required to publish your work. Approximately 10 days after your paper is formally accepted, you will receive an email in providing you with a link to complete the grant of rights. If your paper is eligible for Open Access, our Author Services team will also be in touch regarding any additional information that may be required to arrange payment for your article.

Please note that *Nature Human Behaviour* is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

For information regarding our different publishing models please see our Transformative Journals page. If you have any questions about costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com.

Please use the following link for uploading these materials:
[REDACTED]

If you have any further questions, please feel free to contact me.

Best regards,
Alex McKay
Editorial Assistant
Nature Human Behaviour

On behalf of

Giacomo Ariani
Editor
Nature Human Behaviour

Reviewer #1:
Remarks to the Author:
The Authors have addressed all my concerns and thus I think the manuscript is ready for publication. I just notice that the last sentence of the abstract is still the same as in the original submission, despite the Authors declared in the rebuttal letter they had revised it to better reflect the article conclusions.

Signed,
Luca Ronconi

Reviewer #2:
None

Reviewer #3:
Remarks to the Author:
In this revised version, the authors have tried hard to address the concerns, comments and suggestions from the reviewers with additional analysis of data and modeling, clarification/corrections of description, adding of information and improvements of presentations. I found the response and revision to my comments are adequate.

Author Rebuttal, first revision:

We thank the reviewers for approving our manuscript for publication. We respond here to a remaining point by Reviewer #1

Reviewer #1: "the last sentence of the abstract is still the same as in the original submission, despite the Authors declared in the rebuttal letter they had revised it to better reflect the article conclusions". Please fix this.

Our apologies for this, we have now changed the last sentence to read: These findings provide evidence for a link between image features, time perception, and memory, that can be further explored with models of visual processing.

Final Decision Letter:

Dear Dr Wiener,

We are pleased to inform you that your Article "Memorability shapes perceived time (and vice versa)", has now been accepted for publication in Nature Human Behaviour.

Please note that *Nature Human Behaviour* is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see http://www.nature.com/nathumbehav/info/gta). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Human Behaviour as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Human Behaviour logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

We look forward to publishing your paper.

With best regards,

Giacomo Ariani
Editor
Nature Human Behaviour