Peer Review Information

Journal: Nature Human Behaviour
Manuscript Title: Mental speed is high until age 60 as revealed by analysis of over a million participants
Corresponding author name(s): Mischa von Krause & Stefan T. Radev

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:

12th May 2021

Dear Mr von Krause,

Thank you once again for your manuscript, entitled "Cognitive processing speed is high until age 60: Insights from Bayesian modeling in a one million sample (with a little help of deep learning)," and for your patience during the peer review process.

Your manuscript has now been evaluated by 4 reviewers, whose comments are included at the end of this letter. Although the reviewers find your work to be of interest, they also raise some important concerns. We are very interested in the possibility of publishing your study in Nature Human Behaviour, but would like to consider your response to these concerns in the form of a revised that addresses all of the reviewer comments before we make a decision on publication.

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.

In sum, we invite you to revise your manuscript taking into account all reviewer and editor comments. 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.

We hope to receive your revised manuscript within 3 months. We understand that the COVID-19 pandemic is causing significant disruption for many of our authors and reviewers. If you cannot send your revised manuscript within this time, please let us know - we will be happy to extend the submission date to enable you to complete your work on the revision.

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. 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.

We look forward to seeing the revised manuscript and 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 these revisions further.

Sincerely,
Jamie

Dr Jamie Horder
Senior Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #1: cognitive aging

Reviewer #2: Project Implicit dataset

Reviewer #3: BayesFlow

Reviewer #4: DDMs

REVIEWER COMMENTS:

Reviewer #1:

Remarks to the Author:
The manuscript reports on analysis of IAT data from over a million participants to look at how different aspects of processing speed change across the lifespan. In particular, they use a drift diffusion model (DDM) of decision-making to combine both accuracy and reaction time in the IAT task (treating the task as an arbitrary two choice decision task rather than a measure of implicit associations) and estimate the standard DDM parameters. These include drift rate (standardly interpreted as decision time), nondecision time (perception/encoding, motor response time, and incorporating potential speed-accuracy trade-offs), and boundary separation (decision caution). They show that the linear increases in mean response time that have been reported for decades in longitudinal and cross-sectional studies (offered described as declines in processing speed with age) may be entirely explained by increases in nondecision time and in decision caution and not in drift rate / decision time, which was comparatively stable.

I thought there were some real strengths to this manuscript. It was well motivated, clear, systematic in approach, and the results were very clear. I think the use of this large dataset was important and allowed for a really comprehensive look at changes across the lifespan that would not be possible with a typical sample size. I can imagine some reviewers might quibble with the use of the IAT to look at basic processing speed, but I didn't find this to be problematic (and it would be very difficult to find a dataset of this size for another cognitive task that would be appropriate for DDM parameter estimation across the lifespan). The fact that the authors were able to replicate findings for incongruent and congruent conditions satisfied any concerns I had about the task. The conclusions they draw are also consistent with other (mostly smaller) studies, but with the added advantage that here they were able to look at generalizability by gender, education, and nationality. The sharing of the full analysis code was also strength, improving transparency and reproducibility. Finally, the patterns across the lifespan were very, very clear - assuming the underlying parameters were estimated correctly, there was no question that the cross-sectional differences in decision time look nothing like the classic findings for mean RT and processing speed (or the IAT mean RT data presented here).

With all that said, I had one major concern that is hopefully easily addressable but makes the central claim of the paper much less controversial. The main conclusion (and title) of the paper is that processing speed does not, in fact, decrease substantially across the adult lifespan. Given that declines in processing speed across the lifespan are one of the most well replicated findings in the lifespan literature, this is a provocative conclusion! But this conclusion is dependent on the authors' very narrow definition of processing speed as being the same as the drift rate parameter (representing decision time). While this may be a definition adopted in certain, narrower literatures, it is not the dominant conceptualization of what processing speed is in the aging or neuropsychological literature (where the concept is most frequently invoked). The authors state "meant RTs are not pure measures of processing speed, but instead represent the sum total of disparate cognitive processes". Based on the aging/neuropsychological literature, I would amend this statement to say that "processing speed represents the sum total of disparate cognitive processes". The idea that processing speed and decision time are the same thing is not how most neuropsychologists or aging researchers would define it. Perceptual/encoding processes and motor processes are usually considered part of processing speed - these are captured by the nondecision time parameter of the DDM model, which the authors find decreases linearly across the lifespan. The literature has also shown that nondecision time is associated with general intelligence (I believe the authors' own work supports this), so the idea that the only relevant parameter for understanding processing speed is drift rate / decision time is not well supported.

However, processing speed is a frustrating concept precisely because it lacks any underlying mechanism and collapses across so many different cognitive processes. Well-powered and well-executed efforts like this one that aim to separate different components of processing speed and understand how they change over the lifespan are still very useful. The conclusion that linear declines in processing speed across young and middle adulthood are mostly due to differences in decision caution and perceptual/encoding/motoric processes is not particularly novel, but the strength of the study (to me) is the ability to demonstrate it in such a large sample, with age treated continuously, and addressing questions of generalizability. To address this concern, the authors should make the conclusions of the manuscript more specific (and therefore clearer), using a more specific term than processing speed (or cognitive processing speed - i would consider perceptual speed and encoding speed to be cognitive!) in the title, abstract, y axis labels, and throughout the manuscript.

Other concerns were more minor:
-The difference between incongruent and congruent trials (and why they are analyzed separately) is not really explained. I am familiar with the IAT, but others may not be -- these task conditions should be explained and the decision briefly justified.
-Numbers should be given for the large number of exclusions - this is pretty typical for these sorts of datasets, but it would be good to know how many were excluded for various reasons (e.g. 100% correct, parameter estimates outside prior ranges). I didn't see this anywhere
-Reliability estimates for the various DDM parameters would be really nice to see - it was unclear to me whether higher vs lower correlations between parameters (from supplement) were associated with differences in reliability. I know trial number is already low for DDM, but if it's possible to get some sort of split half reliability for these parameters that would be useful (The individual differences DDM literature should include internal reliability estimation much more often)
-IAT has some very very strange treatment of error RTs that would shock most cognitive psychologists. I assume for this analysis that error RT was just the RT from the first (incorrect) try of a given trial (trials are often repeated until a correct answer), but authors should note in the manuscript that this was the case.

Reviewer #2:
Remarks to the Author:
This manuscript applies the diffusion model (DM) to a very large dataset to provide process-level insight into age-related differences in cognitive processing. In doing so, this manuscript reveals that the well-known positive relationship between age and response times (as a proxy for processing speed) reflects a variety of countervailing effects: a positive relationship between age and decision caution, a positive relationship between age and encoding/motor responses, and a curvilinear relationship between age and processing speed.
To position myself, I am familiar enough with the DM to understand the theory behind, and theoretical contributions made by, this manuscript. That said, I am not familiar enough with the Bayesean and deep learning estimation methods used in this manuscript to critically evaluate them, so I will leave that task to other reviewers. I am, however, highly familiar with the Project Implicit dataset used in this manuscript, and with social cognitive theory more generally. I can say with confidence that the authors' use of these data is novel and interesting. As they note, and in contrast to their use in this manuscript, IAT data such as these are generally interpreted to reflect intergroup biases. This is true, but at the same time this framing overlooks a fairly extensive literature on the process modeling of implicit social cognition (e.g., Payne, 2001; Conrey et al., 2005; Meissner & Rothermund, 2013; Stahl

& Degner, 2007; Nadarevic & Erdfelder, 2011) – though one of this manuscript's authors cites his own paper that is adjacent to this literature (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007). On the one hand, the authors have framed this manuscript as contributing to the DM literature, and I understand that word limits are tight. However, on the other hand, this manuscript is also positioned to contribute to the social cognition literature, and I think it would benefit from a brief discussion of that. For example, all of the social cognitive process modeling cites I included above (besides the Klauer paper) rely on multinomial process tree modeling, which only incorporate response frequency to the exclusion of latency. Until the very recent development of RT-MPT modeling (Heck & Erdfelder, 2016; Klauer & Kellen, 2018), DM was the only approach that could incorporate both response frequency and latency. So this manuscript makes a nice contribution to that relatively small literature on DM in social cognition.
I think this is a very well-done manuscript that relies on a gigantic sample to fill a gap in the existing literature and, thus, will make a nice contribution to multiple literatures. That said, I do have a minor point of contention with one claim the authors make. Specifically, they describe the drift rate as reflecting "a process-pure measure" (line 204). To be sure, compared to response times, drift rate is a relatively more pure measure of processing speed because the DM separately accounts for speed-accuracy and non-decision processes. However, drift rates are not process pure because they reflect the contributions of other processes besides speed-accuracy and non-decision. This is evident, for example, in Figure 2: drift rates largely converge for men and women in the congruent blocks, but diverge in the incongruent blocks. Do the authors have a theory for why drift rates should differ between men and women, but only in the incongruent blocks of the IAT, in a process pure way? I don't. However, other theoretical perspectives posit the joint contribution of multiple processes to responses on the IAT. For example, Conrey et al. (2005) posits that attitudes alone can produce responses in congruent blocks, but that attitudes and inhibition interact to produce responses in incongruent blocks. Klauer et al. (2007) suggests that drift rates are related to attitudes. Thus, known gender differences in attitudes but not inhibition may explain the gender differences in drift rates depicted in the right panels of Figure 2, which undercuts the authors' claims of process purity. I know that this is a relatively minor point in a nice manuscript, and I'm sure that the authors don't have space within the word limit to get into the nuances between evidence-accumulation and dual-process models of cognition. So I'm not asking for any of that. But even process models like the DM can't support claims that their parameters are pure measures of any specific cognitive process, and I think this is an important point to get right.

Reviewer #3:
Remarks to the Author:
My comments should be read by considering that I am not an expert of the applied area for this work (human cognition).

Key results: The paper analyzes a large dataset aimed at understanding certain characteristics of human cognition and consisting of around 1.2 million observations.
The authors challenge the majority of previous studies which use mean response times (RT) as a proxy for processing speed in human cognition.
The main finding from the authors statistical analyses is that, without using RT but instead disentangling different cognitive components, processing speed is nonlinearly associated with age, whereas previous study found that this was similar between young and older subjects.

The statistical computation components for this study are a key point for this paper. In fact, the analysed dataset is considered to be extremely large, given the typical size in previous studies (line 88: "this sample is multiple orders of magnitude larger than the samples used in all previous diffusion model studies combined"), which usually consist of around 1500 observations per study (around 60 subjects per study).
They argue that for most of these studies results are severely limited, since these are often based on comparisons between groups of subjects having very different age,
say college students vs 65-75 old subjects, thus ignoring the large age span of subjects having age 20-65.

I will comment mostly on the statistical part of their study. The methodology that makes it possible to statistically process such a large dataset is fully justified in a previous publication of the second and third author, who introduced
the so-called Bayesflow method in Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L., & Köthe, U. (2020). BayesFlow: Learning complex stochastic models with invertible neural networks. IEEE Transactions on Neural Networks and Learning Systems.
Their Bayesflow produces parameter samples from an approximate posterior distribution, by using neuronal networks to learn features (summary statistics) of a diffusion model generating arbitrary data, and then sample model parameters conditionally on the
learned summary statistics when these are computed on the observed data. This is not unlike so-called approximate Bayesian computation (ABC), but following a different logic, in that during the parameter sampling stage no costly simulated data are
produced. Therefore Bayesflow does not return exact Bayesian inference, but in exchange it makes it possible to produce inference using a very large dataset (I am not an expert of the diffusion model implemented for the data so I cannot comment here,
but in line 360 they write "Performing Bayesian estimation on hundreds of thousands participants is not feasible with current gold-standard Markov chain Monte Carlo (MCMC) methods.").
The fact that their analyes required 8 hrs on a laptop (line 379) makes Bayesian inference a viable option. The fact that the inference via Bayesflow is "amortised" (which means that the trained networks can be reused for future datasets, thus making
posterior sampling a fast procedure) is particularly appealing.

However, all in all, this work seems to be much more than just a showcase for their Bayesflow method, as by processing data orders of magnitude larger than usual they manage to uncover patterns that were previously hidden due
to limitations in studies designs (too limited age-spans used in previous studies) and the RT-based measures.

Within my own expertise in computational statistics, I find the paper well written and captivating. The authors have adhered to good recommendations for a robust "Bayesian workflow" which is to be commended, as they
evaluated the importance of checks for model sensitivity, prior assumptions and posterior inferences using appropriate tools. The analyses appear adequate.

Raw data (1.8 million observations): the author mentions that these data are available at the Project Implicit OSF page (https://osf.io/y9hiq/) which is fairly complex and I haven't verified this myself. The actually processed data
and code are found at one of the author's github page

https://github.com/stefanradev93/DataSizeMatters.

Here follow minor comments for improvement.

MAIN PAPER

- line 87: define "N"
- line 144: "increased decline" looks a bit unfortunate as an expression. Perhaps "accelerated decline" ?
- lines 308-320: maybe it's just me but I find it unpleasant and possibly confusing to put spaces between commas when writing large numbers: that is "1 , 804 , 325" looks like "1 and 804 and 325". I would write it as 1,804,325. And
similarly for other cases.
- line 374: please do not use the italic when writing "true" in "true posterior", as the reader can be (visually) tricked in thinking that you sample from the true posterior (even though you do write "approximate the true posterior").
- line 388: what do you mean with "Given the massive data set available for data mining, inferential statistics were of minor importance to our analyses"...but you definitely do give quite a lot of importance to inference in your paper! Please clarify what you meant.
- "Curve fitting section" starting at line 388 and onward: to me, as a reader, it was not really clear why you needed to do this part. I mean you explain quite a bit why you do the other analyses at the core of your paper, but to me it is a bit
more obscure (because it is less well explained) why you perform these ridge regression analyses, change point regression etc. This section appears to just come up out of the blue. Please motivate.
- some references seem to be poorly detailed in the bibliography, eg [54]-[55] but there may be more. Please recheck the list of references.

SUPPLEMENTARY

- figure 2: please set the x-axes for the 4 plots to have the same range to ease comparisons.
- figure 9: the caption means an "alpha level". What is this?

Reviewer #4:
Remarks to the Author:
Krause et al study processing speed in a binary decision-making task using and unprecedently large number of participants (more than one million) spanning a large range of ages. These data have been recorded before and is publicly posted online. The authors fit diffusion models with 3 free parameters (drift, bound and non-decision time) using a novel deep learning method and found that slowing with age is mostly due to increase in bounds and non-decision time, while cognitive processing speeds remains constant within the range 25-60 years old.

The study is very relevant, but the data are not novel, and the modelling approaches are not fully complete, leaving many unanswered questions and missing critical controls.

One major problem is the lack of a detailed description of the diffusion model being used. There are many versions that has been used in the past, and it would be critical to describe the model in detail in the modelling section by providing equations.

Another issue is that it is unclear what parameters being described (drift, bound heights and non-decision times) are free to vary across trials and conditions within a participant, and which ones are constrained. For instance, it is unclear whether the drift rate is assumed the be the same for all trials in the task within a participant, or whether the drift can be different in each trial. As inputs vary across trials, it would be critical to allow for this flexibility, but the Methods section do not seem to describe what versions of the above is being used. In addition, other parameters of the model are supposed to be the same across trials within a participant, such as the bound and the non-decision times. Again, it is unclear whether these values are constrained or unconstrained in the fit.

It is unclear whether the fitted parameters values lead to a good fit of the data or not, as no measures for the quality of the fit are shown. Therefore, it is unclear whether the fit parameters are interpretable or not.

Many versions of diffusion models allow for collapsing bounds (e.g., Drugowitsch et al, JN, 2012). In a model with collapsing bounds, does the slope of the collapsing bound depend on age? This is an important, question as wrong model assumptions can lead to wrong parameter interpretations, including processing speeds. Given that collapsing bounds and drift have somehow similar effects, it will be critical to separate the two competing interpretations.

The authors claim that given the massive amount of data, other MCMC methods do not work well. Although this can be true, it needs to be justified further. In addition, the authors could use MCMC methods to fit model parameters in smaller subsets of the data and show that MCMC and the Bayesian deep learning methods agree.

Fig 1 could be complemented with accuracy and incorrect RTs as a function of age. Would the diffusion model account for those just based on the fits using correct RTs data?

It is unclear why non-decision times are assumed to be different for correct and incorrect responses. This seems problematic, as a correct or incorrect decision only depends on what happens on the accumulation epoch, not in the motor response -as no lapses are introduced in the model, as far as I can see.

Author Rebuttal to Initial comments

Editor

Thank you for inviting us to hand in a revised version of our manuscript. We found the reviewers' comments to be particularly constructive and helpful. We have also adjusted our manuscript according to the editorial policy and formatting checklist provided by you. In addition, we have adjusted and shortened our title and abstract.

Reviewer 1

The manuscript reports on analysis of IAT data from over a million participants to look at how different aspects of processing speed change across the lifespan. In particular, they use a drift diffusion model (DDM) of decision-making to combine both accuracy and reaction time in the IAT task (treating the task as an arbitrary two choice decision task rather than a measure of implicit associations) and estimate the standard DDM parameters. These include drift rate (standardly interpreted as decision time), nondecision time (perception/encoding, motor response time, and incorporating potential speed-accuracy trade-offs), and boundary separation (decision caution). They show that the linear increases in mean response time that have been reported for decades in longitudinal and cross-sectional studies (offered described as declines in processing speed with age) may be entirely explained by increases in nondecision time and in decision caution and not in drift rate / decision time, which was comparatively stable.

I thought there were some real strengths to this manuscript. It was well motivated, clear, systematic in approach, and the results were very clear. I think the use of this large dataset was important and allowed for a really comprehensive look at changes across the lifespan that would not be possible with a typical sample size. I can imagine some reviewers might quibble with the use of the IAT to look at basic processing speed, but I didn't find this to be problematic (and it would be very difficult to find a dataset of this size for another cognitive task that would be appropriate for DDM parameter estimation across the lifespan). The fact that the authors were able to replicate findings for incongruent and congruent conditions satisfied any concerns I had about the task. The conclusions they draw are also consistent with other (mostly smaller) studies, but with the added advantage that here they were able to look at generalizability by gender, education, and nationality.

The sharing of the full analysis code was also strength, improving transparency and reproducibility. Finally, the patterns across the lifespan were very, very clear - assuming the underlying parameters were estimated correctly, there was no question that the cross-sectional differences in decision time look nothing like the classic findings for mean RT and processing speed (or the IAT mean RT data presented here).

Thank you for this very positive evaluation of our manuscript.

1.1 *With all that said, I had one major concern that is hopefully easily addressable but makes the central claim of the paper much less controversial. The main conclusion (and title) of the* *paper is that processing speed does not, in fact, decrease substantially across the adult lifespan. Given that declines in processing speed across the lifespan are one of the most well replicated findings in the lifespan literature, this is a provocative conclusion! But this conclusion is dependent on the authors' very narrow definition of processing speed as being the same as the drift rate parameter (representing decision time). While this may be a definition adopted in certain, narrower literatures, it is not the dominant conceptualization of what processing speed is in the aging or neuropsychological literature (where the concept is most frequently invoked). The authors state "meant RTs are not pure measures of processing speed, but instead represent the sum total of disparate cognitive processes". Based on the aging/neuropsychological literature, I would amend this statement to say that "processing speed represents the sum total of disparate cognitive processes". The idea that processing speed and decision time are the same thing is not how most neuropsychologists or aging researchers would define it. Perceptual/encoding processes and motor processes are usually considered part of processing speed - these are captured by the nondecision time parameter of the DDM model, which the authors find decreases linearly across the lifespan. The literature has also shown that nondecision time is associated with general intelligence (I believe the authors' own work supports this), so the idea that the only relevant parameter for understanding processing speed is drift rate / decision time is not well supported.*

However, processing speed is a frustrating concept precisely because it lacks any underlying mechanism and collapses across so many different cognitive processes. Well-powered and well-executed efforts like this one that aim to separate different components of processing speed and understand how they change over the lifespan are still very useful. The conclusion that linear declines in processing speed across young and middle adulthood are mostly due to differences in decision caution and perceptual/encoding/motoric processes is not particularly novel, but the strength of the study (to me) is the ability to demonstrate it in such a large sample, with age treated continuously, and addressing questions of generalizability. To address this concern, the authors should make the conclusions of the manuscript more specific (and therefore clearer), using a more specific term than processing speed (or cognitive processing speed - i would consider perceptual speed and encoding speed to be cognitive!) in the title, abstract, y axis labels, and throughout the manuscript.

Thank you for raising the very important point of defining processing speed. In our view, it is a great advantage of computational models in general that they provide clear procedural definitions of constructs that otherwise often are defined in a somewhat fuzzy manner. We agree that response times have often been used as a proxy for processing time. However, we think that it is problematic to consider cognitive processing speed to be the same as reaction time. In our view and following the general rationale of the diffusion model, we define processing speed as the rate at which the available evidence in the task informs the decision process. Please note, that in the diffusion model framework, response times are decomposed into a decision time and a non-decision time. Neither of these are exclusively indicators of processing speed, as decision time also depends largely on threshold settings. For example, a person that processes information quickly still can produce relatively long decision times, if she adopts very conservative thresholds (i.e., she responses only when she is sure about the correct response).

As the term "processing speed" might be misleading in the context of aging studies, we propose the use of the term "mental speed". In a classic definition, Jensen declared mental speed to be "the actual time taken to process information of different types and degrees of complexity" (Jensen, 2006, p. ix). In a similar way, Oliver Wilhelm (2015) stated that the term "[m]ental speed, however, is also used as a tool to express mental work, for example time required per correct response or number of correct responses per time unit" (Wilhelm, 2015, in the Special Issue Information of the Journal of Intelligence Special Issue "Mental Speed and Response Times in Cognitive Tests"). In both these descriptions, "mental speed" is understood as reflecting the time it takes to mentally process a certain amount of information or, put differently, the amount of information processed in a certain time frame. The term "mental" puts the focus on the extraction and processing of relevant information, that is, accumulation of evidence - this can be separated conceptually from processes of basal visual encoding and motor response execution. In this way, we feel that the term "mental speed" better reflects what is captured by the drift rate parameter than the old term "processing speed", since the latter exhibits, as you pointed out, a number of misleading connotations in the context of the cognitive aging literature.

As you have proposed, we have now adjusted our title, abstract, plot labels, and the term used to signify drift rates throughout the manuscript. In addition, we now explain in greater detail how the drift rate parameter in the diffusion model is interpreted by writing in the **Introduction:**

"By employing the DM, it is possible to obtain a model-based estimate of mental speed through the model's drift rate parameter. It is important to note, however, that drift rates do not reflect the whole chain of information processing, but rather specifically the speed of evidence accumulation. Mental speed, as measured by drift rates, is independent of decision caution (boundary separation) and the time required for encoding and motor processes (non-decision time)." (p.2)

1.2 The difference between incongruent and congruent trials (and why they are analyzed separately) is not really explained. I am familiar with the IAT, but others may not be -- these task conditions should be explained and the decision briefly justified.

We now describe the different conditions of the IAT task at several points in our paper and further justify our decision to estimate different drift rates and boundary separations for the two conditions by writing in **Methods**:

"Across the two different main blocks of the experiment, the mappings of the categories to the same response button change. "Good" might share a common response key (e.g., left) with "Black person" in the first condition (typically called "incongruent"), and then be paired with "White person" in the second condition (typically called "congruent")." (p. 8)

"We estimate separate drift rates and boundary separations for the congruent and incongruent conditions, because these parameters have been shown to differ across the IAT conditions in previous studies [51]. In the congruent condition (where, for example, the response categories "European American" and "Good" are mapped to the same response button), participants have been found to show higher drift rates and lower boundary separations than in the incongruent condition (mapping "African American" and "Good" to the same response button). " (p. 9)

1.3 *Numbers should be given for the large number of exclusions - this is pretty typical for these sorts of datasets, but it would be good to know how many were excluded for various reasons (e.g. 100% correct, parameter estimates outside prior ranges). I didn't see this anywhere*

We now report the exact numbers of cases excluded at each step in the Methods section:

"1,804,325 people started the study and provided information on their age. 1,519,257 people provided response time data. 1,313,275 people provided both their age and response time data (this sample constitutes 100% in the following comparisons). Because of potential data saving issues, we excluded cases reporting any latencies of zero, leaving us with 1,303,715 participants (99.27%), and excluded people with a different number of trials recorded than the planned 120 – 1,281,462 people (97.57%) remained. We then excluded participants with below chance (< 50%) accuracy in the response time task (1,280,075 people or 97,47% remaining). To be able to estimate error non-decision times for all participants, we also excluded cases without any errors in the two experimental blocks, with 1,201,355 participants (91.47%) remaining. Further, we excluded cases with more than 12 response times (= 10% of a person's trials) faster than 300 ms, because such an answering pattern indicates careless responding (1,189,105 people or 90.54% remaining). For our analyses, we further excluded cases with any parameter estimates beyond the bounds of our respective (very broad) uniform priors (see below: The diffusion model; 1,186,460 people or 90.34% remaining) or a reported age of more than 80 years (1,185,882 participants or 90.29% remaining) due to the low sample size in very high age. These 1,185,882 people constitute our final sample used in all analyses." (p.8)

1.4 *Reliability estimates for the various DDM parameters would be really nice to see - it was unclear to me whether higher vs lower correlations between parameters (from supplement) were associated with differences in reliability. I know trial number is already low for DDM, but if it's possible to get some sort of split half reliability for these parameters that would be useful (The individual differences DDM literature should include internal reliability estimation much more often)*

This is an important suggestion and we agree that the inclusion of measures of internal reliability would be very beneficial for the diffusion modeling literature. Unfortunately, as you pointed out yourself, the trials numbers in the IAT data are already low for obtaining diffusion model parameters. Thirty (30) trials would thus be largely insufficient for meaningful parameter estimation, as has been shown in previous studies and simulations
(Lerche & Voss, 2017). We thus cannot compute reliable split-half reliabilities, as the interpretability of the respective parameter values would be very questionable. However, we now acknowledge this limitation in the **Discussion** section:

"Finally, trial numbers for each person were quite low for diffusion modeling standards. While 60 trials (per condition) are sufficient for obtaining adequate estimates of the main model parameters according to previous studies and simulations (Lerche & Voss, 2017), we could not compute internal measures of (split-half) reliability due to the low number of trials." (p.7)

Regarding the interpretation of the across-person correlations of the parameters, we would argue that the slightly lower across-condition correlation of drift rates might reflect more specific influences across conditions on drift rates than on boundary separations, maybe due to the setup of the IAT task (i.e., aiming to assess implicit attitudes, see 2.2 below). Moreover, previous studies have, at the very least, shown that the re-test reliability is comparably high for drift rates and boundary separations (Lerche & Voss, 2017b).

1.5 *IAT has some very very strange treatment of error RTs that would shock most cognitive psychologists. I assume for this analysis that error RT was just the RT from the first (incorrect) try of a given trial (trials are often repeated until a correct answer), but authors should note in the manuscript that this was the case.*

Since the latencies of error responses in the IAT data set are unfortunately only recorded for the self-corrected answer, we decided to model this experimental peculiarity explicitly by estimating different non-decision-times (NDTs) for correct and incorrect trials and specifying a much broader prior for the incorrect NDT parameter. In this way, the incorrect NDT parameter should capture the time required to correct one's response by pressing the second response key. To back up this decision, our posterior predictive checks indicate that the model setup does indeed reflect the important data characteristics. You are right that this was not well described in the manuscript, so we now write in section **The diffusion model**:

"We estimate separate non-decision time parameters for correct and incorrect trials due to the way error response times were recorded in the race IAT (i.e., trials do not terminate immediately following a wrong response but require participants to correct their response; in our model, we incorporate the time taken for this additional response into the error non-decision time parameter, as it is unrelated to the actual decision process)." (p.9)

Reviewer 2

This manuscript applies the diffusion model (DM) to a very large dataset to provide process-level insight into age-related differences in cognitive processing. In doing so, this manuscript reveals that the well-known positive relationship between age and response times (as a proxy for processing speed) reflects a variety of countervailing effects: a positive relationship between age and decision caution, a positive relationship between age and encoding/motor responses, and a curvilinear relationship between age and processing speed. To position myself, I am familiar enough with the DM to understand the theory behind, and theoretical contributions made by, this manuscript. That said, I am not familiar enough with the Bayesean and deep *learning estimation methods used in this manuscript to critically evaluate them, so I will leave that task to other reviewers. I am, however, highly familiar with the Project Implicit dataset used in this manuscript, and with social cognitive theory more generally. I can say with confidence that the authors' use of these data is novel and interesting.*

Thank you for this very positive evaluation of our manuscript.

2.1 As they note, and in contrast to their use in this manuscript, IAT data such as these are generally interpreted to reflect intergroup biases. This is true, but at the same time this framing overlooks a fairly extensive literature on the process modeling of implicit social cognition (e.g., Payne, 2001; Conrey et al., 2005; Meissner & Rothermund, 2013; Stahl & Degner, 2007; Nadarevic & Erdfelder, 2011) – though one of this manuscript's authors cites his own paper that is adjacent to this literature (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007). On the one hand, the authors have framed this manuscript as contributing to the DM literature, and I understand that word limits are tight. However, on the other hand, this manuscript is also positioned to contribute to the social cognition literature, and I think it would benefit from a brief discussion of that. For example, all of the social cognitive process modeling cites I included above (besides the Klauer paper) rely on multinomial process tree modeling, which only incorporate response frequency to the exclusion of latency. Until the very recent development of RT-MPT modeling (Heck & Erdfelder, 2016; Klauer & Kellen, 2018), DM was the only approach that could incorporate both response frequency and latency. So this manuscript makes a nice contribution to that relatively small literature on DM in social cognition.

Thank you for pointing this out. We agree that the context of process modeling of social cognition should be referred to more clearly in our text. Accordingly, we have now added a new section in the **Discussion**:

"It should be noted that our study not only contributes to the literature on cognitive aging and diffusion modeling, but also represents an example of using mathematical modeling to analyze IAT data. Most previous model-based analyzes of IAT data have relied on multinomial processing trees (MPTs) fitted on empirical accuracy rates (e.g., Payne, 2001; Conrey et al., 2005; Meissner & Rothermund, 2013; Stahl & Degner, 2007; Nadarevic & Erdfelder, 2011). For example, the so-called quad model (Conrey et al., 2005) provides estimates of the likelihood that implicit biases are activated or overcome, but also of the likelihood that a correct answer can even be determined or that guessing occurs. Recently, a new family of MPT models incorporating response times has been introduced (Heck & Erdfelder, 2016; Klauer & Kellen,

2016; Hartmann & Klauer, 2020). Such models offer an alternative to analyze IAT data and present interesting avenues for future research in the context of cognitive aging. However, the main focus of our study was on the IAT task as an instance of a cognitive two-choice decision task, not on the processes that might be specific to the context of implicit social cognition. Still, the notable mean differences in drift rates between the congruent and incongruent IAT condition provide additional evidence for the interpretation of drift rates as further reflecting association strength in case of the IAT."
(p.7)

2.2 *I think this is a very well-done manuscript that relies on a gigantic sample to fill a gap in the existing literature and, thus, will make a nice contribution to multiple literatures. That said, I do have a minor point of contention with one claim the authors make. Specifically, they describe the drift rate as reflecting "a process-pure measure" (line 204). To be sure, compared to response times, drift rate is a relatively more pure measure of processing speed because the DM separately accounts for speed-accuracy and non-decision processes. However, drift rates are not process pure because they reflect the contributions of other processes besides speed-accuracy and non-decision. This is evident, for example, in Figure 2: drift rates largely converge for men and women in the congruent blocks, but diverge in the incongruent blocks. Do the authors have a theory for why drift rates should differ between men and women, but only in the incongruent blocks of the IAT, in a process pure way? I don't. However, other theoretical perspectives posit the joint contribution of multiple processes to responses on the IAT. For example, Conrey et al. (2005) posits that attitudes alone can produce responses in congruent blocks, but that attitudes and inhibition interact to produce responses in incongruent blocks. Klauer et al. (2007) suggests that drift rates are related to attitudes. Thus, known gender differences in attitudes but not inhibition may explain the gender differences in drift rates depicted in the right panels of Figure 2, which undercuts the authors' claims of process purity. I know that this is a relatively minor point in a nice manuscript, and I'm sure that the authors don't have space within the word limit to get into the nuances between evidence-accumulation and dual-process models of cognition. So I'm not asking for any of that. But even process models like the DM can't support claims that their parameters are pure measures of any specific cognitive process, and I think this is an important point to get right.*

We agree that drift rates, while being more specific measures of mental speed than raw RTs, are not process-pure. Consequently, we have weakened this notion in our revised manuscript by writing in the **Introduction**:

“By employing the DM, it is possible to obtain a model-based estimate of mental speed through the model’s drift rate parameter. It is important to note, however, that drift rates do not reflect the whole chain of information processing, but rather specifically the speed of evidence accumulation. Mental speed as measured by drift rates is independent of decision caution (boundary separation) and the time required for encoding and motor processes (non-decision time).” (p.2)

We also discuss this aspect, especially in relation to the gender differences in the incongruent condition you pointed out, in a new paragraph added to the **Discussion**:
“In addition, it is important to note that drift rates, while constituting a more direct measure of mental speed than mean RTs, still represent both general speed and task-specific aspects - the latter being, in our example, the association strength between IAT categories and corresponding attitudes. While this does not change the interpretation of our results on age differences in model parameters, it might help to explain some of our more specific findings, for instance, the gender differences in drift rates found specifically for the incongruent condition.” (p.7)

Reviewer 3

My comments should be read by considering that I am not an expert of the applied area for this work (human cognition).

Key results: The paper analyzes a large dataset aimed at understanding certain characteristics of human cognition and consisting of around 1.2 million observations. The authors challenge the majority of previous studies which use mean response times (RT) as a proxy for processing speed in human cognition. The main finding from the authors statistical analyses is that, without using RT but instead disentangling different cognitive components, processing speed is nonlinearly associated with age, whereas previous study found that this was similar between young and older subjects.

The statistical computation components for this study are a key point for this paper. In fact, the analysed dataset is considered to be extremely large, given the typical size in previous studies (line 88: "this sample is multiple orders of magnitude larger than the samples used in all *previous diffusion model studies combined"), which usually consist of around 1500 observations per study (around 60 subjects per study). They argue that for most of these studies results are severely limited, since these are often based on comparisons between groups of subjects having very different age, say college students vs 65-75 old subjects, thus ignoring the large age span of subjects having age 20-65.*

I will comment mostly on the statistical part of their study. The methodology that makes it possible to statistically process such a large dataset is fully justified in a previous publication of the second and third author, who introduced the so-called Bayesflow method in Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L., & Köthe, U. (2020). BayesFlow: Learning complex stochastic models with invertible neural networks. IEEE Transactions on Neural Networks and Learning Systems.

Their Bayesflow produces parameter samples from an approximate posterior distribution, by using neuronal networks to learn features (summary statistics) of a diffusion model generating arbitrary data, and then sample model parameters conditionally on the learned summary statistics when these are computed on the observed data. This is not unlike so-called approximate Bayesian computation (ABC), but following a different logic, in that during the parameter sampling stage no costly simulated data are produced. Therefore Bayesflow does not return exact Bayesian inference, but in exchange it makes it possible to produce inference using a very large dataset (I am not an expert of the diffusion model implemented for the data so I cannot comment here, but in line 360 they write "Performing Bayesian estimation on hundreds of thousands participants is not feasible with current gold-standard Markov chain Monte Carlo (MCMC) methods.").

The fact that their analyes required 8 hrs on a laptop (line 379) makes Bayesian inference a viable option. The fact that the inference via Bayesflow is "amortised" (which means that the trained networks can be reused for future datasets, thus making posterior sampling a fast procedure) is particularly appealing.

However, all in all, this work seems to be much more than just a showcase for their Bayesflow method, as by processing data orders of magnitude larger than usual they manage to uncover patterns that were previously hidden due to limitations in studies designs (too limited age-spans used in previous studies) and the RT-based measures.

Within my own expertise in computational statistics, I find the paper well written and captivating. The authors have adhered to good recommendations for a robust "Bayesian workflow" which is to be commended, as they evaluated the importance of checks for model sensitivity, prior assumptions and posterior inferences using appropriate tools. The analyses appear adequate.

Raw data (1.8 million observations): the author mentions that these data are available at the Project Implicit OSF page (https://osf.io/y9hiq/) which is fairly complex and I haven't verified this myself. The actually processed data and code are found at one of the author's github page https://github.com/stefanradev93/DataSizeMatters.

Thank you for this very positive evaluation of our manuscript.

3.1. *(MAIN PAPER) Here follow minor comments for improvement. line 87: define "N"*

Now defined.

3.2 *line 144: "increased decline" looks a bit unfortunate as an expression. Perhaps "accelerated decline" ?*

Changed accordingly.

3.3 *lines 308-320: maybe it's just me but I find it unpleasant and possibly confusing to put spaces between commas when writing large numbers: that is "1 , 804 , 325" looks like "1 and 804 and 325". I would write it as 1,804,325. And similarly for other cases.*

Changed accordingly.

3.4 *line 374: please do not use the italic when writing "true" in "true posterior", as the reader can be (visually) tricked in thinking that you sample from the true posterior (even though you do write "approximate the true posterior").*

We removed the italic emphasis.

3.5 *line 388: what do you mean with "Given the massive data set available for data mining, inferential statistics were of minor importance to our analyses"...but you definitely do give quite a lot of importance to inference in your paper! Please clarify what you meant.*

We removed the misleading sentence.

3.6 *"Curve fitting section" starting at line 388 and onward: to me, as a reader, it was not really clear why you needed to do this part. I mean you explain quite a bit why you do the other analyses at the core of your paper, but to me it is a bit more obscure (because it is less well explained) why you perform these ridge regression analyses, change point regression etc. This section appears to just come up out of the blue. Please motivate.*

We agree that this section was not well-motivated. We now justify the section by stating our aims pursued by the piecewise regressions:

“Once we had obtained parameter estimates for each participant, we aimed to represent the non-linear relationships between age and the cognitive parameters statistically. Due to the presence of non-linear relationships, we computed separate piece-wise Bayesian ridge regressions of each quantity of interest (mean correct RT and DM parameters) on age as the simplest and yet reasonable approximation of the observed age trends. With this analysis, we pursued the following two goals: i) a more principled account of the observed age-related trend changes (e.g., stability vs. decline), and ii) accurate uncertainty quantification to account for the high variability of cognitive parameters within each age group.” (p.10)

3.7 - *some references seem to be poorly detailed in the bibliography, eg [54]-[55] but there may be more. Please recheck the list of references.*

Thank you for pointing this out. We re-checked all references and made corrections where necessary.

3.8 *(SUPPLEMENTARY) figure 2: please set the x-axes for the 4 plots to have the same range to ease comparisons.*

Changed accordingly.

3.9 *figure 9: the caption means an "alpha level". What is this?*

We clarified that alpha signifies the opacity of the color.

Reviewer 4

Krause et al study processing speed in a binary decision-making task using and unprecedently large number of participants (more than one million) spanning a large range of ages. These data have been recorded before and is publicly posted online. The authors fit diffusion models with 3 free parameters (drift, bound and non-decision time) using a novel deep learning method and found that slowing with age is mostly due to increase in bounds and non-decision time, while cognitive processing speeds remains constant within the range 25-60 years old. The study is very relevant, but the data are not novel, and the modelling approaches are not fully complete, leaving many unanswered questions and missing critical controls.

Thank you for providing us with very important suggestions on how the manuscript should be improved.

4.1 *One major problem is the lack of a detailed description of the diffusion model being used. There are many versions that has been used in the past, and it would be critical to describe the model in detail in the modelling section by providing equations.*

We agree that the presentation of the diffusion model needed improvement. We have now extended the description in section The diffusion model by writing:

"The mechanistic model we use in this work has the basic form

$$dx = vdt + \xi dt^{1/2} \text{ with } \xi \sim N(0, 1)$$

where dx denotes accumulated evidence, v denotes the average speed of information accumulation (drift rate), and ξ represents a stochastic component.

A core assumption of the DM is that task-relevant information is integrated at multiple neurocognitive levels in which perceptual evidence for one of the alternatives is dynamically accumulated at a constant rate (v). A categorical decision for one of the alternatives is determined as soon as a pre-defined threshold (a) is reached. The basic DM also assumes an additive constant factor (τ) accounting for non-decision processes, such as encoding or motor responses." (p.8)

4.2 *Another issue is that it is unclear what parameters being described (drift, bound heights and non-decision times) are free to vary across trials and conditions within a participant, and which ones are constrained. For instance, it is unclear whether the drift rate is assumed the be the same for all trials in the task within a participant, or whether the drift can be different in each trial. As inputs vary across trials, it would be critical to allow for this flexibility, but the Methods section do not seem to describe what versions of the above is being used. In addition, other parameters of the model are supposed to be the same across trials within a participant, such as the bound and the non-decision times. Again, it is unclear whether these values are constrained or unconstrained in the fit.*

The reviewer is right that more clarification was necessary regarding free model parameters. In addition to the extended description of the diffusion model, we now explicitly write in section The **diffusion model**:

$$\boldsymbol{\theta} = (v_1, v_2, a_1, a_2, \tau_c, \tau_n)."$$

We did not include inter-trial variability parameters (that is, inter-trial variation in parameter estimates), since the trial numbers in the IAT are rather sparse and inter-trial variability parameters require many trials for a reliable and meaningful estimation. Moreover, these additional parameters are not readily interpretable with regard to neurocognitive processes implicated in aging. We now acknowledge and justify our decision by writing in section **The diffusion model**:

"Note, that some versions of the basic DM allow for parameters to vary randomly across trials, which introduces additional so-called inter-trial variability parameters.
Since these parameters are notoriously hard to estimate [41] and the number of trials in the IAT is rather low, we fix all inter-trial variability parameters to zero in the current application. Moreover, this decision harmonizes with our aim to keep the model as simple and as interpretable as possible." (p.9)

4.3 *It is unclear whether the fitted parameters values lead to a good fit of the data or not, as no measures for the quality of the fit are shown. Therefore, it is unclear whether the fit parameters are interpretable or not.*

We include posterior predictive checks (PPCs) typical of RT data in the Supplementary Information as part of our principled Bayesian workflow. The PPCs indicate that, overall, the parameters can reproduce the observed data quite well given the limited amount of information (60 trials per condition). Since we fit a parsimonious version of the DM including the most well-studied parameters, we believe that our parameters are interpretable within the context of evidence accumulation theories.

4.4 *Many versions of diffusion models allow for collapsing bounds (e.g., Drugowitsch et al, JN, 2012). In a model with collapsing bounds, does the slope of the collapsing bound depend on age? This is an important, question as wrong model assumptions can lead to wrong parameter interpretations, including processing speeds. Given that collapsing bounds and drift have somehow similar effects, it will be critical to separate the two competing interpretations.*

This is indeed an intriguing and important point. As suggested by the reviewer, we have implemented and fit a popular diffusion model allowing for collapsing bounds, as formulated by Voskuilen, Ratcliff, and Smith (2016). The model has the same generative form (stochastic ODE) as the basic diffusion model, but assumes that the upper ($a1$) and lower ($a2$) boundaries collapse as a function of time t:

$$a_1(t) = \frac{a}{2}\left(1 - \kappa \frac{t}{t + t_{0.5}}\right)$$

$$a_2(t) = -\frac{a}{2}\left(1 - \kappa \frac{t}{t + t_{0.5}}\right)$$

where κ denotes the amount of collapse and $t0.5$ represents a semi-saturation constant. Thus, the model embodies two additional free parameters to estimate, over which we place sufficiently broad priors. All implementation details and a reproducible Bayesian workflow for this model are included in the notebook *IAT_BayesFlow_TrainingPhase_Collapsing.ipynb* in the accompanying code repository.

The first important observation we made is that the new parameters κ and $t0.5$ are hardly (if any) identifiable given 120 trials (i.e., posterior equals prior in most cases). This observation runs in favor of the more parsimonious basic diffusion model we used in the main paper, since the data appears to contain little to no information for the recovery of the collapsing bounds parameters. Moreover, our observation confirms that of Voskuilen et al. (2016), who reported that the collapsing boundary model generates response times distributions with the same shape as those produced by the fixed boundary model and that the additional parameters are rather difficult to estimate.

The second important observation we made is that the age-related patterns of the main diffusion parameters obtained on the basis of the collapsing bounds model are essentially the same as when using the more parsimonious basic diffusion model. These results are now available in the Supplementary Information (p. 16) and they represent a further robustness check for our reported relationships between age and cognitive parameters.

4.5 *The authors claim that given the massive amount of data, other MCMC methods do not work well. Although this can be true, it needs to be justified further. In addition, the authors could use MCMC methods to fit model parameters in smaller subsets of the data and show that MCMC and the Bayesian deep learning methods agree.*

As suggested by the reviewer, we also apply HMC-MCMC (as implemented in Stan) to fit a corresponding diffusion model to a random sub-sample of 6400 participants (200 from each of the 32 time-ordered original data files). We provide the code for reproducing these results on our GitHub page.

Our MCMC results show that parameter estimates obtained via MCMC sampling were strongly correlated to parameter estimates obtained via BayesFlow. The across-method correlations in parameters were .96 and .97 for drift rates and .96 and .93 for boundary separation for the congruent and incongruent condition, respectively. Non-decision times in correct trials correlated at .91, while error non-decision times correlated at .90. These results were obtained using the very strict criterion of excluding any cases that showed any divergent transitions during MCMC sampling. When not excluding these, correlations were even higher. We report these analyses in the **Supplementary Information** (p. 18).

Most importantly, MCMC chains (using as few as 1000 samples and a burn-in period of another 1000 samples) took on average 33 seconds to run on a multi-core workstation. Thus, obtaining parameter estimates for 1,189,555 people would have taken approximately 456 days (1,189,555*33 seconds / 3600 (scale to hours) / 24 (scale to days)) when using MCMC sampling, strengthening our account that such an approach would be practically infeasible.

4.6 *Fig 1 could be complemented with accuracy and incorrect RTs as a function of age. Would the diffusion model account for those just based on the fits using correct RTs data?*

We now include figures showing accuracy rates and incorrect RTs (in the two experimental conditions) as a function of age in the Supplementary Information (p. 13 – 15). Our diffusion model fits are based on both correct and error RTs. As our posterior predictive checks show (see point 4.3 above), the model accounts for both correct and error RTs fairly well.

4.7 *It is unclear why non-decision times are assumed to be different for correct and incorrect responses. This seems problematic, as a correct or incorrect decision only depends on what happens on the accumulation epoch, not in the motor response -as no lapses are introduced in the model, as far as I can see.*

As we describe above in our answer to Reviewer 1 (1.5), we include the two different non-decision time parameters for correct and error responses to account for the particular way error response times are recorded in the IAT task. As error response times not only include the time taken until the erroneous response is given, but also the time needed to correct one's answer, we model this additional time as a non-decisional process specific to error trials. In this way, we can account for the recorded error RTs within the model. As our posterior predictive checks show, our approach worked reasonably well in providing adequate model fit also for error RTs. We still do acknowledge that our approach might be deemed sub-optimal from a modeling perspective. However, the focus of the present work lies on age-related patterns of the model's parameters, not on perfectly reproducing the patterns in error response times found for this particular task. Given the very robust age differences we found, we are confident that this modeling choice did not influence the interpretation of our results.

References

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Elsevier.

Lerche, V., & Voss, A. (2017). Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological research*, 81(3), 629-652.

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior research methods*, 49(2), 513-537.

Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of mathematical psychology*, 73, 59-79.

Wilhelm, O. (2015). Special Issue "Mental Speed and Response Times in Cognitive Tests". *Journal of intelligence*. Retrieved from http://www.mdpi.com/journal/jintelligence/special_issues/mentalspeed

Decision Letter, first revision:

3rd September 2021

Dear Dr von Krause,

Thank you once again for your revised manuscript, entitled "Mental speed is high until age 60 as revealed by a model-based neural network analysis of big data," and for your patience during the peer review process.

Your manuscript has now been evaluated again by 2 reviewers (the other 2 did not raise any major concerns last time, so we did not ask them to re-review), whose comments are included at the end of this letter.

You will see that the reviewers find your manuscript to be much improved, however, Reviewer #4 has some remaining comments that must be addressed. Therefore, we'd like to invite you to submit another revised version.

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.

In sum, we invite you to revise your manuscript taking into account all reviewer and editor comments. 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.

We hope to receive your revised manuscript within four to eight weeks. We understand that the COVID-19 pandemic is causing significant disruption for many of our authors and reviewers. If you cannot send your revised manuscript within this time, please let us know - we will be happy to extend the submission date to enable you to complete your work on the revision.

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. 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.

We look forward to seeing the revised manuscript and 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 these revisions further.

Sincerely,
Jamie

Dr Jamie Horder
Senior Editor
Nature Human Behaviour

REVIEWER COMMENTS:

Reviewer #1:
Remarks to the Author:
The authors have been very responsive to my concerns. I believe the manuscript in its current form would make an important contribution to the literature.

Laura Germine

Reviewer #4:
Remarks to the Author:

The authors have substantially improved the paper, but I still have one important request:

It is still unclear whether the model can fit well the accuracy and the error RTs data. In the new version of the manuscript, the authors plot the raw data for accuracy and error RTs in the SI, but no fit is provided, as far as I can see. I would recommend using the format of Fig. 1 to show that the model fits to accuracy and error RTs. Do the model predictive means provide a good fit?

In addition, a reference to the paper Drugowitsch et al, JN, 2012 could be added to the paper in relation to the new additions on collapsing bound. Just a minor detail: I understand that the parameters of the collapsing bounds are very variable and difficult to interpret, but is the model with collapsing bounds better than the model without them? Did the authors provide a model comparison? Maybe I missed it, but it was not obvious to me.

Author Rebuttal, first revision:

Reviewer 4

The authors have substantially improved the paper, but I still have one important request:
It is still unclear whether the model can fit well the accuracy and the error RTs data. In the new version of the manuscript, the authors plot the raw data for accuracy and error RTs in the SI, but no fit is provided, as far as I can see. I would recommend using the format of Fig. 1 to show that the model fits to accuracy and error RTs. Do the model predictive means provide a good fit?

Thank you for acknowledging our revision. We now provide extended analyses of model fit to accuracy data, error response times, and correct response times, separated by experimental conditions. These can be found in the SI, Figures 9, 10, and 11. Model fit is adequate also for accuracy and error response times data. We describe our posterior predictive check procedure on page 6 in the SI.

In addition, a reference to the paper Drugowitsch et al, JN, 2012 could be added to the paper in relation to the new additions on collapsing bound.

We added the reference.

Just a minor detail: I understand that the parameters of the collapsing bounds are very variable and difficult to interpret, but is the model with collapsing bounds better than the model without them? Did the authors provide a model comparison? Maybe I missed it, but it was not obvious to me.

We did not perform a formal model comparison between the standard diffusion model and the collapsing boundaries model. The main reason for this decision is that we only opted to show that the two versions of the diffusion model considered in our study yield identical age-related patterns regarding the three core model parameters. Indeed, we deem the study of these age-related patterns the main purpose of our paper.

Accordingly, we treated the question of which model provides a better fit to the current data as beyond the scope of the present paper. Furthermore, in a Bayesian framework, there are multiple approaches to address the question of model choice:

1. Prior predictive approaches (i.e., Bayes factors, model posterior probabilities)
2. Posterior predictive approaches (i.e., cross-validation, WAIC)
3. Metric-based approaches (i.e., predictive performance on new data with respect to a metric)

A systematic model comparison study would thus require a thorough consideration of the utility of all three approaches and would introduce a lengthy discussion which goes beyond the purpose of the current work.

Note, that we provide a detailed analysis of model fit for the collapsing boundaries model on our project's GitHub page but prefer to leave relative model preference for future work designed especially for the task (and considering more than just the two models applied in our study).

Decision Letter, second revision:

Our ref: NATHUMBEHAV-210314680B

16th September 2021

Dear Dr. von Krause,

Thank you for submitting your revised manuscript "Mental speed is high until age 60 as revealed by a model-based neural network analysis of big data" (NATHUMBEHAV-210314680B). It has now been seen by Reviewer #4 again. As you can see, the reviewers find that the paper has improved in revision.

I am therefore happy to say that we are willing in principle to publish it in Nature Human Behaviour, pending minor revisions to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements shortly. Please do not upload the final materials and make any revisions until you receive this additional information from us.

However, you may wish to start making changes to the manuscript now. One important issue you will need to address is the reporting of statistics. Nature Human Behaviour requires key results to be supported by statistics reported in the text, not just in figures alone. You will need to add:

1) the results of the linear Bayesian change-point regression (this can be reported in the SI but it must be in text)
2) more importantly, the results of the piece-wise Bayesian ridge regression (the key statistics must be given in the main text; more details can be given in Tables or Extended Data Figures).

Essentially, you will need to fully report the statistics underlying each panel of Figure 1, in the main text. All other results will also need to be supported by full statistics, but these could be located in the SI.

Please do not hesitate to contact me if you have any questions.

Sincerely,
Jamie

Dr Jamie Horder
Senior Editor
Nature Human Behaviour

Reviewer #4 (Remarks to the Author):

Thanks for the changes

Decision letter, final requests:

** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors. **

Our ref: NATHUMBEHAV-210314680B

5th November 2021

Dear Dr. von Krause,

Thank you for your patience as we’ve prepared the guidelines for final submission of your Nature Human Behaviour manuscript, "Mental speed is high until age 60 as revealed by a model-based neural network analysis of big data" (NATHUMBEHAV-210314680B). 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 check and comment on any additional marked-up edits we have proposed within the text. 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 "Mental speed is high until age 60 as revealed by a model-based neural network analysis of big data". For those reviewers who give their assent, we will be publishing their names alongside the published article.

Cover suggestions

As you prepare your final files we encourage you to consider whether you have any images or illustrations that may be appropriate for use on the cover of Nature Human Behaviour.

Covers should be both aesthetically appealing and scientifically relevant, and should be supplied at the best quality available. Due to the prominence of these images, we do not generally select images featuring faces, children, text, graphs, schematic drawings, or collages on our covers.

We accept TIFF, JPEG, PNG or PSD file formats (a layered PSD file would be ideal), and the image should be at least 300ppi resolution (preferably 600-1200 ppi), in CMYK colour mode.

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 you will not receive your proofs until the publishing agreement has been received through our system.

Please note that <i>Nature Human Behaviour</i> 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

<B>Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. For submissions from January 2021, 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 our standard licensing terms will need to be accepted, including our self-archiving policies. Those standard 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,
Chloe Knight
Editorial Assistant

Nature Human Behaviour

On behalf of

Jamie

Dr Jamie Horder
Senior Editor
Nature Human Behaviour

Reviewer #4:
Remarks to the Author:
Thanks for the changes

Final Decision Letter:

Dear Dr von Krause,

We are pleased to inform you that your Article "Mental speed is high until age 60 as revealed by analysis of over a million participants", has now been accepted for publication in Nature Human Behaviour.

I'm very sorry for the delay in this final step of the acceptance process.

Please note that *Nature Human Behaviour* is a Transformative Journal (TJ). Authors whose manuscript was submitted on or after January 1st, 2021, 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. IMPORTANT NOTE: Articles submitted before January 1st, 2021, are not eligible for Open Access publication. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. For submissions from January 2021, 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 our standard licensing terms will need to be accepted, including our self-archiving policies. Those standard licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

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.

You will not receive your proofs until the publishing agreement has been received through our system.

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,
Jamie

Dr Jamie Horder
Senior Editor
Nature Human Behaviour