ARTICLES
https://doi.org/10.1038/s41562-021-01282-7

nature human behaviour

Mental speed is high until age 60 as revealed by analysis of over a million participants

Mischa von Krause[1,2] ✉, Stefan T. Radev[1,2] ✉ and Andreas Voss[1]

Response speeds in simple decision-making tasks begin to decline from early and middle adulthood. However, response times are not pure measures of mental speed but instead represent the sum of multiple processes. Here we apply a Bayesian diffusion model to extract interpretable cognitive components from raw response time data. We apply our model to cross-sectional data from 1.2 million participants to examine age differences in cognitive parameters. To efficiently parse this large dataset, we apply a Bayesian inference method for efficient parameter estimation using specialized neural networks. Our results indicate that response time slowing begins as early as age 20, but this slowing was attributable to increases in decision caution and to slower non-decisional processes, rather than to differences in mental speed. Slowing of mental speed was observed only after approximately age 60. Our research thus challenges widespread beliefs about the relationship between age and mental speed.

Mental speed is a fundamental property of cognitive agents and an important prerequisite for timely and adequate responses in complex environments. Older people are often assumed to be slower thinkers than younger people—a notion that has notable consequences in work life[1,2] and that has seemingly found strong empirical support. Over the past few decades, a large body of research has consistently reported a negative relation between mental speed and age; that is, older people tend to be slower than younger people across a wide variety of cognitive tasks and contexts[3,4]. This approximately linear trend starts in young adulthood, at ages 20 to 30 (refs. [3,5–7]), and has been reported in both cross-sectional and longitudinal studies[3,7–9]. The notion that mental speed declines as early as young and middle adulthood has important implications for the study of human cognition. Moreover, since developmental patterns of cognitive abilities are linked to changes in the brain[10], studying these patterns can also provide insights into the neurophysiological basis of cognition.

The vast majority of findings on age and mental speed rely on mean response times (RTs) in elementary cognitive tasks (for example, comparison of two letters) as a measure of basic speed of information processing[4,5,11]. However, this approach has two major shortcomings. First, the solitary use of mean RTs does not utilize the full information contained in empirical RT distributions and ignores accuracy data that are also obtainable from experimental paradigms. Second, mean RTs are far from pure measures of mental speed but instead represent the sum total of disparate cognitive processes[12]. For instance, speed–accuracy trade-offs (that is, different settings of response caution that affect both the speed and accuracy of responses) and the time taken for encoding and motor processes contribute to mean RTs, although they are unrelated to mental speed. Thus, the extent to which mean RTs reflect mental speed is, at the very least, debatable[13–15].

To avoid reliance on aggregate data, mathematical models of cognition strive to decompose observed behaviour into interpretable and neurophysiologically plausible parameters. One of the most popular process models for analysing RT data is the diffusion model (DM[16–20]; see Methods for a more detailed description of the model). 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; rather, they specifically reflect 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). Moreover, the parameters of the DM have been extensively validated both experimentally[21–23] and neurophysiologically[24–26].

In the past two decades, a growing number of diffusion modelling studies on age differences in a great variety of experimental environments has been published[14,23,27–38]. Most of these studies compared groups of young adults (around age 20) with old adults (aged 60 and older), with respect to the model's parameters. Interestingly, it has often been reported that mental speed exhibits no differences between young and old adults. Conversely, decision caution and non-decision times were often markedly increased in old age.

Although model-based analyses of cognitive ageing have many advantages over the direct analysis of raw data, many model-based studies have two serious shortcomings, both related to the samples used. First, sample sizes were small in most studies, which is especially problematic for research on individual differences seeking to increase reliability through larger samples. For instance, a recent meta-analysis summarizing 25 studies had a total sample size of only 1,503 observations, indicating an average sample size of 60 participants per study[39]. Second, most studies compared only two age groups, typically college-age students and older adults aged 60 to 75. Taken together, these two aspects severely limit the generalizability of previous results, especially with regard to the age span between 25 and 60 years—that is, large parts of young and middle adulthood.

There are two main reasons for the small sample sizes common in diffusion modelling studies. First, data collection for such studies is tedious, given the large number of trials per person that were long thought to be required for diffusion modelling[17]. However, such requirements are now considered to be largely overstated[40,41]. Second, and more importantly, fitting the DM to observed data is computationally expensive, especially when employing sampling-based Bayesian estimation methods. Obtaining individual parameters even from moderately large samples is thus often

[1]Institute of Psychology, Heidelberg University, Heidelberg, Germany. [2]These authors contributed equally: Mischa von Krause, Stefan T. Radev.
✉e-mail: mischa.vonkrause@psychologie.uni-heidelberg.de; stefan.radev@psychologie.uni-heidelberg.de

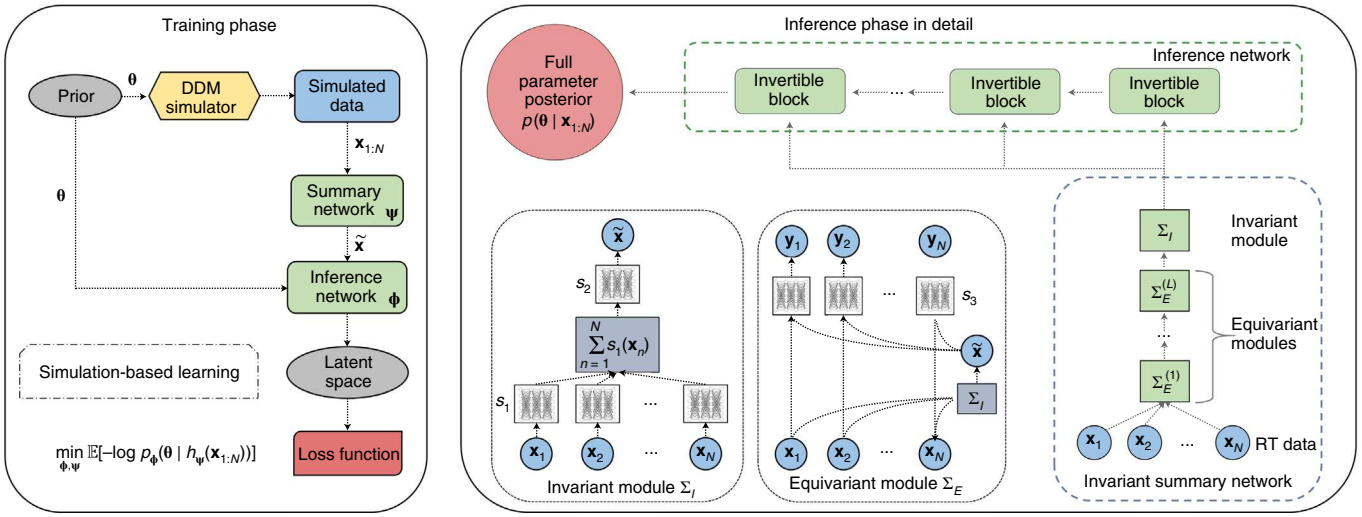

Fig. 1 | The BayesFlow framework used for individual parameter estimation on more than a million datasets. During training (left), the computational model serves as an 'instructor', which, by means of simulations, guides the summary network ($\boldsymbol{\psi}$) and the inference network ($\boldsymbol{\phi}$) to become 'experts' in inverting the model and recovering plausible estimates of cognitive parameters ($\boldsymbol{\theta}$). During inference (right), the trained networks efficiently process all observed data ($\boldsymbol{x}_{1:N}$) and estimate the full posterior over the parameters of interest. The training effort thus 'amortizes' over multiple estimation passes, as no further training of the networks is required[65]. Specialized invariant networks (s_1, s_2)[64] process the independent and identically distributed (i.i.d.) RTs and accuracy data to obtain a fixed-size vector representation $\tilde{\mathbf{x}}$. Layers of equivariant networks (s_3) increase the expressiveness of the architecture by transforming each data point $\boldsymbol{x}_n$ into an intermediate representation $\boldsymbol{y}_n$. See Methods for further details.

infeasible for practical reasons. Yet, to provide a robust analysis of individual differences in mental speed in relation to age, a rather large dataset including participants across the entire lifespan seems imperative.

In recent years, Bayesian methods have become the gold standard for model-based inference in cognitive modelling[42]. Bayesian methods allow for principled uncertainty quantification in the form of full posterior distributions over quantities of interest (for example, the parameters of a cognitive model). Once estimated, the posterior distribution can be used to extract credibility intervals or to perform posterior predictions to assess the quality of model fit. Moreover, posterior correlations between model parameters can be extracted and used as measures of (linear) disentanglement between parameters at an individual level. However, a major disadvantage of standard Bayesian methods for cognitive models (for example, Markov chain Monte Carlo methods) is their computational slowness, which makes them impractical or even impossible to apply in data-rich contexts. In this work, we therefore demonstrate the utility of a deep learning framework developed to scale up model-based Bayesian inference to millions of datasets[43].

We present an analysis of cross-sectional age differences in DM parameters estimated from a massive dataset of more than 1,000,000 participants, using RTs and accuracy rates collected in an online implicit association test (IAT[44]). Notably, this sample is multiple orders of magnitude larger than the samples used in all previous DM studies combined. Our deep learning architecture for parameter estimation is based on a two-stage inference framework, which is illustrated in Fig. 1 and described in the Methods[43]. Regarding chronological age, our sample covers childhood till late adulthood (ages 10 to 80), with a sufficient depth for fine-grained and robust year-by-year analysis.

In our study, we derive substantial insights into individual differences in cognitive parameters by applying Bayesian diffusion modelling to a large sample with the help of modern deep learning methods. Accordingly, our approach yields robust findings on age-related patterns of different aspects of cognition, separating mental speed, decision caution and non-decision parts of RTs.

We observe a clear nonlinear association between drift rate as an index of mental speed and age, which is strikingly different from the association implied by mean RTs and more informative than the age differences found in previous DM studies. Our model-based analysis thus reveals a picture of age differences in cognitive parameters yielding a radically different implication than the one based on analyses of raw RT data.

Results

Table 1 shows descriptive statistics of age, mean correct RTs in both experimental conditions (incongruent and congruent; see Methods for a description of the experimental setup) and the posterior means of all estimated DM parameters. Figure 2 depicts our main findings. Mean correct RTs, mental speed, decision caution and correct non-decision time are plotted against age in years. The figure shows the results for one of the two experimental conditions (incongruent trials). The other condition (congruent trials) yields very similar patterns, which are presented in the Supplementary Information (Supplementary Fig. 12). Each dot represents the mean of the individual posterior parameter means for one year of age. The vertical bars represent one standard deviation within each year of age. To better describe the age-related patterns we found, we estimated linear Bayesian change-point models combined with piecewise Bayesian ridge regressions (Methods). The estimated change points and piecewise regression lines together with their respective uncertainties are also depicted in Fig. 2.

Mean RTs. As evident from Fig. 2, cross-sectional mean correct RTs decrease sharply from age 10 to about 20, with the change point of the age trend estimated at age 19 (mean model-implied change per year, $\bar{b} = -0.024$; 95% highest density interval (HDI), (−0.025, −0.023); change-point posterior mean, 19.0; 95% HDI, (18.2, 19.8)). After that, mean correct RTs show a quasi-linear increasing trend until the estimated change point at age 62 ($\bar{b} = 0.006$; 95% HDI, (0.006, 0.006); change-point posterior mean, 62.1; 95% HDI, (60.5, 63.7)). Thereafter, the average increase in RTs per year accelerates ($\bar{b} = 0.022$; 95% HDI, (0.021, 0.023)), although the data become more sparse when approaching age 80 (for example, $N = 169$ for age 80).

Table 1 | Descriptive statistics

	Mean	s.d.	Minimum	Maximum
Age	27.42	12.33	10.00	80.00
Mean correct RT (incongruent)	1.00	0.31	0.36	5.75
Mean correct RT (congruent)	0.86	0.24	0.36	5.16
Mental speed (incongruent)	1.58	0.64	0.10	6.99
Mental speed (congruent)	2.10	0.86	0.10	6.99
Decision caution (incongruent)	1.91	0.51	0.24	4.00
Decision caution (congruent)	1.83	0.53	0.46	4.00
Non-decision time (correct)	0.38	0.07	0.10	2.89
Non-decision time (incorrect)	1.29	0.84	0.10	7.00

Mental speed is indicated by drift rate; decision caution is indicated by boundary separation. Age was computed as the year of data collection minus the year of birth.

Mental speed. Drift rates—that is, our proxy for measuring mental speed—increase notably from age 10 to 30 in our cross-sectional data ($\bar{b} = 0.034$ until the first change point; 95% HDI, (0.033, 0.034)). After this, the mean values of drift rates remain fairly stable until age 60, showing little age-related difference during middle adulthood ($|\bar{b}| < 0.001$; 95% HDI, (−0.001, 0.000)). Around age 60, an accelerated negative trend in mental speed commences, which holds until age 80 ($\bar{b} = -0.020$; 95% HDI, (−0.021, −0.018)). Importantly, this inverted U-shaped pattern does not mirror the age patterns found for the other DM parameters and mean RTs. Our change points are estimated at ages 24 (posterior mean, 24.4; 95% HDI, (22.8, 26.2)) and 60 (posterior mean, 59.9; 95% HDI, (56.9, 62.8)). The change-point model misses the minor increase in drift rates that continues until age 30, as well as the slight decrease in drift rates starting at age 50.

Decision caution. Boundary separation—that is, estimates of decision caution—decreases from age 10 to about age 20 ($\bar{b} = -0.025$; 95% HDI, (−0.026, −0.023)), after which it shows a quasi-linear increase until age 65 ($\bar{b} = 0.011$; 95% HDI, (0.011, 0.011)). Thereafter, the average increase in boundary separation per year accelerates ($\bar{b} = 0.021$; 95% HDI, (0.018, 0.023)). Change points are estimated at ages 18 (posterior mean, 17.8; 95% HDI, (17.3, 18.4)) and 65 (posterior mean, 64.9; 95% HDI, (62.9, 66.8)). It should be noted that in the congruent experimental condition (Supplementary Fig. 12), the second change point for boundary separation is estimated at age 40 (posterior mean, 40.2; 95% HDI, (37.4, 43.0)), and the subsequent increasing trend is less pronounced there ($\bar{b} = 0.005$; 95% HDI, (0.004, 0.005)).

Non-decision times. Non-decision time estimates—that is, the time taken for encoding and motor response—decrease from age 10 to the estimated change point of age 15 ($\bar{b} = -0.006$; 95% HDI, (−0.008, −0.005); change-point posterior mean, 15.0; 95% HDI, (13.4, 16.7)), after which they show a quasi-linear increase until age 80 ($\bar{b} = 0.002$; 95% HDI, (0.002, 0.002)). The age differences for decision caution and non-decision times closely mirror the pattern found for RTs, suggesting that these components could have a large impact on the mean levels of response latencies over the life course.

Additional analyses and robustness checks. As can further be observed, variability in the mean correct RTs increases across the lifespan (correlation of age and standard deviation in mean correct RT in the incongruent condition, $r = 0.78$; 95% HDI, (0.70, 0.85); in the congruent condition, $r = 0.76$; 95% HDI, (0.68, 0.83)). This trend is paralleled by the increase in variance found for correct non-decision times ($r = 0.75$; 95% HDI, (0.67, 0.83)), error non-decision times ($r = 0.92$; 95% HDI, (0.90, 0.95)) and boundary separations (incongruent boundary separations, $r = 0.48$; 95% HDI, (0.33, 0.63); congruent boundary separations, $r = 0.90$; 95% HDI, (0.87, 0.94)). Conversely, the between-person variability in drift rates shows no clear pattern of age-related increase (incongruent drift rates, $r = 0.17$; 95% HDI, (0.00, 0.34); congruent drift rates, $r = -0.56$; 95% HDI, (−0.67, 0.42)).

To ensure that our findings hold across a wide range of conditions, we conducted several robustness checks. Figure 3 shows that the mean-level pattern for drift rates is robust across genders, levels of education and experimental conditions (congruent versus incongruent). However, the accelerated decline in drift rates after age 60 is more pronounced for the incongruent condition, and women show higher mean levels of drift rates also in the incongruent condition. The vertical bars in Fig. 3 indicate standard errors of the means. Due to the very large sample size, the standard errors are very small for all age groups except for the very old participants. This guarantees that the differences in mental speed across the lifespan were assessed very accurately. We performed additional robustness checks by comparing the trends in age effects across different subsamples. For this purpose, we first divided the sample into four almost evenly sized subsamples. Across these subsamples, the mean-level patterns were virtually identical. The same was true when comparing participants born in the United States with those originating from other countries, as well as when comparing participants working on tasks with different classes of stimuli (that is, 'Black/White' or 'African American/European American'). All these additional analyses can be found in the Supplementary Information (Supplementary Figs. 13–16), where we also report correlations between the different DM parameters, both across participants and within each person—the latter by utilizing the individual posterior distributions (Supplementary Tables 1 and 2).

Discussion

In this work, we presented a cross-sectional study of age differences in mean RTs and cognitive processes as measured by the DM. We applied the DM to a massive dataset containing RT and accuracy data from the IAT. Our sample covers large parts of the human lifespan (ages 10 to 80) in sufficient depth for a fine-grained analysis of age differences at a year-specific level. Given the sample size, our analyses would have been infeasible using standard parameter estimation procedures. Thus, our deep learning method for parameter estimation was both necessary and efficient for the task at hand. Moreover, our findings stand in pronounced contrast to previous findings on age differences in mental speed. We will now discuss the implications of our findings.

Our results replicate the age-related decline in mean RTs previously reported[3–7]. In our sample, mean RTs followed a negative trend during the teenage years, were fastest around age 20 and showed a nearly linear increase thereafter, which further accelerated after age 60. It is important to note that these findings are in line with earlier RT studies. This indicates that the diverging patterns found for the DM parameters are not based on data that are qualitatively different from those typically obtained in the field.

In the DM, decision caution (that is, the amount of information sampled before making a decision) is represented by the boundary separation parameter[18]. Our results suggest that, on average, boundary separation declines from age 10 to approximately age 18, indicating that people at college age were the least cautious in our

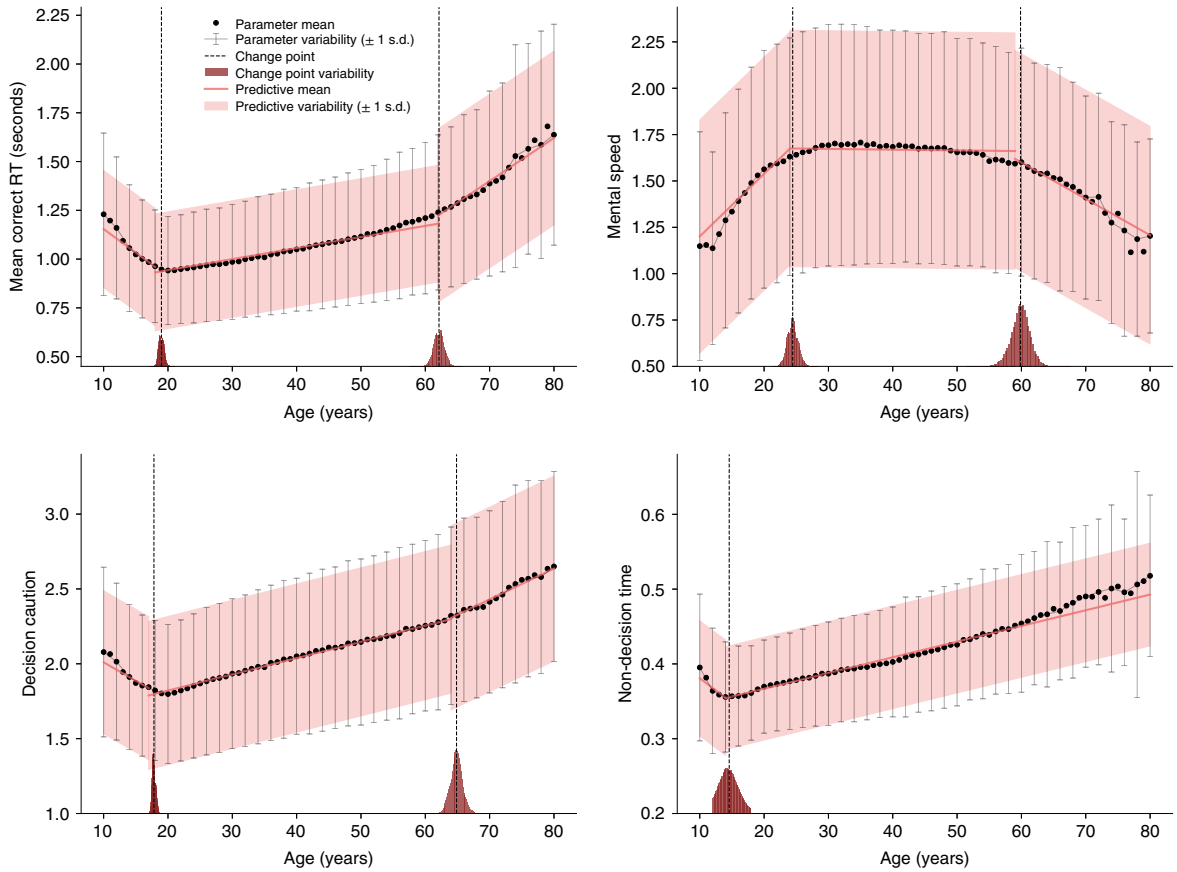

Fig. 2 | Mean correct RTs and DM parameters as functions of age. The black points indicate the means computed separately for each year of age. The bars indicate the standard deviations (shown only for every second year for better clarity). The red lines denote the Bayesian piecewise ridge regression model's mean predictions, which describe the observed means fairly well. The shaded red regions denote the uncertainty (standard deviation) of the piecewise model's predictions. The dashed lines indicate the mean change points estimated from the per-age-group averaged data, with the full posterior distributions (scaled for readability) of the change points shown at the bottom of each plot. Both the data- and model-implied standard deviations highlight the great variability within each year of age. Nevertheless, the year-specific means suggest a clear and consistent pattern for mean correct RT and for each parameter. The figure depicts drift rates and boundary separations for the incongruent condition and non-decision times obtained from correct responses. Very similar trends for the congruent condition and non-decision times from incorrect responses can be found in the Supplementary Information (Supplementary Fig. 12). For all panels, $N = 1{,}185{,}882$.

sample—they were the most willing to trade off accuracy for speed. After age 18, decision caution increases linearly until about age 65 in the incongruent condition, with a greater increase per year thereafter until age 80. In the congruent condition, the increasing trend in decision caution during old age was less pronounced, and the change in age-related trends was estimated to occur at age 40. Both findings might be attributable to the lower task difficulty in the congruent condition. Moreover, the trend towards higher decision caution becomes noticeable very early in adult life. The increase in the amount of information sampled before making a decision thus provides a first explanation for the age-related increase in RTs starting in young adulthood.

Non-decision time is the DM parameter that represents all processes beyond information sampling and evidence accumulation in a decision task. These processes are typically thought to encompass the time taken for the encoding of stimuli and motor response execution[22,45]. Interestingly, in our sample, non-decision times were, on average, fastest around age 14 to 16, with people outside this range needing more time for non-decision processes. It seems that these processes, should they represent a trait-like ability, reach their peak earliest among all cognitive abilities typically studied in the literature[7]. After age 16, non-decision times exhibit a linear increase that continues until age 80. The increase in the time needed for non-decision processes thus provides a second explanation for the slower RTs found with increasing age, as early as young and middle adulthood.

Our most remarkable finding concerns the drift rate—that is, the parameter representing mental speed in the DM framework. The drift rate denotes the average rate of information sampling per time unit and theoretically represents a more precise measure of mental speed than mean RTs, because speed–accuracy trade-offs and non-decision aspects are controlled for by other model parameters. During early adulthood, drift rates showed, on average, a continuous positive age trend—that is, mental speed became faster from age 10 to age 30. Mental speed thus peaks notably later than the lowest points of decision caution (around age 20) and non-decision times (around age 15). This result partly mirrors previous findings reporting that mental speed is still high around age 30 (ref. [6]). Yet, in our sample, mental speed showed a slight increase from age 20 to 30, which is in contrast to previous findings based on the analysis of mean RTs. It should be noted that our change-point analysis indicated that the positive age trend in drift rates is weaker from age 25 to age 30 than in the years before that, with the corresponding change point estimated at age 25.

Most importantly, our analyses suggest that the average levels of mental speed remain roughly stable across all of middle adulthood

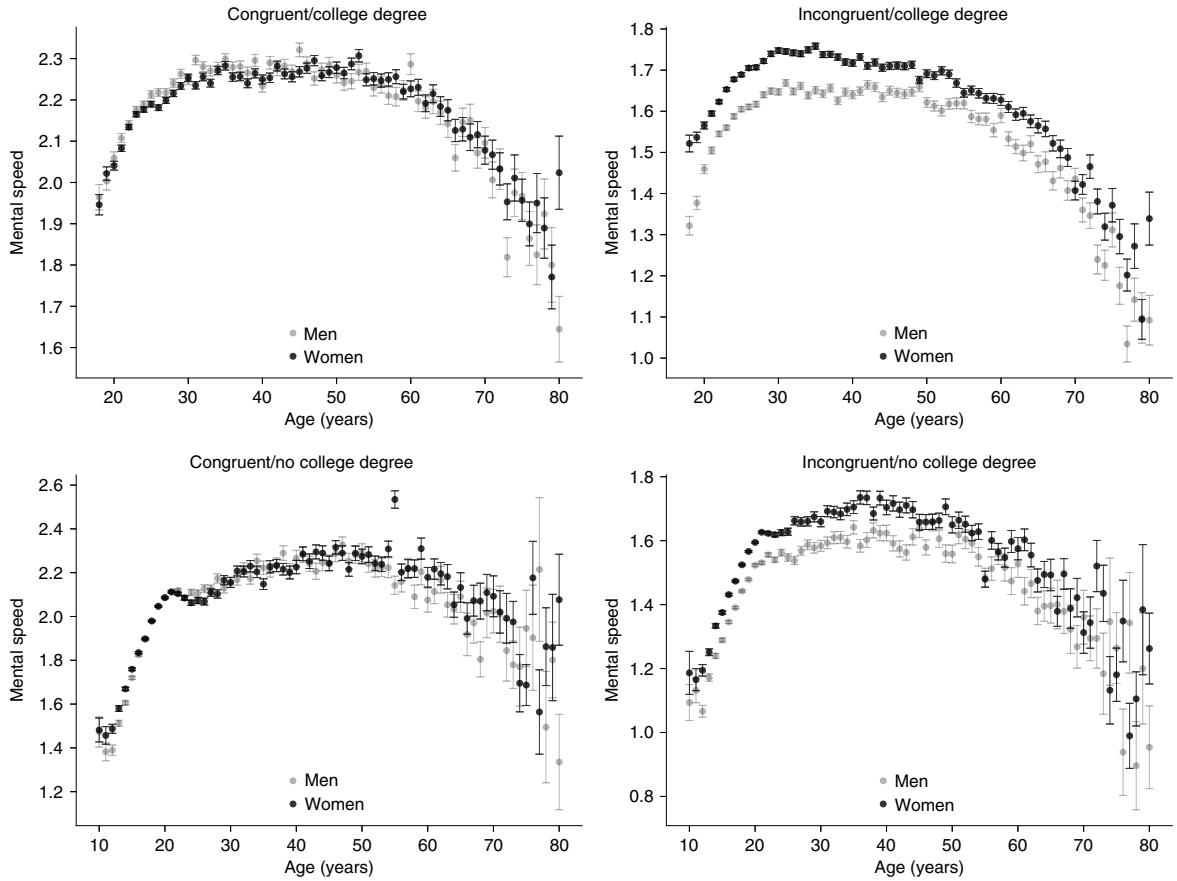

Fig. 3 | Mental speed as a function of age, experimental condition and demographic variables. Separate analyses of age differences in drift rates are shown for subgroups split by gender, level of education and experimental condition (congruent or incongruent IAT trials). Across subgroups, we observe the same inverted U-shape as in our main analysis. The vertical bars denote standard errors of the mean. Total $N = 1{,}185{,}882$, 38.69% female, 46.89% with at least college-level education.

(age 30 to 60), with only slight decreases from age 50 on. This surprising finding remains hidden if only mean RTs are analysed, as these do not reflect a pure measure of mental speed but are heavily influenced by decision caution and time required for motor processes. The pattern was robust across different stimuli, experimental conditions and several demographic factors. Accordingly, we conclude that the age-related increase of RTs in early and middle adulthood can be attributed exclusively to differences in decision caution and non-decision time, not to differences in mental speed. Only after about age 60 do drift rates start to show an accelerating negative age-related decline, with the lowest mean values found for the oldest participants. These age-related declines in mental speed in old age are in line with what has been reported in previous studies on cognitive ageing. However, our analysis suggests that the decline starts much later in life than has typically been assumed.

The higher boundary separations, higher non-decision times and lower drift rates found for people aged 60 and older jointly explain the accelerated age-related increase in mean RTs among the oldest participants. From about age 60 on, these three components contributing jointly to mean RTs all show age trends that lead to slower RTs. In other words, older people display higher decision caution, slower non-decision time and slower mental speed.

Our key findings also explain the age-related findings reported in previous diffusion modelling studies. Typically, these studies have compared two groups of participants: college-aged students and people aged 60 and older[14,23,27–35,38]. A consistent result of these studies is that older participants show higher boundary separations and non-decision times, but comparable drift rates.

When looking at our data, it is plausible that the linear age trends from age 20 onwards we found for boundary separation and non-decision times are consistent with the effects found in previous two-group studies. However, previous studies reporting no differences in drift rates between young and late adulthood might have overlooked the nonlinear age trend and the peak in drift rates from age 30 to age 50, because this group was not represented in the samples.

Our results are also in line with recently reported results on age differences in DM parameters using a continuous assessment of age[36]. In that study, a peak in mental speed around age 30 was also observed in a wide variety of different tasks. However, the sample size across later young adulthood and middle adulthood was too small to reveal clear age trends.

Another finding emerging from our study is the fact that DM parameters showed different cross-sectional patterns of across-person variability over the lifespan. While the variances of drift rates remained roughly the same or even decreased into old age, non-decision times and decision caution showed an increase in variability. The latter pattern is also present for mean RTs. It thus seems plausible that the greater spread in mean RTs observed for older people is attributable to greater interindividual differences in encoding and motor processes or decision caution, not in mental speed.

Finally, we report age differences in DM parameters in late childhood and adolescence (see also ref. [46]), thus allowing the study of differentiable temporal patterns in these age periods. Most notably, the fastest non-decision times were observed at ages 14 to 16, with mean RTs, mental speed and decision caution all showing much later turning points.

The differing age-related patterns of the DM parameters become more plausible when viewed in the context of the literature linking changes in cognitive abilities with changes in their neurophysiological basis[10]. According to the scaffolding theory of aging and cognition[47], people differ in their use of different compensatory techniques (for example, the activation of additional neural networks), all of which aim to counter the detrimental effects of age-related changes in brain structure. While such compensatory strategies might be well suited to keeping the level of mental speed in simple decision-making tasks high across large parts of the lifespan, more basal processes (such as the ones captured by non-decision time) might be less adaptable[36].

It should be noted that our study not only contributes to the literature on cognitive ageing and diffusion modelling but also represents an example of using mathematical modelling to analyse IAT data. Most previous model-based analyses of IAT data have relied on multinomial processing trees fitted on empirical accuracy rates[48–52]. For example, the so-called quad model[49] provides estimates of the likelihood that implicit biases are activated or overcome, as well as of the likelihood that a correct answer can even be determined or that guessing occurs. Recently, a new family of multinomial processing tree models incorporating RTs has been introduced[53–55]. Such models offer an alternative to analyse IAT data and present interesting avenues for future research in the context of cognitive ageing. 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 conditions provide additional evidence for the interpretation of drift rates as further reflecting association strength in the case of the IAT.

Our study has a number of advantages over previous studies of cognitive ageing, the most prominent being (1) the massive sample size, allowing for detailed age-related analyses, and (2) the use of Bayesian diffusion modelling to disentangle different components of the decision process in a robust and theoretically grounded way. However, we must also note some limitations of this study.

First, our data stem from only one particular type of decision-making task—namely, the race IAT. One might thus question whether our results generalize to other experimental paradigms or real-life scenarios. Regarding this limitation, it should be noted that our results (1) replicated across different experimental conditions and types of stimuli and (2) were in line with the findings reported in a number of studies on age differences in DM parameters. These previous studies spanned a vast variety of experimental tasks and paradigms, although with much smaller sample sizes. It thus seems plausible that our results, albeit based on a single type of task, should generalize to many other typical decision-making contexts.

A second limitation concerns the cross-sectional nature of our findings. It remains an open question whether the age differences and trends found in our data represent within-person developmental processes. We did not study longitudinal change, and neither did we account for cohort effects. However, given the clear age trends (with the majority of means almost perfectly aligned across age groups) found for the cognitive parameters of interest, we argue that our data provide as clear a picture of developmental patterns as is reasonably achievable using cross-sectional data. We also note that the IAT data made publicly available by Project Implicit[44] include datasets from the years 2002 to 2020, making it possible to study cohort effects, and also participant IDs, making it possible to study longitudinal change in participants taking the task several times. Such analyses were beyond the scope of this paper but might be well worthwhile in future endeavours.

Finally, the trial numbers for each person were quite low for diffusion modelling standards. While 60 trials (per condition) are sufficient for obtaining adequate estimates of the main model parameters according to previous studies and simulations[41], we could not compute internal measures of (split-half) reliability due to the low number of trials.

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 the model parameters, it might help explain some of our more specific findings—for instance, the gender differences in drift rates found specifically for the incongruent condition.

To conclude, according to our model-based analysis of a very large dataset of human RTs, mental speed increases until age 30, remains stable until around age 60 and declines thereafter. Furthermore, the slowdown in mean RTs in young and middle adulthood seems attributable to age-related changes in decision caution and non-decision times. Only in old age do the cumulative effects of all three cognitive parameters—mental speed, decision caution and non-decision time—contribute to an accelerated slowdown that is also evident from the raw RT data. Thus, for large parts of the human lifespan and typical work careers, our results challenge the widespread notion of an age-related slowdown in mental speed.

Methods

Our analyses are based on publicly available race IAT data provided by Project Implicit[44]. All participants provided informed consent; for details on the ethics procedure, see ref. [44]. The procedure was approved by the Institutional Review Board at the University of Virginia (institutional board review protocol number 2186). We extracted raw RT and accuracy data, as well as demographics, all of which were collected from September 2016 to December 2018. All data are openly available on the Project Implicit Open Science Framework (OSF) page: https://osf.io/y9hiq/. In addition, all analysis scripts for reproducing the results are available at https://github.com/stefanradev93/DataSizeMatters.

Participants. A total of 1,804,325 people started the study and provided information on their age; 1,519,257 people provided RT data, and 1,313,275 people provided both their age and RT 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 or 97.57% remaining). We then excluded participants with below-chance (<50%) accuracy in the RT 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. Furthermore, we excluded cases with more than 12 RTs (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 further analyses, we excluded cases with any parameter estimates beyond the bounds of our respective (very broad) uniform priors ('The DM'; 1,186,460 people or 90.34% remaining) or with 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. Of these, 38.69% were female, and 61.30% were male (the question asked concerned the sex assigned at birth). The mean age was 27.42 years (s.d. = 12.33), with a robust sample size across the entire age span of 10 to 80 years. About half of the participants (46.89%) had completed at least college-level education. The majority (84.06%) of the participants indicated that they were born in the United States, with the rest reporting different countries of origin.

Task. The race IAT is a quasi-standard cognitive task originally designed to measure implicit racial bias[56]. In a series of binary decisions, people have to classify words and images as belonging to one of two categories—for example, 'good/bad' or 'Black person/White person'. 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 (for example, 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'). Sixty trials are completed in each of the two conditions. The difference in mean RTs is then used to obtain a measure of implicit bias[57]. The exact procedure and materials can be found on the Project Implicit OSF page (https://osf.io/y9hiq/ and ref. [44]). We did not use the IAT as an instrument to study implicit cognition; instead, we used it as an example of a simple binary decision task.

The DM. In the present work, we employ the DM, a prominent mechanistic model of neurocognitive dynamics designed to explain human performance in simple decision-making tasks[18]. The DM is embedded in the larger model class of evidence accumulator models, which conceptualize information processing as a gradual, temporally ordered, and noisy process[58]. The mechanistic model we use in this work has the basic form

$$\mathrm{d}x = v\mathrm{d}t + \xi \mathrm{d}t^{1/2} \text{ with } \xi \sim \mathcal{N}(0, 1), \quad (1)$$

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 predefined 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. Importantly, the key parameters of the DM and its mechanistic formulation are well validated in experimental settings[21–23] and well grounded in biological neural-network theory[58].

To decompose performance in the race IAT into meaningful cognitive parameters, we formulate and fit a DM with six free parameters: $\boldsymbol{\theta} = (v_1, v_2, a_1, a_2, \tau_c, \tau_n)$. Here, v_1 and v_2 denote the speed of information processing (drift rates) in the two experimental conditions, a_1 and a_2 denote the decision thresholds (boundary separation), and τ_c and τ_n denote the additive non-decision time constants for correct and incorrect responses. 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[59]. 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).

We estimate separate non-decision time parameters for correct and incorrect trials due to the way error RTs were recorded in the race IAT. Trials do not terminate immediately following a wrong response but require the 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.

Note that some versions of the basic DM allow for parameters to vary randomly across trials, which introduces additional so-called intertrial 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 intertrial 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.

Our choice of Bayesian priors for the DM parameters reflects the goal to cover meaningful parameter ranges, as known from previous studies[60]. However, we also place uniform priors over the plausible numerical ranges to render the data maximally informative for posterior inference. We place broad uniform priors over both drift rates—that is, $v \sim \mathcal{U}(0.1, 7)$—which we deem sufficient to cover the entire range of realistically observable drift rates (that is, mental speed indicators). On the basis of similar considerations, we place a broad uniform prior over the boundary separation parameters, $a \sim \mathcal{U}(0.1, 4)$. For the non-decision constants, we use $\tau_c \sim \mathcal{U}(0.1, 3)$ and $\tau_n \sim \mathcal{U}(0.1, 7)$, incorporating our expectation of longer non-decision times for incorrect responses in the particular task.

Parameter estimation. Performing Bayesian estimation on hundreds of thousands of participants is not feasible with current gold-standard Markov chain Monte Carlo methods. We therefore resort to amortized Bayesian inference via specialized neural networks, which nevertheless guarantee correct posterior inference under perfect convergence[43]. The term 'amortized inference' refers to an approach that reduces the computational cost of Bayesian estimation by splitting the analysis into a costly upfront training phase, followed by an extremely efficient inference phase[43].

Basically, the BayesFlow method comprises a summary network h and an inference network f, which are trained jointly via simulations from the full Bayesian model:

$$p(\boldsymbol{\theta}, \mathbf{x}_1, \ldots, \mathbf{x}_N) = p(\boldsymbol{\theta}) \prod_{n=1}^{N} p(\mathbf{x}_n \mid \boldsymbol{\theta}) \quad (2)$$

Simulations are realized via a Monte Carlo simulation program, which efficiently samples from the prior and runs the DM with the sampled parameter configurations to generate synthetic datasets. The outputs of the simulation program are then fed to the neural networks, and the networks' parameters are optimized via standard backpropagation. The role of the summary network is to reduce datasets of arbitrary size to fixed-size vector representations in a completely end-to-end manner. The role of the inference network is to generate samples from an approximate posterior, $p_{\boldsymbol{\phi}}$, via a conditional invertible neural network, $f_{\boldsymbol{\phi}}$. Thus, once trained, the two networks are able to efficiently approximate the true posterior, $p(\boldsymbol{\theta} \mid \mathbf{x}_{1:N})$, given any possible dataset arising from the model.

Denoting the inference network parameters as $\boldsymbol{\phi}$ and those of the summary networks as $\boldsymbol{\psi}$, the two networks are trained to minimize the following Kullback–Leibler divergence criterion:

$$\min_{\boldsymbol{\phi},\boldsymbol{\psi}} \mathbb{E}_{p(\boldsymbol{\theta},\mathbf{x})} \left\{ -\log p_{\boldsymbol{\phi}}[\boldsymbol{\theta} \mid h_{\boldsymbol{\psi}}(\mathbf{x}_{1:N})] \right\} \quad (3)$$

which corresponds to minimizing the discrepancy between the true and the approximate amortized posterior induced by the networks. To train the networks, we performed approximately 50,000 simulations from the DM model with the priors for the parameters as described in the previous paragraph. Training the networks took approximately eight hours on a GPU-accelerated laptop. Inference on the entire dataset took approximately 24 hours on a machine without GPU acceleration.

Bayesian workflow. To further enhance the transparency and trustworthiness of our Bayesian pipeline, we follow the steps pertaining to a principled Bayesian workflow, as advocated by Schad et al.[61]. Accordingly, we partition our pipeline into the following steps: (1) prior predictive checks, (2) checks of computational faithfulness, (3) checks of model adequacy/sensitivity and (4) posterior predictive checks. These validation results, along with other robustness analyses, are described and visualized in the Supplementary Information (Supplementary Figs. 1–11).

Curve fitting. Once we had obtained parameter estimates for each participant, we aimed to represent the nonlinear relationships between age and the cognitive parameters statistically. Due to the presence of nonlinear relationships, we computed separate piecewise 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: (1) a more principled account of the observed age-related trend changes (for example, stability versus decline) and (2) accurate uncertainty quantification to account for the high variability of cognitive parameters within each age group.

Accordingly, our statistical analyses followed a two-step approach. First, we performed a linear Bayesian change-point regression on the age-group-averaged data using the R package for multiple change points mcp[62]. Note that this step ignores all variability within an age group and thus focuses on fast change-point detection, which otherwise turned out to be infeasible when executed on the full dataset. In the second step, we extracted the posterior distribution of each change point and used the corresponding posterior means for a piecewise Bayesian ridge regression on the full dataset. In this way, the piecewise model's predictive means and uncertainty account for the full variability in the estimated parameters.

We placed the following priors over change points to broadly reflect the trends visible in the data: for mean correct RTs, $t_1 \sim \mathcal{U}(15, 25)$ and $t_2 \sim \mathcal{U}(35, 70)$; for drift rates, $t_1 \sim \mathcal{U}(20, 40)$ and $t_2 \sim \mathcal{U}(35, 70)$; for boundary separations $t_1 \sim \mathcal{U}(15, 25)$ and $t_2 \sim \mathcal{U}(35, 70)$; and for non-decision times, $t_1 \sim \mathcal{U}(12, 18)$, where the scales of measurement correspond to chronological age. For the Bayesian ridge regression, we used the default priors available through the scikit-learn implementation in the Python programming language[63,64].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw data are available on the Project Implicit OSF page (https://osf.io/y9hiq/). The processed data, including the DM parameter estimates, can be found on our GitHub page (https://github.com/stefanradev93/DataSizeMatters).

Code availability

We provide open-source code for replicating all analyses and pretrained neural networks for preprocessing and obtaining the Bayesian diffusion model parameter estimates on our GitHub page (https://github.com/stefanradev93/DataSizeMatters).

Received: 18 March 2021; Accepted: 15 December 2021;
Published online: 17 February 2022

References

1. *National Prevalence Survey of Age Discrimination in the Workplace* (Australian Human Rights Commission, 2015).
2. Erber, J. T. & Long, B. A. Perceptions of forgetful and slow employees: does age matter? *J. Gerontol. B* **61**, 333–339 (2006).
3. Salthouse, T. A. Selective review of cognitive aging. *J. Int. Neuropsychol. Soc.* **16**, 754–760 (2010).
4. Jensen, A. R. *Clocking the Mind: Mental Chronometry and Individual Differences* (Elsevier, 2006).
5. Salthouse, T. A. The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* **103**, 403–428 (1996).

6. Salthouse, T. A. What and when of cognitive aging. *Curr. Dir. Psychol. Sci.* **13**, 140–144 (2004).
7. Hartshorne, J. K. & Germine, L. T. When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychol. Sci.* **26**, 433–443 (2015).
8. Schaie, K. W. What can we learn from longitudinal studies of adult development? *Res. Hum. Dev.* **2**, 133–158 (2005).
9. Zimprich, D. & Martin, M. Can longitudinal changes in processing speed explain longitudinal age changes in fluid intelligence? *Psychol. Aging* **17**, 690–695 (2002).
10. Oschwald, J. et al. Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. *Rev. Neurosci.* **31**, 1–57 (2019).
11. Frischkorn, G. T. & Schubert, A.-L. Cognitive models in intelligence research: advantages and recommendations for their application. *J. Intell.* **6**, 34 (2018).
12. Pachella, R. G. *The Interpretation of Reaction Time in Information Processing Research* Technical Report (Michigan Univ. Ann Arbor Human Performance Center, 1973).
13. Schubert, A.-L. & Frischkorn, G. T. Neurocognitive psychometrics of intelligence: how measurement advancements unveiled the role of mental speed in intelligence differences. *Curr. Dir. Psychol. Sci.* **29**, 140–146 (2020).
14. Ratcliff, R., Thapar, A. & McKoon, G. Individual differences, aging, and IQ in two-choice tasks. *Cogn. Psychol.* **60**, 127–157 (2010).
15. Lerche, V. et al. Diffusion modeling and intelligence: drift rates show both domain-general and domain-specific relations with intelligence. *J. Exp. Psychol. Gen.* **149**, 2207–2249 (2020).
16. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **85**, 59–108 (1978).
17. Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).
18. Ratcliff, R. & Rouder, J. N. Modeling response times for two-choice decisions. *Psychol. Sci.* **9**, 347–356 (1998).
19. Voss, A., Nagler, M. & Lerche, V. Diffusion models in experimental psychology: a practical introduction. *Exp. Psychol.* **60**, 385–402 (2013).
20. Fudenberg, D., Newey, W., Strack, P. & Strzalecki, T. Testing the drift–diffusion model. *Proc. Natl Acad. Sci. USA* **117**, 33141–33148 (2020).
21. Lerche, V. & Voss, A. Experimental validation of the diffusion model based on a slow response time paradigm. *Psychol. Res.* **83**, 1194–1209 (2019).
22. Voss, A., Rothermund, K. & Voss, J. Interpreting the parameters of the diffusion model: an empirical validation. *Mem. Cogn.* **32**, 1206–1220 (2004).
23. Arnold, N. R., Bröder, A. & Bayen, U. J. Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychol. Res.* **79**, 882–898 (2015).
24. McGovern, D. P., Hayes, A., Kelly, S. P. & O'Connell, R. G. Reconciling age-related changes in behavioural and neural indices of human perceptual decision-making. *Nat. Hum. Behav.* **2**, 955–966 (2018).
25. Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L. & Segraves, M. A. Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J. Neurophysiol.* **97**, 1756–1774 (2007).
26. Kühn, S. et al. Brain areas consistently linked to individual differences in perceptual decision-making in younger as well as older adults before and after training. *J. Cogn. Neurosci.* **23**, 2147–2158 (2011).
27. Ball, B. H. & Aschenbrenner, A. J. The importance of age-related differences in prospective memory: evidence from diffusion model analyses. *Psychon. Bull. Rev.* **25**, 1114–1122 (2018).
28. Dully, J., McGovern, D. P. & O'Connell, R. G. The impact of natural aging on computational and neural indices of perceptual decision making: a review. *Behav. Brain Res.* **355**, 48–55 (2018).
29. Janczyk, M., Mittelstädt, P. & Wienrich's, C. Parallel dual-task processing and task-shielding in older and younger adults: behavioral and diffusion model results. *Exp. Aging Res.* **44**, 95–116 (2018).
30. McKoon, G. & Ratcliff, R. Aging and IQ effects on associative recognition and priming in item recognition. *J. Mem. Lang.* **66**, 416–437 (2012).
31. Ratcliff, R., Thapar, A. & McKoon, G. The effects of aging on reaction time in a signal detection task. *Psychol. Aging* **16**, 323–341 (2001).
32. Ratcliff, R., Gomez, P. & McKoon, G. A diffusion model account of the lexical decision task. *Psychol. Rev.* **111**, 159–182 (2004).
33. Thapar, A., Ratcliff, R. & McKoon, G. A diffusion model analysis of the effects of aging on letter discrimination. *Psychol. Aging* **18**, 415–429 (2003).
34. Spaniol, J., Madden, D. J. & Voss, A. A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* **32**, 101–117 (2006).
35. Spaniol, J., Voss, A., Bowen, H. J. & Grady, C. L. Motivational incentives modulate age differences in visual perception. *Psychol. Aging* **26**, 932–939 (2011).
36. von Krause, M., Lerche, V., Schubert, A.-L. & Voss, A. Do non-decision times mediate the association between age and intelligence across different content and process domains? *J. Intell.* **8**, 33 (2020).
37. Schubert, A.-L., Hagemann, D., Löffler, C. & Frischkorn, G. T. Disentangling the effects of processing speed on the association between age differences and fluid intelligence. *J. Intell.* **8**, 1 (2020).
38. McKoon, G. & Ratcliff, R. Aging and predicting inferences: a diffusion model analysis. *J. Mem. Lang.* **68**, 240–254 (2013).
39. Theisen, M., Lerche, V., von Krause, M. & Voss, A. Age differences in diffusion model parameters: a meta-analysis. *Psychol. Res.* **85**, 2012–2021 (2020).
40. Ratcliff, R. & Childers, R. Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision* **2**, 237–279 (2015).
41. Lerche, V., Voss, A. & Nagler, M. How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behav. Res. Methods* **49**, 513–537 (2017).
42. Lee, M. D. & Wagenmakers, E.-J. *Bayesian Cognitive Modeling: A Practical Course* (Cambridge Univ. Press, 2014).
43. Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L. & Köthe, U. BayesFlow: learning complex stochastic models with invertible neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 1–15 (2020).
44. Xu, K., Nosek, B. & Greenwald, A. Psychology data from the race implicit association test on the Project Implicit demo website. *J. Open Psychol. Data* **2**, e3 (2014).
45. Ratcliff, R. Modeling aging effects on two-choice tasks: response signal and response time data. *Psychol. Aging* **23**, 900–916 (2008).
46. Ratcliff, R., Love, J., Thompson, C. A. & Opfer, J. E. Children are not like older adults: a diffusion model analysis of developmental changes in speeded responses. *Child Dev.* **83**, 367–381 (2012).
47. Reuter-Lorenz, P. A. & Park, D. C. How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychol. Rev.* **24**, 355–370 (2014).
48. Payne, B. K. Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. *J. Pers. Soc. Psychol.* **81**, 181–192 (2001).
49. Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K. & Groom, C. J. Separating multiple processes in implicit social cognition: the quad model of implicit task performance. *J. Pers. Soc. Psychol.* **89**, 469–487 (2005).
50. Meissner, F. & Rothermund, K. Estimating the contributions of associations and recoding in the implicit association test: the real model for the IAT. *J. Pers. Soc. Psychol.* **104**, 45–69 (2013).
51. Stahl, C. & Degner, J. Assessing automatic activation of valence: a multinomial model of EAST performance. *Exp. Psychol.* **54**, 99–112 (2007).
52. Nadarevic, L. & Erdfelder, E. Cognitive processes in implicit attitude tasks: an experimental validation of the trip model. *Eur. J. Soc. Psychol.* **41**, 254–268 (2011).
53. Heck, D. W. & Erdfelder, E. Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychon. Bull. Rev.* **23**, 1440–1465 (2016).
54. Klauer, K. C. & Kellen, D. RT-MPTs: process models for response-time distributions based on multinomial processing trees with applications to recognition memory. *J. Math. Psychol.* **82**, 111–130 (2018).
55. Hartmann, R. & Klauer, K. C. Extending RT-MPTs to enable equal process times. *J. Math. Psychol.* **96**, 102340 (2020).
56. Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
57. Greenwald, A. G., Nosek, B. A. & Banaji, M. R. Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* **85**, 197–216 (2003).
58. Usher, M. & McClelland, J. L. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* **108**, 550–592 (2001).
59. Klauer, K. C., Voss, A., Schmitz, F. & Teige-Mocigemba, S. Process components of the implicit association test: a diffusion-model analysis. *J. Pers. Soc. Psychol.* **93**, 353–368 (2007).
60. Matzke, D. & Wagenmakers, E.-J. Psychological interpretation of the ex-Gaussian and shifted Wald parameters: a diffusion model analysis. *Psychon. Bull. Rev.* **16**, 798–817 (2009).
61. Schad, D. J., Betancourt, M. & Vasishth, S. Toward a principled Bayesian workflow in cognitive science. *Psychol. Methods* **26**, 103–126 (2020).
62. Lindeløv, J. K. mcp: an R package for regression with multiple change points. Preprint at *OSF Preprints* https://doi.org/10.31219/osf.io/fzqxv (2020).
63. Van Rossum, G. & Drake Jr, F. L. Python Tutorial (Centrum voor Wiskunde en Info rmatica, 2006).
64. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
65. Bloem-Reddy, B. & Teh, Y. W. Probabilistic symmetries and invariant neural networks. *J. Mach. Learn. Res.* **21**(90), 1–61 (2020).

Acknowledgements

This research was supported by a grant from the German Research Foundation to the Graduate School 530 SMiP (GRK 2277; Statistical Modeling in Psychology; to all authors). The funders had no role in study design, data collection and analysis, decision

to publish or preparation of the manuscript. We thank Project Implicit for openly sharing their data.

Author contributions

M.v.K. conceived the research idea and studied the literature. S.T.R. conceived the simulation-based inference method. M.v.K. and S.T.R. wrote the code and scripts for all methodological steps, performed the analyses, and visualized the results. M.v.K and S.T.R. wrote and prepared the original draft. M.v.K., S.T.R. and A.V. wrote, reviewed and edited the final manuscript. All authors have read and agreed to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-021-01282-7.

Correspondence and requests for materials should be addressed to Mischa von Krause or Stefan T. Radev.

Peer review information *Nature Human Behaviour* thanks Laura Germine and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Mischa von Krause

Last updated by author(s): 2021/11/16

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Our manuscript relies on data collected and provided by Project Implicit (K. Xu, B. Nosek, and A. Greenwald, "Psychology data from the race implicit association test on the project implicit demo website,"Journal of Open Psychology Data, vol. 2, no. 1, 2014).
Data analysis	All code is available at https://github.com/stefanradev93/DataSizeMatters. We used the R package mcp (Version 0.3.0) and the the Python implementation scikit-learn for our analyses. Parameter estimated was conducted using BayesFlow (S. T. Radev, U. K. Mertens, A. Voss, L. Ardizzone, and U. Köthe, "Bayesflow: Learning complexstochastic models with invertible neural networks,"IEEE Transactions on Neural Networks andLearning Systems, pp. 1–15, 2020..

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

We analyse data provided by Project Implicit and publicly available at OSF (https://osf.io/y9hiq/; K. Xu, B. Nosek, and A. Greenwald, "Psychology data from the race implicit association test on the project implicit demo website,"Journal of Open Psychology Data, vol. 2, no. 1, 2014)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative cross-sectional
Research sample	Existing Implicit Association Test dataset provided by Project Implicit (https://osf.io/y9hiq/; K. Xu, B. Nosek, and A. Greenwald, "Psychology data from the race implicit association test on the project implicit demo website,"Journal of Open Psychology Data, vol. 2, no. 1, 2014).
Sampling strategy	We used convenience data collected between September 2016 and December 2018, with a total initial N of over 1,800,000. This is a very large sample for diffusion model analyses, and allows us to provide detailed analyses of age trends. More data would not have provided a significant advantage, and including data collected at different time points (e.g., 2002) might have introduced confounding effects.
Data collection	See the procedure description on the OSF site and the corresponding paper. (https://osf.io/y9hiq/; K. Xu, B. Nosek, and A. Greenwald, "Psychology data from the race implicit association test on the project implicit demo website,"Journal of Open Psychology Data, vol. 2, no. 1, 2014).
Timing	September 2016 to December 2018
Data exclusions	Our original sample contained 1,804,325 people. We excluded cases that did not complete the task, did not provide their year of birth, were older than 80 or younger than 10 years at time of data collection, had more than 10% response times under 300 ms, or had below chance accuracy. In order to be able to obtain full data-informed parameter sets for each person, including the error non-decision time, we excluded all participants with 100% accuracy. Further, we excluded trials faster than 300 ms (psychologically implausible) or slower than 10 seconds (inattention). After fitting the diffusion models, we excluded cases with estimates for drift rates, boundary separations, or non-decision times beyond the bounds of our respective (very broad) prior ranges (see manuscript and appendix). This left us with a final sample of 1,185,882 people. See Methods section for further details.
Non-participation	See https://osf.io/y9hiq/ and K. Xu, B. Nosek, and A. Greenwald, "Psychology data from the race implicit association test on the project implicit demo website,"Journal of Open Psychology Data, vol. 2, no. 1, 2014).
Randomization	Participants were randomly assigned to different types of stimuli (Black/White, African American/European American).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☐ Antibodies
☐	☐ Eukaryotic cell lines
☐	☐ Palaeontology and archaeology
☐	☐ Animals and other organisms
☐	☒ Human research participants
☐	☐ Clinical data
☐	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☐ ChIP-seq
☐	☐ Flow cytometry
☐	☐ MRI-based neuroimaging

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above: we re-analysed existing data.
Recruitment	See above: we re-analysed existing data.
Ethics oversight	See above: we re-analysed existing data. The procedure was approved by the Institutional Review Board at the University of Virginia (institutional board review protocol number 2186).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*

Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*

Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g.*

Normalization template

original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.