

Parsing Memory and Non-Memory Contributions to Age-Related Declines in Mnemonic Discrimination Performance: A Hierarchical Bayesian Diffusion Decision Modeling Approach

Caroline Chwiesko^{*1,2}, John Janeczek^{1,2}, Stephanie Doering^{1,2}, Martina Hollearn^{1,2}, Liv McMillan^{1,2}, Joachim Vandekerckhove³, Michael D. Lee^{2,3}, Roger Ratcliff⁴, Michael A. Yassa^{1,2*}

¹ Department of Neurobiology and Behavior, University of California, Irvine, CA, USA

² Center for the Neurobiology of Learning and Memory, University of California, Irvine, CA, USA

³ Department of Cognitive Science, University of California, Irvine, CA, USA

⁴ Department of Psychology, The Ohio State University, Columbus, OH, USA

Running Head:

DIFFUSION DECISION MODEL AND MEMORY IN AGING

Article Type:

Research Paper

Keywords:

Memory, aging, retrieval, pattern separation, hippocampus, drift diffusion, discrimination, generalization, decision-making

*** Correspondence to:**

Caroline Chwiesko, Ph.D. | 1410 Biological Sciences 3, Irvine CA 92697 | cchwiesk@uci.edu | 949-824-3919

Michael A. Yassa, Ph.D. | 320 Qureshey Research Lab, Irvine, CA 92697 | myassa@uci.edu | 949-824-1687

ABSTRACT

The mnemonic discrimination task (MDT) is a widely used cognitive assessment tool. Performance in this task is believed to indicate an age-related deficit in episodic memory stemming from a decreased ability to pattern separate among similar experiences. However, cognitive processes other than memory ability might impact task performance. In this study, we investigated whether non-mnemonic decision-making processes contribute to the age-related deficit in the MDT. We applied a hierarchical Bayesian version of the Ratcliff Diffusion Model to the MDT performance of 26 younger and 31 cognitively normal older adults. It allowed us to decompose decision behavior in the MDT into different underlying cognitive processes, represented by specific model parameters. Model parameters were compared between groups and differences were evaluated using the Bayes factor. Our results suggest that the age-related decline in MDT performance indicates a predominantly mnemonic deficit rather than differences in non-mnemonic decision-making processes. In addition, this mnemonic deficit might also involve a slowing in processes related to encoding and retrieval strategies, which are relevant for successful memory as well. These findings help to better understand what cognitive processes contribute to the age-related decline in MDT performance and may help to improve the diagnostic value of this popular task.

INTRODUCTION

By 2050, the global population over the age of 65 is projected to nearly double from 12% to 22% (World Health Organization 2021). While memory decline is one of the hallmark cognitive changes observed with aging, episodic memory, in particular, shows a clear and steady decline with increasing age (Nilsson 2003). An even more severe decline in episodic memory is a feature of the early stages of Alzheimer's disease (AD) (Weintraub et al. 2012) which affects more than 6.5 million Americans today (Alzheimer's Association 2022). To ensure successful early detection and diagnosis, it is crucial to understand and distinguish episodic memory changes that are part of the normal aging process from those that may be an indication of pathological aging.

The Mnemonic Discrimination Task (MDT), a modified object recognition memory task, has become a widely used digital cognitive assessment tool in the learning and memory as well as the aging and AD fields, due to its sensitivity to detect early and subtle changes in episodic memory (Leal and Yassa 2018). The utility of this task is further indicated by its recent inclusion as a key outcome measure in large-scale clinical trials in amnesic mild cognitive impairment, a prodrome for AD (Rosenzweig-Lipson et al. 2021) as well as asymptomatic AD (Sperling et al. 2014).

In the MDT, participants undergo incidental encoding of object stimuli and, after a brief delay, are administered an old/new recognition test with exact repetitions (targets), novel items (foils), and items perceptually similar to those viewed during the study phase (lures). Participants' capacity to correctly reject similar lure items as "New" is taken as a measure of their mnemonic discrimination performance. This performance is thought to reflect their capacity for *pattern separation*, a neural computation that reduces mnemonic interference among similar experiences by creating and storing non-overlapping neural representations (Yassa and Stark 2011). A large body of evidence reviewed elsewhere (Leal and Yassa 2018) has demonstrated empirically that this ability is highly dependent on the hippocampus and, in particular, the dentate gyrus (DG) and CA3 subfields.

A large body of literature has now demonstrated that performance on the MDT, specifically the *lure discrimination* performance, is reliably diminished in older adults, in comparison to unaffected performance on repeated targets and novel foils. In other words, older adults are generally more likely to falsely recognize similar lure items and identify them as "Old" rather than "New". This is in the absence of dementia or other comorbidities, which could additionally affect performance in these other conditions. The performance deficit in lure discrimination is believed to stem mainly from a decreased ability to pattern separate among similar experiences, owing to age-related circuit alterations in the DG and CA3 regions of the hippocampus that have been investigated across species (see reviews by Leal and Yassa 2015; Wilson et al. 2006). Neuroimaging studies in humans have demonstrated that lure discrimination deficits in older adults are linked to structural and functional alterations in the DG/CA3 region and its input pathways in the entorhinal cortex (Sinha et al. 2018; Reagh et al. 2018; Yassa et al. 2011b, 2011a).

Despite the abundance of work demonstrating performance deficits in the MDT in older adults and linking this to neurobiological underpinnings, an area that has received much less attention is dissecting the cognitive processes that underlie the reported age-related performance decline. It is increasingly appreciated that the MDT is likely not a process-

pure task and that cognitive processes other than memory ability might impact the outcome of task performance (i.e., the learning-performance distinction; Cahill et al. 2001). Non-mnemonic processes that contribute to MDT performance may also decline with age and thus may obscure what changes are truly driven by memory decline.

Earlier work attempted to systematically exclude other confounds that could contribute to the age-related decline in lure discrimination performance in the MDT, by assessing various alterations in task design, instructions, attentional and intentional strategies during encoding (Stark et al. 2015). Consistency of the age-related deficit across all manipulations led to the general conclusion that it is likely neurobiological changes in the hippocampus that result in age-related deficits in mnemonic lure discrimination. However, these manipulations in task design still did not allow for the direct measurement and evaluation of the mnemonic component. In addition, other processes that might not have been challenged by the implemented task manipulations could still contribute to performance differences. Therefore, it is still difficult to conclude from this study whether it is a mnemonic deficit in lure discrimination that underlies the age-related decline in lure discrimination performance. A few recent studies have further attempted to shed light on the possible contribution of other component processes to MDT performance. For example, one study showed that visual perceptual deficits in older adults are associated with mnemonic lure discrimination deficits (Davidson et al. 2019). Another study showed that depending on the test format, perceptual ambiguity, and executive function could play significant roles in mnemonic discrimination performance in older adults (Gellersen et al. 2021). A significant limitation of these studies is that they still leave the possibility of confounding factors as they all rely on linking performance either to standard neuropsychological assessments of memory (e.g., verbal list recall) or other aspects of cognitive function like attention, perception, or executive function.

In the current study, we take a different approach to investigate the potential contributions of non-mnemonic processes to the age-related decline in MDT performance. We recognize that a participant's timed "Old"/"New" endorsement of an image is the result of a decision-making process. This decision is based on several cognitive processes that altogether contribute to the final decision. A key contributor is the memory ability itself. However, the time that the participant takes to respond also contributes to a decision. This is termed **response caution** and is a key consideration in speeded decision making due to the *speed-accuracy trade-off* – the phenomenon in which the more time a decision-maker takes to make a decision, the more likely it is to be accurate, and vice versa (Wickelgren 1977; Bogacz et al. 2010). Another process that can contribute to decision making is the presence of a **response bias** – the general tendency to respond more often with one of the response alternatives (Stanislaw and Todorov 1999). Thus, individual differences in response caution or response bias, or both can confound interpretations of task performance.

To investigate the extent to which age-related decline in mnemonic lure discrimination performance is driven by differences in these non-mnemonic decision-making processes, we use the Ratcliff Diffusion Decision Model (Forstmann et al. 2016; Ratcliff 1978). The Diffusion Decision Model (DDM) is a popular mathematical process model that decomposes decision behavior in speeded two-choice response tasks into the different underlying cognitive processes. It, therefore, allows us to measure directly the

psychological processes that underlie MDT performance. In contrast to previous MDT analysis methods that reduce performance to a singular metric (e.g., “lure discrimination index (LDI)”, “pattern separation bias”, or “d-prime”) (Stark et al. 2019a; Yassa et al. 2011a; Stark et al. 2015), this model presents a more sophisticated and comprehensive analysis of decision behavior. Although the previous measures attempt to account for a response bias as a potential confound to overall performance, they cannot identify other underlying cognitive processes that contribute to task performance. This is chiefly because they are entirely based on decision accuracy but do not consider the time a participant takes to make a response (reaction time - RT). Decision accuracy and reaction time are related to one another but carry important and distinct information about the underlying cognitive processes. By disregarding the reaction time, summary measures provide an incomplete, possibly even inaccurate picture of behavior. For example, they do not consider the speed-accuracy tradeoff and are therefore confounded by response caution.

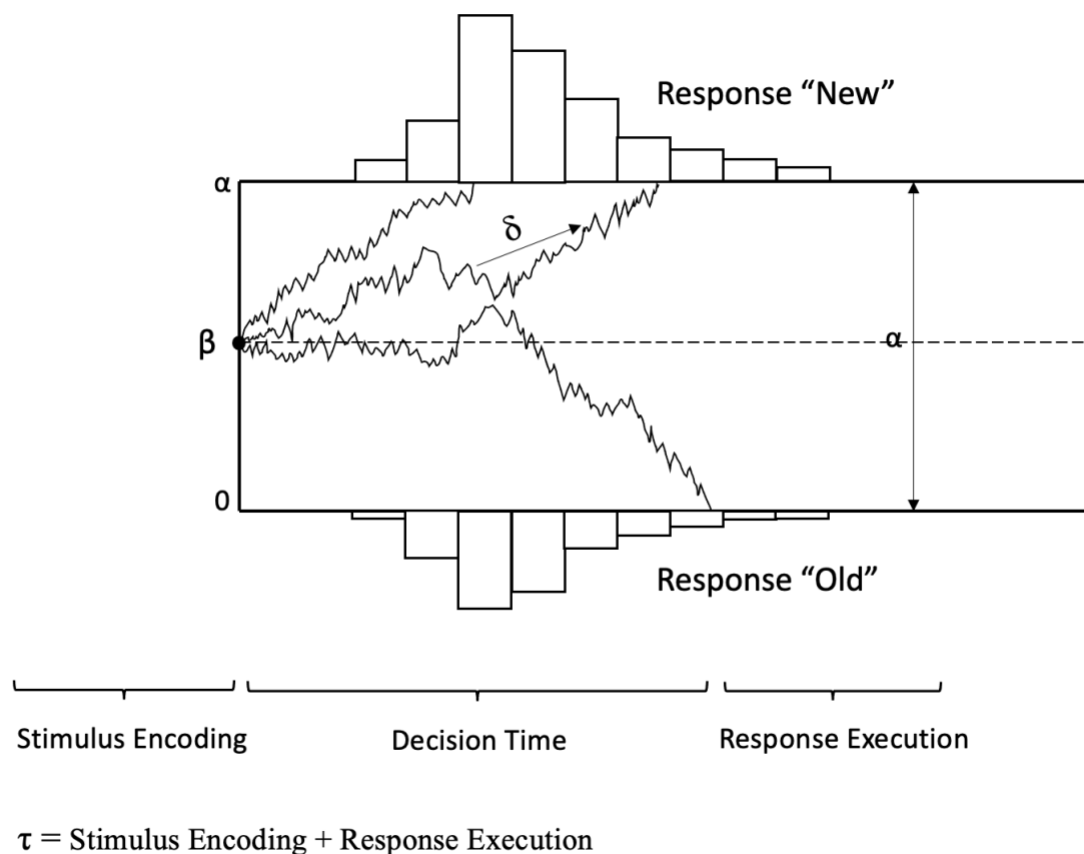


Figure 1. Graphical Illustration of the main parameters in the Ratcliff Diffusion Model. δ = drift rate indicating the average amount of information that can be extracted from a stimulus and memory and accumulated across time, β = starting point indicating an initial bias toward one of the two response alternatives. A starting point at 0.5 indicates no response bias, α = boundary separation indicating the amount of evidence that one needs to accumulate before a response can be made, τ = non-decision time. Encompasses all non-decision processes during encoding and response execution.

In contrast, the DDM considers all aspects of the behavioral data jointly -accuracy and the shapes of the RT distributions for correct and incorrect responses. Another key advantage of the model is that it inherently accounts for the speed-accuracy trade-off and

response caution (Forstmann et al. 2016). The model assumes that two-choice decisions are based on the noisy accumulation of information about a stimulus over time. Accumulation of information starts from a specific starting point and accumulates toward one of two decision boundaries. Each boundary represents one of two possible responses (e.g., “Old”/“New”). Once a boundary is reached a decision is made with the corresponding response (Ratcliff et al. 2016; Ratcliff and McKoon 2008) (**Figure 1**).

The DDM estimates four latent parameters that can be interpreted in terms of specific cognitive processes that underlie the decision behavior (Ratcliff 2002; Voss et al. 2004): drift rate (δ), boundary separation (α), starting point (β) and non-decision time (τ). The boundary separation parameter (α) represents response caution in the model. It defines the amount of evidence that one needs to accumulate before a response can be made. The starting point parameter (β) represents an a priori response bias toward one of the two response alternatives. We will consider both of these parameters as representing non-mnemonic contributions to the decision-making in the MDT. The drift rate parameter (δ), on the other hand, is the key parameter that is driving the mnemonic decision and is therefore representing the mnemonic contribution to performance. In the context of the MDT, the drift rate can be understood as an index of mnemonic discrimination ability. It reflects the quality of evidence that a participant can extract from a currently observed lure image and from a stored memory and accumulate over time to identify a lure as “New”. Lastly, the non-decision time parameter (τ) in the model represents the non-decision component. It is not part of the decision process and therefore does not make any contribution to the choice behavior in the model. It encompasses the mean duration of all non-decision processes during encoding and motor response execution. To improve the fit of the DDM to experimental data, Roger Ratcliff included an across-trial variability of the parameters δ , τ and β (Ratcliff and Tuerlinckx 2002; Ratcliff 1978; Ratcliff and Rouder 1998; Laming 1968). Therefore, the DDM model is sometimes also called the Ratcliff Diffusion model.

In the current study, we apply the Ratcliff Diffusion Model to the decision behavior of older and younger adults performing the MDT. We estimate for each of the model parameters (α , β , δ , τ) that represent different cognitive processes, the group difference between younger and older adults. We then evaluate the strength of the evidence for each parameter group difference based on the Bayes factor by comparing the ratio of the likelihood that the difference is 0. A series of models are compared, and the best-performing model is chosen based on the Deviance Information Criterion (DIC). Due to the many benefits of a hierarchical extension of the model, we implemented here a Hierarchical Bayesian version of the DDM (Vandekerckhove et al. 2011). The hierarchical approach has been shown to be superior to standard methods when the number of data points per participant is low, as typically found in memory studies (Ratcliff and Childers 2015). Instead of using any of the existing diffusion-model fitting packages we used a custom-designed model that gave us full flexibility regarding parameter specification and model assumptions.

MATERIALS AND METHODS

Participants

The study sample consisted of 26 young adults (ages 18-29 years) and 31 older adults (ages 60-91 years). These numbers account already for 5 participants that had to be excluded as described later. See **Table 1** for demographics and neuropsychological test scores. Individuals in the young group were recruited from the community and screened for major neurological or psychiatric conditions, as well as a history or current substance use disorder using an in-house short medical screening questionnaire. If participants responded with “yes” to any of the substance use questions, a follow-up phone interview was performed where the National Institute on Drug Abuse (NIDA) Quick Screen (<http://www.drugabuse.gov/nidamed/screening/>) was administered to assess the severity of substance involvement. Only subjects that were categorized as low or moderate risk (cutoff < 26) were included in the study. All older adults in this study were recruited from the community as well and enrolled in a large longitudinal study of aging and preclinical Alzheimer’s disease (BEACoN Study, NIA R01AG053555 PI: Yassa). They were thoroughly screened for history of major health conditions including but not limited to neurological disorders, psychiatric disorders, chronic illnesses, medications, history of recreational drug use, smoking and alcohol use, and questions about the intactness of visual and auditory abilities (e.g.: “Do you wear a hearing aid?”). Additionally, the determination of “clinically normal” status was based on several levels of evaluation that allowed us to exclude those with mild cognitive impairment (MCI) or dementia. All participants had a Clinical Dementia Rating (CDR) score of 0, and a Mini-Mental State Exam (MMSE) score of 27 or above. Performance on standardized neuropsychological assessments was all within 1 standard deviation of age-based norms. All participants (young and old) were screened for depression using the Beck Depression Inventory (BDI cut-off < 14, Beck et al. 1961) and insomnia using the Pittsburgh Sleep Quality Index (PSQI Global Score cut-off < 8, Buysse et al. 1989). They had all normal or corrected-to-normal vision and were not color blind. All participants gave written informed consent prior to participation in the study. After completing the study, participants were debriefed and compensated with electronic gift card payments. The study was approved by the Institutional Review Board at the University of California, Irvine.

GROUP	OLD	YOUNG
N	31	26
Age (years)	73.3 (7.3)	24.1 (3.2)
Sex	21F	17F
Education (years)	16.1 (2.4)	16.1 (2.6)
RAVLT-Delayed (A7)	10 (3.3)	
MMSE	28.1 (1.3)	

Table 1. Characteristics of all participants included in this study. All data are reported as mean (sd). RAVLT, Rey-Auditory Verbal Learning Test; MMSE, Mini-Mental State Examination.

Testing Procedures

The Mnemonic Discrimination Task was administered to all participants in this study in a testing session that lasted approximately 15 minutes. The task was programmed in PsychoPy 3 (Peirce et al. 2019) and presented on an Apple computer with a 24-inch monitor that was placed approximately 27 inches away from participants. A keyboard, placed at a comfortable distance, was used to enter responses. Paper index cards were placed between the keyboard and the monitor to orient the participants to the correct response button press.

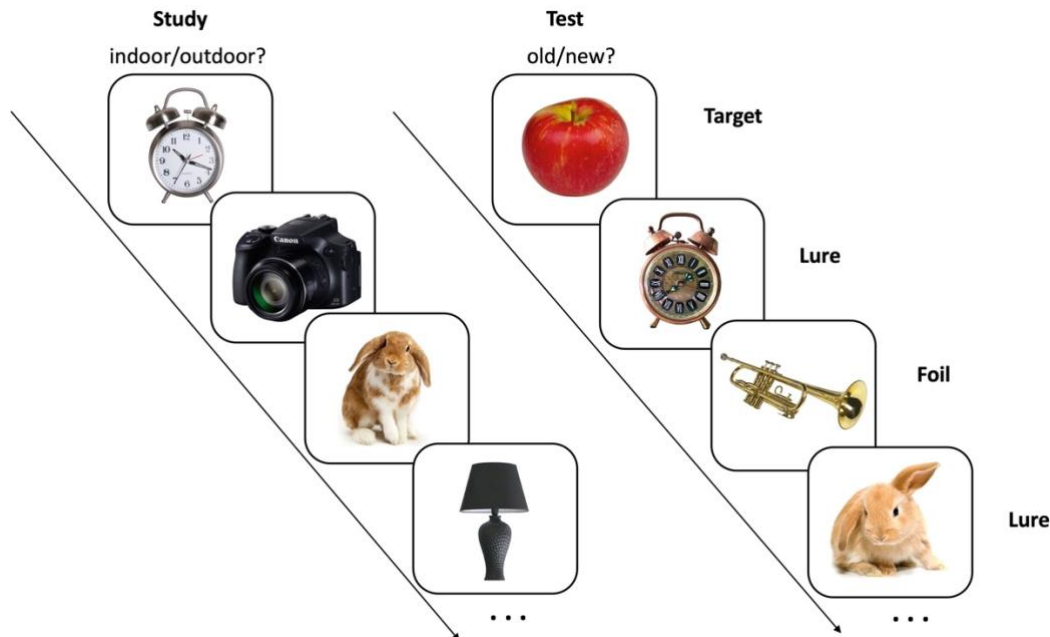


Figure 2. Illustrative diagram of the Mnemonic Discrimination Task (MDT). Pictures were shown one at a time. During the study phase, participants were asked to give indoor/outdoor judgments. The test phase followed right after the study phase. In the test phase participants were asked to give old/new judgments.

The MDT procedure consisted of two phases (**Figure 2**). During the incidental study phase, participants were presented with 120 colored images of everyday objects displayed at the center of the screen. Participants were instructed to quickly make “Indoor” or “Outdoor” judgments for each image using the ‘F’ and ‘J’ keys, respectively. Each image was presented for 2000 msec with a 500 msec inter-stimulus interval (fixation cross). The test phase followed immediately after and included 160 images: 40 trials were exact repetitions of previously seen images (targets), 40 trials were novel images (foils), and 80 trials were images that were perceptually similar but not identical to previously seen images (lures). Lure trials were distributed into 40 high similarity lures and 40 low similarity lures (with respect to the original object items). Each item was presented for 2000 msec with a 500 msec inter-stimulus interval (fixation cross). Similarity rankings were based on a prior study (Lacy et al. 2011). For each image participants had to quickly judge whether it was “Old” or “New” using again the ‘F’ and ‘J’ keys, respectively. Importantly, participants were instructed to answer “Old” only for any image that they thought was the *exact* image from the study session, and to answer “New” to all other images, including any image that might have looked similar to a study image but was not exactly the same. Participants were further instructed to always keep their fingers on the

keys during the test and to respond as fast and as accurately as possible. Only responses made within the 2000 msec display time were recorded and in both study and test phase, the images remained on the screen for the full 2000 msec regardless of response time to ensure that exposure time was matched across trials.

To ensure that participants understood task instructions, they were first administered a short practice run prior to the actual task, which included 4 trials, one from each category. A performance of 75% correct had to be achieved to continue to the actual task. If the performance was below 75% task instructions were explained again and the practice trial was repeated.

Quality Control Procedures

Prior to fitting the DDM, trials with no response were removed from the data, resulting in the removal of 220 “no response” trials in the older group (7 trials per participant) and 68 “no response” trials in the younger group (2.6 trials per participant). Note that Bayesian models consider the number of data points while approximating parameter estimations. A lower number of trials would be expressed in more uncertainty of the estimated parameters (see next section). Next, RT data was preprocessed to remove contaminants. Contaminants that were suspected to be fast guesses were removed by using an exponentially weighted moving average (EWMA) control method (see Vandekerckhove and Tuerlinckx 2007). This method identifies the minimal RT where responses start to deviate from what would be expected when guessing. This minimal RT is then used as a lower cut-off and all RTs below this cut-off are censored. This led to a removal of 5% of the data for older adults and 5.55% for younger adults. An upper cut-off was already implemented during the experiment by not recording any RTs that were slower than 2000 msec.

To ensure that only data from participants that understood and followed the task instructions was used, we excluded participants whose recognition memory performance for the foils and targets was two or more standard deviations below the respective group mean. Recognition memory performance for the foils was assessed based on the correct rejection rate for the foil stimuli (number of “New” responses for foils divided by the total number of foil responses). Recognition memory for targets was assessed based on the target hit rate (number of “Old” responses for targets divided by the total number of target responses). Based on those criteria two older and three younger participants were excluded.

Hierarchical Bayesian Diffusion Decision Model

Accuracy and response time for each trial for each participant were fit to a Hierarchical Bayesian Diffusion Decision model by using custom-designed models in JAGS version 4.3.0, an open source program for the analysis of Bayesian models using Markov Chain Monte Carlo (MCMC) (<https://sourceforge.net/projects/mcmc-jags/>). A series of models with increasing complexity in terms of the number of hierarchical levels and the permitted variation of parameters across those levels were specified and fit to the data. There are two general characteristics of Bayesian hierarchical models that are also important

features of the models specified in the current study. They are hierarchical because lower levels are nested in higher levels. This means that each parameter estimate at a given level is constrained by a normal distribution of that parameter at the higher order level. Each normal distribution is defined by a mean μ and a standard deviation σ . The Bayesian aspect of the model refers to the fact that each parameter is estimated by defining a prior distribution of the parameter's likelihood, which is then updated to a posterior distribution after observing the data (Vandekerckhove et al. 2011). Accordingly, each parameter estimate in the model is represented by such a posterior distribution to account for the uncertainty associated with the estimation. More uncertainty is reflected in a wider posterior.

The Deviance Information Criterion (Spiegelhalter et al. 2014, for a discussion of DIC see Spiegelhalter et al. 1994), which evaluates relative model performance based on a balance between model complexity and model fit, was used to determine the model with the best architecture. The model architecture with the smallest DIC (indicating the best fitting model) was chosen for diffusion model parameter estimation (Model 7). A list of all models and their DICs can be found in **Supplementary Table S1**. We will refer to Model 7 below as the Full Model since it is comprehensive and exploratory. We also estimated DDM parameters for the old and young group with two standard diffusion model fitting methods -the two chi-square method (for description of the method, see Ratcliff and Tuerlinckx 2002) and the publicly available Diffusion model fitting package fast-dm (Voss and Voss 2007) (see **Supplementary Table S2**). A comparison and discussion of these and other fitting methods can be found in Ratcliff and Childers (2015).

The Full Model included four levels of hierarchy: a group level g , a participant level s , a condition level c (targets, foils, high similarity lures, low similarity lures), and a trial level t . All four DDM parameters (δ , β , τ , α) were allowed to randomly vary across groups and participants. Accordingly, each parameter was estimated for each group and each participant. The parameters δ , β , τ were also allowed to vary across trials and were therefore estimated for each trial. Furthermore, the drift rate parameter δ representing mnemonic discrimination ability in our model was also allowed to vary across conditions as the discrimination ability was expected to differ depending on the image type (foil, target, lure).

Although DDM parameters were estimated at each of the four hierarchical levels our primary objective was to determine the cognitive decision-making processes in which the older and younger groups differed. We, therefore, focused on comparing the means of the DDM parameters (α , β , τ , δ_{Target} , δ_{Foil} , δ_{HighSim} , δ_{LowSim}) at the group level. To facilitate a comparison of the group means, we specified the model to directly estimate group difference parameters. One group difference parameter was estimated for each of the DDM parameters and represented the difference in the group means between the older and younger adults (α_{Diff} , β_{Diff} , τ_{Diff} , $\delta_{\text{DiffTarget}}$, δ_{DiffFoil} , $\delta_{\text{DiffHighSim}}$, $\delta_{\text{DiffLowSim}}$). For δ , a difference parameter was estimated for each condition. To ensure comparable variance of the estimated parameters in both groups, the group means for each DDM parameter were derived from the following linear combination in the model:

$$\text{GroupMean}_{\text{xOLD}} = \text{GroupMid}_{\text{x}} - 0.5 * \text{Group Difference parameter}_{\text{x}}$$

$$\text{GroupMean}_{x\text{YNG}} = \text{GroupMid}_x + 0.5 * \text{Group Difference parameter}_x$$

$x = \text{DDM parameters } (\alpha, \beta, \tau, \delta_{\text{Target}}, \delta_{\text{Foil}}, \delta_{\text{HighSim}}, \delta_{\text{LowSim}})$

The estimated GroupMid parameter was the same for both groups and represented a DDM parameter value that was in between the two group means.

Lastly, the upper boundary in our model was designated as the “New” response and the lower boundary as the “Old” response. For the interpretation of the drift rate values, it follows that positive values of the estimated δ group mean reflect an average accumulation of evidence toward the “New” response and negative values reflect an average accumulation of evidence toward the “Old” response. A graphical representation of the Full Model and the nesting of its levels is shown in **Figure 3**. The JAGS code for this and all other models can be found at https://github.com/Yassa-TNL/diffusion-model-mdto/tree/master/jags_files

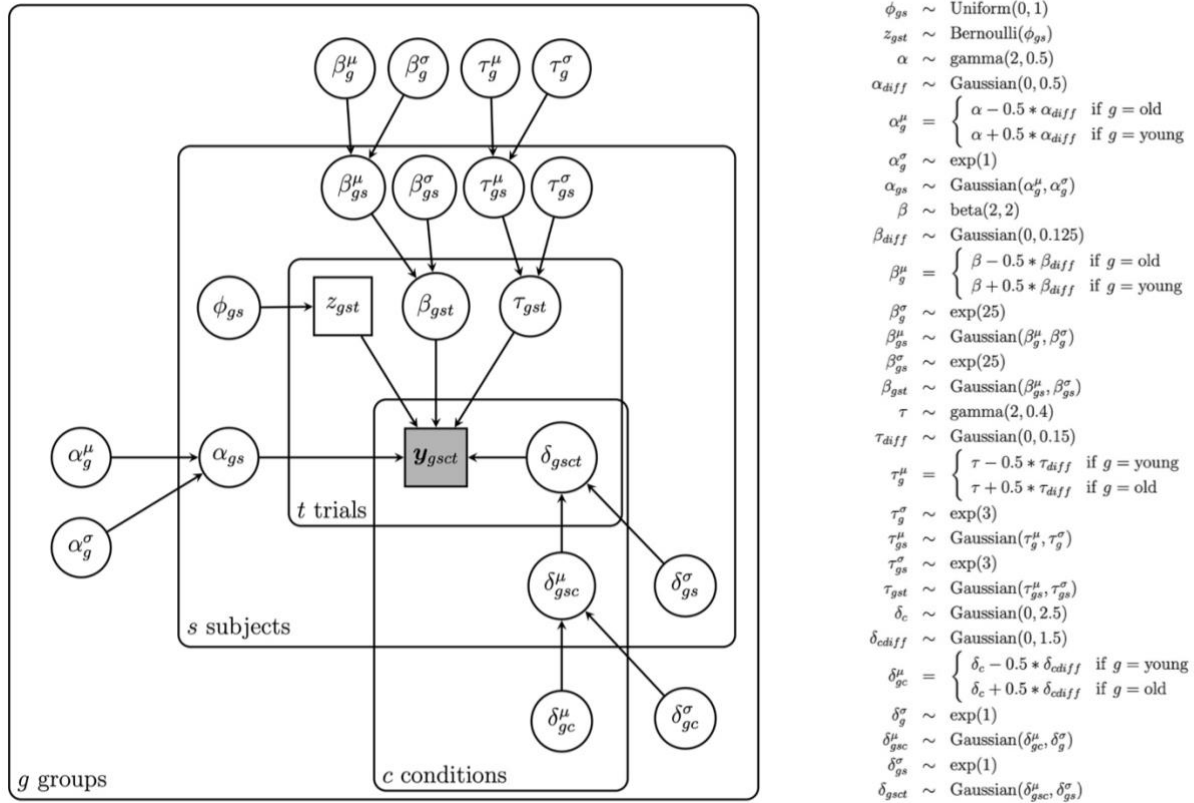


Figure 3. A graphical representation of our Hierarchical Bayesian Diffusion Model (Full Model) to investigate the difference in model parameters between older and younger adults in the MDT. The shaded node indicates the observed (bivariate) data (response type (correct/incorrect) and response time). Unshaded nodes indicate parameters that are being estimated in the model. The parameters β and τ are indexed for group (g), participant (s), and trial (t) meaning that we allowed those parameters to be different for each group, participant, and trial. δ is additionally indexed for condition (c) meaning that δ was also allowed to differ across conditions. α was allowed to differ across the two groups and participants.

DDM Parameter Estimation

Parameters were estimated with the joint probability of each choice response (correct or incorrect) and its response time being distributed according to a Wiener distribution with four parameters $Y(\text{gsct}) \sim \text{Wiener}(\alpha(\text{gs}), \delta(\text{gsct}), \beta(\text{gsct}), \tau(\text{gsct}))$. Estimation was performed according to the Bayesian approach. First, prior distributions were defined for all parameters, which were then updated to posterior distributions. Accordingly, each parameter estimate is described by a posterior distribution which is a probability distribution that quantifies the uncertainty about the estimated model parameter after having observed the data.

Posterior distributions were estimated simultaneously for all model parameters through Markov-chain Monte Carlo (MCMC) methods. Data were sampled using 6 independent chains each containing 250,000 samples. The first 80,000 samples were discarded as burn-in to ensure convergence of the chains. The R-hat statistic was used to confirm convergence in all chains (Gelman and Rubin 1992). Parameters with values below a threshold of 1.05 across all chains were considered as successfully converged. The defined prior distributions for all parameters are listed in **Figure 3**. All priors were uninformative but with theoretically informed limits on the possible range of each parameter. For the GroupMid and the Difference parameters, we used uninformative priors that were centered at 0 but still allowed for values other than 0.

Evaluating Evidence for Group Differences Using the Bayes Factor

To evaluate the difference between the older and younger groups for each of the DDM parameters individually in the Full Model, we used the Savage-Dickey density ratio (Dickey and Lientz 1970) as a simple method to compute the Bayes factor (BF). We calculated the BF for each Group Difference parameter by taking the ratio of a Difference parameter's prior distribution value at 0 to the Difference parameter's posterior distribution value at 0. This BF thus compares how the likelihood of the Group Difference being 0 in a given Difference parameter changes after having observed the data. A BF of k , when greater than 1, indicates that it is k times less likely that a difference parameter equals 0 after seeing the data than before (prior). This can be interpreted as evidence in favor of the hypothesis that there is a group difference. On the other hand, a BF of m , when lower than 1, indicates that it is $1/m$ more likely that a difference parameter equals 0 after seeing the data than before. This can be interpreted as evidence in favor of the hypothesis that there is no group difference. Lastly, a BF around 1 indicates that observing the data did not provide any additional information about how the likelihood of a Group Difference being 0 changes and we therefore have no evidence in favor of either hypothesis.

Model Optimization and Selection of Final Reduced Model

The Full Model allowed for group differences in all DDM parameters and the evidence for a group difference was evaluated for each parameter individually. Therefore, this model can be viewed as an exploratory model aiming to identify for which parameters there is evidence for a group difference. Accordingly, evidence in the Full Model considers the likelihood of a group difference for each parameter individually and independent of the

other parameters' group difference likelihood and in the context of all DDM parameters being allowed to differ between groups. We then wanted to re-estimate model parameters and re-evaluate the evidence for a group difference from the exploratory model in a more proper model context. We used the exploratory model to discover the underlying data structure (e.g. in what parameters group differences exist) and then used the reduced model to run parameter estimations in the more accurate data structure. We, therefore, specified a last model (Final Reduced Model) that was informed by the results of the Full Model. Different than in the Full Model, in the Final Reduced Model, we allowed group differences only in those parameters that we found evidence for in the Full Model (we will refer to these parameters as "final parameters") while group differences in the remaining parameters were set to 0. Evidence for the likelihood of a group difference in the Final Reduced Model is considered therefore in the context of no other group differences being allowed. Another important difference is that in the Reduced Model the evidence for the likelihood of the group differences was evaluated for all parameters jointly as opposed to individually. It is important to note that the Bayes Factor that is calculated in the Final Reduced Model is still valid, even when we test in this case a hypothesis that was inspired by previously looking at the data in the Full Model. The Full Model and the Final Reduced Model did not differ in any other respect.

To evaluate the joint likelihood of a group difference for the final parameters in the Final Reduced Model, we calculated a multidimensional version of the Savage-Dickey Density ratio. For this, we defined one multivariate normal distribution based on the prior distributions for the final parameters. This resulted in a multidimensional prior distribution with a mean at (0, 0, 0) (according to the mean of each individual prior distribution at 0) and the covariance matrix corresponding to a diagonal matrix that consists of the variances of each of the individual priors. We then calculated the value of this 3D prior distribution at (0, 0, 0). Next, we used the sampled data from the posteriors of the final parameters (which were sampled simultaneously) to create a 3D posterior distribution and to calculate the value at (0, 0, 0) in that distribution as well. Our 3D BF was then calculated by taking the ratio of the 3D prior distribution value at (0, 0, 0) to the 3D posterior distribution value at (0, 0, 0). A general overview of the steps of the modeling procedure can be found in **Figure 4**.

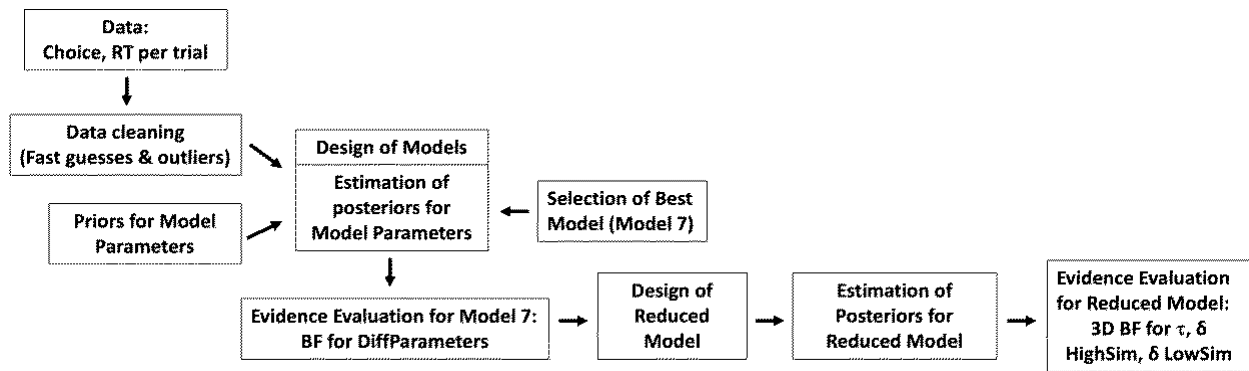


Figure 4. General scheme of the steps of the modeling procedure. Model 7 (Full Model) allowed group differences in all parameters. The Reduced Model was designed based on evidence from the Full Model and allowed group differences only in DDM parameters that we found evidence for in the Full Model (final parameters). BF: Bayes Factor (calculated as the Savage-Dickey Density ratio).>

Evaluating Model Performance

To assess the quality of the model fit we modeled reaction time distributions with correct and incorrect responses based on the estimated model parameters. Responses were modeled for 40 trials for each condition and participant and compared to the observed data.

The modeled data was obtained by sampling from the posterior distributions of each diffusion model parameter (δ , β , τ from posteriors at trial level; α from the posterior at participant level) and entering the sampled diffusion model parameter values into the Wiener Distribution. The sampling procedure was repeated 250,000 times for each trial. This generated for each of the 40 trials a posterior predictive distribution of 250,000 responses with each response being represented by a reaction time and accuracy (correct or incorrect).

We first compared the modeled and observed accuracy for each participant and condition. For this we calculated for every trial in the modeled dataset the percentage of “New” responses (corresponding to the correct response) in the samples of the posterior predictive. We then averaged the accuracies across all trials in each condition and participant and compared it to the accuracies in the observed dataset.

We additionally calculated the percentage of overlap between the modeled and the observed response types. For this, we obtained the modeled accuracy by determining the response type (correct or incorrect) of a given trial based on a majority vote for one of the two boundaries (corresponding to a response type) in the samples of the posterior predictive. Trials that were ‘no response’ trials in the observed data set were not modeled. We then calculated for each condition and participant the percentage of trials where responses were matching between the modeled and the observed trials.

Lastly, we compared how well the modeled reaction time distributions describe the observed ones. For this, we approximated the reaction time distributions by calculating the 0.1, 0.5, and 0.9 quantiles and compared the modeled and observed quantiles. The observed quantiles were obtained by using the reaction time distribution of all trials within a given response category (correct and incorrect), condition, and participant and calculating the density function. The density function was then used to calculate a cumulative distribution. The points where the probability was 0.1, 0.5, and 0.9 represent the corresponding reaction time quantiles.

The modeled quantiles were calculated by first determining the response category (correct and incorrect) of a modeled trial based on the response category in the observed dataset of that corresponding trial. Then only RTs of samples in the posterior distribution that predicted the same boundary (corresponding to response type) as in the observed data were used. Lastly, the RTs of all modeled trials that belong to the same response category, condition, and participant were put together, and quantiles were calculated as in the observed data set described above.

RESULTS

Quality Check of Model Performance

The comparisons between the modeled and the observed data for the Full Model and the Final Reduced Model indicate that the models fitted the empirical data well. **Figures 5a** and **5b** show the modeled and observed accuracy and the overlap in the modeled and observed response types for the Full Model. **Figure 6** shows the overlap between the modeled and observed quantiles of the reaction time distributions for the Full Model. The corresponding graphs for the Final Reduced Model can be found in the **Supplementary Figures S2 and S3**.

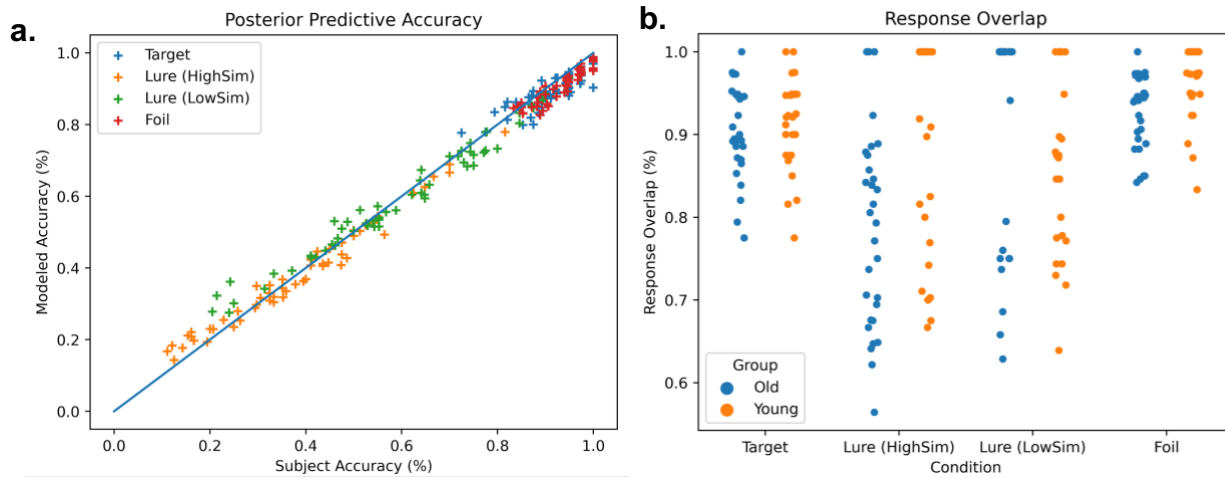
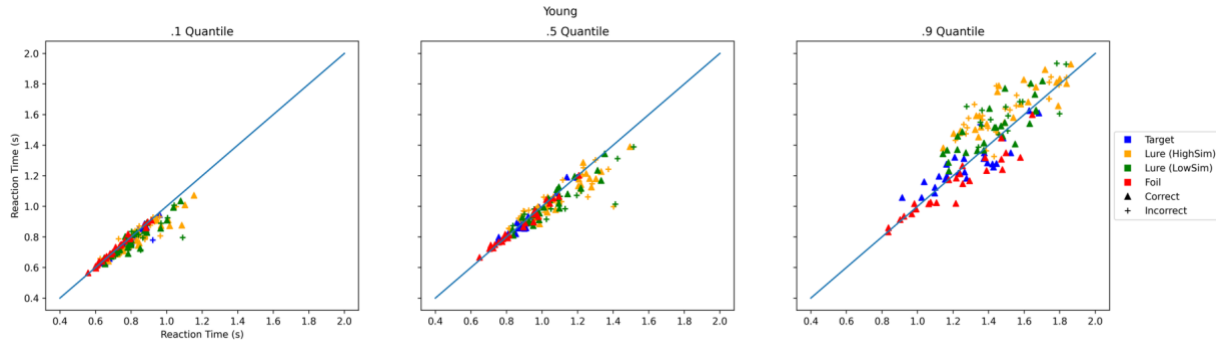


Figure 5. Quality check of our model performance. **a.** The modeled and observed accuracy for the Full Model. The modeled accuracy data fit the observed accuracy data well. Each cross represents the accuracy of one participant averaged across all trials within each condition. **b.** Overlap between the modeled and observed response type (“Old”/“New”). Individual data points represent the percentage of agreement in response type between the observed and modeled dataset for each participant and condition.



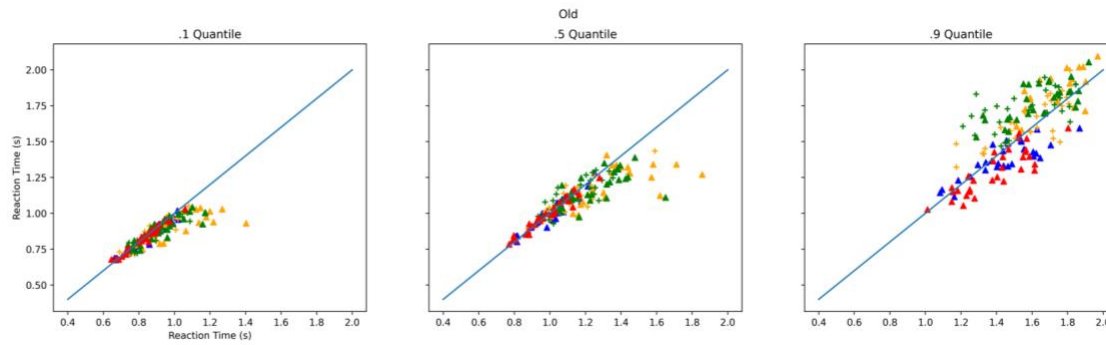


Figure 6. Quality check of model performance. Overlap between the modeled and observed 0.1, 0.5, and 0.9 quantiles of the modeled and observed reaction time distributions for the Full Model. Each data point represents the RT of one participant per condition and response type (“Old”/“New”).

Group Differences in Diffusion Model Parameters (Full Model)

RTs in each condition for the young and older group and the lure discrimination index (LDI) are presented in the Supplementary Figure S1. Similar to previous studies, older in comparison to younger adults showed a significantly lower LDI in the two lure conditions indicating a poorer mnemonic lure discrimination performance in older adults.

Table 2 shows the posterior group mean estimates for each of the Diffusion model parameters for the young and the old group and the Bayes Factors for the corresponding Group Difference parameters for the Full Model.

We found evidence for a group difference in the non-decision time parameter τ . A difference between the old and young group in the non-decision time was approximately 4 times more likely after having observed the data (BF for τ Diff = 4.44). The posterior group mean estimates showed a 90 ms longer non-decision time in older adults in comparison to younger adults (0.80 ms vs. 0.71 ms).

Model parameter	YOUNG	OLD	Bayes Factor Group Diff
Response caution (α)	1.42 [1.37, 1.47]	1.43 [1.37, 1.49]	0.08
Response bias (β)	0.48 [0.45, 0.52]	0.48 [0.46, 0.50]	0.16
Encoding/Motor Response (τ)	0.71 [0.67, 0.77]	0.80 [0.75, 0.84]	4.44
Discrimination ability (δ)			
Target	-2.56 [-2.99, -2.23]	-2.35 [-2.69, -2.07]	0.23
Lure (HighSim)	0.00 [-0.29, 0.28]	-0.91 [-1.18, -0.68]	> 10 000
Lure (LowSim)	1.05 [0.76, 1.37]	0.06 [-0.17, 0.30]	> 10 000
Foil	3.85 [3.45, 4.61]	3.30 [2.96, 3.93]	0.94

Table 2. Posterior group mean estimates (means of the posterior distributions at the group level) of the Diffusion Model Parameters for old and young and the corresponding 95% Bayesian credibility intervals (95% probability that the population parameter lies in the interval) of the Full Model. The last column shows the Bayes Factor (Savage-Dickey Density ratio). A BF of k , when > 1 , indicates that it is k times less likely that a difference parameter equals 0 after seeing the data. This is interpreted as evidence in favor of the hypothesis that there is a group difference. A BF of m , when < 1 , indicates that it is $1/m$ more likely that a difference parameter equals 0 after seeing the data. This is interpreted as evidence in favor of the hypothesis that there is no group difference. For drift rate interpretation: positive values are linked to the “New” response which is represented by the upper threshold in the model. Negative values are linked to the “Old” response which is represented by the lower threshold. A drift rate value around 0 indicates that no evidence could be extracted, and the participant is guessing.

More interestingly, for both lure conditions we found evidence for a group difference in the drift rate parameter δ which is indexing mnemonic discrimination ability in our task. The likelihood for a group difference between old and young in drift rate for the Lure (HighSim) and the Lure (LowSim) condition was $> 10\,000$ times more likely respectively (BF for $\delta\text{Diff}_{\text{HighSim}} = 10\,310$, BF for $\delta\text{Diff}_{\text{LowSim}} = 42\,509$) after having observed the data. A benefit of the drift rate parameter is that it provides information about the quality of evidence that a participant can extract from the stimulus and from memory and hence it allows us to analyze the nature of the age-related differences in more detail. A positive drift rate indicates that on average the information accumulation drifts toward the upper boundary corresponding to the “New” response. A negative drift rate indicates information accumulation toward the lower boundary corresponding to the “Old” response. The numerical value of the drift rate on the other hand informs about the magnitude of this drift. Higher drift rate values indicate a faster accumulation of information toward the boundary and hence a faster approaching of a boundary. Higher drift rate values are typically observed in easier task conditions and result in higher accuracy and faster RT. Accordingly, we found that in the lure (LowSim) condition, the drift rate in younger adults indicated that younger adults accumulated evidence toward the correct “New” response while a drift rate close to 0 in the older adults group indicated that older adults could not reliably accumulate evidence toward either one of the response boundaries (group mean estimates δ Lure (LowSim): 1.05 (young) vs. 0.06 (old)). When the similarity of the lures increased in the lure (HighSim) condition, the drift rate close to 0 in younger adults indicated that now younger adults could not accumulate evidence toward either of the response boundaries. However, for the older adults group the drift rate indicated now that older adults accumulated on average information toward the incorrect “Old” response (group mean estimates δ Lure (HighSim): 0.00 (young) vs. -0.91 (old)). We did not learn anything new from the data in regard to a group difference in drift rate for the foil condition as indicated by the BF for $\delta\text{Diff}_{\text{Foil}}$ around 1. However, the likelihood of no group difference in the drift rate parameter for the target condition was approximately 4 times more likely after having observed the data (BF for $\delta\text{Diff}_{\text{Target}} = 0.23$) (see **Figure 7** for a comparison of all drift rates across conditions and participant groups).

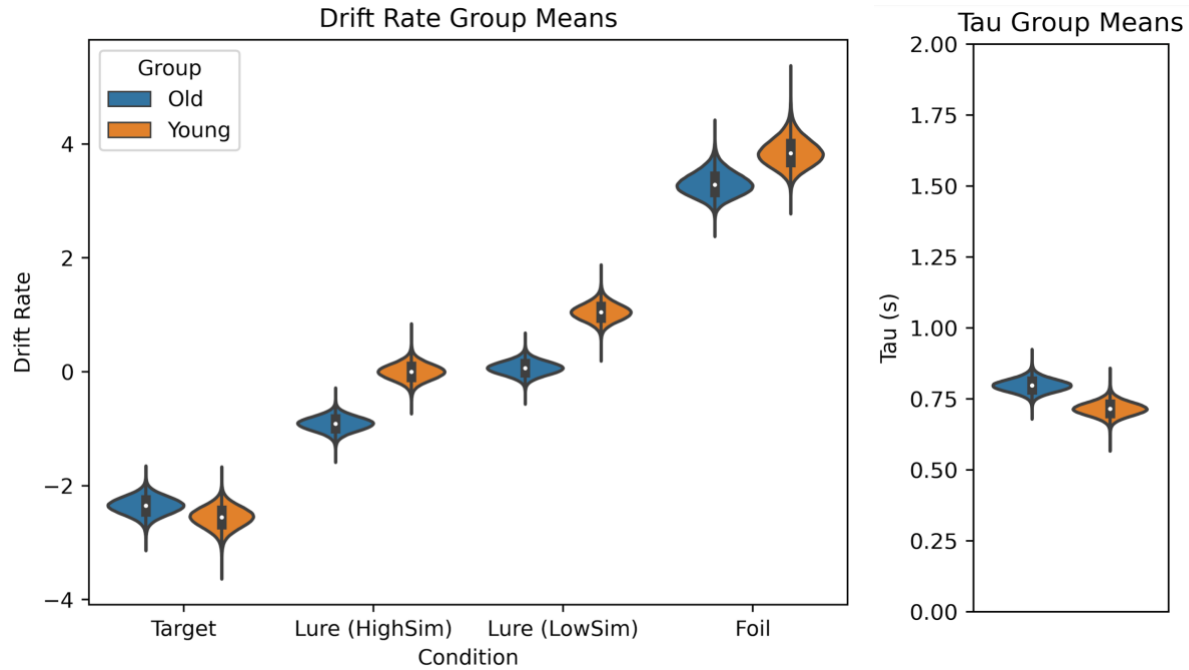


Figure 7. A. Estimated drift rate (δ) group means (means of the posterior distributions at the group level) for older and younger adults for all four conditions. A negative drift rate represents on average “Old” responses. A positive drift rate represents on average “New” responses. Older adults show a drift rate around 0 for the lure (LowSim) condition indicating on average that they were not able to accumulate evidence towards either of the response boundaries (similar to younger adults in the lure (HighSim) condition). A negative drift rate in the lure (HighSim) condition indicates on average that older adults accumulate evidence towards the incorrect “Old” response for highly similar lures. B. Estimated non-decision time (τ) group means for older and younger adults. Older adults have a 90ms longer non-decision time.

We further found evidence that the old and young groups did not differ in the response caution parameter α and the response bias parameter β . The likelihood of no difference between groups in response caution was approximately 13 times more likely (BF for α Diff = 0.08) and in response bias approximately 6 times more likely (BF for β Diff = 0.16). Although the posterior group mean estimates for the older and younger adults indicated a slight bias toward the “Old” response in both groups (posterior group mean estimate β for old and young: < 0.5), the 95% Bayesian credible intervals included 0.5 (β at 0.5 indicates no response bias) hence the probability of no response bias in both age groups was still 95%. The two chi-square and the fast-dm fitting methods replicated the group differences in τ , δ Lure (LowSim), and δ Lure (HighSim) (see **Supplementary Table S2**).

Group Differences in Diffusion Model Parameters (Final Reduced Model)

Based on the results from the Full Model we investigated in the Final Reduced Model evidence for the joint likelihood of a group difference in the δ parameters in the two lure conditions and τ (δ Diff_{HighSim}, δ Diff_{LowSim}, τ Diff). **Table 3** shows the posterior group mean estimates of the Final Reduced Model. The drift rate and non-decision time estimates differed only marginally from the estimates in the Full Model confirming once again the group differences we describe in the Full Model. The 3D Bayes Factor for the joint

likelihood of a group difference in δ (HighSim), δ (LowSim) and τ was $>10\ 000$ indicating that the likelihood of a group difference in all 3 parameters is more than 10 000 times more likely after having observed the data.

Model parameter	YOUNG	OLD	3D Bayes Factor for Group Diff
Response caution (α)	1.43 [1.39, 1.46]	1.43 [1.39, 1.46]	
Response bias (β)	0.48 [0.46, 0.50]	0.48 [0.46, 0.50]	
Encoding/ Motor Response (τ)	0.71 [0.66, 0.76]	0.80 [0.76, 0.85]	
Discrimination ability (δ)			
Target	-2.45 [-2.71, -2.23]	-2.45 [-2.71, -2.23]	
Lure (HighSim)	0.00 [-0.28, 0.27]	-0.94, [-1.23, -0.70]	
Lure (LowSim)	1.01 [0.74, 1.30]	0.07 [-0.18, 0.31]	
Foil	3.55 [3.26, 3.91]	3.55 [3.26, 3.91]	
δLure (HighSim), δLure (LowSim), τ			$> 10\ 000$

Table 3. Posterior group mean estimates (means of the posterior distributions at the group level) of the Diffusion Model Parameters for old and young and the corresponding 95% Bayesian credibility intervals (95% probability that the population parameter lies in the interval) of the Final Reduced Model. The last column shows the 3D Bayes Factor (Savage-Dickey Density ratio) for the joint likelihood of a group difference for the parameters δ Lure (HighSim), δ Lure (LowSim), and τ . The BF of 10 000 indicates that the likelihood of a group difference in all three parameters is more than 10 000 times more likely after seeing the data. This is interpreted as evidence in favor of the hypothesis that there is a group difference in δ Lure (HighSim), δ Lure (LowSim), and τ . For drift rate interpretation: positive values are linked to the “New” response which is represented by the upper threshold in the model. Negative values are linked to the “Old” response which is represented by the lower threshold. A drift rate value around 0 indicates that no evidence could be extracted, and the participant is guessing.

Lastly, we decided to evaluate the magnitude of the group difference in the δ (HighSim), δ (LowSim) and τ parameters. For this, we examined how much the parameter values would differ between a representative old participant and a representative young participant based on the estimated group mean differences and group standard deviations in the Final Reduced Model. We generated a representative distribution for an old and a young participant for δ (HighSim), δ (LowSim) and τ by sampling from the group mean and group standard deviation posterior distributions of the respective parameter (we had different standard deviation distributions for each parameter. But the standard deviation for delta were the same across conditions but different for old and young). Using these representative participant posterior distributions, we calculated the probability that a representative old participant would have a higher τ , a lower δ (HighSim), and a lower δ (LowSim) than a representative young participant, respectively. The probability was

calculated by dividing the number of samples that satisfy the criterion out of all the samples from the representative posterior distributions. Representative distribution parameters were sampled at the same time while all other parameters during the model estimation were sampled.

The representative participant distribution indicated that the likelihood of a representative older adult having a lower drift rate in the lure (LowSim) condition or lure (HighSim) condition was 87% respectively. The likelihood of a representative older adult having a higher τ than a representative young adult was 70%. These likelihoods suggest that despite individual variability within each age group, the differences in group means, which we found evidence for based on the BFs, are also meaningful at the individual participant level.

DISCUSSION

In the current study, we used the Ratcliff Diffusion Model to investigate whether differences in non-mnemonic decision-making processes underlie the well-documented age-related decline in lure discrimination performance in the MDT or if this decline is solely based on a deficit in mnemonic ability. Our results from the Full and the Reduced Model provide supporting evidence that the age-related decline in lure discrimination performance is largely driven by differences in mnemonic discrimination abilities. In comparison to younger adults, older adults showed in both lure conditions a poorer mnemonic discrimination ability as measured by the drift rate parameter. In addition, we also found a slowing in the non-decision time in older adults, indicating that older adults are slower in processes like the speed of encoding and executing the motor response. Differences in non-mnemonic decision processes like response caution and response bias did not contribute to the decline in performance in older adults.

Non-mnemonic decision-making processes do not contribute to age-related decline in MDT performance

It has been proposed that in addition to a decline in hippocampal pattern separation, an age-related decline in lure discrimination performance in the MDT could be related to age-related changes in decision-making processes during memory judgments (Pishdadian et al. 2020). For example, a recent meta-analysis reported a more liberal response bias (higher tendency toward “Old” responses) in older adults in recognition memory tasks (Fraundorf et al. 2019). However, the presence of a response bias in older adults was not consistently observed across different studies and varied as a function of study material, study task, and test task. To our knowledge, the presence of an age-related response bias in the MDT has never been systematically evaluated. The focus was rather on comparing performance measures like d-prime and LDI across groups that were assumed to not be confounded by a response bias. In the current study, we show that there is no difference in response bias between older and younger adults that contributes to the age-related decline in mnemonic lure discrimination performance. This is also interesting because we have a higher proportion of stimuli in our task for which the response “New” is the correct response. Past studies have found

evidence for a response bias when manipulating stimulus proportion (Leite & Ratcliff, 2011, Criss, 2010). While speculative, the absence of a response bias could be related to the specific task instructions, in which participants attention was drawn to the similarity aspect of the task. This may have made them look at each stimulus more carefully and consequently mitigate the impact of a response bias.

It is important to note, that bias manipulation can be modeled in two different ways in the Diffusion Model – by changes to the response bias or the zero point of the drift rate. In the study from Leite & Ratcliff manipulation of stimulus proportion caused changes to the response bias parameter but manipulations in the decision cut-off in their task were reflected in changes to the zero-point position of the drift rate. Manipulating both simultaneously, lead to changes in both parameters. Similarly, Criss (2010) reported in her recognition memory study effects of stimulus proportion increase on the response bias parameter and effects of list-strength increase (words were presented either one time or 5 times at study) on the zero point of the drift rate. However, a simultaneous manipulation of both aspects was not investigated in this study.

In addition to the stimulus proportion difference in our task, we also have potentially another bias manipulation through the different similarity-levels of the stimuli (e.g. a higher bias to respond with “New” for high-similarity lures than for low-similarity lures). It is possible that in our specific task the response bias was reflected in changes to the zero point position of the drift rate, rather than by changes in the response bias parameter. The exact effects of different bias manipulations might be specific to a task and task instructions. It will need to be further investigated what model parameters are sensitive to what kind of bias manipulation in memory tasks.

Response caution is another decision process that could contribute to differences in mnemonic lure discrimination performance in the MDT. Studies that employed a Diffusion Model analysis have often reported a higher response caution in older adults in perceptual decision making and standard recognition memory (McGovern et al. 2018; Spaniol et al. 2006; Ratcliff et al. 2004). A higher response caution typically results in slower but more accurate decisions. Considering older adults are less accurate, more often incorrectly labeling lures “Old”, we would rather expect a lower response caution in older adults. Previous MDT performance measures were not able to take the influence of age-related differences in response caution to performance into account because they did not consider the decision time. Using the DDM, we show here for the first time that, at least in our study, older and younger adults do not differ in their response caution during MDT performance. Studies have shown that response caution can be increased in an experiment when accuracy is emphasized in the task instructions (Thapar et al. 2003; Voss et al. 2004). It is therefore possible that we did not observe any group differences in response caution due to our specific task instructions, which ask participants to respond as fast and as accurately as possible. Similarly, in a recognition memory task used by Spaniol et al. 2006 in which participants received instructions that emphasized both accuracy and speed, no age-related difference in response caution was found (Spaniol et al. 2006).

Drift rate reveals a memory bias toward “Old” in older but not in younger adults

In contrast to previous MDT performance measures that were used to infer mnemonic lure discrimination ability, the drift rate measure from the DDM represents likely a more accurate index of mnemonic lure discrimination ability. This is because, unlike previous measures that were based solely on accuracy data, the drift rate parameter is derived from a model that considers the accuracy and the response time (including the shape of the RT distributions) and it accounts for both dependent measures simultaneously. This allows for a consideration of the speed-accuracy trade-off and allows us to assess mnemonic ability that is not contaminated by differences in response bias or response caution (White et al. 2010). Accordingly, we were able to show here with a more accurate measure, the drift rate, that the age-related decline in mnemonic lure discrimination performance is driven by a deficit in mnemonic abilities. Furthermore, because the drift rate informs about the quality of the evidence that participants can extract from the current lure and from a previously formed memory representation, we were also able to analyze the nature of the age-related deficit in more detail. We found different mechanisms of older adults' mnemonic lure discrimination deficit across the two lure conditions. For lures that shared only low similarity with previously seen images, older adults showed poorer mnemonic discrimination ability than younger adults because they were not able to extract meaningful evidence from the lure images and from memory. The drift rate indicated that decisions about the “Old”/“New” status of low similarity lures in older adults were largely driven by noisy fluctuations of information resulting in correct decisions only at chance level. However, for lures that had a high similarity to previously seen images the drift rate showed that the mnemonic discrimination deficit in older adults was predominantly characterized by extracting information that indicated incorrectly that a lure is an old image. The tendency of older adults to extract evidence from the current image and memory that makes them recognize a highly similar lure as an old image is characteristic of a memory bias toward “Old”. Note that a memory bias toward “Old” is distinct from a response bias toward “Old”. A memory bias indicates that the underlying reason for incorrect “Old” judgments to lures is a memory effect, while a response bias toward “Old” indicates an effect at the response level -a tendency to respond more often with “Old” (Spaniol et al. 2008; White et al. 2010). Due to the methodological shortcomings of previous MDT performance measures as discussed earlier, previous measures were not able to provide clear evidence for a memory bias. These measures are more confounded by other processes, and are not a direct measure of the quality of evidence that participants are able to extract.

The observation of a memory bias toward “Old” is in agreement with the prevailing hypothesis that older adults show a decline in mnemonic lure discrimination due to an increased propensity toward over-generalization. The underlying reason of this over-generalization is thought to lie in an age-related shift in hippocampal network dynamics away from pattern separation and toward pattern completion. Pattern completion describes a neural computation that contributes to memory function by retrieving a previously stored pattern based on a partial cue. In line with this theory, we have previously shown, based on BOLD signal activity changes, that signal to increases in lure similarity (consistent with pattern separation) in the DG/CA3 subfields is attenuated in older adults (Yassa et al. 2011b). For low similarity lures, pattern separation signals were observed similarly in young and older adults. However, for high similarity lures, pattern

separation signals diminished in older adults (suggesting pattern completion) while remaining high in younger adults. These findings supported a theoretical prediction of how the input/output transfer functions of the DG/CA3 in old and young adults are shifted (Yassa et al. 2011b). The transfer function describes pattern separation and completion behavior in the hippocampus as a function of change in input. It is suggested that in older adults the transfer function is shifted in a way where larger amounts of change in input are required for the DG/CA3 to switch from pattern completion to pattern separation. The requirement for larger changes was termed representational rigidity and re-capitulated findings in studies of aged rats (Wilson et al. 2006).

There are some important differences between the current study and the aforementioned neuroimaging work, which do not allow us to directly relate our DDM behavioral results to the above-described pattern separation signal behavior in the DG/CA3. In contrast to the explicit instructions in the current study where participants were asked to give overt memory judgments, participants in the Yassa et al. (2011) study were instructed to give “Indoor” “Outdoor” judgements while viewing repeated, novel and lure images in a continuous presentation and hence encoding of the “Old”/“New” status of an image was incidental. It has been shown that BOLD signal activity patterns in the hippocampus differ between the overt and incidental task designs (Motley and Brock Kirwan 2012; Stark et al. 2019b), likely due to varying demands on top-down feedback and decision-making processes. Nevertheless, our findings are consistent with the theoretical prediction of an age-related shift in the transfer function of the DG/CA3 and a higher tendency for pattern completion resulting in over-generalization in older adults.

In our study, the discrimination ability in older adults for low-similarity lures was comparable to the discrimination ability for high-similarity lures in younger adults. This suggests that older adults need indeed larger amounts of change to show a comparable mnemonic discrimination ability to younger adults. Furthermore, our results suggest a higher tendency toward over-generalization in older adults. Older adults went from poor discrimination ability for low similarity lures to an incorrect memory bias toward “Old” (consistent with the idea of pattern completion) for high similarity lures. In contrast, no indication of a memory bias toward “Old” and therefore no indication for over-generalization was observed in the younger adults for the same similarity level. Instead, younger adults transitioned from having good discrimination ability for low similarity lures to poor discrimination ability for high similarity lures.

Slower non-decision time in older adults: relevant for poor mnemonic ability?

In addition to an age-related decline in mnemonic ability as indicated by lower drift rates, we also found that older adults had a longer non-decision time compared to young adults. An age-related slowing in the non-decision time parameter is a reliable finding which has been previously reported (Ratcliff et al. 2004; Spaniol et al. 2006). It is typically interpreted as reflecting an age-related slowing in processes related to encoding and response execution. As an extra-decisional component, the non-decision parameter is assumed to not influence the response choice and hence accuracy (in the model it mostly determines the location of the leading edge of the reaction time distributions) (Ratcliff and Tuerlinckx 2002; Wagenmakers 2009). However, we argue that the age difference that we found

here in the non-decision component, could also contribute to the decreased mnemonic discrimination ability in older adults that is reflected in the lower drift rate parameter.

This is because the non-decision parameter is not well specified in the model as it encompasses the sum of the duration of any other process that does not contribute to the decision process. This is not limited to encoding and response execution processes but can also include processes related to accessing and evaluating memory representations (Ratcliff and McKoon 2008; Wagenmakers 2009). From a theoretical point of view, a slower non-decision time when reflecting a slowing in encoding and memory access and evaluation could contribute to a decrease in mnemonic lure discrimination ability as these processes play an important role in the successful formation and retrieval of memories. This idea is also plausible from the model's perspective as the DDM parameters are not independent and correlations between the non-decision parameter and the drift rate have been reported (Ratcliff and Smith 2004; Thapar et al. 2003).

It has been indeed shown that strategic retrieval processes relevant to memory access and evaluation are contributing to successful lure discrimination performance in the MDT. During overt memory judgments, participants likely employ a recall-to-reject strategy (Kirwan and Stark 2007; Stark et al. 2019b; Trelle et al. 2017). This strategy involves retrieving the originally stored representation from memory and comparing it with the currently presented image to correctly identify the lure as an item not previously viewed. The recall-to-reject strategy is thought to place significant demands on cognitive control and strategic retrieval processes which include accessing the details of a memory representation, constructing a mental image, maintaining this image in working memory, and mentally evaluating the details with the currently presented image (Van der Linden et al. 2009; Badre and Wagner 2007). These control processes are considered part of the executive functions that are supported by the prefrontal cortex (PFC), a region that is also affected by aging. Older adults have been shown to be impaired in PFC-mediated executive functions and the recall-to-reject strategy (Trelle et al. 2017; Cohn et al. 2008; Buckner 2004). A recent study by Gellersen et al. (2021) showed that strategic retrieval ability indexed by tests of executive function best-explained performance in a yes/no version of the MDT in older adults (Gellersen et al. 2021). Another study by Wais et al. 2018 used transcranial magnetic stimulation to perturb the normal neural function of the mid-ventrolateral prefrontal cortex, a region implicated in cognitive control of high-fidelity long-term memory retrieval. The perturbation led to a subsequent diminished discrimination of similar lures in the MDT indicating that PFC-mediated strategic retrieval processes are crucial for the retrieval of high-fidelity memory representations in the MDT (Wais et al. 2018). Consistent with these reports, a recent investigation in aged rats using a LEGO object-based MDT, showed that temporary inactivation of the prefrontal cortex (PFC) during test using the GABA agonist muscimol impaired mnemonic discrimination performance, suggesting that PFC-mediated retrieval processes play a critical role in this process (Johnson et al. 2021).

The cognitive control and strategic retrieval processes that are required for a successful recall-to-reject strategy are likely part of the non-decision time in the DDM because, like general encoding and response execution processes, they do not depend on the nature of the currently presented stimulus (Wagenmakers 2009). In line with this idea, Spaniol et al. (2006) reported that the non-decision component was affected by the type of

retrieval that participants engaged in (Spaniol et al. 2006). The non-decision time was slower for episodic retrieval in comparison to semantic retrieval. This effect was more pronounced in older adults. Considering the lack of perceptual differences between episodic and semantic stimuli during the study phase in this study, the differences in non-decision time were likely driven by differential processes during the test phase. The authors speculated that it could indicate an adaptive response to perceived changes in task difficulty (e.g., a slowing of motor operations during the more difficult episodic task). Alternatively, the slowing in the non-decision time for episodic retrieval could also indicate the recruitment of additional cognitive control and retrieval processes more necessary for episodic than semantic retrieval (Moscovitch and Melo 1997). In summary, the age-related slowing in the non-decision component in our study could indicate an age-related slowing in cognitive control and strategic retrieval processes, in addition to a slowing in encoding and response execution. This slowing in cognitive control and strategic retrieval processes could put older adults at a disadvantage. When given the same amount of time as the younger adults, it could make it more difficult for older adults to successfully implement the recall-to-reject strategy, thereby contributing to a poorer mnemonic discrimination ability.

The slowing in the non-decision time could also reflect a slowing in encoding processes during the test phase, which could also contribute to poorer mnemonic lure discrimination ability in older adults. Recent evidence suggests that in addition to hippocampus-dependent processes like pattern separation, successful mnemonic lure discrimination in older adults is also affected by processes relevant to the perceptual processing and encoding of a high-fidelity representation. For example, it has been shown that visual perception and the ability to form detailed representations correlate with MDT performance in older adults (Davidson et al. 2019). In addition, poorer mnemonic lure discrimination performance in older adults can be at least partially accounted for by an age-related slowing in processing speed during the study phase of the MDT (Foster and Giovanello 2020). This could lead to older adults being able to encode fewer details of an image when given the same amount of time as younger adults. It stands to reason that the construction and encoding of a high-fidelity representation at study phase and at test is a prerequisite for a correct endorsement of a lure as “New”. The current analysis focused on data during the test phase of the MDT. Therefore, the slower non-decision time reported in this work would rather indicate slower encoding processes for the test stimuli. However, it is reasonable to assume that older adults would be also slower at encoding of the stimuli during the study phase. This would put them at a time disadvantage in being able to encode high-fidelity representations, which are needed for a correct judgment in the test phase.

Study Limitations

A limitation of this study is that while the DDM identifies the component processes that contribute to a memory judgment decision as a whole, it cannot break down the distinct processes that contribute to the mnemonic ability component. The use of the Diffusion Model’s drift rate only informs whether a participant has a good mnemonic ability (indicated by being able to extract a good quality of evidence) or poor mnemonic ability (indicated by extracting a poor quality of evidence). However, the drift rate cannot inform

about what specific processes determine a good or bad mnemonic ability. Although our drift rate findings are in line with the idea of age-related changes in the hippocampus and a bias toward over-generalization, it has not been investigated whether the drift rate measure in the context of the MDT is directly linked to hippocampus function and presumable pattern separation and completion. As discussed above, other processes related to retrieval strategies and encoding of high-fidelity representations have been shown to be relevant to successful mnemonic lure discrimination and a slowing in these processes could be contributing to the poor mnemonic ability and lower drift rates in older adults. Although the slower non-decision time in older adults suggests that these processes show an age-related slowing, we are not able to identify to which individual degree the encoding and retrieval processes are slowed down because we are not able to tease the individual non-decision processes in the DDM apart. In sum, while we were able to show that the age-related decline in mnemonic lure discrimination performance is predominantly based on an age-related deficit in mnemonic ability, we cannot identify with the DDM which memory-relevant processes exactly underlie the poorer mnemonic ability in older adults. Another limitation is that we cannot make conclusions about what brain regions are involved in the age-related decline in mnemonic ability. We cannot provide any evidence that the poorer mnemonic ability in older adults is driven by changes in the hippocampus as we did not investigate brain regions or patterns of neural activity. As mentioned above it has not been investigated whether the drift rate is linked to hippocampal function and what brain regions make contributions to the drift rate in the MDT. A meta-analysis of perceptual decision-making studies identified a fronto-parietal network that was related to the drift rate (Mulder et al. 2014). However, most likely, the neural substrate of the drift rate will be specific to the task and the type of evidence that is accumulated. It is therefore not clear whether this extends to mnemonic tasks and requires further study.

Future Directions

Recently, the drift rate was suggested as a novel cognitive marker of preclinical AD due to its higher sensitivity in comparison to other more traditional performance measures. The drift rate has been shown to distinguish healthy from pathological aging. Older cognitively healthy adults with a family history of AD showed a decreased drift rate in comparison to older cognitively healthy adults who did not have a family history of AD in an episodic recognition memory task (Aschenbrenner et al. 2016). Another study found lower drift rates in memory-disordered patients in comparison to unimpaired controls using an item recognition and lexical decision task. Drift rates in mild Alzheimer's disease patients were also lower than in Mild Cognitive impairment (MCI) (Ratcliff et al. 2022). Using statistical and machine learning methods the same study found that the drift rate together with the other model parameters was 83% accurate at distinguishing memory-disordered adults from unimpaired controls. This demonstrates that the consideration of the performance in all component processes together in the Diffusion model might be particularly beneficial to the diagnostics of pathological aging. Moreover, it is likely that the relative sensitivity of the drift rate and the other diffusion model parameters to preclinical AD processes will also depend on the particular task that is administered. Considering the high sensitivity of the MDT, drift rate measures derived from this task

may provide an even more sensitive cognitive marker to identify early signs of pathological aging. Therefore, future studies applying hierarchical Bayesian diffusion modeling to the MDT in older adults at risk for AD may be very informative.

In summary, our study provides evidence that the underlying source of the age-related decline in MDT performance is predominantly a deficit in mnemonic ability in older adults rather than differences in non-mnemonic decision-making processes like response caution and response bias. Furthermore, an age-related slowing in encoding or strategic retrieval processes could be involved in this deficit. Further work is needed to identify what specific processes that are relevant to successful memory are contributing to this mnemonic ability deficit and link it with neural circuit changes with aging.

References

- Aschenbrenner AJ, Balota DA, Gordon BA, Ratcliff R, Morris JC. 2016. A diffusion model analysis of episodic recognition in preclinical individuals with a family history for Alzheimer's disease: The adult children study. *Neuropsychology* **30**: 225–238.
- Badre D, Wagner AD. 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* **45**: 2883–2901.
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. 1961. An Inventory for Measuring Depression. *Arch Gen Psychiatry* **4**: 561–571.
<https://jamanetwork.com/journals/jamapsychiatry/fullarticle/487993> (Accessed November 2, 2022).
- Bogacz R, Wagenmakers EJ, Forstmann BU, Nieuwenhuis S. 2010. The neural basis of the speed–accuracy tradeoff. *Trends Neurosci* **33**: 10–16.
- Buckner RL. 2004. Memory and Executive Function in Aging and AD: Multiple Factors that Cause Decline and Reserve Factors that Compensate. *Neuron* **44**: 195–208.
- Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. 1989. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res* **28**: 193–213.
- Cahill L, McGaugh JL, Weinberger NM. 2001. The neurobiology of learning and memory: some reminders to remember. *Trends Neurosci* **24**: 578–81.
- Cohn M, Emrich SM, Moscovitch M. 2008. Age-Related Deficits in Associative Memory: The Influence of Impaired Strategic Retrieval. *Psychol Aging* **23**: 93–103.
[/doiLanding?doi=10.1037%2F0882-7974.23.1.93](https://doi.org/10.1037%2F0882-7974.23.1.93) (Accessed November 16, 2022).
- Davidson PSR, Vidjen P, Trincao-Batra S, Collin CA. 2019. Older Adults' Lure Discrimination Difficulties on the Mnemonic Similarity Task Are Significantly Correlated With Their Visual Perception. *J Gerontol B Psychol Sci Soc Sci* **74**: 1298–1307.
- Dickey J, Lientz BP. 1970. *The Weighted Likelihood Ratio, Sharp Hypotheses about Chances, the Order of a Markov Chain*.
- Forstmann BU, Ratcliff R, Wagenmakers E-J. 2016. Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annu Rev Psychol* **67**: 641–666.

- Foster CM, Giovanello KS. 2020. Domain general processes moderate age-related performance differences on the mnemonic similarity task. *Memory* **28**: 528–536.
- Fraundorf SH, Hourihan KL, Peters RA, Benjamin AS. 2019. Aging and recognition memory: A meta-analysis. *Psychol Bull* **145**: 339–371.
- Gellersen HM, Trelle AN, Henson RN, Simons JS. 2021. Executive function and high ambiguity perceptual discrimination contribute to individual differences in mnemonic discrimination in older adults. *Cognition* **209**.
- Gelman A, Rubin DB. 1992. *Inference from Iterative Simulation Using Multiple Sequences*.
- Johnson SA, Zequeira S, Turner SM, Maurer AP, Bizon JL, Burke SN. 2021. Rodent mnemonic similarity task performance requires the prefrontal cortex. *Hippocampus* **31**: 701–716. <https://onlinelibrary.wiley.com/doi/full/10.1002/hipo.23316> (Accessed November 9, 2022).
- Kirwan CB, Stark CEL. 2007. Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory* **14**: 625–633.
- Lacy JW, Yassa MA, Stark SM, Muftuler LT, Stark CEL. 2011. Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning & Memory* **18**: 15–18. <http://learnmem.cshlp.org/content/18/1/15.full> (Accessed November 2, 2022).
- Laming DRJ. 1968. Information theory of choice-reaction times. *Academic Press*.
- Leal SL, Yassa MA. 2018. Integrating new findings and examining clinical applications of pattern separation. *Nat Neurosci* **21**: 163–173.
- Leal SL, Yassa MA. 2015. Neurocognitive Aging and the Hippocampus across Species. *Trends Neurosci* **38**: 800–812.
- Leite FP, Ratcliff R. 2011. What cognitive processes drive response biases? A diffusion model analysis. *Judgm Decis Mak* **6**: 651–687.
- McGovern DP, Hayes A, Kelly SP, O’Connell RG. 2018. Reconciling age-related changes in behavioural and neural indices of human perceptual decision-making. *Nat Hum Behav* **2**: 955–966.
- Moscovitch M, Melo B. 1997. Strategic retrieval and the frontal lobes: Evidence from confabulation and amnesia. *Neuropsychologia* **35**: 1017–1034.
- Motley SE, Brock Kirwan C. 2012. A parametric investigation of pattern separation processes in the medial temporal lobe. *Journal of Neuroscience* **32**: 13076–13084.
- Mulder MJ, van Maanen L, Forstmann BU. 2014. Perceptual decision neurosciences - a model-based review. *Neuroscience* **277**: 872–884.
- Nilsson L-G. 2003. Memory function in normal aging. *Acta Neurol Scand Suppl* **179**: 7–13.
- Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK. 2019. PsychoPy2: Experiments in behavior made easy. *Behav Res Methods* **51**: 195–203.
- Pishdadian S, Hoang N V., Baker S, Moscovitch M, Rosenbaum RS. 2020. Not only memory: Investigating the sensitivity and specificity of the Mnemonic Similarity Task in older adults. *Neuropsychologia* **149**.
- Ratcliff R. 2002. A diffusion model account of response time and accuracy in a brightness discrimination task: fitting real data and failing to fit fake but plausible data. *Psychon Bull Rev* **9**: 278–291. <https://pubmed.ncbi.nlm.nih.gov/12120790/> (Accessed November 3, 2022).
- Ratcliff R. 1978. A theory of memory retrieval. *Psychol Rev* **85**: 59–108.

- Ratcliff R, Childers R. 2015. Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making. *Decision (Wash D C)* **2015**: 237–279. <https://pubmed.ncbi.nlm.nih.gov/26236754/> (Accessed November 2, 2022).
- Ratcliff R, McKoon G. 2008. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput* **20**: 873–922. <https://pubmed.ncbi.nlm.nih.gov/18085991/> (Accessed November 3, 2022).
- Ratcliff R, Rouder JN. 1998. *MODELING RESPONSE TIMES FOR TWO-CHOICE DECISIONS*. *Psychological science* **9**(5).
- Ratcliff R, Scharre DW, McKoon G. 2022. Discriminating Memory Disordered Patients From Controls Using Diffusion Model Parameters From Recognition Memory. *J Exp Psychol Gen* **151**: 1377–1393.
- Ratcliff R, Smith PL. 2004. A comparison of sequential sampling models for two-choice reaction time. *Psychol Rev* **111**: 333–367.
- Ratcliff R, Smith PL, Brown SD, McKoon G. 2016. Diffusion Decision Model: Current Issues and History. *Trends Cogn Sci* **20**: 260–281.
- Ratcliff R, Thapar A, McKoon G. 2004. A diffusion model analysis of the effects of aging on recognition memory. *J Mem Lang* **50**: 408–424.
- Ratcliff R, Tuerlinckx F. 2002. Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review* **2002 9:3 9**: 438–481. <https://link.springer.com/article/10.3758/BF03196302> (Accessed November 8, 2022).
- Reagh ZM, Noche JA, Tustison NJ, Delisle D, Murray EA, Yassa MA. 2018. Functional Imbalance of Anterolateral Entorhinal Cortex and Hippocampal Dentate/CA3 Underlies Age-Related Object Pattern Separation Deficits. *Neuron* **97**: 1187–1198.e4.
- Rosenzweig-Lipson S, Barton R, Gallagher M, Mohs R. 2021. HOPE4MCI trial: First trial targeting reduction of hippocampal overactivity to treat mild cognitive impairment due to Alzheimer’s disease with AGB101. *Alzheimer’s & Dementia* **17**.
- Sinha N, Berg CN, Tustison NJ, Shaw A, Hill D, Yassa MA, Gluck MA. 2018. APOE ε4 status in healthy older African Americans is associated with deficits in pattern separation and hippocampal hyperactivation. *Neurobiol Aging* **69**: 221–229. <https://pubmed.ncbi.nlm.nih.gov/29909179/> (Accessed November 2, 2022).
- Spaniol J, Madden DJ, Voss A. 2006. 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. /record/2006-01955-009 (Accessed November 3, 2022).
- Spaniol J, Voss A, Grady CL. 2008. Aging and emotional memory: cognitive mechanisms underlying the positivity effect. *Psychol Aging* **23**: 859–872. <https://pubmed.ncbi.nlm.nih.gov/19140656/> (Accessed November 3, 2022).
- Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, Aisen P. 2014. The A4 study: stopping AD before symptoms begin? *Sci Transl Med* **6**: 228fs13.
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. 2014. *The deviance information criterion: 12 years on*.
- Spiegelhalter DJ, Freedman LS, Parmar MKB. 1994. *Bayesian Approaches to Randomized Trials*. <https://www.jstor.org/stable/2983527>.
- Stanislaw H, Todorov N. 1999. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers* **1999 31:1 31**: 137–149. <https://link.springer.com/article/10.3758/BF03207704> (Accessed November 3, 2022).

- Stark SM, Kirwan CB, Stark CEL. 2019a. Mnemonic Similarity Task: A Tool for Assessing Hippocampal Integrity. *Trends Cogn Sci* **23**: 938–951.
- Stark SM, Kirwan CB, Stark CEL. 2019b. Mnemonic Similarity Task: A Tool for Assessing Hippocampal Integrity. *Trends Cogn Sci* **23**: 938–951.
- Stark SM, Stevenson R, Wu C, Rutledge S, Stark CEL. 2015. Stability of age-related deficits in the mnemonic similarity task across task variations. *Behavioral neuroscience* **129**: 257–268.
- Thapar A, Ratcliff R, McKoon G. 2003. A diffusion model analysis of the effects of aging on letter discrimination. *Psychol Aging* **18**: 415–429. /record/2003-07824-007?doi=1 (Accessed November 3, 2022).
- Trelle AN, Henson RN, Green DAE, Simons JS. 2017. Declines in representational quality and strategic retrieval processes contribute to age-related increases in false recognition. *J Exp Psychol Learn Mem Cogn* **43**: 1883–1897.
- Van der Linden MU de L-UI> D des sciences cognitives > P cognitive, Meulemans TU de L-UI> D des sciences cognitives > N, Marczewski PU de L-UI> S cognitives > N, Collette FU de L-UI> D des sciences cognitives > N. 2009. The relationships between episodic memory, working memory, and executive functions: The contribution of the prefrontal cortex. *Psychol Belg* **40**. <https://orbi.uliege.be/handle/2268/18821> (Accessed November 17, 2022).
- Vandekerckhove J, Tuerlinckx F. 2007. Fitting the ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review* **14**: 1011–1026. <https://link.springer.com/article/10.3758/BF03193087> (Accessed November 2, 2022).
- Vandekerckhove J, Tuerlinckx F, Lee MD. 2011. Hierarchical diffusion models for two-choice response times. *Psychol Methods* **16**: 44–62.
- Voss A, Rothermund K, Voss J. 2004. Interpreting the parameters of the diffusion model: An empirical validation. *Mem Cognit* **32**: 1206–1220.
- Voss A, Voss J. 2007. *Fast-dm: A free program for efficient diffusion model analysis*.
- Wagenmakers EJ. 2009. Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology* **21**: 641–671.
- Wais PE, Montgomery O, Stark CEL, Gazzaley A. 2018. Evidence of a Causal Role for mid-Ventrolateral Prefrontal Cortex Based Functional Networks in Retrieving High-Fidelity Memory. *Sci Rep* **8**.
- Weintraub S, Wicklund AH, Salmon DP. 2012. The neuropsychological profile of Alzheimer disease. *Cold Spring Harb Perspect Med* **2**.
- White CN, Ratcliff R, Vasey MW, McKoon G. 2010. Using diffusion models to understand clinical disorders. *J Math Psychol* **54**: 39–52.
- Wickelgren WA. 1977. Speed-accuracy tradeoff and information processing dynamics. *Acta Psychol (Amst)* **41**: 67–85.
- Wilson IA, Gallagher M, Eichenbaum H, Tanila H. 2006. Neurocognitive aging: prior memories hinder new hippocampal encoding. *Trends Neurosci* **29**: 662–670. <https://pubmed.ncbi.nlm.nih.gov/17046075/> (Accessed November 2, 2022).
- World Health Organization. 2021. Ageing and health. *Ageing and Health Factsheet*. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> (Accessed November 2, 2022).
- Yassa MA, Lacy JW, Stark SM, Albert MS, Gallagher M, Stark CEL. 2011a. Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus* **21**: 968–79.

Yassa MA, Mattfeld AT, Stark SM, Stark CEL. 2011b. Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proc Natl Acad Sci U S A* **108**: 8873–8878.

Yassa MA, Stark CEL. 2011. Pattern separation in the hippocampus. *Trends Neurosci* **34**: 515–25.

2022. 2022 Alzheimer's disease facts and figures. *Alzheimer's and Dementia* **18**: 700–789.

Parsing Memory and Non-Memory Contributions to Age-Related Declines in Mnemonic Discrimination Performance: A Hierarchical Bayesian Diffusion Decision Modeling Approach

Chwiesko et al.

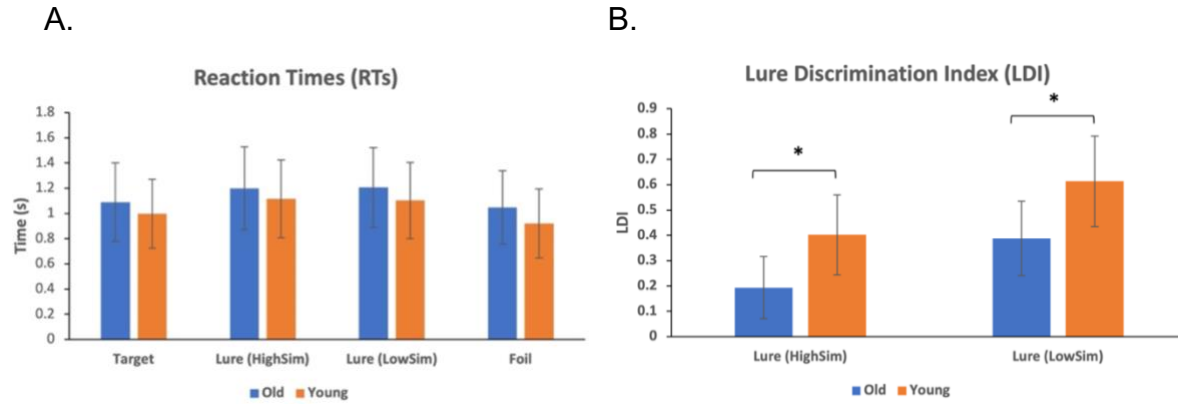
SUPPLEMENTARY MATERIALS

Model	Model structure	DIC
Model 0	$\alpha (g), \beta (g), \tau (g), \delta (g, c)$	10688.25
Model 1	$\alpha (g), \beta (g), \tau (g), \delta (g, c, s)$	9567.33
Model 2	$\alpha (g), \beta (g), \tau (g, s), \delta (g, c, s)$	6573.04
Model 3	$\alpha (g), \beta (g, s), \tau (g, s), \delta (g, c, s)$	6496.65
Model 4	$\alpha (g, s), \beta (g, s), \tau (g, s), \delta (g, c, s)$	6390.63
Model 5	$\alpha (g, s), \beta (g, s), \tau (g, s, t), \delta (g, c, s)$	2553.64
Model 6	$\alpha (g, s), \beta (g, s, t), \tau (g, s, t), \delta (g, c, s)$	2344.74
Model 7 (Full model)	$\alpha (g, s), \beta (g, s, t), \tau (g, s, t), \delta (g, c, s, t)$	438.61

Supplementary Table S1. Models of different complexity (number of hierarchical levels (g: Group, c: Condition, s: Subject, t: Trial) and permitted variation of model parameters (α : Boundary separation, β : Response Bias, τ : Non-decision time, δ : drift-rate) across these levels) that were fitted to the data and their Deviance Information Criteria (DICs). Based on the lowest DIC, Model 7 was chosen for analysis.

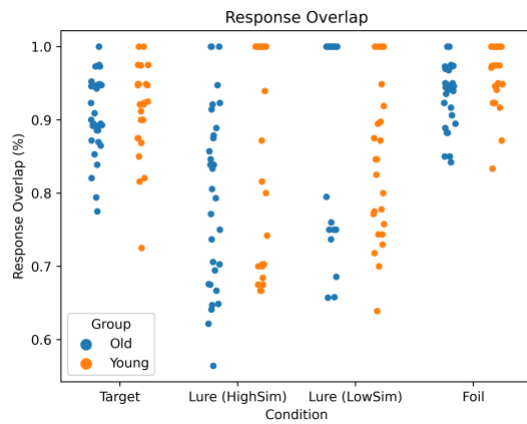
	OLD			YOUNG			Group difference		
	Full Model	Fastdm	ChiSquare	Full Model	Fastdm	ChiSquare	Full Model (BF)	Fastdm (p-value)	ChiSquare
Response caution (α)	1.43 (1.37,1.49)	1.56	0.19	1.42 (1.37,1.47)	1.5	0.18	0.08	0.12	0.07
Response bias (β)	0.48 (0.46, 0.5)	0.49	0.52	0.48 (0.45,0.52)	0.5	0.5	0.16	0.31	0.2
Encoding/ Motor Response (τ)	0.80 (0.75,0.84)	0.75	0.82	0.71 (0.67,0.77)	0.68	0.74	4.44	<0.01	<0.01
Discrimination ability (δ)									
Target	-2.35 (-2.69,-2.07)	-2.24	-0.45	-2.56 (-2.99,-2.23)	-3.15	-0.56	0.23	<0.05	<0.01
Lure (HighSim)	-0.91 (-1.18,-0.68)	-0.11	0	0.00 (-0.29,0.28)	-0.91	-0.15	>10 000	<0.001	<0.001
Lure (LowSim)	0.06 (-0.17,0.3)	0.67	0.11	1.05 (0.76,1.37)	-0.08	-0.01	>10 000	<0.001	<0.001
Foil	3.3 (2.96,3.93)	1.91	0.36	3.85 (3.45,4.61)	2.36	0.4	0.94	<0.05	0.14

Supplementary Table S2. Estimated group means for Diffusion Decision Model Parameters based on our Hierarchical Bayesian model (Full model) using Markov-chain Monte Carlo (MCMC), and two standard fitting methods -the publicly available fast-dm package (outliers removed) and the two chi-square fitting method. Group differences in Model parameters are evaluated based on the Bayes factor (Savage-Dickey Density ratio) for our Full Model, and t-test for independent groups for the latter two fitting methods. The last two columns list the p-values for the t-tests. The group differences in the non-decision time (τ), and the drift rate parameter for the two lure conditions (δ (HighSim) and (δ (LowSim)) that we found evidence for in our Hierarchical Bayesian model, were also found using the fast-dm and the two chi-square fitting methods. However, note that the exact values of the group mean estimates differ across the three fitting methods. For a discussion and analysis of the two standard fitting methods mentioned here, and others, including a Hierarchical Bayesian method, and how they compare in their performance of recovering model parameters see Ratcliff and Childers (2015). This work showed that for a small number of 40 observations per condition, the Hierarchical Bayesian method provided superior parameter recovery in contrast to the standard two chi-square method. The authors concluded that a Hierarchical Bayesian model approach is the method of choice when the number of observations is small.

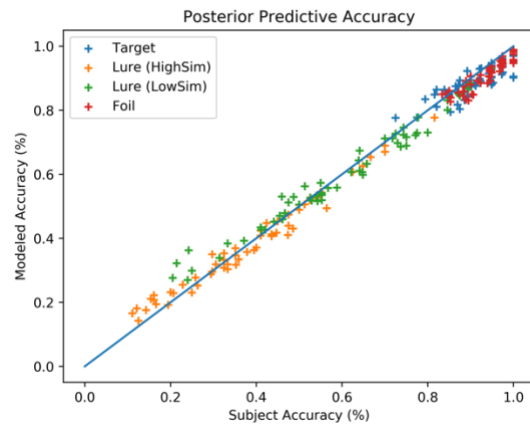


Supplementary Figure S1. A. Reaction Times for the younger and older adults group and each task condition. Error bars show the standard deviation. B. Lure Discrimination Index for younger and older adults for the two lure conditions. The LDI assesses lure discrimination performance by calculating the difference between the probability of giving a “similar” response to the lure items minus the probability of giving a “similar” response to the foils. Groups were compared using an independent t-test. * $p < 0.0001$

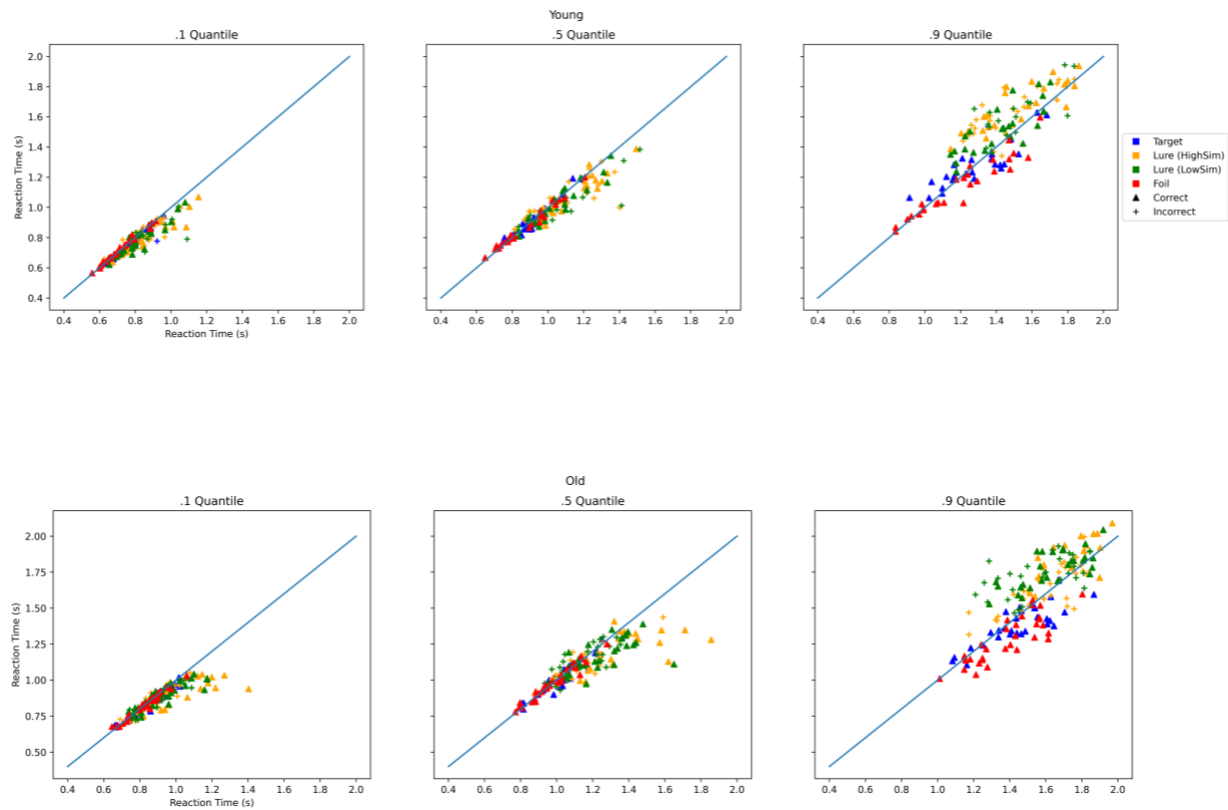
A.



B.



Supplementary Figure S2. Quality check of model performance. A. The modeled and observed accuracy for the Reduced Model. The modeled accuracy data fit the observed accuracy data well. Each cross represents the accuracy of one participant averaged across all trials within each condition. B. Overlap between the modeled and observed response type ("Old"/ "New"). Individual data points represent the percentage of agreement in response type between the observed and modeled dataset for each participant and condition.



Supplementary Figure S3. Quality check of model performance. Overlap between the modeled and observed 0.1, 0.5, and 0.9 quantiles of the modeled and observed reaction time distributions for the Reduced Model. Each data point represents the RT of one participant per condition and response type (“Old”/ “New”).