Optimizing the Color Shapes Task for Ambulatory Assessment and Computational Cognitive Feature Extraction via Drift Diffusion Modeling

Sharon H. Kim[1,2,7], Jonathan G. Hakun[3,4,5,6], Yanling Li[7], Karra D. Harrington[6], Daniel B. Elbich[3], Martin J. Sliwinski[1,6], Joachim Vandekerckhove[8], and Zita Oravecz[1,2,6,7]

[1]Department of Human Development and Family Studies, The Pennsylvania State University, University Park, Pennsylvania, USA
[2]Department of Social Data Analytics, The Pennsylvania State University, University Park, Pennsylvania, USA
[3]Department of Neurology, The Pennsylvania State University, College of Medicine, Hershey, Pennsylvania, USA
[4]Department of Psychology, The Pennsylvania State University, University Park, Pennsylvania, USA
[5]Department of Public Health Sciences, The Pennsylvania State University, College of Medicine, Hershey, Pennsylvania, USA
[6]Center for Healthy Aging, The Pennsylvania State University, University Park, Pennsylvania, USA
[7]Quantitative Developmental Systems Methodology Core, The Pennsylvania State University, University Park, Pennsylvania, USA
[8]Department of Cognitive Sciences, University of California, Irvine, Irvine, California, USA

Abstract

Background: Recent advances in cognitive digital assessment methodology, including high-frequency, ambulatory assessments, have shown promise to improve the detection of subtle cognitive changes. The use of computational modeling approaches may further improve the sensitivity of the digital cognitive assessments to detect subtle cognitive changes by capturing features that reflect core cognitive processes from non-cognitive/non-decision-related processes.

Objective: We explored the validity of a brief, smartphone-based adaptation of an associative visual working memory change detection task that has shown sensitivity in the early detection of preclinical AD-related cognitive impairment. We aimed to optimize the task for computational cognitive feature extraction with drift diffusion modeling.

Methods: We analyzed data from a sample of 68 participants (69% women; 81% White; age range 24-80 years, mean 49 years, SD 14), who completed 60 trials for each of the 16 variations of a Visual Working Memory Binding task (‘Color Shapes’) on smartphones, over an 8-day period. A drift diffusion model was fit to Color Shapes response time and accuracy data to dissociate features of the decision-making process. We experimentally manipulated three properties of the Color Shapes task (study time, probability of change, choice urgency) to test how these constraints yield differences in key computational features (rate of evidence accumulation process, initial bias towards a response option,

caution in decision making). We also evaluated how an additional task property, the test array size (whole display vs. single probe), impacted responses across all conditions.

Results: Overall accuracy was high (>80%) and minimal missingness (1.3% of expected data) was observed. Three primary observations were made regarding the task property manipulations and drift diffusion model fit: (1) increasing the probability of change was associated with higher 'initial bias' toward a "different" response, (2) increasing the choice urgency during the test phase was associated with decreased caution in decision-making ('boundary separation'), and (3) contrary to expectation, longer study times did not affect evidence accumulation rate ('drift rate'). In addition, we observed that individual differences in evidence accumulation rate (drift rate) and caution (boundary separation) were associated with age.

Conclusions: We identified a version of the Color Shapes task that is ideal for smartphone-based, repeated cognitive assessments in real-world settings, especially when the resulting data are analyzed through computational cognitive modeling. Our proposed approach can advance the development of tools and programs for an efficient and effective early detection and monitoring of early risk for Alzheimer's disease.

Keywords: subtle cognitive decline; computational cognitive markers; smartphone-based cognitive testing; drift diffusion model; Bayesian multilevel modeling

Correspondence concerning this article should be addressed to Sharon Ha Eun Kim, 235 Health and Human Development Building, University Park, PA, 16802 or by e-mail at skim@psu.edu / shaeunkim@gmail.com.

Introduction

Cognitive testing is widely used to diagnose cognitive deficits from various conditions, including Alzheimer's Disease (AD)[1]. Many of these conditions have a long preclinical phase, characterized by cognitive changes that are more subtle than what most cognitive diagnostic tests are designed to detect[2,3]. Ratcliff and McKoon[1] have suggested that cognitive (and neuropsychological) testing could capitalize on advances in cognitive modeling in order to improve such cognitive assessments. At the same time, repeated ambulatory cognitive assessments have also shown promise for improving assessment accuracy.[4] Here, we test a smartphone-based implementation of a cognitive task that has the potential to detect cognitive changes present in the early phases of AD. Our proposed approach relies on extracting features via state-of-the-art computational cognitive modeling from repeated ambulatory assessments cognitive performance data.

Recent evidence suggests sensitive measures of subtle cognitive change may be achieved from high-frequency cognitive assessment study designs,[5] especially when combined with computational modeling.[6] For example, high-frequency ecological momentary assessments (EMA)[7] of cognitive functioning across days produces a rich data stream across time and natural daily contexts,[8] providing a representative sample of daily cognition. These repeated assessments in natural contexts also mitigate reliability concerns,[5] while preserving ecological validity,[9] and permit the implementation of computational methods to effectively explore subtle within-person cognitive processes.[10] Computational modeling can identify subtle latent processes underlying manifest cognitive performance data and explore individual differences therein to identify patterns of risk for AD.[11]

When exploring latent cognitive processes, a number of cognitive manifestations of early AD-related pathological processes have been identified, including deficits in working memory "binding" abilities.[12,13] In this project, we tested a mobile adaptation of a visual working memory binding task developed by Parra et al.,[14] We refer to this task here as the "Color Shapes" task.[4] Previous implementations of the Color Shapes task for remote and mobile administration have shown promise in generating comparable scores as in-person testing,[15,16] and differentiating amongst potential AD risk factors such as mild cognitive impairment,[17] AD-biomarker status,[18] and social.[19] However, these studies focused on analyzing aggregated mean performance data across all administrations and did not explore the cognitive processes through computational modeling.

We propose that computational modeling of repeated administrations of Color Shapes task performance data can further enhance our ability to capture subtle cognitive change processes. Computational modeling[20,21] offers tools to quantify the subtle cognitive processes underlying observed performance behavior from high-density data streams. Parameters of computational models map onto substantive constructs of interest. Additionally, this approach can formally assess the validity of the model by evaluating fit with observed data.[11] One such model, the drift diffusion model (DDM)[22], characterizes decision-making as a noisy process of evidence accumulation towards response options, and can disentangle into hypothesized generative cognitive processes. The DDM captures the latent processes underlying observed performance of response time and accuracy.[1,23,24]

By fitting the DDM to response time and accuracy data, we can disentangle subtle underlying cognitive processes such as rate of evidence accumulation, caution in decision making, and response biases.[22,25] For instance, individual differences in the speed of evidence accumulation and caution can elucidate important idiosyncrasies in decision making so that we can better identify those at risk for AD.[26-28]

Objectives

The aim of the current study was twofold: (1) to conduct an experimental validation of the DDM analytic approach in the context of RT and accuracy data from the Color Shapes task; (2) to identify a Color Shapes task version that is most optimal for DDM-based cognitive feature extraction in real life settings. Overall, we propose that the DDM-based computational modeling approach can extract useful signal that can serve as novel digital cognitive markers of risk for subtle cognitive decline. Our study intended to lay the groundwork for this approach and illustrate the corresponding statistical tools for data analysis.

Methods

Participant Recruitment

We recruited 69 adults from the Pennsylvania area (47 females, 69%; 21 males, 31%) via online strategies (e.g., enrolled in ResearchMatch, a non-profit NIH-funded program for medical study participation). Participants were screened via telephone to assess eligibility. Eligible participants were over 18, fluent in English, had access to reliable internet connection, did not have a motor or visual impairment that would interfere with operating a smartphone, or a history of neurological injury or disease (e.g., stroke, seizures). All participants provided written informed consent prior to participation in the study. After an initial remote onboarding session (conducted via video conference), they were asked to complete cognitive tasks (Color Shapes trials, see below) for 8 days for approximately 20 minutes each day. Participants were compensated up to $100 for completion. This study was approved by the Pennsylvania State University's Institutional Review Board. More information on Design and Procedure are provided after a brief description of the task and the computational modeling framework chosen to analyze the data.

Measures

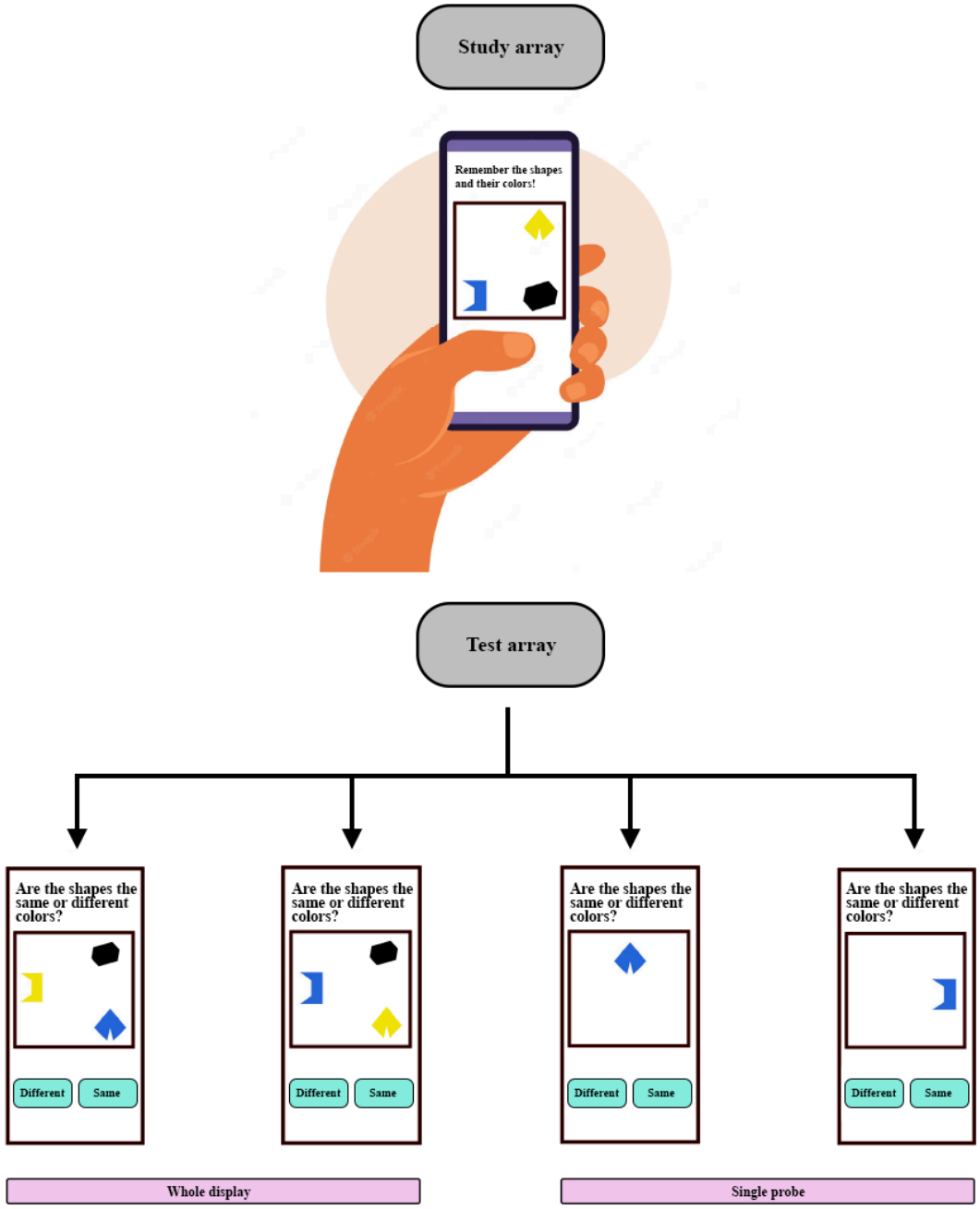
Demographics. Participants responded to gender, age, race/ethnicity, ethnic status, and level of education. Participants were asked whether they had ever tested positive for COVID-19 (data were not analyzed). Basic health information was provided on ResearchMatch upon sign up (e.g., BMI, medical conditions, medication, etc.).

The Color Shapes task. The Color Shapes task is a visual array change detection task where participants were asked to determine whether the combination of features (color and shape) among visual objects distributed throughout the array change between study and test arrays. In our study, participants were asked to study a set of three abstract shapes, each with a unique color. After a brief 900ms delay, they were presented with a test array and asked to determine if the stimuli contained in the test array contained the

same combination of shape and color as the study array. To identify an optimal combination of task parameters for mobile administration and drift diffusion modeling, we experimentally manipulated four features of the task in a within-persons 2 [study duration] × 2 [probability of change] × 2 [response duration] × 2 [probe type] full factorial experimental design. Specifically, each trial of the Color Shapes task began with a fixation cross displayed for 500ms after which 3 colored abstract shapes appeared for either a short (500ms) or longer (2000ms) period of time (study duration). Participants were asked to study the specific combinations of shape and color presented in the study array objects. After a brief 900ms delay period, either 3 colored shapes or 1 colored shape reappeared (probe type) at different locations throughout the test array, as illustrated in Figure 1. The colored shapes at test either maintained the same combinations of color and shape (a "SAME" trial) or 2 of the shapes swapped colors (a "DIFFERENT" trial). Participants indicated whether they believed the combination of shapes and colors presented in the test array match the combinations presented in the study array by pressing SAME or DIFFERENT buttons located at the bottom of the screen. Test array stimuli remained on the screen for either a short (3000ms) or long (10000ms) period of time (response duration). Each trial has either a 50%/50% chance or 20%/80% chance (probability of change) of being a SAME/DIFFERENT trial.

Debrief Survey. Upon completing the final session, participants completed the exit survey consisting of three items. First, participants were asked how typical their routine, activities, and experiences were during the 8 study days on a 5-point scale, ranging from 1=very unusual, unusual, neutral, typical, 5=very typical, with a prefer not to respond option). Second, participants were asked whether they were impacted by any unusual circumstances or stressful events over the study duration. There were 13 response options ranging from nothing unusual to various stressful events (e.g., negative social interaction, health issues, financial issues, world events, etc.). Lastly, participants were asked "over time, did you develop any strategies for better performance on the brain games?" (yes/no). If they responded yes, they were asked to provide more details about the strategies they used. This was an unstructured free response question. We only analyzed responses from the third item on cognitive strategy use in the present study.

Figure 1. Study and test arrays of the Color Shapes task



An illustration of the Color Shapes task showing the study array and test array with four configurations from left to right: 1) DIFFERENT trial with whole display (3 colored shapes) shown in the test array, 2) SAME trial with whole display (3 colored shapes) shown in the test array, 3) DIFFERENT trial with single probe (1 colored

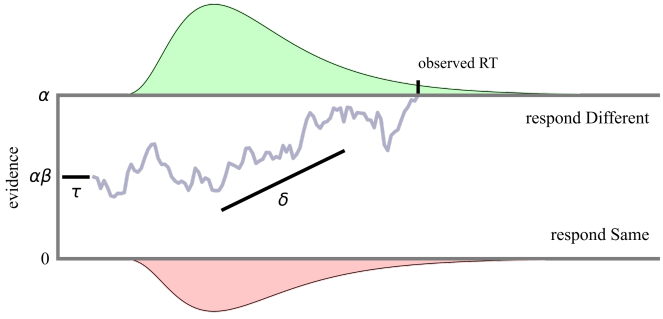
shape) shown in the test array, and 4) SAME trial with single probe (1 colored shape) shown in the test array.

Computational modeling with the Drift Diffusion Model

The DDM falls under the family of sequential sampling models with continuous time and continuous evidence (Ratcliff, 1978). A graphical illustration of the model is shown in Figure 2. This model assumes that each decision-making task starts a stochastic information accumulation process of incoming evidence towards one of the two response options, indicated by upper and lower bounds in Figure 2, labeled with the 'respond Different' and 'respond Same' response options of the Color Shapes task. The distance between these bounds (the boundary separation) captures the threshold amount of evidence needed for a response to be chosen. Mathematically, the information accumulation process is modeled as a one-dimensional or Wiener diffusion process that terminates at the absorbing response boundary. With evidence towards the upper boundary represented by positive values and lower boundary represented by negative values, decisions are based on this Wiener diffusion process in continuous time as a single total that integrates both antagonistic types of evidence.

The four core parameters of the DDM are presented in Table 1 (adapted from [22,29]). Formally, the drift rate (δ) describes the speed of information accumulation by capturing the average amount of evidence that is accumulated from the stimulus over a small time period, the boundary separation (α) describes the level of evidence required to make a decision (capturing speed-accuracy trade-off), the initial bias (β) describes the starting status towards one or the other boundary prior to the evidence accumulation process, and the non-decision time (τ) represents the component of response time that is not related to the decision-making process. When an individual is presented with 3 colored shapes in the study array across repeated trials, the DDM models the evidence accumulation process toward a DIFFERENT or SAME decision using the individual's initial bias, non-decision time, boundary separation, and evidence accumulation rate. More details on the mathematical formulation of the model are provided in Appendix A. Given a participant's choice response time data, we can estimate these parameters and interpret them in the context of the Color Shapes task.[29]

Figure 2. An illustration of the drift diffusion model



In a decision-making process, evidence is accumulated over time at an average drift rate of δ. The decision process terminates if the cumulative evidence value reaches 0

(lower boundary for ‘respond Same’) or α (upper boundary for ‘respond Different’). At the onset of a trial, the decision process can be biased due to the amount of preliminary evidence given by αβ. The non-decision time τ reflects the duration of perceptual processes such as stimulus encoding and response execution, which occur outside the actual decision-making process. Equation 1 in the Appendix describes the response time distributions that follow from these model assumptions. Figure generated via [30] .

Table 1. Summary of the parameters of the Drift Diffusion Model

Symbol	Parameter	Definition	Interpretation
δ	Drift rate	Average quality of the stimulus and information accumulation process	Higher δ indicates faster accumulation of evidence
β	Initial bias	Starting point bias for either response	$\beta > 0.5$ indicates bias towards DIFFERENT response
α	Boundary separation	Distance quantifying evidence required to make a decision; Speed-accuracy trade-off	Higher α indicates more caution (higher accuracy)
τ	Non-decision time	Motor response time, encoding time	High τ indicates slow encoding and/or motor response

Design and Procedure

After providing informed consent and enrolling in the study, participants were sent smartphones configured with the Color Shapes cognitive assessments. Next, a remote meeting with a study administrator was scheduled where they were onboarded onto the study and trained on the study protocol (e.g., smartphone instructions). Over the next 8 days, participants were instructed to complete a self-initiated daily session consisting of overall 16 versions of the Color Shapes task (see description below), with each daily session taking under 20 minutes. Two versions of the task were completed once per day over eight days, and the order of the presented 16 versions were drawn randomly from 3 different schedules. Each version consisted of 60 ultra-brief trials per task, with a maximum of 960 total possible trials. The participants could only complete the testing once per day, the testing had to be completed in a single setting, which timed out after 30 minutes. Reminder alerts were provided every hour until 9 pm. Upon completing the final session, participants completed a final exit survey on the smartphone. Afterwards they were debriefed by a study coordinator via video conference and were provided with instructions to return the smartphone.

We experimentally manipulated features of the Color Shapes task with two objectives in mind: to test the sensitivity for capturing changes in terms of the latent

parameters of the DDM, and to compare the 3-colored-shapes test array and 1-colored-shape test array versions of the task shown in Figure 1. Specifically, we manipulated across four task features each with two levels:

(1) short (500ms) vs. long (2000ms) study array time,
(2) low (50% DIFFERENT trials and 50% SAME trials) vs. high (80% DIFFERENT trials and 20% SAME trials) proportion of DIFFERENT trials, capturing different probabilities of change in the properties of the shapes from the study array to the test array,
(3) minimal choice urgency (100000ms to respond) vs. high choice urgency (3000ms to respond), and
(4) whole display (3 colored shapes in the test array) vs. single probe (1 colored shape in the test array).

Manipulations (1)-(3) aimed at influencing parameters of the drift diffusion model: (1) drift rate [δ] with the study array, (2) initial bias [β] with the different probabilities of change, and (3) boundary separation [α] with the different time urgency to respond. The non-decision time [τ] was estimated for each person but was not manipulated. An illustration of these three cognitive task manipulations is shown in Figure 3 for the single probe trials.

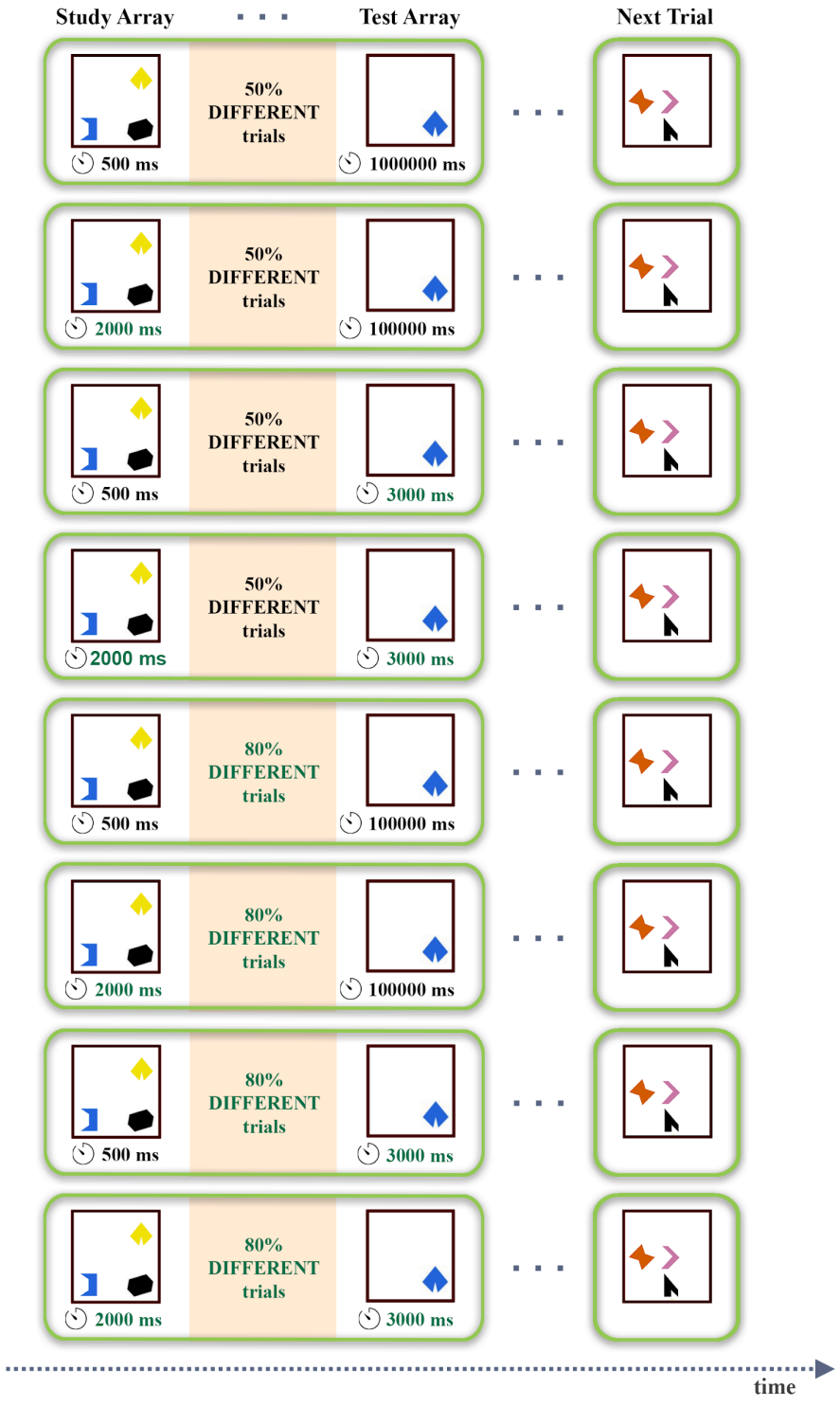
We expected the following effects: 1) faster drift rates with longer study times, that is more time for encoding leading to faster accumulation of evidence, 2) initial bias towards DIFFERENT responses with higher probability of DIFFERENT trials, and 3) lower (closer) boundary separation, that is less caution in decision-making with increased urgency to choose a response. With feature (4), we studied how the first three manipulation would work with two visually different versions of the task, where the test array with single probe version has the additional advantage of reducing unwanted strategy use. The experimental manipulations are summarized in Table 2.

Table 2. Drift diffusion model experimental manipulations

Name	Condition	Parameter	Levels
ST	Study time	Drift rate (δ)	500ms study time (*Short study time*)
			2000ms study time (*Long study time*)
PC	Probability of change	Initial bias (β)	50% DIFFERENT / 50% SAME (*Low change*)
			80% DIFFERENT / 20% SAME (*High change*)
CU	Choice urgency	Boundary separation (*a*)	100000ms response time window (*Minimal urgency*)
			3000ms response time window (*High urgency*)
Probe	Probe type	–	Test array with 3 colored shapes (*Whole display*)

Test array with 1 colored shape (*Single probe*)

Figure 3. Eight out of 16 Color Shapes task versions



An illustration of eight out of 16 Color Shapes task versions (all with test arrays with 1 colored shapes) with example of DIFFERENT trial types shown here. Manipulations for task arrays are emphasized in green font and baseline conditions in black font. Versions have been ordered to flow from least manipulated to most manipulated versions.

Data Processing. We set the RT-based outliers attending to the real-world constraints of the Color Shapes task. The DDM was intended to assess decision-making for fast one-shot processes. Trials with RTs faster than 200ms and slower than 7000ms were excluded. Ultra-fast RTs suggested technical errors and long RTs was assessed as not following speeded task instructions or not paying attention to the task.

Data Analysis. We fit a tailored, multilevel DDM to response time and accuracy to derive latent cognitive parameters of interest. The multilevel DDM was cast in a Bayesian framework. We estimated drift rates, initial bias, and boundary separation parameters for each condition and individual, and non-decision time parameters for each individual. Finally, the response time data was scaled from milliseconds to seconds for the analysis. A detailed description of our multilevel drift diffusion model is provided in Appendix A. All analysis scripts and data are provided on OSF.[31]

Results

Participant characteristics

We recruited 69 participants for the study, but excluded 1 participant due to missing 99.1% of data (9/960 trials completed). Our final sample size therefore is N = 68, and demographics for these participants are displayed in Table 3. Participants were on average 49 years of age, 69% (47/68) were women, 31% were men (21/68), and none identified differently. Participants' ages spanned across younger to older adulthood (14 years *SD*; range 24-80 years). Regarding race, 81% (55/68) identified as White, 4% (3/68) as Black or African American, 6% (4/68) as Asian, 1.5% (1/68) American Native or Alaska Native, 1.5% (1/68) as Native Hawaiian or Pacific Islander, 3% (2/68) as multi-ethnic, and 3% (2/68) as other. As for ethnicity, 96% (65/68) identified as non-Hispanic and 4% (3/68) identified as Hispanic. In terms of highest education received, 5% (3/68) reported high school, 14% (9/68) reported vocational or some college, 12% (8/68) received an Associate's degree, and 68% (46/68) reported at least a Bachelor's degree or higher.

Descriptive Statistics

41 participants completed all possible 960 trials (16 versions by 60 trials). 22 participants completed at least 75% of the trials. Four participants completed 63% and only one participant completed only 25% of the trials. More detailed trial-level summary data are shown in Table 4.

There was 1.22% missingness in the Color Shapes task trial responses (n=730). Overall, only 0.09% of trials (n=54) were excluded using pre-determined RTs thresholds (e.g., unrealistically fast RTs<200ms or extended RTs>7000ms). This RTs threshold was used to assess the quality of data and potential non-engagement with the task.

Table 5 shows that accuracy rates by test array condition (single probe vs. whole display). The mean overall accuracy rate of correct trials among all trials was 85.83%. Accuracy rate was higher in the whole display (91.57%) than in the single probe (80.13%) condition.

Table 3. Demographics characteristics of the sample (*N*=68)

Variable	Category	Value	Percent
Gender	Woman	47	69
	Man	21	31
Race	White	55	81
	Black/African American	3	4
	Asian	4	6
	American Native/Alaska Native	1	1.5
	Native Hawaiian/Pacific Islander	1	1.5
	Multi-ethnic	2	3
	Other	2	3
Ethnicity	Non-Hispanic	65	96
	Hispanic	3	4
Education	High school	3	5
	Vocational/Some college	9	14
	Associate's degree	8	12
	Bachelor's degree	19	29
	Post-bachelor	2	3
	Master's degree	19	29
	Doctoral degree	6	9
Age	Mean	49	
	SD	14	
	Min-Max	24-80	

Group-level summaries of DDM parameters

Summaries of DDM parameter estimates are shown for the condition that was aimed for their manipulations in Table 6. Results are separated for the different Probe Type test array conditions: whole display (WD) and single probe (SP). Mean and 95% credible interval (CI) estimates are based on the posterior distribution of the corresponding parameters. The CI is a Bayesian statistic that can quantify the uncertainty that the true parameter value lies within the CI. A 95% CI indicates that the interval has a 95% probability of containing the true value. If 0 is included in the 95% CI, we conclude there is a 95% probability that the true parameter value lies near 0. For clarity, findings are presented similarly in structure across the three DDM manipulations: 1) drift rate is displayed with the two conditions related to Study Time, 2) initial bias is shown with Probability of Change conditions, and 3) boundary separation is shown with Choice Urgency conditions.

For drift rate, higher absolute δ values represent faster accumulation of evidence. By definition, the initial bias ranges between 0 and 1, with a value of 0.5 representing no bias. For the ease of regression model (initial bias estimates were regressed on task conditions), we worked with logit transformed initial bias estimates (see details in Appendix A), which meant that positive estimated β values expressed bias towards DIFFERENT response, while negative β values for SAME response. When the corresponding 95% CI contained 0, we concluded that there was no credible bias in either way, as for example, as expected, with the low change conditions in both probe types. Finally, as the boundary separation parameter by definition could only take positive values, we applied a log-transformation to this parameter to ease regression modeling. Higher α values simply relate to higher response boundaries in this scale as well, just like in the original scale.

Table 4. Summary on participant engagement during the study

Completion status	Trials completed	Versions completed	Participant *N*	Percent
100%	960	16	41	60%
94%	900	15	2	3%
88%	840	14	13	19%
75%	720	12	7	10%
63%	600	10	4	6%
25%	240	4	1	2%

Table 5. Accuracy rates across all trials and between test array probe types

Probe Type	Correct Trials	Total Trials	Accuracy
Overall	50,623	58,978	85.83%
Whole display	26,945	29,427	91.57%
Single probe	23,678	29,551	80.13%

While we can see clear differences in Table 6 between the conditions in both probe types, we need to test whether these differences are credible. Therefore, for the manipulated features listed in (1)-(3), we estimated contrast parameters by taking the difference between each condition (Manipulation - Baseline) to examine differences on expected cognitive processes. Specifically, we estimated 3 contrast parameters: *Study-Time-Contrast* (difference in long study time - short study time), *Probability-of-Change-Contrast* (high change - low change), and *Choice-Urgency-Contrast* (high urgency – minimal urgency). The Bayesian framework allowed us to derive these contrasts as regular model parameters with posterior distributions, which allows for principled testing of our hypothesis. Table 7 summarizes the estimates for these below. All parameters were estimated simultaneously within a single-step model estimation process. The strength of this approach is the correct

carryover of uncertainty and avoiding bias in sequential estimations. We discuss results for each DDM parameter manipulation next.

Study Time on Drift Rates

First, we examined the differences in drift rate estimates (δ) between the study time conditions, broken down by probe type conditions (upper part of Table 7). For the posterior means of the Study-Time-Contrast parameter (difference in long study time – short study time), positive difference values indicate faster drift rates in the manipulated long study time (2000ms) than in the control short study time (500ms), while negative difference values indicate slower drift rates in the manipulated condition compared to the baseline condition. As can be seen from Table 7, the drift rate was credibly higher in the control condition in the WD probe version of the task. While the manipulation had no credible effect in the SP version of the task.

Table 6. Posterior summaries of the DDM parameters per condition

Parameter	Label	Condition	Probe type	Posterior mean	95% CI	
					Lower	Upper
Drift rate (δ)	*ST+*	*Long study time*	WD	1.36	1.13	1.57
			SP	1.04	0.70	1.69
	ST-	Short study time	WD	1.62	1.36	1.93
			SP	0.97	0.66	1.39
Initial bias (β)	*PC+*	*High change*	WD	0.16	0.09	0.24
			SP	0.12	0.03	0.20
	PC-	Low change	WD	0.07	-0.02	0.15
			SP	-0.06	-0.02	0.15
Boundary separation (α)	*CU+*	*High urgency*	WD	0.89	0.72	1.03
			SP	0.70	0.59	0.86
	CU-	Minimal urgency	WD	0.90	0.70	1.12
			SP	0.66	0.53	0.76

Note. Manipulation conditions are italicized.

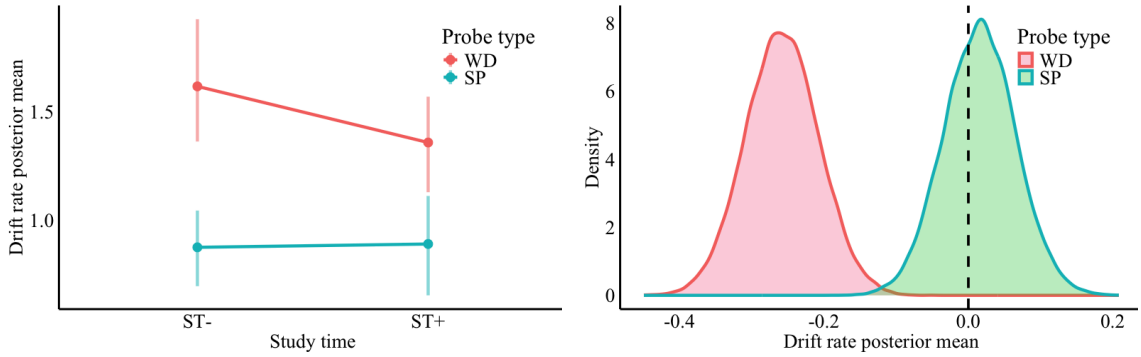
Table 7. Posterior differences of manipulated condition on DDM parameters

Parameters	Condition	Probe Type	Posterior mean	95% CI	
				Lower	Upper
Drift rate (δ)	ST+ − ST-	WD	-0.26	-0.36	-0.16
		SP	0.02	-0.08	0.11

Initial rias (β)	PC+ – PC-	WD	0.09	0.04	0.13
		SP	0.18	0.13	0.23
Boundary separation (α)	CU+ – CU-	WD	-0.01	-0.05	0.03
		SP	-0.05	-0.08	-0.01

The findings on the drift rate manipulation diverged from expectations. Figure 4(a) illustrates the posterior mean summaries of the estimated average drift rate between ST conditions for the different probe types (see values in Table 6). Drift rates were credibly faster on WD trials with shorter time to inspect the study array (ST-; mean 1.62; 95% CI 1.36-1.93) than when given more time to study (ST+; mean 1.36; 95% CI 1.13-1.57). This is also illustrated in Figure 4(b), which shows that the red posterior distribution of the drift rate differences (Study-Time-Contrast parameter) between ST and probe type conditions (mean −0.26; 95% CI −0.36 to −0.16) does not overlap with 0. This finding was unexpected, as we expected that longer time to study the test array would result in faster drift rates. Drift rates on SP trials did not differ credibly between shorter study times (ST-, mean 0.97, 95% CI 0.66-1.39) and longer study times (ST+; mean 1.04; 95% CI 0.70, 1.69). As shown in Figure 4(b), the green posterior distribution of the Study-Time-Contrast parameter overlaps with 0 (mean 0.05; 95% CI −0.08 to 0.11). That is, our drift rate manipulation simply did not influence SP trials.

Figure 4. Posterior means and posterior differences of drift rate parameters for study time conditions.



(a) Drift rate posterior group means

(b) Distribution of drift rate posterior differences

Probability of Change Manipulation on Initial Bias

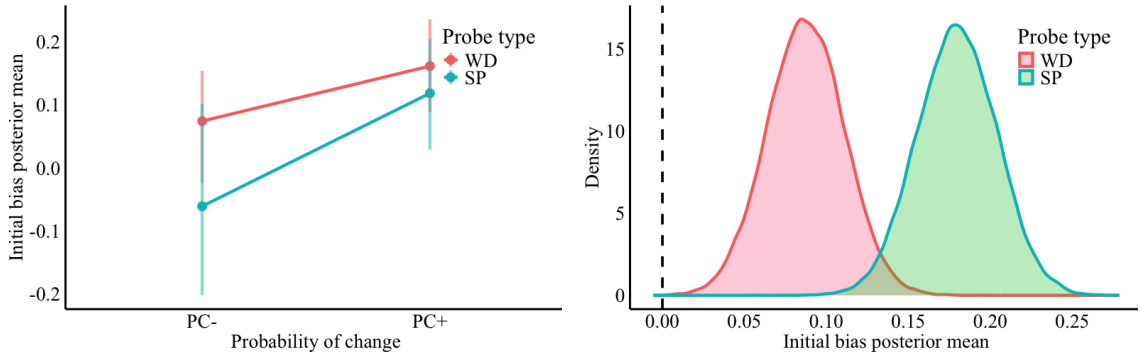
Second, we examined the estimated *Probability-of-Change-Contrast* parameter between probe types. This manipulation was intended to influence the initial bias parameter (β) that captures the degree of a priori bias towards DIFFERENT/SAME response options. As mentioned earlier, this parameter was logit-scaled, with values at 0 representing no systematic bias towards both binary responses, positive values indicating greater bias towards the DIFFERENT response, and negative values indicating bias towards the lower SAME response. Therefore, in Table 7, for the calculated contrast parameter taking the difference in initial bias estimates between the PC conditions (PC+

− PC-), the positive values indicated greater a priori biases towards the DIFFERENT response boundary in the manipulated high change condition.

We observed positive changes in initial bias with higher PC manipulation in both probe types– all contrast values were credibly positive. Figure 5(a) shows the posterior mean summaries of the initial bias parameter between PC conditions and probe types. When tasks were manipulated to have higher probabilities of DIFFERENT trials (80% of trials) than SAME trials (20% of trials), estimates of the initial bias were sensitive to the manipulation and took positive values overall, reflecting a tendency towards selecting the DIFFERENT boundary, both on WD trials, mean 0.16; 95% CI 0.09-0.24, and in SP trials, mean 0.12; 95% CI 0.03-0.20. As expected, trials with even presentation of DIFFERENT and SAME stimuli indicated negligible response bias in both WD (mean 0.07, 95% CI −0.02 to 0.15), and in SP trials (mean −0.06; 95% CI −0.02 to 0.15), as shown in Table 6.

The initial bias manipulation effect is further illustrated in Figure 5(b), where the distribution of posterior differences in β per PC conditions are shown. The range of the contrast parameter's distributions for both probe types do not include 0, suggesting credible effects of this manipulation on the initial bias parameter across probe types. This effect is more pronounced in the single probe condition (green distribution, mean 0.18; 95% CI 0.13-0.23), compared to the whole display condition (red distribution, mean 0.09, 95% CI 0.04-0.13).

Figure 5. Posterior means and posterior differences of initial bias parameter for probability of change conditions



(a) Initial bias posterior group means

(b) Distribution of initial bias posterior differences

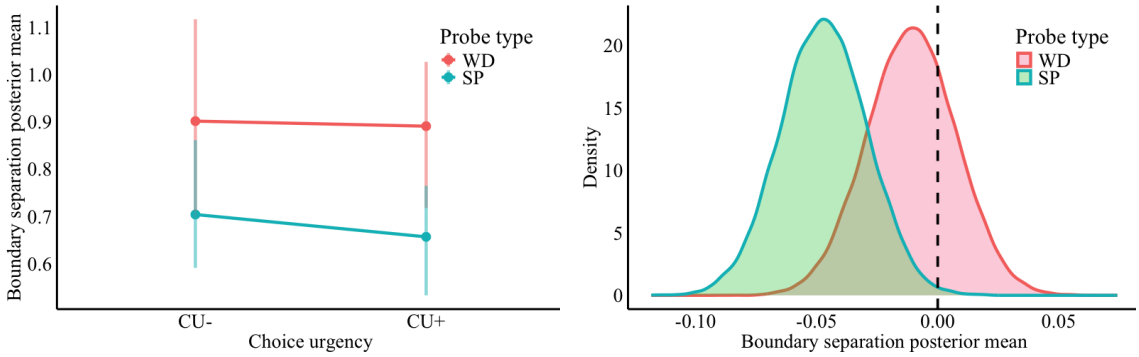
Choice Urgency Manipulation on Boundary Separation

Thirdly, we examined differences in the boundary separation parameter by choice urgency and probe type conditions. This manipulation was intended to influence the response boundary separation: the amount of evidence required to make a decision. Larger boundary separation corresponds to higher threshold values (α). We expected the boundary separations to be lower in the manipulated condition with higher choice urgency to respond (3000ms) than in the baseline condition with minimal urgency and more time to respond (100s). The boundary separation parameter was log-transformed to ease the regression modeling, but this scale remained easy to interpret as higher boundary separation values still indicated more information needed before making a decision (with

values closer to 0 indicating less information acquired before making a decision). In Table 7, negative Choice-Urgency-Contrast values (CU+ − CU-) indicated higher boundary separation in the manipulated condition (high urgency). Positive differences represented higher response boundary separation in the manipulated high urgency condition than the control minimal urgency condition.

This manipulation worked as expected for the SP version of the task. The posterior mean summaries of the boundary separation parameter between CU conditions and probe types are shown in Figure 6(a), based on values from Table 6. For the SP trials, the manipulation was successful as boundary separation parameters were sensitive to the different CU conditions (mean −0.05; 95% CI −0.08 to −0.01), with participants exhibiting lower boundary separation in the High CU to respond trials. This contrast parameter for SP trials is shown in Figure 6(b), with almost all of the green posterior distribution being larger than 0. Comparatively, looking at the red distribution corresponding to the WD trials, we saw that 0 is almost in the middle of that distribution (mean −0.01; 95% CI −0.05 to 0.03), signifying no credible effect for this manipulation in the WD condition.

Figure 6. Posterior means and posterior differences of the boundary separation parameter for choice urgency conditions



(a) Boundary separation posterior group means

(b) Distribution of boundary separation posterior differences

Bayesian Correlations

We explored associations between age and the three person-specific parameters of the DDM (drift rate, boundary separation, and initial bias) by running Bayesian correlation analysis in JASP (Version 0.17.1).[32] This analysis yields a regular *Pearson correlation coefficient* for each tested link and a corresponding *Bayes Factor* that in our case summarizes the amount of evidence for the coefficient to be different from 0.[33] For instance, Bayes Factors (BF_{10}) of 3−10 indicate moderate evidence for the correlation, 10−30 indicates strong evidence, 30−100 indicates very strong evidence, and beyond. Based on previous studies,[34] we predicted that older participants would have slower drift rates and higher boundary separation in decision making, while we did not expect age to be credibly linked to initial bias. We examined the parameters in three manners: overall

trials (WD+SP), WD trials, and SP trials. Correlation coefficients and their corresponding Bayes Factors are shown in Table 8.

Age and drift rate. The relationship between age and drift rate was of particular interest, as the declines in cognitive processing may be related to early indicators of pre-ADRD. Associations were explored across all trials (WD+SP) then stratified by probe types. As expected, the Pearson's *r* coefficients of age with overall drift rate indicated a negative association with age, r=−0.45; 95% CI −0.61 to −0.23, suggesting that older participants have slower drift rates. Results indicated extreme evidence for this association and that the data was at least 216 times more likely under the alternative hypothesis than the null hypothesis (BF_{10}=210.56). We found the similar links when we broke down this association by probe type. The results indicated very strong evidence for negative association between drift rate and age in SP trials (r=−0.42; 95% CI −0.59 to −0.19; BF_{10}=67.58) and WD trials (r=−0.40; 95% CI −0.57 to −0.17; BF_{10}=39.06).

Age and initial bias. As expected, we did not find evidence for association between initial bias and age across WD+SP trials (r=0.21; 95% CI −0.03 to 0.42, BF_{10} =0.66). People's age was not expected to be linked with whether they are more likely to choose DIFFERENT or SAME on a trial. However, when we separated the probe types, we found some weak evidence of association in the WD condition (r=0.31; 95% CI 0.07, 0.50, BF_{10}=3.77). Importantly, the SP condition showed evidence for the lack of correlation (r=−0.03; 95% CI −0.26 to 0.21; BF_{10}=0.16), which we can see by taking the reciprocal of the corresponding Bayes Factor in Table 8 (BF_{01}=1/.16=6.25), which represented 6.25 more evidence for the lack of correlation.

Age and boundary separation. The Pearson's *r* coefficients reported a positive association with age, r=0.33; 95% CI 0.10-0.52. This suggested that advancing age is linked with higher boundary separation. For WD+SP trials, the computed Bayes Factor (BF_{10}=6.00) indicated moderate evidence and that the data was at least 6 times more likely under the alternative hypothesis than the null hypothesis. Importantly, this boundary separation association was accentuated in the SP. Boundary separation in SP trials, r=0.42; 95% CI 0.19-0.59, indicated very strong evidence (BF_{10}=69.61) for the alternative hypothesis. The Bayes Factor indicated that the data are 69 times more likely under the alternative hypothesis than the null hypothesis. For boundary separation in WD, we did not find a credible association and the Bayes Factor provided no evidence (BF_{10} ≈1) for the alternative hypothesis, r=0.24; BF_{10}=0.99; 95% CI 0.00-0.44.

Strategy Use

Subsequently, we conducted a post hoc qualitative analysis of cognitive strategy use in light of the unexpected effect of Study Time on drift rates. As part of the debrief and exit survey, participants were presented with a brief series of unstructured questions. Summary of strategy utilization is presented in Table 9 and all the responses are reported in Appendix B. From the 68 participants, 79% (n=54) reported using some strategy whereas 15% (n=10) reported no strategy utilized. Individuals who responded "Yes" were presented with an open-ended response to elaborate on specific strategies. One common

theme emerged upon reviewing the responses. Many strategies involved using some tactic involving two or more shapes. Responses were categorized based on whether individuals reported any mention of two shapes or not. 61% (*n*=33) reported some strategy using two shapes, whereas the other 39% (*n*=21) individuals reported a different strategy that did not involve two shapes. For instance, participants reported using mnemonic strategies such as organizing objects by size, creating heuristics with objects, or integrating the array into a single object.

Table 8. Bayesian correlations of DDM parameters and age

				95% CI	
Parameters	Probe type	Statistics	Value	Lower	Upper
Drift rate (δ)	WD+SP	Pearson's *r*	-0.45 ***	-0.61	-0.23
		BF_{10}	216.31		
	WD	Pearson's *r*	-0.40 ***	-0.57	-0.17
		BF_{10}	39.89		
	SP	Pearson's *r*	-0.42 ***	-0.59	-0.19
		BF_{10}	68.63		
Initial bias (β)	WD+SP	Pearson's *r*	0.21	-0.03	0.42
		BF_{10}	0.66		
	WD	Pearson's *r*	0.31	0.07	0.50
		BF_{10}	3.77		
	SP	Pearson's *r*	-0.03	-0.26	0.21
		BF_{10}	0.16		
Boundary separation (α)	WD+SP	Pearson's *r*	0.33 *	0.10	0.52
		BF_{10}	5.91		
	WD	Pearson's *r*	0.24	-0.00	0.44
		BF_{10}	0.98		
	SP	Pearson's *r*	0.42 ***	0.19	0.59
		BF_{10}	69.04		

a) Bayes Factors (BF_{10}) indicate represents ratio of likelihood of evidence in favor of alternative model (1) than null model (0). * $BF_{10} > 3$, ** $BF_{10} > 10$, *** $BF_{10} > 30$.

Discussion

We used a contemporary approach to task validation by showing how features extracted from cognitive task performance scores can experimentally be manipulated.[35] Integrating theory-based working memory measures with computational cognitive methodology, we focused on different methodological strengths to capture cognitive ability in daily life settings. This approach was powered by the Bayesian statistical engine that allowed for simultaneous estimation of all latent features, as well as person – and condition-specific variations in it.

Principal Results

We presented a novel ambulatory adaptation of the Color Shapes task. The Color Shapes task has previously been established in literature as a sensitive and specific measure of symptomatic and asymptomatic preclinical-AD.[1,36] Response time and accuracy data from this task have not been analyzed in a drift diffusion modeling framework before. Our approach focused on disentangling subtle cognitive processes underlying manifest Color Shapes task performance with the drift diffusion model.

Specifically, we aimed at optimizing the Color Shapes task for drift diffusion modeling in ambulatory assessment settings. Models such as the DDM can incorporate speed and accuracy characteristics in RT data that may offer novel insights into cognitive changes that account for the speed-accuracy trade-off.[23] For this, we tested whether we are able to experimentally manipulate three parameters of the DDM by changing features of the Color Shapes task. We compared whole display (WD) and single probe (SP) versions across these manipulations. We also tested whether individual-level DDM parameters were meaningfully associated with participant's age, given previous studies showed that older participants tended to have declines in speed of cognitive processing[37] and monitor responses to prioritize accuracy over swift responses.[38] Finally, we looked at whether participants would be likely to use strategies for performing better on the Color Shapes task.

Table 9. Summary of responses related to cognitive strategy use

Cognitive strategy reported	*N*	Percent
Used some strategy	54	79%
Description includes two shapes	33	61%
Any strategy that does not include two shapes	21	39%
Did not report any strategy	10	15%
No response	4	6%

Post hoc analysis of reported cognitive strategy usage. Raw responses are reported in the appendix.

Our results showed Color Shapes task feature manipulations affected cognitive processes, as captured by the DDM parameters. When we changed the proportion of DIFFERENT vs. SAME trials, the initial bias shifted, indicating subtle preference towards the DIFFERENT response in both SP and WD versions. However, when we manipulated the boundary separation feature, we observed the expected effect only in the SP version that tests 1 colored shape. Increasing the time pressure to choose (i.e., higher choice urgency), lowered the cautiousness in responding to the task; and vice versa. Lastly, increasing the amount of time to study the array did not increase the drift rate, or speed of evidence accumulation as expected. Increasing the study time did not impact the drift rate in the SP version. Alternatively, increasing the study time had the inverse effect than expected, with longer study times associated with slower drift rates in WD trials.

We propose that the SP version of the Color Shapes task is more optimal for collecting data for DDM-based cognitive processes disassociation in ambulatory settings than the WD version. This is supported by the successful DDM parameter manipulation for this version for 2 out of 3 DDM features. Moreover, the correlations analysis between individual-level DDM parameters and age was more consistent with the SP version: 1) There was evidence for links between age and drift rate, which is in line with previous findings, and 2) evidence for the lack of correlation between age and initial bias, as expected. The WD version did not show evidence for correlation between age and boundary separation and had some evidence for unexpected weak correlation between age and initial bias. Moreover, post hoc qualitative data analysis showed that most of the unwanted strategy use was related to encoding only two shapes out of the three in the study array. This strategy would not be feasible for participants in the SP version of the task.

Our results indicated that this ambulatory task version (SP version of the Color Shapes task with short study time (500ms), low probability of change (50%-50% DIFFERENT/SAME), and minimal choice urgency) responded well to the intended manipulations and was associated with convergent and divergent constructs as theorized. Giving 500ms to study has been commonly used in clinical settings and might also prevent verbal encoding of the stimuli. We found that drift rates were higher in the 500ms condition, which might support a more one-shot decision-making style, which adheres better with the assumptions of the DDM. The other two features of the task, that is even presentation of stimuli and no strict limit on time to respond are also typical in the literature.

This optimization of an ambulatory version of the Color Shapes task for accessibility and computational cognitive modeling advances the development of digital tools that extract novel digital markers to empower individuals to initiate health services earlier. By implementing the DDM to disentangle ambulatory cognitive performance data into more nuanced processes underlying performance, we can open opportunities to integrate cognitive aging theoretical frameworks with early detection advances ushered in by technological advances in computational tools[41] and multi-model data.[42] Our study integrated the Color Shapes task linked with preclinical AD-related cognitive changes with computational cognitive methodology and focused on their respective methodological strengths for capturing cognitive processes in daily life settings. This approach is powered by the Bayesian statistical engine that allows for a person-centered approach for quantifying individual-specific risk probabilities, for example in terms of DDM parameter estimates.

Taken together with drift diffusion modeling, the results found that this approach was effective in measuring subtle differences in latent cognitive processes from observed performance. This approach has the potential to extract sensitive digital-based cognitive indicators that can reveal subtle cognitive decline at earlier stages of AD/ADRD. The longitudinal implementation of this approach holds promise for being able to monitor these digital cognitive features over short and long periods of time. On a broader scale,

this approach can be effectively tailored for designing person-centered treatment plans for monitoring cognitive health status over adulthood.

Limitations and Future Directions

Finally, the current study had some limitations. First, there may be underlying bias related to participants and study selection. For instance, the population of individuals that register for research participation may be qualitatively compared to individuals that do not sign-up. Second, cognitive health status was assessed via self-reported information and may be bolstered with biological/clinical validation. Third, there may be relevant reasons for missing data on cognitive performance that can be important to address in future computerized assessment designs.

As this study was conducted entirely remotely, it may require additional consideration to control for unintended cognitive strategy use. We opted to ask participants at the end regarding any cognitive strategy utilization, rather than providing explicit instructions that may potentially bias performance data. As the Color Shapes task is intended to assess cognitive processes underlying visual working memory,[4,13] using unwanted cognitive strategies utilizing different cognitive domains (e.g., verbal encoding), the Color Shapes task could fail to measure changes in visual short-term memory system and bias results. Designs to assess subtle changes associated with cognitive impairment may benefit from conducting a priori technical optimization to identify best strategies to control for unwanted cognitive strategy use. By optimizing the task for unsupervised remote testing, this aids in designing valid and reliable ambulatory cognitive instruments.

AD has a long preclinical phase marked by neuropathological changes.[44] While early stages of preclinical AD can be detected using biomarkers of amyloidosis and neurodegeneration (e.g., cerebrospinal fluid, brain imaging, blood[45]), mobile cognitive testing offers an easily accessible early monitoring tool. Due to the nature of subtle progressive cognitive decline in AD, individuals may conflate pathological cognitive decline with normative age-related changes, and consequently delay seeking timely medical care[40] or wait until symptoms worsened.[39] One overarching goal of this study was to develop an optimized version of the Color Shapes task that is effective for unsupervised, high-frequency cognitive testing. Our results showed evidence that mobilizing smartphone technology to administer ambulatory cognitive assessments in this manner is a viable approach for sampling a representative measure of cognitive status variability in daily contexts. A recent survey reported that 70% of Americans want to know if they were at risk for preclinical AD for seeking earlier treatment, however, only 60% of Americans were reluctant to discuss MCI symptoms and early signs with their healthcare providers.[39] Commonly cited reasons included concerns with receiving incorrect diagnosis, treatment, or waiting until symptoms did not resolve. Our results were promising as SP trials may be less cognitively loading and may support long-term adherence. This study demonstrates an effective and feasible design to address related to effective screening and earlier intervention.

Conclusion

Collecting data with the Color Shapes task in ambulatory settings and analyzing it using a drift diffusion modeling approach yield digital markers of key cognitive processes. We found that the single probe (test array with 1 colored shape) version of the Color Shapes task with 500ms (short) study time, even presentation (50%-50%) of DIFFERENT/SAME trials, and minimal response urgency exhibited expected performance and associations. The novel digital cognitive features that could be effective for sensitive and earlier screening of individuals for secondary AD/ADRD prevention, earlier intervention, and targeting modifiable risk factors.[1,2]

Appendix A

Drift Diffusion Model

Specification of the Drift Diffusion model

To analyze response time and accuracy data from the Color Shapes task, we applied a Bayesian hierarchical drift diffusion model.[48] The underlying assumption of the DDM is that during these tasks, participants engage in a noisy information accumulative process that stops when the cumulative evidence reaches a predetermined response boundary separation for the decision. The DDM describes such data into four parameters that capture the decision process: drift rate δ, boundary separation α, initial bias β, and non-decision time τ.

More specifically, the process assumptions of the diffusion model are that a single evidence counter accumulates towards one of two decision boundaries, with a starting point that may be closer to one boundary than the other. Figure 2 illustrates the process above. Given the freedom of two decision bounds, the model accounts for two distinct types of bias in the response process. In addition to differences in the speed of evidence accumulation process (which is reflected in δ), the diffusion model allows for an a-priori bias that is prior to and independent of the information accumulation process (here parameterized as a proportion, so that an initial bias $\beta = 0.5$ implies a-priori indifference).

Overview of basic model parameters

The drift rate δ quantifies the ease of evidence processing and is generally considered not under subjective control. Higher absolute values $|\delta|$ indicate faster and accurate decisions. Conversely, lower $|\delta|$ values indicate decisions slower information accumulation, thus longer RTs and with responses closer to chance. The absolute value can hence be interpreted as the ability of the person to perform the task.

The boundary separation parameter α captures the amount of information needed to execute a response. It is high for slow and deliberative responses, and low for fast and error-prone responses. This parameter can be interpreted as the speed-accuracy trade-off or the response caution of the participant during the task. It is generally considered to be under subjective control: participants can choose to prioritize speed or accuracy.

The initial bias parameter β is the amount of information for or against either choice alternative that the participant holds before seeing a stimulus. If it is close to 1, the participant has an a-priori bias towards responding DIFFERENT; if it is close to 0, the

participant has an a-priori bias towards responding SAME; if it is 0.5, the participant has no a-priori bias. Participants can choose to prefer one response over another.

Finally, the non-decision time τ parameter captures all time that is taken up from the start of the task to the initiation of the information accumulation process. Commonly this includes the encoding and response processes, and it also includes other perceptual and cognitive processes depending on the individual and task.

Specification of the multilevel Drift Diffusion model with condition-specific means

We extended the model hierarchically to account for between- and within-subject variability in these cognitive parameters (Vandekerckhove et al., 2011). Broadly, when conducting multilevel modeling, there are two kinds of parameters. On the one hand, there are the group-level (population) parameters that capture the nomothetic patterns between groups. The individual-level parameters capture the idiographic patterns that are constrained by group-level parameters.[49] The group-level parameters can account for the extent of similarity between individuals. Group- and person-level parameters influence each other via information pooling.[50]

In Equation 1 below, the drift diffusion model likelihood function is governed by the Wiener process, with parameters at the person-level and experimental condition-level. $Y_{p,c_x,t}$ represents the observed cognitive task performance for each person, p, for each of the x conditions (c), c_x being manipulated or not (e.g., $c_1 = 1$ means the first condition being manipulated,[1] as opposed to 0, $c_1 = 0$), on each trial, j. The observed data include both the RT and accuracy dimensions, and follow a Wiener distribution specified by the parameters α, β, τ, and δ, which were allowed to differ for each individual. Parameters α, β, and δ also differed across the experimental conditions. Additionally, the δ, drift rates changed sign between SAME and DIFFERENT trials, captured by the l index ($l = 1$ for different, $l = 2$ for same). Finally, person-specific τ_p were estimated.

$$Y_{p,c_x,t} \sim Wiener(\alpha_{p,c_x}, \beta_{p,c_x}, \tau_p, \delta_{p,c_x,l}) \quad (1)$$

Hyper-prior and prior distributions

Given that the boundary separation parameters of the DDM can only take positive values, person- and trial-specific boundary separation α parameters were assigned log-normal distributions as population-level distributions, specified as:

$$\alpha_{p,c_x} \sim Lognormal(\mu_{\alpha_{c_x}}, \sigma^2_{\alpha}) \quad (2)$$

where $\mu_{\alpha_{c_x}}$ is the group-level mean for the boundary separation parameter, allowed to vary between conditions c_x and σ^2_{α} captures the group-level variance.

[1] Respectively, they index the four experimental conditions: (1) study time, (2) probability of change, (3) choice urgency, and (4) probe type.

The initial bias parameter could only take values between 0 and 1, therefore we specified a normal hyperprior on its logit-transformation:

$$logit(\beta_{p,c_x}) \sim Normal(\mu_{\beta_{c_x}}, \sigma^2_{\beta}) \quad (3)$$

where $\mu_{\beta_{c_x}}$ is the group-level mean for the initial bias parameter, allowed to vary between conditions c_x and σ^2_{β} captures the group-level variance.

For individual-level drift rate parameter estimates for the DIFFERENT trials (l = 1), δ_{p,c_x,l_1}, the population distribution was specified as normal:

$$\delta_{p,c_x,l_1} \sim Normal(\mu_{\delta_{c_x}}, \sigma^2_{\delta}) \quad (4)$$

where $\mu_{\delta_{c_x}}$ is the group-level mean for the drift rate parameter, allowed to vary between conditions c_x and σ^2_{δ} captures the group-level variance. The drift rate for SAME trials (l = 2) were parameter estimates from DIFFERENT trials that were multiplied by −1 (based on encoding RTs with SAME responses as negative) to represent processes towards the lower (SAME) decision boundary with absolute drift rate.

$$\delta_{p,c_x,l_2} = -1 \times \delta_{p,c_x,l_1} \quad (5)$$

For the non-decision time we assigned a uniform prior distribution, specified as a beta distribution with shape and rate parameters 1 and 1:

$$\tau_p \sim Beta(1, 1) \quad (6)$$

Finally, we specified prior distributions for the means for the three DDM parameters ($\mu_{\alpha_{c_x}}$, $\mu_{\beta_{c_x}}$, $\mu_{\delta_{c_x}}$) to be standard normal [Normal(0, 1)], and for the standard deviations to be uniformly distributed between 0 and 100 [Uniform(0,100)].

Bayesian implementation of the multilevel DDM

All analyses were conducted in R, Version 4.2.2.[51] The Bayesian model was implemented in Just Another Gibbs Sampling (JAGS)[52] interfaced with R via the rjags package.[53] The Wiener distribution was specified using a custom JAGS module.[54] The analytical scripts are available on Open Science Framework.[31]

Using Markov chain Monte Carlo (MCMC) algorithms implemented in JAGS. We ran 4 chains drawing 15,000 samples each with 500 burn-in samples and 500 adaptation samples per chain, resulting in 60,000 total posterior samples for each parameter.

We checked the quality of MCMC samples drawn from the posterior distribution. The effective sample size (*ESS*) measures and $\hat{R}$ statistics provided numerical measures of the representativeness and stability of the MCMC chains. We met general guidelines with all $\hat{R}$ values being below 1.1,[55] indicating adequate convergence within and between chains. Additionally, we had sufficient *ESS* (measure of independent information in autocorrelated chains) of over 1,000.[50]

Acknowledgements

ZO, MS, JH and KH were supported by the NIH (R01 AG074208, R56 AG074208, and U2 CAG060408). JH was supported by the NIH under award numbers R00AG056670 and R33AG078084. JV was supported by NSF grant #2051186. KH was supported by T32 AG049676 training grant. This work utilized assessments developed under NIA Grant U2 CAG060408.

Conflicts of Interest

KH is a full-time employee of CogState Ltd., which develops cognitive assessments. This study did not utilize any CogState products. No other conflicts of interest to declare.

References

1. Ratcliff, R., & McKoon, G. (2022). Can neuropsychological testing be improved with model-based approaches? *Trends in Cognitive Sciences, 26*(11), 899–901.
2. Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., . . . others (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia, 7(3), 280–292.
3. Tsoi, K. K., Chan, J. Y., Hirai, H. W., Wong, S. Y., & Kwok, T. C. (2015). Cognitive tests to detect dementia: a systematic review and meta-analysis. JAMA internal medicine, 175(9), 1450-1458.
4. Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2018). Reliability and validity of ambulatory cognitive assessments. Assessment, 25(1), 14-30.
5. Nicosia, J., Aschenbrenner, A. J., Balota, D. A., Sliwinski, M. J., Tahan, M., Adams, S., ... & Hassenstab, J. (2023). Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease. Journal of the International Neuropsychological Society, 29(5), 459-471.
6. Oravecz, Z., Harrington, K. D., Hakun, J. G., Katz, M. J., Wang, C., Zhaoyang, R., & Sliwinski, M. J. (2022). Accounting for retest effects in cognitive testing with the Bayesian double exponential model via intensive measurement burst designs. Frontiers in Aging Neuroscience, 14, 897343.
7. Stone, A. A., Shiffman, S. S., & DeVries, M. W. (1999). Ecological momentary assessment.
8. Patrick, K., Griswold, W. G., Raab, F., & Intille, S. S. (2008). Health and the mobile phone. American Journal of Preventive Medicine, 35(2), 177–181.
9. Koo, B. M., & Vizer, L. M. (2019). Mobile technology for cognitive assessment of older adults: a scoping review. Innovation in aging, 3(1), igy038.
10. Milner, T., Brown, M. R. G., Jones, C., Leung, A. W. S., & Brémault-Phillips, S. (2024). Multidimensional digital biomarker phenotypes for mild cognitive impairment: Considerations for early identification, diagnosis and monitoring. Frontiers in Digital Health, 6, 1265846.
11. Wiecki, T. V., Poland, J., & Frank, M. J. (2015). Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. Clinical Psychological Science, 3(3), 378-399.
12. Didic, M., Barbeau, E. J., Felician, O., Tramoni, E., Guedj, E., Poncet, M., & Ceccaldi, M. (2011). Which memory system is impaired first in Alzheimer's disease? Journal of Alzheimer's Disease, 27(1), 11–22.

13. Parra, M. A., Calia, C., Pattan, V., & Della Sala, S. (2022). Memory markers in the continuum of the Alzheimer's clinical syndrome. Alzheimer's Research & Therapy, 14(1), 1–16.
14. Parra, M. A., Abrahams, S., Logie, R. H., Méndez, L. G., Lopera, F., & Della Sala, S. (2010). Visual short-term memory binding deficits in familial Alzheimer's disease. Brain, 133(9), 2702–2713.
15. Butler, J., Owobowale, A. S., Watermeyer, T., Danso, S. O., & Parra-Rodriguez, M. (2024). Developing an AI algorithm to detect predictors of poor performance in a self-administered, web-based digital biomarker for Alzheimer's disease: Proof of concept.
16. Butler, J., Watermeyer, T. J., Matterson, E., Harper, E. G., & Parra-Rodriguez, M. (2024). The development and validation of a digital biomarker for remote assessment of Alzheimer's diseases risk. Digital Health, 10, 20552076241228416.
17. Cerino, E. S., Katz, M. J., Wang, C., Qin, J., Gao, Q., Hyun, J., ... & Sliwinski, M. J. (2021). Variability in cognitive performance on mobile devices is sensitive to mild cognitive impairment: Results from the Einstein Aging Study. Frontiers in Digital Health, 3, 758031.
18. Thompson, L. I., Strenger, J. R., Emrani, S., De Vito, A., Harrington, K., Roque, N., ... & Correia, S. (2023). 91 Remote Smartphone-Based Assessment Predicts Standard Neuropsychological Test Performance and Cerebral Amyloid Status in Cognitively Normal Older Adults. Journal of the International Neuropsychological Society, 29(s1), 493-494.
19. Zhaoyang, R., Scott, S. B., Martire, L. M., & Sliwinski, M. J. (2021). Daily social interactions related to daily performance on mobile cognitive tests among older adults. PLoS One, 16(8), e0256583.
20. Lövdén, M., Li, S. C., Shing, Y. L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. Neuropsychologia, 45(12), 2827-2838.
21. Patzelt, E. H., Hartley, C. A., & Gershman, S. J. (2018). Computational phenotyping: using models to understand individual differences in personality, development, and mental illness. Personality Neuroscience, 1, e18.
22. Ratcliff, R. (1978). A theory of memory retrieval. Psychological review, 85(2), 59.
23. De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. Frontiers in Psychology, 10, 102.
24. Stafford, T., Pirrone, A., Croucher, M., & Krystalli, A. (2020). Quantifying the benefits of using decision models with response time and accuracy data. Behavior Research Methods, 52, 2142–2155.

25. Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. Neural Computation, 20(4), 873–922.
26. Aschenbrenner, A. J., Balota, D. A., Gordon, B. A., Ratcliff, R., & Morris, J. C. (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(2), 225.
27. Karimi, H., Marefat, H., Khanbagi, M., Karami, A., & Vahabi, Z. (2021). Drift Diffusion Model of Animacy Categorization Task Can Detect Patients with Mild Cognitive Impairment and Mild Alzheimer's Disease. Frontiers in Biomedical Technologies.
28. Ratcliff, R., Scharre, D. W., & McKoon, G. (2022). Discriminating memory disordered patients from controls using diffusion model parameters from recognition memory. Journal of Experimental Psychology: General, 151(6), 1377.
29. Vandekerckhove, J. (2009). Extensions and applications of the diffusion model for two-choice response times. Unpublished doctoral dissertation, University of Leuven.
30. Villarreal, M., Chávez De la Peña, A. F., Mistry, P. K., Menon, V., Vandekerckhove, J., & Lee, M. D. (2024). Bayesian graphical modeling with the circular drift diffusion model. Computational Brain & Behavior, 7(2), 181-194.
31. Kim, S.H., Li, Y., Oravecz, Z., Data and code for: Digital biomarkers. OSF. Link here. Published [August 28, 2024.]
32. JASP Team. (2024). JASP (Version 0.17.3) [Computer software]. Retrieved from https:// jasp-stats.org/
33. Lee, M. D., & Wagenmakers, E.-J. (2014). Bayesian cognitive modeling: A practical course. Cambridge University Press.
34. Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. Psychology and Aging, 18(3), 415.
35. Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. Psychological review, 111(4), 1061.
36. Parra, M. A., Abrahams, S., Logie, R. H., Méndez, L. G., Lopera, F., & Della Sala, S. (2010). Visual short-term memory binding deficits in familial Alzheimer's disease. Brain, 133(9), 2702–2713.
37. Salthouse, T. A. (2019). Trajectories of normal cognitive aging. Psychology and Aging, 34(1), 17.
38. Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. British Journal of Psychology, 70(2), 305-311.
39. Alzheimer's Association. (2022). Alzheimer's disease facts and figures. https://www.alz .org/alzheimers-dementia/facts-figures. ([Accessed 20-Jul-2023])

40. Harada, C. N., Love, M. C. N., & Triebel, K. L. (2013). Normal cognitive aging. Clinics in Geriatric Medicine, 29(4), 737–752.
41. Vrahatis, A. G., Skolariki, K., Krokidis, M. G., Lazaros, K., Exarchos, T. P., & Vlamos, P. (2023). Revolutionizing the early detection of Alzheimer's disease through non-invasive biomarkers: the role of artificial intelligence and deep learning. Sensors, 23(9), 4184.
42. Vrahatis, A. G., Vlamos, P., Avramouli, A., Exarchos, T., & Gonidi, M. (2021). Emerging machine learning techniques for modelling cellular complex systems in Alzheimer's disease. In GeNeDis 2020: Computational Biology and Bioinformatics (pp. 199-208). Springer International Publishing.
43. Rajan, K. B., Weuve, J., Barnes, L. L., McAninch, E. A., Wilson, R. S., & Evans, D. A. (2021). Population estimates of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020–2060). Alzheimer's & Dementia, 17(12), 1966-1975.
44. Price, J. L., & Morris, J. C. (1999). Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 45(3), 358-368.
45. Blennow, K., & Zetterberg, H. (2018). The past and the future of Alzheimer's disease fluid biomarkers. Journal of Alzheimer's Disease, 62(3), 1125-1140.
46. Alzheimer's Association. (2023). Alzheimer's disease facts and figures. https://www.alz .org/alzheimers-dementia/facts-figures. ([Accessed 20-Jul-2023])
47. Lam, B., Masellis, M., Freedman, M., Stuss, D. T., & Black, S. E. (2013). Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. Alzheimer's Research & Therapy, 5(1), 1–14.
48. Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. Psychological Methods, 16(1), 44.
49. Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. Journal of Mathematical Psychology, 55(1), 84–93.
50. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2013). Bayesian data analysis. Chapman and Hall/CRC.
51. R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/
52. Plummer, M., et al. (2003). Jags: A program for analysis of Bayesian graphical models using gibbs sampling. In Proceedings of the 3rd International Workshop on Distributed Statistical Computing (Vol. 124, pp. 1–10).

53. Plummer, M., Stukalov, A., & Denwood, M. (2016). rjags: Bayesian graphical models using MCMC. R package version, 4(6).
54. Wabersich, D., & Vandekerckhove, J. (2014). Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). Behavior Research Methods, 46, 15-28.
55. Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics, 7(4), 434–455.