UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Dissociable Contributions of Goal-Relevant Evidence and Goal-Irrelevant Familiarity to Individual and Developmental Differences in Conflict Recognition

Weigard, A.; Suzuki, T.; Skalaban, L.J.; Conley, M.; Cohen, A.O.; Garavan, H.; Heitzeg, M.M.; Casey, B.J.; Sripada, C.; Heathcote, A.

DOI
10.1111/cogs.70019

Publication date
2024

Document Version
Final published version

Published in
Cognitive Science

License
CC BY-NC

Link to publication

Citation for published version (APA):
Weigard, A., Suzuki, T., Skalaban, L. J., Conley, M., Cohen, A. O., Garavan, H., Heitzeg, M. M., Casey, B. J., Sripada, C., & Heathcote, A. (2024). Dissociable Contributions of Goal-Relevant Evidence and Goal-Irrelevant Familiarity to Individual and Developmental Differences in Conflict Recognition. *Cognitive Science*, *48*(11), Article e70019. https://doi.org/10.1111/cogs.70019

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.
UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:16 Oct 2026

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 48 (2024) e70019
© 2024 The Author(s). *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).
ISSN: 1551-6709 online
DOI: 10.1111/cogs.70019

Dissociable Contributions of Goal-Relevant Evidence and Goal-Irrelevant Familiarity to Individual and Developmental Differences in Conflict Recognition

Alexander Weigard,[a] Takakuni Suzuki,[b] Lena J. Skalaban,[c] May Conley,[d] Alexandra O. Cohen,[e] Hugh Garavan,[f] Mary M. Heitzeg,[a] B. J. Casey,[g] Chandra Sripada,[a] Andrew Heathcote[h]

[a]*Department of Psychiatry, University of Michigan*
[b]*Department of Psychology, University of Tulsa*
[c]*Department of Psychology, Temple University*
[d]*Department of Psychology, Yale University*
[e]*Department of Psychology, Emory University*
[f]*Department of Psychiatry, University of Vermont*
[g]*Department of Neuroscience and Behavior, Barnard College of Columbia University*
[h]*Department of Psychological Methods, University of Amsterdam*

Received 8 February 2024; received in revised form 20 September 2024; accepted 4 November 2024

Abstract

Recent studies using the diffusion decision model find that performance across many cognitive control tasks can be largely attributed to a task-general efficiency of evidence accumulation (EEA) factor that reflects individuals' ability to selectively gather evidence relevant to task goals. However, estimates of EEA from an n-back "conflict recognition" paradigm in the Adolescent Brain Cognitive Development[SM] (ABCD) Study, a large, diverse sample of youth, appear to contradict these findings. EEA estimates from "lure" trials—which present stimuli that are familiar (i.e., presented previously) but do not meet formal criteria for being a target—show inconsistent relations with EEA estimates from other trials and display atypical v-shaped bivariate distributions, suggesting many individuals are responding based largely on stimulus familiarity rather than goal-relevant stimulus features. We present a new formal model of evidence integration in conflict recognition tasks that distinguishes individuals'

Correspondence should be sent to Alexander Weigard, Department of Psychiatry, University of Michigan, Rachel Upjohn Building, 4250 Plymouth Road, Ann Arbor, MI 48109, USA. E-mail: asweigar@med.umich.edu

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

EEA for goal-relevant evidence from their use of goal-irrelevant familiarity. We then investigate developmental, cognitive, and clinical correlates of these novel parameters. Parameters for EEA and goal-irrelevant familiarity-based processing showed strong correlations across levels of n-back load, suggesting they are task-general dimensions that influence individuals' performance regardless of working memory demands. Only EEA showed large, robust developmental differences in the ABCD sample and an independent age-diverse sample. EEA also exhibited higher test-retest reliability and uniquely meaningful associations with clinically relevant dimensions. These findings establish a principled modeling framework for characterizing conflict recognition mechanisms and have several broader implications for research on individual and developmental differences in cognitive control.

Keywords: Diffusion model; Evidence accumulation; Inattention; Working memory; n-back

1. Introduction

Individual differences in "cognitive control" functions that facilitate goal-directed behavior have been a central focus of basic cognitive science (Botvinick & Cohen, 2014; Friedman & Robbins, 2022; Mashburn, Tsukahara, & Engle, 2020), research on human development (Davidson, Amso, Anderson, & Diamond, 2006; Somerville & Casey, 2010), and clinical research on the etiology of psychiatric disorders, including attention-deficit/hyperactivity disorder (Alderson, Hudec, Patros, & Kasper, 2013; Kofler, Rapport, Bolden, Sarver, & Raiker, 2010; Rapport et al., 2008; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005), substance use disorders (Khurana, Romer, Betancourt, & Hurt, 2017; Squeglia & Gray, 2016; Stevens et al., 2014), and psychosis (Bozikas & Andreou, 2011; Gold et al., 2019; Mesholam-Gately, Giuliano, Goff, Faraone, & Seidman, 2009). Such work heavily relies on behavioral task-based measures that are thought to index an individual's ability to maintain goal-relevant information and use that information to guide behavior.

A popular framework for studying individual differences in cognitive control involves "conflict recognition" paradigms (Oberauer, 2005), in which participants are asked to maintain "target" stimuli that meet specific goal-relevant criteria in memory and are then asked to distinguish them from stimuli that are familiar (i.e., have also been presented recently), but do not meet these criteria. The n-back task, in which individuals are presented with a sequence of stimuli and asked to respond as to whether these stimuli meet specific order-based target criteria (e.g., being the stimulus presented two items previously in a "2-back" task), is a principal example. Conflict arises in this paradigm from having two different kinds of nontarget stimuli: "novel" stimuli, which have not been presented previously, and "lure" stimuli, which have been presented previously during the trial block but do not meet explicit criteria for being a target (e.g., a stimulus presented one trial back during the 2-back task). Although novel stimuli can be easily rejected as nontargets due to their lack of familiarity, lure stimuli present a greater challenge for rejection because their familiarity, which indicates potential status as a target, directly conflicts with the goal-relevant evidence indicating that they must be rejected. Lures are, therefore, essential for ensuring that participants are actively maintaining goal-relevant information about stimulus-context bindings rather than just responding on the basis of stimulus familiarity (Oberauer, 2005; Szmalec, Verbruggen,

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Vandierendonck, & Kemps, 2011). The n-back has emerged as one of the most commonly used paradigms in studies of goal-directed cognition because it effectively taxes an individual's ability to actively maintain and apply goal-relevant information and produces robust and well-replicated functional activation patterns in neuroimaging (Jaeggi, Buschkuehl, Perrig, & Meier, 2010; Lamichhane, Westbrook, Cole, & Braver, 2020; Owen, McMillan, Laird, & Bullmore, 2005; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009).

It is, therefore, unsurprising that the n-back has been included as one of the neuroimaging paradigms in the Adolescant Brain Cognitive Development (ABCD)® Study, an open access, multisite study that has recruited over 11,000 youth from 21 sites across the United States and follows them throughout their adolescent years (Casey et al., 2018). ABCD provides a significant resource for developmental cognitive neuroscience and allied fields due to its diverse sample (Garavan et al., 2018; Gard, Hyde, Heeringa, West, & Mitchell, 2023), comprehensive measures (spanning the domains of neuroimaging, cognition, mental and physical health, and social and environmental context), and large sample size, which is critical for identifying replicable and meaningful effects (Dick et al., 2021; Marek et al., 2022; Owens et al., 2021).

The ABCD n-back task includes two conditions that are completed during functional magnetic resonance imaging (fMRI) data collection. In the 2-back condition, participants are presented with a sequence of stimuli (pictures of faces making various affective expressions and pictures of places) and are asked to respond as to whether the stimulus is a "target," defined as being the same stimulus as was presented two trials previously. In the "0-back" task, participants are presented with similar stimulus sequences, but the criterion for a stimulus being a target is defined as that stimulus being the same as a target stimulus that is presented at the start of 0-back trial blocks (i.e., a classic recognition memory task). The more difficult 2-back task is assumed to require greater maintenance of goal-relevant information than the 0-back task, and the contrast between the 2-back and 0-back is, therefore, typically interpreted as isolating the construct of "working memory," although this characterization has been challenged (Kane, Conway, Miura, & Colflesh, 2007; Schmiedek et al., 2009).

Indeed, traditional cognitive constructs like "working memory" are often rooted in verbal theories that are loosely defined, which can make these theories difficult to test and the relevant constructs difficult to distinguish from one another (Oberauer & Lewandowsky, 2019; Verbruggen, McLaren, & Chambers, 2014; Weigard, Huang-Pollock, Brown, & Heathcote, 2018). A promising alternative to verbal theories is the use of formal cognitive process models, which provide clear, mathematically specified explanations for how individuals complete cognitive tasks (Heathcote, Brown, & Wagenmakers, 2015; Oberauer & Lewandowsky, 2019; Townsend, 2008; Turner, Forstmann, Love, Palmeri, & Van Maanen, 2017). One class of such models, termed "evidence accumulation models," have been notably successful at describing data from a wide variety of paradigms across the cognitive and neural sciences (Heathcote & Matzke, 2022; Smith & Ratcliff, 2004). The diffusion decision model (DDM) is one of the most commonly applied evidence accumulation models for two-choice tasks. It assumes that noisy evidence gradually informs a decision variable that drifts between two boundaries representing each response option, and that a response option is selected when this variable crosses one of the boundaries (Ratcliff, Smith, Brown, & McKoon, 2016). The DDM's main parameters include the drift rate (v), which indexes the efficiency with which the decision

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

variable drifts toward the correct response boundary, the boundary separation (*a*), which represents an individual's level of response caution, and the nondecision time (*Ter*), which accounts for time spent on peripheral perceptual and motor processes. The DDM and similar EEA models have been used to describe the n-back task by assuming that the decision variable is informed by goal-relevant evidence regarding stimulus-context bindings and that a response option is selected when a critical threshold for either a "target" or "nontarget" response is reached (Evans, Steyvers, & Brown, 2018; Oberauer, 2005; Pedersen et al., 2023; Weigard et al., 2024). The drift rate (v) parameter is typically greatest with novel stimuli and lowest with lure stimuli, with target stimuli intermediate ($v_{novel} > v_{target} > v_{lure}$).

More recently, integrations of evidence accumulation models with factor-analysis and structural equation modeling (SEM) methods have enabled the investigation of higher-order latent factors that drive individual differences in performance across a wide variety of cognitive tasks, including the n-back. These studies have repeatedly found that the DDM's drift rate parameter forms a strong, trait-like general factor that consistently drives individuals' performance across tasks that range from simple decisions to those that involve complex goal-directed processing (Lerche et al., 2020; Schmiedek et al., 2007; Schubert & Frischkorn, 2020; Schubert, Frischkorn, Hagemann, & Voss, 2016; Stevenson et al., 2024; Weigard, Clark, & Sripada, 2021). N-back task performance appears to be largely attributable to the same general factor (Löffler, Frischkorn, Hagemann, Sadus, & Schubert, 2024). This factor has been hypothesized to reflect an individual's task-general efficiency of evidence accumulation (EEA), or the general ability to efficiently gather goal-relevant information with which to make adaptive decisions (Weigard & Sripada, 2021). EEA may be a key contributor to individual differences in general cognitive ability (Lerche et al., 2020; Schubert & Frischkorn, 2020) and lower levels of EEA appear to be relevant to several forms of psychopathology (Heathcote et al., 2015; Sripada & Weigard, 2021; Weigard et al., 2021), especially childhood attention problems (Huang-Pollock, Karalunas, Tam, & Moore, 2012; Wiker et al., 2023; Ziegler, Pedersen, Mowinckel, & Biele, 2016).

The current study was motivated by an attempt to replicate these individual difference findings in the ABCD sample. We start by documenting a surprising pattern of associations between standard DDM drift rate parameters estimated from n-back lure trials and drift rates on other n-back trials that was not only inconsistent with these prior studies, but also suggested that applications of factor analysis or SEM to these data would be inappropriate due to violations of their assumptions. This pattern indicates that a nontrivial portion of the ABCD sample is relying largely on stimulus familiarity rather than goal-relevant evidence to complete the task. To help explain these anomalies, we next turn to a theoretical framework presented by Oberauer (2005), who posited that individuals use two general types of evidence to complete the n-back task: goal-relevant evidence related to stimulus-context bindings (e.g., the binding of the target stimulus to the context of being two spaces back in the 2-back sequence) and goal-irrelevant evidence related to stimulus familiarity (whether the stimulus was presented recently, regardless of whether it meets formal target criteria). Although goal-irrelevant evidence for stimulus familiarity can facilitate performance on trials with novel stimuli, which can be easily rejected as "nontarget" based on their unfamiliarity alone, familiarity evidence cannot be used to distinguish between target and lure stimuli. Oberauer

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(2005) proposed that, as a consequence, individuals strategically adopt an asymmetric evidence criterion: they require stronger familiarity evidence to make a "target" response than to make a "nontarget" response. This strategic shift in evidence criterion effectively reduces drift rates on target trials relative to novel trials. Furthermore, as familiarity evidence on lure trials shifts the decision process toward the incorrect ("target") response boundary, this goal-irrelevant information further reduces drift rates on lure trials relative to target trials, contributing to the typical pattern of observed n-back drift rates ($v_{novel} > v_{target} > v_{lure}$). We use this framework as the basis of a new formal model of evidence integration in conflict recognition paradigms that can be used to separate the influence of: (1) goal-relevant EEA, (2) the degree to which individuals utilize familiarity evidence during the task, and (3) the degree to which individuals adopt a more conservative criterion for making "target" responses than for making "nontarget" responses on the basis of familiarity evidence. The resulting model parameters yield several insights about mechanisms of performance on conflict recognition paradigms, their development, and their relevance to clinical symptom dimensions.

2. Methods

2.1. ABCD sample

Baseline and year 2 follow-up data were drawn from the curated ABCD Study data release version 4.0 (https://nda.nih.gov; DOI 10.15154/1,523,041). ABCD is a largescale consortium study that has recruited 11,878 children, ages 9–10, across 21 study sites and is currently following them throughout their adolescence. Participants were recruited with a sampling strategy (described in detail in: Garavan et al., 2018) designed to recruit a sample that reflects the diversity of the U.S. population as closely as possible. N-back data were collected during neuroimaging at the baseline wave and the task was thereafter repeated at 2-year intervals. The multisite ABCD Study was approved by each site's Institutional Review Board (IRB) as well as a central IRB at the University of California, San Diego. All ABCD participants provide informed consent (parents) or assent (children).

ABCD curated trial-level n-back data were available for 10,042 individuals at the baseline session and 7133 individuals at the year 2 follow-up session. Of these individuals, 8824 (88%) and 6539 (92%) met our basic inclusion criteria for data quality: accuracy rate clearly greater than chance (>0.55) and omission (i.e., nonresponse) rates of <0.25 on both the 0-back and 2-back tasks.

2.2. Independent age-diverse sample

To investigate age effects over a wider age range, we analyzed data from a separate sample of people who completed the same n-back task in a recently published study (Skalaban et al., 2022). For this study, a community sample of 175 participants between the ages of 8 and 30 were recruited from the metro area of a mid-sized U.S. city. Data collection was approved by the local IRB. All adult participants provided written consent, and all minor participants provided written assent. Of these 175 participants, the original study excluded 10 participants for

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

having full-scale IQ <70, four participants who had task performance close to chance (<0.60, following recommendations from ABCD release notes), eight participants due to an experimenter error, and three for dropping out of the study before completing a second appointment. All 150 remaining participants (76 females) met the inclusion criteria we used for the ABCD Study data, and we, therefore, included these 150 participants in the current study's analyses. The sample consisted of individuals from diverse racial and ethnic backgrounds: 39% White, 13% Black, 7% Hispanic, 9% Hispanic/Latinx, 3% Non-Hispanic/Latinx, 12% Asian, 2% American Indian, 6% other/mixed-race, 9% not reported. Additional details of this sample are reported in Skalaban et al. (2022).

2.3. ABCD n-back task

The ABCD n-back was designed to be an "emotional" variant of the traditional n-back task and uses stimuli that, for a given task block, consist of images either of faces displaying different affective expressions or images of places (e.g., buildings or parks). In both the 0-back and 2-back conditions, images are presented in a serial order for blocks of 10 trials each, of which two are target stimuli, 2–3 are lure stimuli, and the remainder are novel stimuli. All stimuli are presented for 2 s followed by a 500 ms fixation cross. Targets for the 2-back task are stimuli that are identical to the stimulus presented two spaces previously in the sequence. Target stimuli for the 0-back task are presented to participants immediately prior to the beginning of each block of 10 trials. Lure stimuli on both the 0- and 2-back tasks are stimuli that were previously presented within the block but that do not meet formal criteria for being a target. An example of a 0-back lure would be a stimulus that is presented for a second time within the block but that does not match the target stimulus presented prior to the block. Examples of 2-back lures include stimuli that were presented one, three, or four spaces back in the sequence. Participants complete eight blocks of each condition, resulting in a total of 80 trials for both the 0-back task and 2-back task. Additional information regarding task timing, stimulus images, and fMRI protocol is described in detail elsewhere (Casey et al., 2018).

2.4. Criterion measures

Given the documented importance of evidence accumulation processes for complex cognitive abilities (Lerche et al., 2020; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007) and childhood attention problems (Weigard et al., 2024; Wiker et al., 2023; Ziegler et al., 2016), we sought to index relations between model parameters and several relevant criterion variables.

General cognitive ability was indexed using raw scores of the overall Cognitive Function Composite from the NIH Toolbox Battery (Akshoomoff et al., 2013; Weintraub et al., 2013). This composite is a measure of performance across all tasks in the battery, spanning a wide array of cognitive functions including processing speed, vocabulary, reading, working memory, episodic memory, and cognitive flexibility. The composite displays evidence of excellent test-retest reliability ($r = .96$) in youth (Akshoomoff et al., 2013). Given that the n-back is considered to primarily index working memory, we also investigated specific associations of n-back model parameters with scores on the Toolbox List Sorting Working Memory Test.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This task assesses participants' ability to manipulate information in working memory by presenting them with pictures of animals or food of different sizes and asking them to repeat the list of items presented in the order of the smallest to largest item. Scores on this individual test also show good reliability (ICC = 0.86) in youth and convergent validity in predicting scores from other working memory measures (Tulsky et al., 2013).

In addition to these traditional cognitive measures, we also sought to determine whether parameters from our evidence accumulation model of the n-back were related to measures of EEA drawn from an entirely separate task: "go" trials of the ABCD stop-signal task (SST). These trials feature a simple decision task in which participants decide whether an arrow is pointing to the right or left side of the screen. We used archival DDM parameter estimates from a prior study (Weigard et al., 2024) that fit a standard DDM to ABCD SST data. The fitting procedures, which are described in detail in this prior study, were nearly identical to the current study's fitting procedures described in the section below (i.e., individual-level Bayesian estimation using priors generated with an independent subset of ABCD participants). The DDM has been widely applied to the SST within ABCD (Epstein et al., 2023) and elsewhere (Fosco, White, & Hawk, 2017; Howlett, Harlé, & Paulus, 2021; Karalunas & Huang-Pollock, 2013) to index DDM parameters in the context of simple decision-making. Here, we focus on the standard DDM's drift rate (v) parameter as an index of EEA.

As parent and teacher reports of attention problems provide complementary information (Cordova et al., 2022; Narad et al., 2015; Weigard et al., 2024), children's attention problems were indexed with both the parent-report data from the Child Behavior Checklist (Achenbach, 2001; Achenbach & Ruffle, 2000) and teacher-report data from the Brief Problem Monitor (Achenbach, McConaughy, Ivanova, & Rescorla, 2011). The "Attention Problems" scale from each of these measures includes items such as "can't concentrate," "inattentive or easily distracted," and "impulsive or acts without thinking" that are rated on a 0−2 scale. Both Attention Problems scales have good internal consistency and test-retest reliability ($\alpha/r > .80$) and parent and teacher ratings on these scales show evidence of construct validity by relating to Attention-Deficit/Hyperactivity Disorder diagnosis (Achenbach et al., 2011; Chen, Faraone, Biederman, & Tsuang, 1994; Edwards & Sigel, 2015).

2.5. DDM analyses of the n-back

All models were estimated using Dynamic Models of Choice, a suite of functions in the R language that allows for simulation and Bayesian estimation of the DDM and other models of choice response time (Heathcote et al., 2019). We created informative priors for DDM parameters in each ABCD condition and wave following prior work (Weigard, Matzke, Tanis, & Heathcote, 2023). Specifically, we removed the data from a subset of 200 singleton ABCD participants who had data that met inclusion criteria at both the baseline and year 2 waves, fit a hierarchical Bayesian model to this independent subsample, and then obtained informative priors for each model parameter by fitting truncated normal distributions to the full distribution of all individual-level posterior samples from these hierarchical models (all priors are displayed in Tables S1 and S2). This prior-generation strategy provides constraints on individuals' parameter estimates without the drawbacks of using a fully hierarchical modeling

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

approach with all participants (e.g., demands on computing resources and nonindependence of individual-level estimates from hierarchical Bayesian models).

Before model estimation, we removed trials with response times < 200 ms as fast guesses. The standard DDM fit to these data included drift rate parameters for each n-back trial type ($v_{novel,}$ v_{target}, v_{lure}), boundary separation (a), start point (z), nondecision time and its between-trial variability (*t0*, *st0)*, as well as a "go failure" (p_{gf}) parameter to account for omissions due to inattention. The p_{gf} parameter assumes that, on a given trial, a participant will fail to respond due to an attentional lapse with probability p_{gf}. For example, a participant with $p_{gf} = 0.08$ will, on average, show omissions due to inattention on 8% of trials. Omissions that alternately occur because participants' responses fall outside of the 2-s response window were also accounted for in the model's likelihood function using methods developed in prior work (Damaso et al., 2021). Specifically, we assumed that if a given set of DDM parameters predicts that a subset of RTs would be greater than 2 s, the likelihood of omissions would be proportional to the size of this subset of RTs. If a greater proportion of omissions is observed than would be predicted by such RTs falling outside of the response window, the p_{gf} parameter would account for these excess omissions. We did not estimate the other DDM between-trial variability parameters (*sz*, *sv*) due to the difficulties of doing so without large numbers of trials (Lerche, Voss, & Nagler, 2017). Similarly, although the ABCD n-back contains stimuli from different types of image categories, including places and facial expressions of different emotions, we did not attempt to model performance differences between these categories because doing so would lead to many experimental conditions having only a handful of trials, which would likely cause substantial difficulties with estimating model parameters. The reparameterized model structure described below was identical to the standard DDM model structure except where described below in the Results section.

Sampling for all model parameters was conducted with the differential evolution Markov chain Monte Carlo method (Turner, Sederberg, Brown, & Steyvers, 2013) and convergence was determined by the Gelman−Rubin statistic falling below 1.10 (Gelman & Rubin, 1992). Model fit was assessed with posterior predictive plots (Supplementary Materials) and the medians of posterior distributions for DDM parameter values were retained as point estimates for subsequent analyses.

2.6. Data visualization and inferential analyses

All data visualization and analysis was conducted within R (R Core Team & others, 2013) and all code for this study is openly available at: https://osf.io/tmzye/. Bivariate relations in the ABCD data were visualized with smoothed scatterplots using the smoothScatter() function, which produces smoothed density representations of data obtained with a two-dimensional kernel density estimate. Smoothed scatter plots are useful when the number of individual participants is large, as this causes points in traditional scatterplots to overlap with one another and obscure trends in the data. 95% confidence intervals (CIs) for relevant Cohen's d and Pearson's r values were estimated for the ABCD sample using a clustered bootstrapping procedure that accounts for the nesting of individuals within families and ABCD sites. For the independent age-diverse sample, a standard (individual-level) bootstrapping procedure was

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

used to estimate 95% CIs for *d* and *r* values. Effects of age in the independent age-diverse sample were estimated by regressing n-back model parameters on both linear and quadratic age. Analyses of model parameter relations with criterion measures in ABCD were conducted with mixed effects models using the *gamm4* and *MiMin* R packages that accounted for nesting by family and ABCD site and included age, sex, family income, highest parental education, parent marital status, race, and ethnicity as covariates. Effect size (marginal r^2) for these relations was estimated using likelihood ratio tests.

3. Results

3.1. Lure drift rates from the ABCD n-back task show atypical bivariate relations that suggest the presence of familiarity-based responding

Fig. 1 displays smoothed scatterplots and correlation coefficients for associations between standard DDM drift rate parameters estimated from target, lure, and novel n-back trial conditions in the baseline (blue, below diagonal) and year 2 (red, above diagonal) data sets. Several patterns in these associations are notable. In general, drift rates for target and novel trials (e.g., second and third rows/columns) show moderate to strong associations with one another and with themselves across levels of n-back load. The scatterplots for these associations also consistently display the shape of a bivariate normal distribution, suggesting that they can be easily described by models that assume simple linear associations with normally distributed errors (e.g., factor analysis models). These observations are consistent with prior work (Lerche et al., 2020; Schmiedek et al., 2009; Schubert & Frischkorn, 2020; Schubert et al., 2016; Stevenson et al., 2024; Weigard et al., 2021), identifying a general drift rate factor that influences performance across many task conditions.

However, drift rate estimates from lure trials (first and fourth rows/columns) show weaker and more variable associations with those from other trial types and most of these associations fail to display the shape of a bivariate normal distribution, instead following an atypical v-shaped pattern. This is most apparent in the baseline session, in which lure drift rates from 0-back and 2-back trials show very weak, or even negative, correlations with 2-back target drift rates. The corresponding associations in the year 2 data are slightly stronger, but the v-shaped pattern is still apparent. Beyond being inconsistent with the hypothesis of a general drift rate factor, the atypical shape of these bivariate associations also suggests that standard factor analysis models would, more generally, be inadequate to describe these data, given that the associations clearly violate the distributional assumptions of these models.

The v-shaped pattern present across lure trial drift rate associations indicates that individuals with lure drift rates greater than 0 generally display the expected positive relation with other drift rates but that individuals with lure drift rates below 0 (indicating a pattern of evidence accumulation that moves, on average, toward the incorrect "target" response boundary on lure trials) generally display the opposite relation. This pattern can plausibly be explained by the use of information about stimulus familiarity (Oberauer, 2005) as opposed to goal-relevant stimulus features (i.e., whether the stimulus meets explicit criteria for being a tar-

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

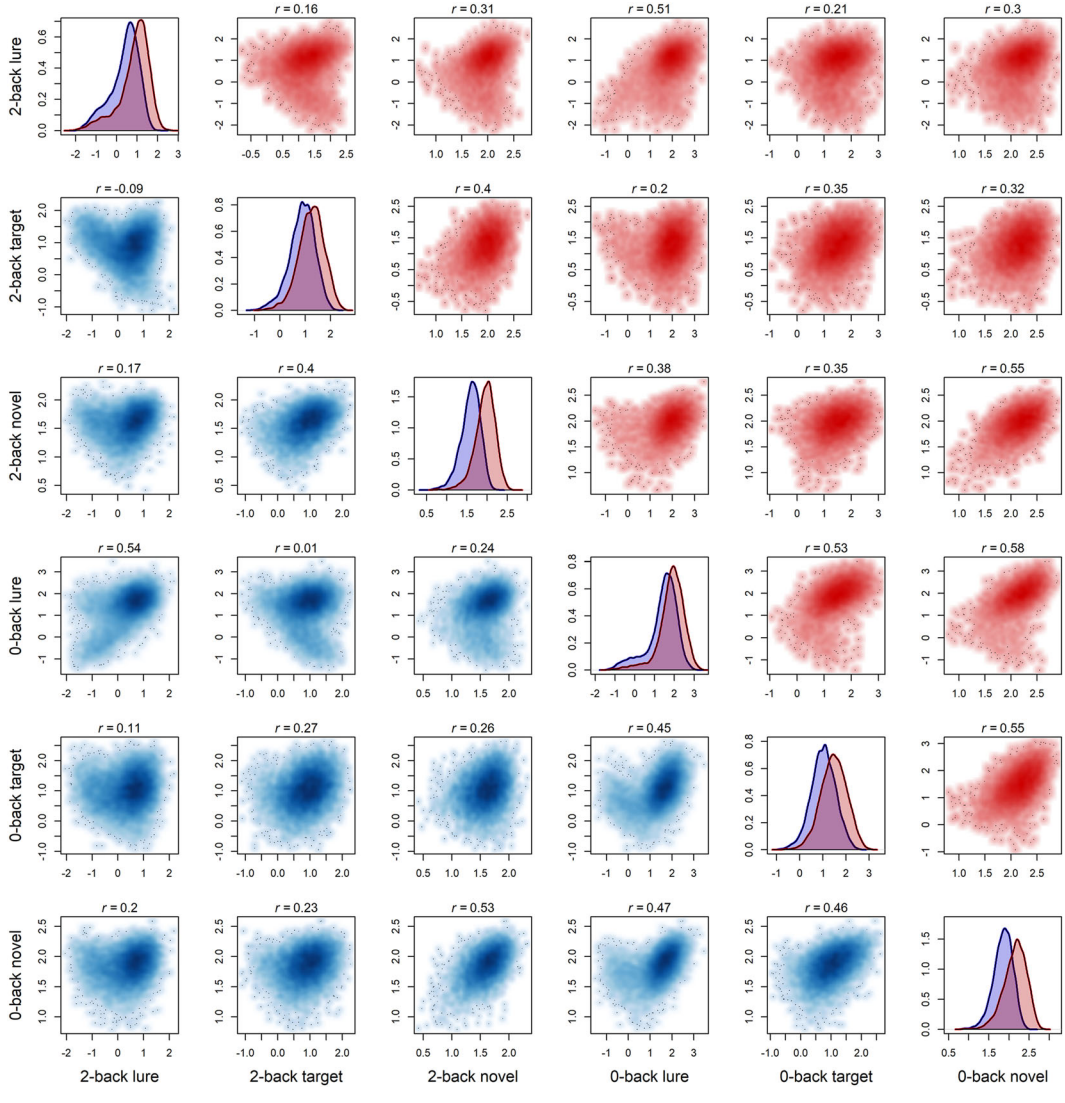

Fig. 1. Smoothed scatterplots of associations between drift rates from ABCD n-back task conditions in the baseline (blue; below the diagonal) and year 2 (red; above the diagonal) data. Kernal density plots within the diagonal compare the density of drift rate parameter estimates for the baseline (blue) and year 2 (red) data. Correlation coefficients (r) for each association are displayed above each scatterplot.

get). If individuals are responding mostly based on familiarity, their lure drift rate would be strongly negative, as lure stimuli would incorrectly be identified as targets, while their target and novel trial drift rates would continue to be positive, as the familiarity strategy is effective for trials in these conditions.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In the following section, we formalize the assumption that individuals' responses on the n-back task are determined by a mixture of the efficiency of their evidence accumulation related to goal-relevant stimulus features (i.e., EEA) and, alternately, their use of goal-irrelevant evidence related to familiarity.

3.2. A novel formal model of evidence integration on the n-back task provides a principled decomposition of individuals' performance

We refit the DDM to the ABCD baseline and year 2 data using all modeling procedures described above, but with one exception: rather than separately estimating the drift rates for each of the three trial conditions, we assumed that these rates were differentially driven by separate parameters with distinct process interpretations. We began with Oberauer's (2005) assumption that individuals use a combination of goal-relevant evidence related to stimulus-context bindings and goal-irrelevant evidence related to stimulus familiarity. We assumed that efficiency of accumulating goal-relevant evidence, represented by parameter e, drives drift rates to the same degree across all three trial types, but that the overall rate for each trial type is differentially impacted by goal-irrelevant evidence related to stimulus familiarity. Oberauer (2005) posits that familiarity evidence from novel stimuli is typically treated as strong evidence for a stimulus being a "nontarget," but that familiarity evidence from repeated (target and lure trial) stimuli necessarily provides more ambiguous evidence for a "target" response as targets and lures cannot be distinguished based on familiarity alone. Hence, estimates of familiarity evidence on trials with novel stimuli (f_n) were allowed to differ from estimates of familiarity evidence on trials with repeated (target and lure) stimuli (f_r). Both types of familiarity parameter derive from the same underlying familiarity evidence dimension and differ only due to a participant's strategic use of a familiarity-evidence criterion.

As "target" is the correct decision for target stimuli, positive values of f_r would enhance EEA toward the correct decision boundary for target trials. However, as "nontarget" is the correct decision for lure stimuli, f_r has an opposing effect on lure trials, driving the decision process toward the incorrect decision boundary. Therefore, an increase in f_r should cause an increase in the observed drift rate on target trials and a corresponding decrease of the same magnitude in the observed drift rate on lure trials. On the basis of these assumptions, we fit a model in which the observed drift rates across novel, target, and lure trials were determined by the following set of equations (see Supplementary Materials for a detailed derivation):

$$v_{target} = e + f_r$$

$$v_{novel} = e + f_n$$

$$v_{lure} = e - f_r$$

The e, f_n, and f_r parameter estimates allow for the theoretical propositions of Oberauer (2005) to be directly measured: accumulation of goal-relevant evidence related to stimulus-context bindings (e), accumulation of the strong evidence for a "nontarget" response on trials with unstudied (novel) stimuli (f_n), and accumulation of the more ambiguous evidence for a "target" response on trials with studied (target and lure) stimuli (f_r). However, we then

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

applied two additional transformations to the f_n and f_r parameters to better index theoretically meaningful processes related to individuals' use of familiarity evidence.

As individuals do not know, at the beginning of a trial, whether a stimulus is novel or repeated, it is implausible that they would be able to deweight familiarity evidence from novel, relative to repeated, stimuli during the processing of the trial. Rather, we contend that the tendency for f_n to be greater than f_r is better characterized as the result of a "stimulus bias" (White & Poldrack, 2014) in which individuals adopt a more conservative evidence criterion (i.e., require much stronger evidence) for making a "target" response on the basis of familiarity evidence than for making a "nontarget" response on the basis of familiarity evidence. Criterion shifts can be measured by subtracting drift rates across opposing stimulus categories (Ging-Jehli, Arnold, Roley-Roberts, & DeBeus, 2022; Kloosterman et al., 2019), and we, therefore, adopt the following transformation to index individuals' familiarity evidence criterion (*fc*):

$$fc = f_n - f_r$$

Positive values of this familiarity criterion (*fc*) parameter indicate the expected adoption of a more conservative criterion for "target" as opposed to "nontarget" responses, whereas negative values indicate the opposite pattern and values close to 0 indicate a neutral familiarity-evidence criterion. In addition to criterion settings, combined f_n and f_r values are also determined by the overall strength of the impact of goal-irrelevant familiarity evidence on an individual's decision process (*fs*), which we index as the average of the familiarity rate parameters:

$$fs = \frac{f_n + f_r}{2}$$

This familiarity strength (*fs*) parameter is likely jointly determined by an individual's ability to accumulate the goal-irrelevant evidence related to stimulus familiarity and by their strategic choice of how strongly to weight familiarity evidence relative to goal-relevant evidence specific to stimulus-context bindings. Although we cannot tease apart these two sources of variation in the current model, we assume that individuals whose strategies place a greater emphasis on goal-irrelevant familiarity-based evidence will show higher levels of *fs*.

In sum, this novel model of evidence integration on the n-back task yields three clearly interpretable psychological-process parameters: the efficiency with which individuals accumulate goal-relevant evidence (*e*), the strength of the impact of goal-irrelevant familiarity evidence on an individual's decision process (*fs*), and shifts in participants' criterion for making a "target" versus "nontarget" response on the basis of familiarity evidence (*fc*). These parameters also make principled predictions about individuals' observed drift rates across n-back conditions. An individual who is effectively utilizing goal-relevant evidence across all trials (high *e*), de-emphasizing the use of goal-irrelevant familiarity evidence (low *fs*), and setting a more conservative familiarity evidence criterion for "target" relative to "nontarget" responses (high *fc*) would be expected to display the commonly observed pattern of $v_{novel} > v_{target} > v_{lure}$ (Oberauer, 2005). At the other end of the spectrum, an individual who is completing the task using familiarity evidence alone ($e = 0$; high *fs*) and makes no criterion shift to account for the ambiguity of familiarity evidence on target and lure trials ($fc = 0$) would have identical

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
Guide to the interpretation of the conflict recognition evidence integration model parameter values

Model parameter	Definition	Value range	Value interpretation
Efficiency (*e*)	Efficiency of accumulating evidence relevant to the explicit task goal	$e > 0$	Individual is utilizing goal-relevant evidence to make choices
		$e = 0$	Individual is not considering goal-relevant evidence
Familiarity Strength (*fs*)	Captures ability to utilize familiarity evidence as well as the strategic choice of how strongly to weight familiarity evidence	$fs > 0$	Individual is utilizing goal-irrelevant familiarity evidence to make choices
		$fs = 0$	Individual is not considering goal-irrelevant familiarity evidence
Familiarity Criterion (*fc*)	The degree to which individuals adopt a more conservative criterion for making "target" relative to "nontarget" responses	$fc > 0$	Individual requires more familiarity evidence to make a "target" response than a "nontarget" response
		$fc < 0$	Individual requires more familiarity evidence to make a "nontarget" response than a "target" response

observed drift rates on target and novel trials and a negative observed drift rate of the same absolute value on lure trials. This latter pattern would characterize the set of individuals who drive the v-shape in the bivariate associations of lure drift rates with target and novel drift rates (Fig. 1): those who have strongly negative rates on lure trials but strongly positive rates on target and novel trials. Table 1 provides a reference guide for interpretation of the values of each parameter.

Posterior predictive plots of model fit are displayed in Figs. S1 and S2 alongside comparable plots for the standard DDM. As detailed in Supplementary Materials, these plots indicated that both the standard DDM and the novel model provided an adequate fit to the data across all samples. Because our novel model of evidence integration is essentially a three-parameter transformation of the original three drift rate parameters, we did not expect the new model to display better fit to individuals' choice and RT data than the standard DDM. Therefore, we do not engage in a formal model comparison using metrics that assess models' fit to these individual-level data features (e.g., information criteria). We emphasize that the reparameterization is not meant to provide a better description of choice and RT data at the individual level. Rather, it is intended as a measurement model that provides a more theoretically principled and interpretable description of individual differences in cognitive mechanisms by positing a set of three underlying parameters that can account for individuals' observed drift rates on target, nontarget, and lure trials and can explain their patterns of interrelations, including the unexpected v-shaped relation between target and lure drift rates.

Group distributions of model parameter estimates for the baseline (blue) and year 2 (red) data sets are displayed in Fig. 2. Absolute values of *e* and *fs* were generally above 0 across n-back load conditions, indicating that most individuals utilize both goal-relevant evidence (*e*) and goal-irrelevant evidence related to stimulus familiarity (*fs*) to generate responses in the n-back task. Most values of *fc* were also positive, suggesting that individuals generally display

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

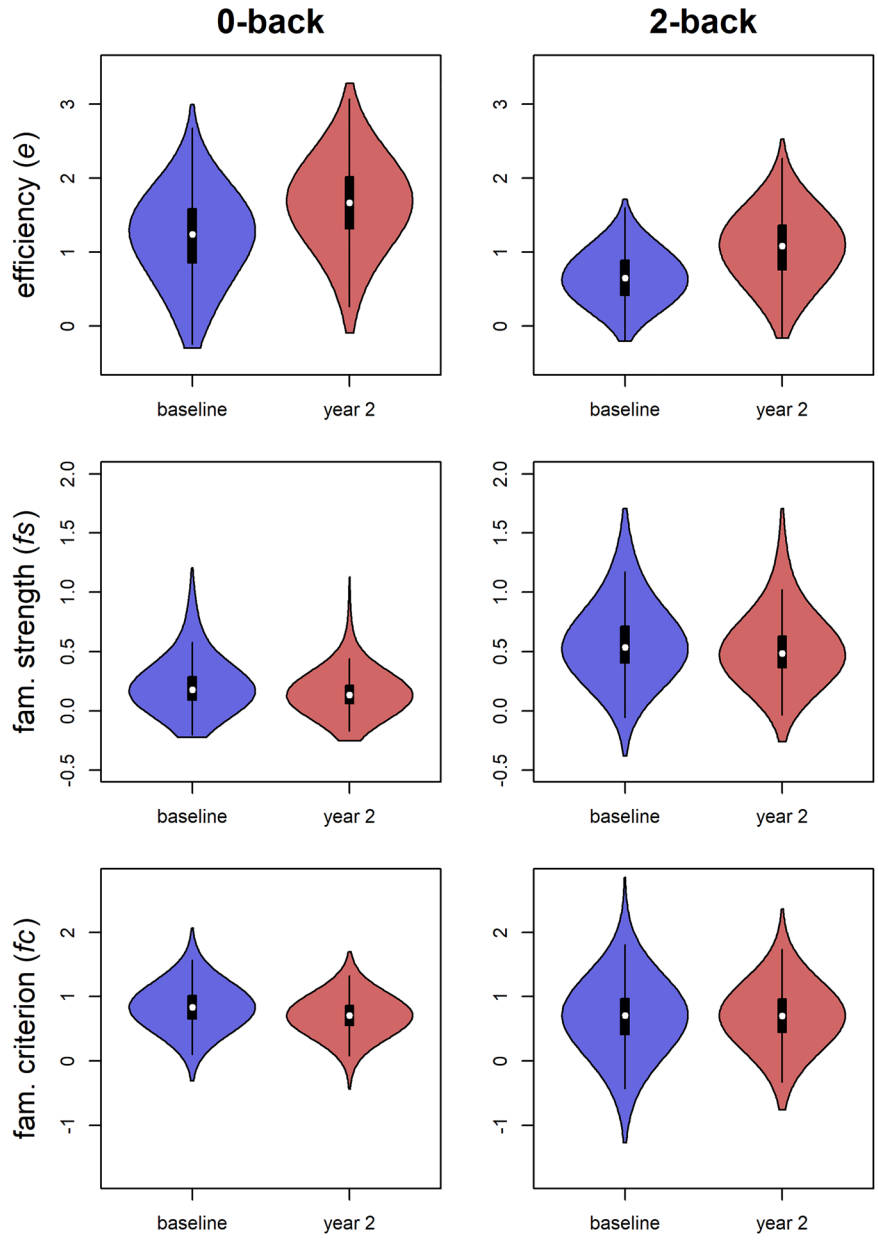

Fig. 2. Distributions of ABCD participants' parameter estimates, including efficiency of evidence accumulation for the goal-relevant choice (*e*), the strength of goal-irrelevant evidence related to familiarity (*fs*), the shifts in the criterion for familiarity evidence (*fc*), across n-back load conditions and across the baseline (blue) and year 2 (red) waves. Distributions are represented as violin plots, which are standard box plots that are embedded within kernel density plots of the same data.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the expected criterion shift in which they require stronger familiarity-based evidence to make a "target" response than to make a "nontarget" response, given the ambiguity of familiarity evidence for discerning target from lure stimuli. The 2-back task displayed consistently lower *e* (baseline $d = -1.18$ CI $= -1.13$ to -1.22; year 2 $d = -1.19$, CI $= -1.14$ to -1.25) and higher *fs* (baseline $d = 1.51$, CI $= 1.44$–1.59; year 2 $d = 1.84$, CI $= 1.77$–1.93) than the 0-back, indicating poorer goal-relevant evidence quality and greater use of goal-irrelevant familiarity evidence on the more difficult task. The 2-back task displayed moderately lower *fc* than the 0-back task during the baseline time point ($d = -0.40$, CI $= -0.35$ to -0.46) but there were no discernible load-related *fc* differences in year 2 ($d = -0.02$, CI $= -0.07$ to 0.10).

Fig. 3 illustrates the associations of the *e*, *fs*, and *fc* parameters with the observed values of v_{lure} from the standard DDM fit. Across both n-back load levels and both study waves, observed v_{lure} was strongly positively correlated with *e*, strongly negatively correlated with *fs*, and positively correlated with *fc*. For the latter two parameters related to the use of familiarity evidence, this relation appears to largely be driven by individuals with $v_{lure} < 0$, who tend to show the highest influence of familiarity-based evidence on the decision process (i.e., higher *fs*) and who are less likely to show an adaptive criterion shift when considering familiarity evidence (i.e., *fc* close to, or below, 0).

We next aimed to assess whether individuals' *e* was associated with the parameters that account for the influence of familiarity evidence and familiarity criterion shifts (Fig. 4). Associations between *e* and *fs* were generally negative across both load conditions and study waves, suggesting that individuals who were worse at accumulating goal-relevant evidence displayed greater use of goal-irrelevant familiarity evidence. Plots of these associations show similarity to an "L" or "inverse T" shape, suggesting that most individuals who rely strongly on familiarity evidence do so because they are not effectively using goal-relevant evidence at all (*e* close to 0). Associations between *e* and *fc* were notably weaker and less consistent.

3.3. Both EEA and the use of goal-irrelevant familiarity evidence display generality across levels of working memory load

We next sought to assess whether parameters from our novel model of n-back evidence integration showed evidence of task-generality by investigating how strongly individual differences in these parameters were correlated across n-back load levels (Fig. 5). The *e* parameter displayed strong relations across n-back load levels during both waves, consistent with a task-general factor for efficiency of goal-relevant evidence accumulation (i.e., EEA). Notably, we found that, in each load condition, *e* estimates were nearly perfectly correlated with drift rate estimates averaged across target, lure, and novel trials in the standard DDM fit (0-back and 2-back r both $> .99$). This close correspondence is directly predicted by the evidence integration model because solving for average *v* indicates that the *e* parameter is the dominant influence (average $v = 3e + f_n$). Therefore, average drift rates from the standard DDM in this sample are likely to index the same general efficiency construct, consistent with prior reports of strong correlations of average drift rates across n-back load conditions in another subset of ABCD data (Weigard et al., 2024).

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

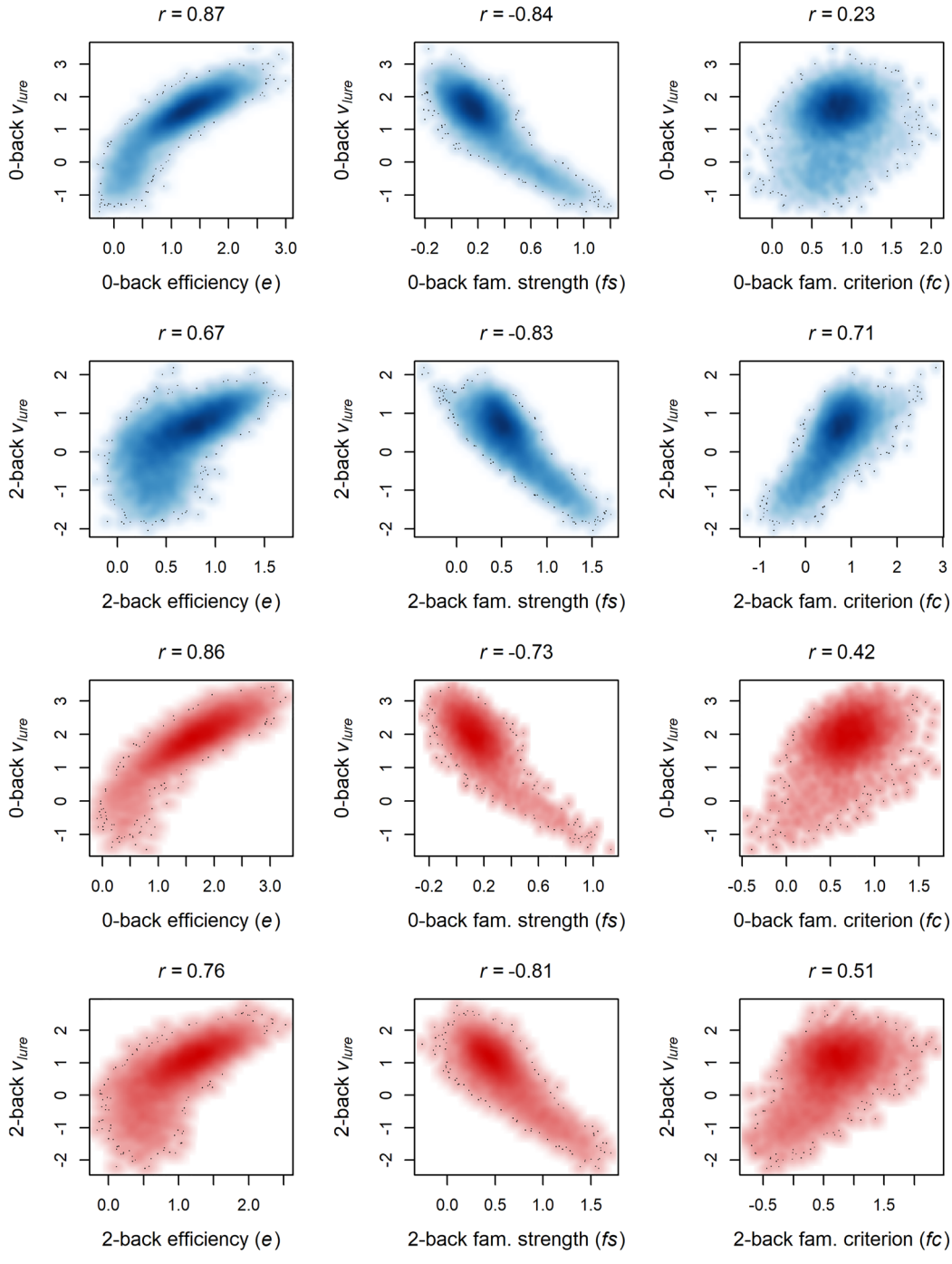

Fig. 3. Relations between parameters from the novel DDM parameterization and observed lure drift rates (v_{lure}) from fits of the traditional DDM for the baseline (blue) and year 2 (red) sample.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

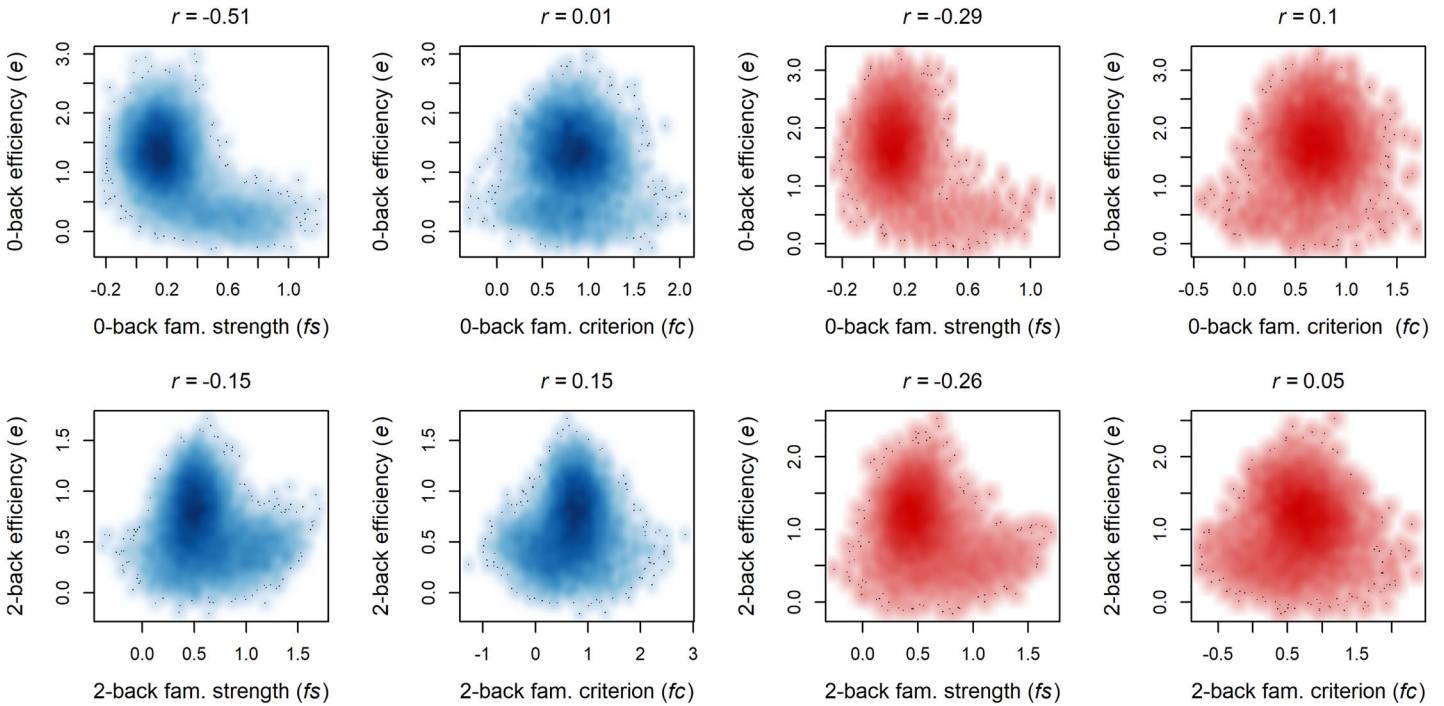

Fig. 4. Associations between individuals' efficiency of accumulation for goal-relevant evidence (*e*) and parameters representing the strength of familiarity evidence (*fs*) shifts in the criterion for familiarity evidence (*fc*) for the baseline (blue) and year 2 (red) sample.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

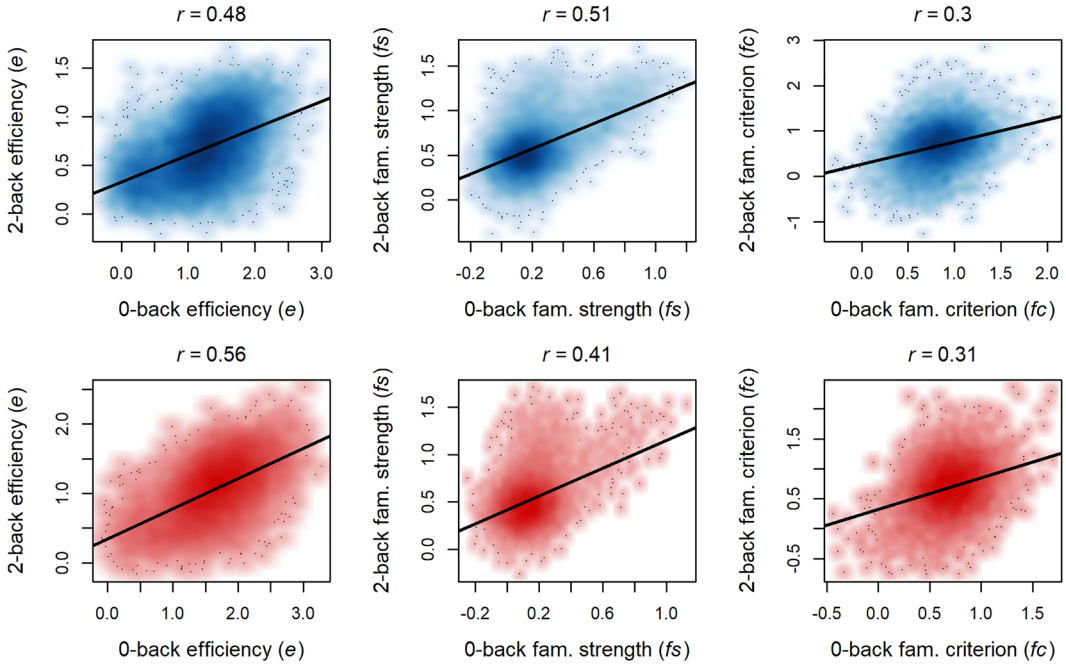

Fig. 5. Plot of parameter correlations across levels of working memory load for the baseline (blue) and year 2 (red) samples. Black lines indicate linear regression lines for the associations of parameters across load levels.

The *fs* and *fc* parameters also consistently displayed moderate-to-strong associations across n-back load levels, suggesting that individuals' utilization of evidence related to stimulus familiarity and their criterion settings for making decisions based on familiarity evidence both show generalization across n-back conditions.

3.4. Only the e *parameter shows evidence for robust developmental effects in ABCD and in an independent sample*

The distributions in Fig. 2 suggest that the *e* parameter shows more apparent differences from the baseline to year 2 waves than the other two parameters. Effect sizes (*d*) for wave effects were estimated for individuals with complete data across both waves ($n = 5284$) to investigate the size of these trends. The *e* parameter showed large increases from the baseline to the year 2 wave (0-back $d = 0.87$, CI = 0.81–0.93; 2-back $d = 1.13$, CI = 1.06–1.20). Small to moderate decreases in year 2 were observed in *fs* (0-back $d = -0.40$, CI = -0.36 to -0.43; 2-back $d = -0.21$, CI = -0.17 to -0.25). The *fc* parameter for the 0-back task displayed a moderate decrease in year 2 ($d = -0.51$, CI = -0.47 to -0.55) and only trivial age-related changes were observed in *fc* for the 2-back ($d = 0.03$, CI = -0.01 to 0.06).

As the year 2 administration was ABCD participants' second exposure to the task, it is difficult to discern whether these observed wave differences are developmental in nature or related to other factors, such as practice effects or the specific contexts of the two visits. In

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Results from regression models regressing each n-back model parameter on linear and quadratic effects of age

Model parameter	Regression coefficient	Est.	SE	*p*
0-back Efficiency (*e*)	Intercept	−0.529	0.484	.276089
	Age	0.251	0.055	1.12E-05
	Age^2	−0.006	0.001	9.48E-05
2-back Efficiency (*e*)	Intercept	−0.883	0.399	.028187
	Age	0.218	0.045	3.72E-06
	Age^2	−0.005	0.001	3.91E-05
0-back Fam. Strength (*fs*)	Intercept	0.177	0.142	.216397
	Age	0.001	0.016	.956735
	Age^2	0.000	0.000	.583043
2-back Fam. Strength (*fs*)	Intercept	0.655	0.233	.005591
	Age	−0.009	0.027	.743035
	Age^2	0.000	0.001	.848997
0-back Fam. Criterion (*fc*)	Intercept	1.111	0.302	.000324
	Age	−0.047	0.034	.172890
	Age^2	0.001	0.001	.196826
2-back Fam. Criterion (*fc*)	Intercept	1.318	0.412	.001711
	Age	−0.073	0.047	.122735
	Age^2	0.002	0.001	.166422

Abbreviations: Est., unstandardized regression coefficient; SE, standard error.

addition, as the ABCD waves covered a narrow age range, ABCD cannot currently be used to gauge the overall developmental trajectory of the process model parameters. To address these limitations, we fit the novel model to an independent age-diverse sample of 150 individuals, spanning middle childhood through young adulthood, who completed the same ABCD version of the n-back task.

Posterior predictive plots (Fig. S3) indicated that the model provided an excellent description of the n-back data in the independent age-diverse sample. Fig. 6 displays scatterplots of 0-back (green) and 2-back (orange) model parameters from this sample by participants' age. Key features of the above results were replicated in this independent sample, including findings of decreased *e* ($d = -1.09$, CI = −0.90 to −1.30) and increased *fs* ($d = 1.82$, CI = 1.61–2.09) in the 2-back condition relative to the 0-back condition. There were no significant load-related differences in *fc* ($d = -0.15$, CI = −0.34 to 0.04). Participants in the independent age-diverse sample also displayed significant and positive cross-load correlations in all parameters (*e* $r = .39$, CI = 0.23–0.52; *fs* $r = .35$, CI = 0.21–0.47; *fc* $r = .27$, CI = 0.15–0.38), providing further evidence for task-generality. Linear regressions (Table 2; lines in Fig. 6) revealed significant linear and quadratic effects of age on *e* across both load levels but no significant effects of age on the *fs* or *fc* parameters. These results suggest that efficiency of goal-relevant EEA shows a strong maturational trend from childhood through individuals' mid-20s but that the observed age-related differences in ABCD *fs* and *fc* estimates are instead likely driven by alternative factors, such as practice effects or experimental context.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

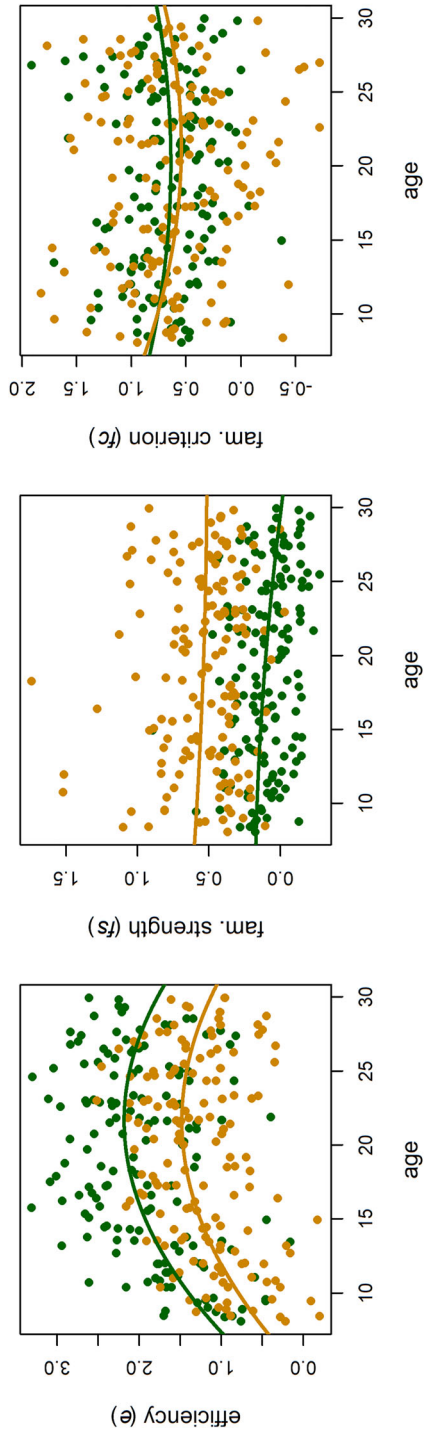

Fig. 6. Plots of 0-back (green) and 2-back (orange) model parameter estimates in the independent age-diverse sample by age. Lines indicate regressions of each parameter on linear and quadratic effects of age.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.5. Parameter e displays the best test-retest reliability and the strongest relations with relevant cognitive and clinical criterion variables

Test-retest reliability correlations (r) across the 2-year gap in ABCD were estimated within the 5284 individuals who had data from both time points to gauge the temporal stability of parameters. Given that all parameters showed evidence of generality across n-back load levels, we estimated reliability for parameters at each load level as well as for the average across levels, which we assumed to index the task-general trait. Reliability for the *e* parameter was close to, or above, the threshold typically considered "acceptable" for reliability (0-back $r = .47$, CI = 0.44–0.49; 2-back $r = .58$, CI = 0.56–0.61; average $r = .60$, CI = 0.57–0.62), which is notable given the 2-year measurement interval and the large degree of developmental change apparent across this interval. Reliability for *fs* (0-back $r = .29$, CI = 0.24–0.34; 2-back $r = .34$, CI = 0.30–0.37; average $r = .40$, CI = 0.34–0.44) and for *fc* (0-back $r = .18$, CI = 0.14–0.21; 2-back $r = .20$, CI = 0.17–0.24; average $r = .23$, CI = 0.19–0.27) was significantly greater than 0 but systematically lower than reliability for *e*. Therefore, although all parameters show some level of stability across time, EEA for goal-relevant information appears to show the greatest evidence of being a relatively stable trait.

To gauge the criterion validity of each parameter, we next investigated their associations with measures of several cognitive and clinical constructs that have been previously associated with n-back task performance and EEA. In the baseline sample of ABCD, we used generalized additive mixed models that accounted for participants' family, study site, and relevant demographics to estimate the size of associations between the n-back model parameters and each measure. As the *e* parameter was correlated with the *fs* and *fc* parameters (Fig. 4), we include all three parameters in each regression simultaneously to ensure that relations between parameters and a given criterion variable were not attributable to parameter intercorrelations. Given the evidence for task-generality in each parameter, we focus here on parameter estimates that are averaged across the 0-back and 2-back conditions as this increases power for detection of associations with the task-general construct. However, we note that the pattern of results is not substantially different when parameters from each n-back load condition were investigated separately (Tables S3 and S4).

Results involving cognitive variables (Table 3a) indicate that the *e* parameter displays moderate-to-large positive associations with general cognitive ability (20.1% of variance explained), performance on a working memory task completed outside of the scanner (8.1%), and EEA measured as drift rate from the simple "go" choice condition of the ABCD SST (11.1%). The *fs* and *fc* parameters did not display substantial relations with either NIH Toolbox measure ($\leq$0.2% of the variance explained) but were both positively related to EEA measured on the SST (*fs* = 4.9%, *fc* = 1.3%), although these relations were both considerably smaller than those of *e*.

Results involving clinical ratings of children's attention problems (Table 3b) indicated that the *e* parameter displays small-to-moderate negative associations with parent- and teacher-reported attention problems (2.2% of parent-report variance explained, 4.6% of teacher-report variance explained), while the *fs* parameter shows statistically significant, but very small (0.1−0.4% of variance explained), associations with attention problems in the same direction as *e*.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3
Results from generalized additive mixed models regressing three relevant cognitive criterion variables (a) and parent- and teacher-reported ratings of attention problems (b) on n-back model parameters averaged across load conditions. Bolded values in parentheses next to each regression parameter estimate indicate the proportion of variance explained (r-squared) for a given variable. All results are from models that accounted for nesting by family and ABCD site and included age, sex, family income, highest parental education, parent marital status, race, and ethnicity as covariates.

(a) Cognitive criterion variables	NIH Toolbox Cognitive Function Composite ($n = 7782$)			NIH Toolbox List Sorting Working Memory Task ($n = 7807$)			Drift Rate (EEA) from the ABCD SST ($n = 7052$)		
	Est. (r^2)	SE	p	Est. (r^2)	SE	p	Est. (r^2)	SE	p
Intercept	46.00	1.26	<.01E-99	75.21	2.02	<.01E-99	.36	0.11	.001261
Efficiency (*e*)	9.43 (**.201**)	0.21	<.01E-99	9.02 (**.081**)	0.34	<.01E-99	.57 (**.111**)	0.02	<.01E-99
Familiarity Strength (*fs*)	1.63 (**.002**)	0.46	.000399	−1.38 (<**.001**)	0.75	.065178	.80 (**.049**)	0.04	4.31E-80
Familiarity Criterion (*fc*)	.50 (<**.001**)	0.29	.088390	−1.13 (**.001**)	0.47	.017250	.26 (**.013**)	0.03	1.89E-22
(b) Clinical ratings of attention problems	Parent-Reported Attention Problems ($n = 7922$)			Teacher-Reported Attention Problems ($n = 2756$)					
	Est. (r^2)	SE	p	Est. (r^2)	SE	p			
Intercept	3.11	0.63	9.54E-07	4.94	0.93	1.26E-07			
Efficiency (*e*)	−1.41 (**.022**)	0.11	1.10E-39	−1.79 (**.046**)	0.15	3.24E-30			
Familiarity Strength (*fs*)	−.63 (**.001**)	0.23	.006561	−1.07 (**.004**)	0.34	.001874			
Familiarity Criterion (*fc*)	−.52 (**.002**)	0.15	.000392	−.36 (**.001**)	0.21	.091783			

Note. Models accounted for study site and family as nested random intercept effects (1|site/family). Likelihood ratio tests were used to estimate marginal r^2 for each n-back model parameter.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Therefore, although the *e* parameter shows robust associations that are consistent with prior work on EEA's task-generality and its importance for complex cognitive abilities (Lerche et al., 2020; Schmiedek et al., 2007) and attention problems (Weigard et al., 2024; Ziegler et al., 2016), the other two parameters show comparatively little evidence for criterion validity. Taken together with evidence of better test-retest reliability for *e*, these results suggest that EEA is a relatively stable trait that shows relevance for individual differences in cognition and clinical dimensions, while the use of goal-irrelevant familiarity-based evidence—despite being essential to describing n-back performance—is less likely to be a stable trait that is relevant to the study of individual differences.

4. Discussion

Formal cognitive process models such as the DDM (Ratcliff et al., 2016) present a promising method for characterizing mechanisms that drive individual and developmental differences in goal-directed cognitive functioning. Previous work in this area has consistently shown that the DDM's drift rate parameter forms a task-general factor that appears to be a key driver of individual differences in cognition (Lerche et al., 2020; Löffler et al., 2024; Schmiedek et al., 2007) and that has been hypothesized to reflect individuals' general efficiency for accumulating goal-relevant evidence, or "EEA" (Weigard & Sripada, 2021). The current study documents an unexpected pattern of associations involving the n-back task, a popular conflict recognition paradigm, from the ABCD Study, a large and diverse sample of youth, that initially appears inconsistent with these findings: drift rates on lure trials show atypical v-shaped associations with other drift rates, suggesting many individuals are responding based primarily on familiarity rather than goal-relevant stimulus features. This pattern represents both a theoretical puzzle and a practical problem for standard methods that characterize individual differences in terms of multivariate normal distributions.

Here, building on Oberauer (2005), we proposed a new model of the cognitive process that combines two different knowledge representations—general stimulus familiarity and the match of a stimulus to the goal of the n-back and other conflict recognition tasks—to provide the input to a DDM. The model solves both the theoretical puzzle and methodological problem in terms of three parameters: efficiency of accumulation for goal-relevant evidence (*e*, equivalent to EEA), the impact of goal-irrelevant evidence related to the stimulus familiarity on the decision process (*fs*), and the degree to which individuals adopt a more conservative criterion for making "target" responses than "nontarget" responses on the basis of familiarity evidence (*fc*). Although all three parameters differ systematically between the 0-back and 2-back conditions, individual differences in each parameter show moderate to strong correlations across conditions, suggesting that these processes generalize across n-back tasks of different difficulty levels. We further found that only the *e* parameter exhibits strong test-retest reliability, robust correlations with relevant cognitive and clinical criterion variables, and clear evidence of maturational change, both in ABCD and in an independent age-diverse data set. Overall, our results demonstrate the centrality of EEA (measured in this model as the *e* parameter) as a reliable, clinically relevant individual difference dimension. They also

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

suggest the need for care in disentangling EEA from concurrently operating familiarity-based processing that is idiosyncratic and does not represent a task-general individual difference dimension.

This work provides a novel extension of the theory first proposed by Oberauer (2005) in an earlier study involving adults and leverages this theory to explain the pattern of data present in ABCD's sample of youth. Oberauer (2005) posits that individuals completing the n-back, and similar "conflict recognition" tasks, rely on a combination of evidence related to bindings between stimuli and relevant contexts ("is this the same item that was presented two spaces back?") and evidence related to familiarity ("have I seen this item recently?"). Novel stimuli can be easily rejected on the basis of their unfamiliarity, but target and lure stimuli cannot be distinguished on the basis of familiarity evidence alone. Individuals, therefore, tend to employ a more conservative criterion for interpreting familiarity as evidence for a "target" response than for interpreting unfamiliarity as evidence for a "nontarget" response, causing their observed drift rates on target trials to be lower than those on novel trials. Familiarity evidence moves the decision process toward the correct "target" boundary on target trials but moves it away from the correct "nontarget" boundary on lure trials, further lowering the observed drift rate on lure trials.

Consistent with this theory, our model's parameter estimates indicate that individuals typically utilize a combination of evidence relevant to the n-back task's goal of responding based on stimulus-context bindings (reflected in *e*) and evidence related to the goal-irrelevant dimension of stimulus familiarity (reflected in *fs*). Parameter estimates also indicate that individuals tend to adopt a more conservative memory criterion for making a "target" response than a "nontarget" response on the basis of familiarity evidence ($fc > 0$), which is a characteristic example of a "stimulus bias" as explained in the DDM framework (Ging-Jehli et al., 2022; Kloosterman et al., 2019; White & Poldrack, 2014). Individuals' use of goal-irrelevant familiarity evidence was greater in the 2-back condition, which could be partially explained by the substantially poorer quality of goal-relevant evidence (*e*) on the 2-back relative to 0-back task (given that the stimulus-context bindings in the 0-back are much easier to maintain). Better quality evidence on the 0-back would reduce participants' need to rely on a familiarity-based strategy. This overall pattern suggests that individuals' performance on tasks that have high working memory demands—or higher "recognition conflict" as described by Oberauer (2005)—displays two key mechanistic differences with performance on tasks that have low working memory demands: poorer quality evidence about goal-relevant stimulus-context bindings (lower *e*) and greater reliance on goal-irrelevant familiarity-based evidence (higher *fs*).

Critically for modeling of individual differences, this novel formalization can account for the atypical v-shaped bivariate associations present in the ABCD sample. Individuals with negative v_{lure} tend to show poorer accumulation of goal-relevant evidence (*e* close to 0), greater reliance on goal-irrelevant familiarity-based evidence (higher *fs*), and a lack of the adaptive criterion shift for familiarity-based evidence (*fc* close to 0). The v-shaped pattern of associations between v_{lure} and the other drift rate parameters is, therefore, driven by the fact that, when *e* and *fc* are close to 0, increases in the *fs* parameter drive v_{target} and v_{novel} higher but

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

drive v_{lure} in the opposite direction, leading v_{lure} to be a negative value of a similar magnitude to v_{target} and v_{novel}.

The *fs* and *fc* parameters are correlated across n-back load levels, suggesting that individuals who rely on familiarity-based strategies tend to do so across multiple conditions of the n-back task. As expected, the *e* parameter was strongly correlated across load levels, consistent with prior work on the task-general factor of EEA (Lerche et al., 2020; Löffler et al., 2024; Schmiedek et al., 2007; Weigard & Sripada, 2021). By explicitly formalizing the assumptions of a general EEA factor and separate parameters related to individuals' use of familiarity-based evidence, our cognitive process model accomplishes what typical factor analysis models alone cannot do because the observed pattern of drift rate data violates their assumption that variables' linear associations can be described by a multivariate normal distribution. The resulting measurement model can be used to index these distinct cognitive mechanisms in isolation or—as the *e* parameter's associations across load levels do meet the assumptions of multivariate normality (Fig. 5)—can be integrated with factor analysis models to index broader factors, such as EEA, that drive performance across tasks beyond the n-back. Indeed, we found evidence that *e* shows robust associations with EEA measured from a simple task without any apparent working memory demands ("go" choices from the ABCD SST) supporting the task-generality of this parameter.

Another key implication of our findings is that they challenge the common assumption that tasks with different levels of working memory demands measure qualitatively different cognitive constructs: for example, that the 0-back measures "sustained attention," whereas the 2-back measures "working memory" (Avery et al., 2020; Kardan et al., 2022; Lin et al., 2023; Rahko et al., 2016). In contrast, the process model decomposition in the current study suggests that performance differences between the 2- and 0-back tasks can be parsimoniously attributed to quantitative differences in the same general cognitive mechanisms that act across n-back load conditions. Furthermore, the size of the cross-load associations of each of the novel parameters (Fig. 5) appears to be similar to, or greater than, the size of their test-retest reliability correlations across the 2-year ABCD measurement interval (Results Section 3.5), suggesting that most, if not all, of the stable trait-like variance in these mechanisms is shared across the 0-back and 2-back tasks. Therefore, attributing individual differences in performance at different n-back load levels to fundamentally distinct cognitive constructs appears unnecessary. Intriguingly, prior work in ABCD shows that performance in 2- and 0-back tasks is differentially predicted by functional connectivity-based neuromarkers derived from predictive models of "working memory" versus "sustained attention" task performance (Kardan et al., 2022), suggesting that there may be qualitative differences in the brain network features that support performance at different levels of n-back load. However, the current study's findings suggest that it is worth considering how variation in dimensional cognitive mechanisms at different n-back load levels may account for these findings.

Furthermore, the current results highlight general efficiency of goal-relevant evidence accumulation (*e*) as the n-back model parameter that has the clearest trait-like properties (relatively high stability across ABCD's 2-year measurement interval), criterion validity, and relevance to neurocognitive maturation. The *e* parameter, measured across both n-back load levels, showed robust relations with individuals' general cognitive ability, working memory

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

performance measured on a separate task, and parent- and teacher-reported inattention that are consistent with prior research on the importance of the broader construct of task-general EEA for cognitive ability (Lerche et al., 2020; Schmiedek et al., 2007) and childhood attention problems (Weigard et al., 2024; Ziegler et al., 2016). The *e* parameter's substantial nonlinear association with age (Fig. 6) suggests that EEA displays rapid maturation in early adolescence that gradually slows in later adolescence and peaks around age 20. Notably, this trend is highly consistent with the maturational trend identified for a task-general factor of executive functioning in recent work that spans several developmental samples (Tervo-Clemmens et al., 2023), suggesting that EEA could be a key contributor to this task-general factor. Overall, these findings underscore the importance of the EEA construct for interpreting associations of n-back task performance with other cognitive measures and with clinical symptom dimensions. More broadly, they suggest that EEA is a key driver of the maturation of goal-directed neurocognitive functioning, a hypothesis that can be more precisely tested in future research.

If the *fs* and *fc* parameters are less trait-like, less clearly linked to development, and less likely to show meaningful relations with criterion variables, what could variability in these parameters be attributed to? As both parameters reflect individuals' use of evidence from the goal-irrelevant dimension of stimulus familiarity, it is possible they could largely be attributed to individuals' idiosyncratic choices about the strategies they use for the task. The generally negative associations of *e* with *fs* (Fig. 4) suggest that at least some of the variance in individuals' use of familiarity evidence can be attributed to people with lower levels of *e* utilizing a more familiarity-based strategy to compensate for their generally poorer performance. However, we also highlight the intriguing finding that *fs* is instead *positively* related to EEA measured on a simple decision-making task (SST "go" choices, Table 3a). As we note when introducing the model, *fs* could be jointly determined by (1) an individual's ability to accumulate goal-irrelevant evidence related to stimulus familiarity and by (2) their strategic choice of how strongly to weight familiarity evidence. We speculate that the *fs* parameter's positive relation with EEA on the SST could reflect the component of *fs* related to gathering evidence for simple decisions, while its negative relation with *e* on the n-back reflects the component of *fs* related to the strategic use of familiarity evidence, although this explanation would have to be more stringently tested using experiments that can dissociate the components of *fs*.

The modest decreases in *fs* and *fc* between the baseline and year 2 ABCD waves, in the absence of any clear maturational trends in the independent age-diverse sample, hint at the possibility of practice effects across study visits. For example, practice-related increases in overall task performance (i.e., in *e*) or in comfort with the task procedures may cause individuals to reduce their reliance on familiarity-based evidence (*fs*). However, the long 2-year interval between sessions, as well as data from an accelerated adult version of the ABCD Study that shows little evidence for practice effects (Rapuano et al., 2022), argue against this possibility. Another possibility is simply that the small age-related increases identified in ABCD were not identified in the smaller age-diverse sample due to differences in statistical power. As the current study only examined a limited set of potential correlates of *fs* and *fc*, further research would be necessary to better characterize the factors that can account for variation in these mechanisms.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Beyond the broader theoretical implications described above, the current findings also have practical implications regarding the measurement of EEA and other cognitive mechanisms on the ABCD n-back task. As predicted by our model, we found that drift rate from the standard DDM, averaged across target, lure, and novel trial conditions, was nearly perfectly correlated with the model's *e* parameter. Hence, the use of the standard DDM without this novel reparameterization model appears sufficient to measure EEA if this is the primary construct of interest. However, the measurement model proposed here offers two key advantages. First, it allows additional parameters that underlie task performance (*fs* and *fc*) to be quantified, which would facilitate research investigations in which individuals' familiarity-based strategy use, or differences in strategy use between n-back load conditions, is of interest. Second, it addresses a critical barrier to integrating cognitive process models of the n-back with factor analysis models aimed at identifying broader cross-task factors. Standard DDM drift rate parameters from different n-back trial conditions are clearly inappropriate for such models because the observed v-shaped associations with v_{lure} violate these models' assumption of simple linear relations between variables with normally distributed errors. In contrast, associations of the *e*, *fs*, and *fc* parameters across load levels appear to meet this assumption (Fig. 5), which would allow our novel reparameterization of the DDM to be integrated with factor analysis and SEM methods. Critically, this is not only true for traditional "two-step" approaches, in which cognitive model parameters are estimated first and then entered into subsequent factor analysis or SEMs (e.g., Lerche et al., 2020; Schmiedek et al., 2007; Weigard et al., 2021), but also for promising new methods that allow for simultaneous estimation of cognitive model parameters and cross-task factor structures (Stevenson et al., 2024; Wall et al., 2021).

Although the current study focused on conflict recognition in a specific working memory paradigm, the current findings should also be considered within the context of the broader debate about the roles of recollection and familiarity processes in recognition memory more generally. A key focus in this area has been evaluating the relative evidence for two influential accounts that largely make similar predictions about empirical benchmark effects from recognition memory tasks (e.g., the shape of receiver-operator characteristics). "Dual-process signal detection" (DPSD) models do so by assuming recognition memory responses can depend on two processes: (1) a recollection process in which stimuli with exceptionally strong memory evidence are accepted as targets, and (2) a signal-detection process in which stimuli for which recollection fails are compared to a criterion along a familiarity evidence dimension on which target and lure stimuli are assumed to have identical variances in their evidence strength (Yonelinas, 1994, 2002). "Unequal-variance signal detection" (UVSD) models account for the same empirical benchmarks by assuming that memory strength is instead compared along a single evidence dimension in a signal-detection process for which variability in the target distribution is larger than variability in the lure distribution (Glanzer, Kim, Hilford, & Adams, 1999; Wixted, 2007). Despite the UVSD model's focus on a single evidence dimension, Wixted (2007) proposed a unifying framework in which recollection and familiarity are distinct sources of memory evidence that are additively combined to determine the memory strength of targets on this single dimension. This framework provides a compelling explanation for the greater variance of the target distribution than the lure distribution in the UVSD framework, as target evidence is impacted by variance from both the familiarity and recollec-

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tion evidence distributions, while lure evidence is impacted by variance in familiarity alone. However, vigorous debate over the relative merits of the DPSD and UVSD models continues (Kwon, Rugg, Wiegand, Curran, & Morcom, 2023; Mickes, Wais, & Wixted, 2009; Parks & Yonelinas, 2007). Relevant to the current study, evidence from applications of the standard DDM to recognition memory paradigms has generally supported predictions of the UVSD model (Osth, Bora, Dennis, & Heathcote, 2017; Starns, Ratcliff, & McKoon, 2012; Starns & Ratcliff, 2014).

Our novel conflict recognition model makes similar assumptions to Wixted's (2007) integrative version of the UVSP model in that it assumes that goal-relevant evidence and familiarity evidence are additively combined to form a single evidence signal. We note that the assumption of additive recollection and familiarity evidence has been similarly adopted in prior models of working memory tasks because its parsimony facilitates the modeling of response time data (Oberauer & Lange, 2009). As in this prior work, it is important to acknowledge that, although our model is consistent with an additive UVSD model, the success of our model does not necessarily provide support for UVSD over DPSD because a formal version of the DPSD model was not explicitly compared. However, we note that the current study's findings suggest that an additive two-process model is sufficient to describe key empirical patterns of individual and developmental differences in conflict recognition performance, which at the very least supports the plausibility of the additive UVSD model proposed by Wixted (2007). Furthermore, the two-process front-end model of drift rates that is proposed in the current study could be easily adapted to more traditional recognition memory tasks and could potentially be altered to reflect the alternate assumptions of the DPSD model in order to allow for tests of these leading recognition memory models' competing predictions.

One potential limitation of our work is that we primarily focused on a single variant of the n-back task that features stimuli with emotional facial expressions on a subset of trials and we were unable to model the potential impact of these emotional stimuli on performance due to the small number of trials available once performance is split across both stimulus type (target, lure, novel) and image category (happy, fearful, and neutral faces and places). Although we cannot definitively claim that our findings from the ABCD emotional n-back task will generalize to other tasks, there are three reasons to believe that these findings are nonetheless highly likely to do so. First, behavioral and neural data from the ABCD emotional n-back task have strong similarities to those from traditional n-backs. Early validation of the ABCD task confirmed that it elicits canonical patterns of neural activation associated with standard n-backs and replicates the expected effects of cognitive load on behavior (Casey et al., 2018). Subsequent work demonstrated that these neural responses to working memory demands show strong convergent validity by predicting working memory performance both within the scanner and in a separate task, whereas neural responses to emotional, versus nonemotional, stimuli showed no relevance to performance (Rosenberg et al., 2020). Another recent study linking EEA to neural activation during both the ABCD n-back and the Human Connectome Project (HCP) n-back paradigm, which has an identical design to the ABCD task but does not feature emotional stimuli, found that all major results were directly replicated across the ABCD and HCP tasks (Weigard et al., 2024). Hence, although the emotional stimuli on the ABCD task elicit unique patterns of neural activation, these stimuli seem to have

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

no discernable impact on the task's construct validity with regard to measuring cognitive performance and its neural basis. Second, the prior literature on EEA demonstrates that drift rate parameter estimates show evidence of a strong general factor across tasks with vastly different types of stimuli and cognitive demands (e.g., verbal vs. spatial tasks, simple decision-making vs. executive functioning, etc.) (Eisenberg et al., 2019; Lerche et al., 2017; Löffler et al., 2024; Schmiedek et al., 2007; Stevenson et al., 2024; Weigard et al., 2021). Therefore, it is highly plausible that the *e* parameter measured on this specific task largely reflects the broader factor of EEA, which appears to generalize to other n-back tasks (e.g., Löffler et al., 2024) and beyond. Finally, consistent with this prior literature, we find that the *e* parameter shows strong associations both with out-of-scanner working memory performance and with EEA measured on the ABCD SST, a relatively simple task with no apparent memory demands. Taken together, the combined findings of this prior literature and the current set of analyses strongly suggest that our results will generalize beyond the ABCD n-back task and that they largely reflect a task-general mechanism necessary for efficient cognitive performance.

In summary, the current study identified an atypical pattern of individual difference associations among measures of performance in different conditions of the ABCD n-back task and proposed a formal process model, inspired by earlier work on the n-back and other conflict recognition paradigms in adults (Oberauer, 2005), that accounts for this pattern. In particular, the model assumes performance is determined by a combination of accumulation of goal-relevant evidence ("EEA") and strategically modulated reliance on goal-irrelevant evidence related to stimulus familiarity. Findings from this process model demonstrate that EEA is a task-general factor that is relevant to neurocognitive development and clinical symptom dimensions, whereas the use of familiarity-based evidence is likely more related to idiosyncratic strategy use, context, or practice effects. Results highlight the benefits of formal cognitive process models for measuring distinct mechanisms of conflict recognition task performance. Moreover, they pave the way for the integration of task-specific process models with broader models of cross-task factor structures in ABCD and other data sets that are critical for characterizing the mechanisms of individual and developmental differences in cognitive control.

Acknowledgments

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development® (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9–10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from ABCD release 4.0 (https://nda.nih.gov; DOI 10.15154/1,523,041). Data specific to the current report can also be accessed in a separate NDA Study (https://nda.nih.gov; DOI 10.15154/rmmn-bx06).

AW was supported by K23 DA051561 and R21 MH130939. MMH was supported by R01 AA025790 and U01 DA041106. CS was supported by R01 MH123458 and U01 DA041106. BJC was supported by U01 DA041174. AOC was supported by K01 DA053438. TS was supported by NIH UL1 TR002240, KL2 TR002241, L30 MH127715, L30 TR004967, and the National Alliance on Mental Illness Unger Research Fellowship.

Conflict of interest statement

The authors declare no conflict of interest.

Data availability statement

All analysis code used in this study is openly available at: https://osf.io/tmzye/. The ABCD data used in the current study can be accessed at the NIMH data archive: https://nda.nih.gov; DOI 10.15154/1,523,041. The New Haven sample data are in the process of being made available at osf.io and, until they are, can be accessed upon request.

Open Research Badges

This article has earned Open Data and Open Materials Research Design badges. Data and materials design are available at https://nda.nih.gov and https://osf.io/tmzye/.

Ethics approval statement

The multisite ABCD Study was approved by each site's Institutional Review Board (IRB) as well as a central IRB at the University of California, San Diego. All ABCD participants provide informed consent (parents) or assent (children). Data collection for the New Haven sample was approved by the Yale IRB. All New Haven adult participants provided written consent, and all minor participants provided written assent.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

References

Achenbach, T. M. (2001). Manual for ASEBA school-age forms & profiles. University of Vermont, Research Center for Children, Youth & Families.

Achenbach, T. M., McConaughy, S., Ivanova, M., & Rescorla, L. (2011). *Manual for the ASEBA brief problem monitor (BPM)*. Burlington, VT: ASEBA.

Achenbach, T. M., & Ruffle, T. M. (2000). The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review*, *21*(8), 265–271.

Akshoomoff, N., Beaumont, J. L., Bauer, P. J., Dikmen, S. S., Gershon, R. C., Mungas, D., Slotkin, J., Tulsky, D., Weintraub, S., Zelazo, P. D., & Heaton, R. K. (2013). VIII. NIH Toolbox Cognition Battery (CB): Composite scores of crystallized, fluid, and overall cognition. *Monographs of the Society for Research in Child Development*, *78*, (4), 119–132.

Alderson, R. M., Hudec, K. L., Patros, C. H., & Kasper, L. J. (2013). Working memory deficits in adults with attention-deficit/hyperactivity disorder (ADHD): An examination of central executive and storage/rehearsal processes. *Journal of Abnormal Psychology*, *122*, (2), 532.

Avery, E. W., Yoo, K., Rosenberg, M. D., Greene, A. S., Gao, S., Na, D. L., Scheinost, D., Constable, T. R., & Chun, M. M. (2020). Distributed patterns of functional connectivity predict working memory performance in novel healthy and memory-impaired individuals. *Journal of Cognitive Neuroscience*, *32*, (2), 241–255.

Botvinick, M. M., & Cohen, J. D. (2014). The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science*, *38*, (6), 1249–1285.

Bozikas, V. P., & Andreou, C. (2011). Longitudinal studies of cognition in first episode psychosis: A systematic review of the literature. *Australian & New Zealand Journal of Psychiatry*, *45*, (2), 93–108.

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., & Thomas, K. M.; ABCD Imaging Acquisition Workgroup. (2018). The Adolescent Brain Cognitive Development (ABCD) Study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54.

Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: A receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology*, *62*, (5), 1017.

Cordova, M. M., Antovich, D. M., Ryabinin, P., Neighbor, C., Mooney, M. A., Dieckmann, N. F., Miranda-Dominguez, O., Nagel, B. J., Fair, D. A., & Nigg, J. T. (2022). Attention-deficit/hyperactivity disorder: Restricted phenotypes prevalence, comorbidity, and polygenic risk sensitivity in the ABCD baseline cohort. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*, (10), 1273–1284.

Damaso, K. A., Castro, S. C., Todd, J., Strayer, D. L., Provost, A., Matzke, D., & Heathcote, A. (2021). A cognitive model of response omissions in distraction paradigms. *Memory & Cognition*, *50*, (5), 962–978.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, (11), 2037–2078.

Dick, A. S., Lopez, D. A., Watts, A. L., Heeringa, S., Reuter, C., Bartsch, H., Fan, C. C., Kennedy, D. N., Palmer, C., Marshall, A., Haist, F., Hawes, S., Nichols, T. E., Barch, D. M., Jernigan, T. L., Garavan, H., Grant, S., Pariyadath, V., Hoffman, E., & Thompson, W. K. (2021). Meaningful associations in the adolescent brain cognitive development study. *NeuroImage*, *239*, 118262.

Edwards, M. C., & Sigel, B. A. (2015). Estimates of the utility of Child Behavior Checklist/Teacher Report Form Attention Problems scale in the diagnosis of ADHD in children referred to a specialty clinic. *Journal of Psychopathology and Behavioral Assessment*, *37*, 50–59.

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*, (1), 2319.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Epstein, J. N., Karalunas, S. L., Tamm, L., Dudley, J. A., Lynch, J. D., Altaye, M., Simon, J. O., Maloney, T. C., & Atluri, G. (2023). Examining reaction time variability on the stop-signal task in the ABCD Study. *Journal of the International Neuropsychological Society*, *29*, (5), 492–502.

Evans, N. J., Steyvers, M., & Brown, S. D. (2018). Modeling the covariance structure of complex datasets using cognitive models: An application to individual differences and the heritability of cognitive ability. *Cognitive Science*, *42*, (6), 1925–1944.

Fosco, W. D., White, C. N., & Hawk, L. W. (2017). Acute stimulant treatment and reinforcement increase the speed of information accumulation in children with ADHD. *Journal of Abnormal Child Psychology*, *45*, 911–920.

Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*, *47*, (1), 72–89.

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., & Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience*, *32*, 16–22.

Gard, A. M., Hyde, L. W., Heeringa, S. G., West, B. T., & Mitchell, C. (2023). Why weight? Analytic approaches for large-scale population neuroscience data. *Developmental Cognitive Neuroscience*, *59*, 101196.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, (4), 457–472.

Ging-Jehli, N. R., Arnold, L. E., Roley-Roberts, M. E., & DeBeus, R. (2022). Characterizing underlying cognitive components of ADHD presentations and co-morbid diagnoses: A diffusion decision model analysis. *Journal of Attention Disorders*, *26*, (5), 706–722.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, (2), 500.

Gold, J. M., Barch, D. M., Feuerstahler, L. M., Carter, C. S., MacDonald, III, A. W., Ragland, J. D., Silverstein, S. M., Strauss, M. E., & Luck, S. J. (2019). Working memory impairment across psychotic disorders. *Schizophrenia Bulletin*, *45*, (4), 804–812.

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An introduction to good practices in cognitive modeling. In Birte U. Forstmann and Eric-Jan Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 25–48).

Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019). Dynamic models of choice. *Behavior Research Methods*, *51*, 961–985.

Heathcote, A., & Matzke, D. (2022). Winner takes all! What are race models, and why and how should psychologists use them? *Current Directions in Psychological Science*, *31*, (5), 383–394.

Heathcote, A., Suraev, A., Curley, S., Gong, Q., Love, J., & Michie, P. T. (2015). Decision processes and the slowing of simple choices in schizophrenia. *Journal of Abnormal Psychology*, *124*, (4), 961.

Howlett, J. R., Harlé, K. M., & Paulus, M. P. (2021). Inhibitory failures in cocaine use disorder: Not paying attention when there is a need to be cautious. *Drug and Alcohol Dependence*, *226*, 108833.

Huang-Pollock, C. L., Karalunas, S. L., Tam, H., & Moore, A. N. (2012). Evaluating vigilance deficits in ADHD: A meta-analysis of CPT performance. *Journal of Abnormal Psychology*, *121*, (2), 360.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*, (4), 394–412.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, (3), 615.

Karalunas, S. L., & Huang-Pollock, C. L. (2013). Integrating impairments in reaction time and executive function using a diffusion model framework. *Journal of Abnormal Child Psychology*, *41*, (5), 837–850.

Kardan, O., Stier, A. J., Cardenas-Iniguez, C., Schertz, K. E., Pruin, J. C., Deng, Y., Chamberlain, T., Meredith, W. J., Zhang, X., Bowman, J. E., Lakhtakia, T., Tindel, L., Avery, E. W., Lin, Q., Yoo, K., Chun, M. M., Berman, M. G., & Rosenberg, M. D. (2022). Differences in the functional brain architecture of sustained attention and working memory in youth and adults. *PLoS Biology*, *20*, (12), e3001938.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Khurana, A., Romer, D., Betancourt, L. M., & Hurt, H. (2017). Working memory ability and early drug use progression as predictors of adolescent substance use disorders. *Addiction*, *112*, (7), 1220–1228.

Kloosterman, N. A., de Gee, J. W., Werkle-Bergner, M., Lindenberger, U., Garrett, D. D., & Fahrenfort, J. J. (2019). Humans strategically shift decision bias by flexibly adjusting sensory evidence accumulation. *eLife*, *8*, e37321.

Kofler, M. J., Rapport, M. D., Bolden, J., Sarver, D. E., & Raiker, J. S. (2010). ADHD and working memory: The impact of central executive deficits and exceeding storage/rehearsal capacity on observed inattentive behavior. *Journal of Abnormal Child Psychology*, *38*, 149–161.

Kwon, S., Rugg, M. D., Wiegand, R., Curran, T., & Morcom, A. M. (2023). A meta-analysis of event-related potential correlates of recognition memory. *Psychonomic Bulletin & Review*, *30*, (6), 2083–2105.

Lamichhane, B., Westbrook, A., Cole, M. W., & Braver, T. S. (2020). Exploring brain−behavior relationships in the N-back task. *NeuroImage*, *212*, 116683.

Lerche, V., von Krause, M., Voss, A., Frischkorn, G. T., Schubert, A.-L., & Hagemann, D. (2020). Diffusion modeling and intelligence: Drift rates show both domain-general and domain-specific relations with intelligence. *Journal of Experimental Psychology: General*, *149*, (12), 2207.

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

Lin, Y.-S., Weibel, J., Landolt, H.-P., Santini, F., Slawik, H., Borgwardt, S., Cajochen, C., & Reichert, C. F. (2023). Brain activity during a working memory task after daily caffeine intake and caffeine withdrawal: A randomized double-blind placebo-controlled trial. *Scientific Reports*, *13*, (1), 1002.

Löffler, C., Frischkorn, G. T., Hagemann, D., Sadus, K., & Schubert, A.-L. (2024). The common factor of executive functions measures nothing but speed of information uptake. *Psychological Research*, *88*, (4), 1092.

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., & Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*, (7902), 654–660.

Mashburn, C. A., Tsukahara, J. S., & Engle, R. W. (2020). Individual differences in attention control. In *Working memory: The state of the science*.

Mesholam-Gately, R. I., Giuliano, A. J., Goff, K. P., Faraone, S. V., & Seidman, L. J. (2009). Neurocognition in first-episode schizophrenia: A meta-analytic review. *Neuropsychology*, *23*, (3), 315.

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science*, *20*, (4), 509–515.

Narad, M. E., Garner, A. A., Peugh, J. L., Tamm, L., Antonini, T. N., Kingery, K. M., Simon, J. O., & Epstein, J. N. (2015). Parent–teacher agreement on ADHD symptoms across development. *Psychological Assessment*, *27*, (1), 239.

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, *134*, (3), 368.

Oberauer, K., & Lange, E. B. (2009). Activation and binding in verbal working memory: A dual-process model for the recognition of nonwords. *Cognitive Psychology*, *58*, (1), 102–136.

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, *26*, 1596–1618.

Osth, A. F., Bora, B., Dennis, S., & Heathcote, A. (2017). Diffusion vs. linear ballistic accumulation: Different models, different conclusions about the slope of the zROC in recognition memory. *Journal of Memory and Language*, *96*, 36–61.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*, (1), 46–59.

Owens, M. M., Potter, A., Hyatt, C. S., Albaugh, M., Thompson, W. K., Jernigan, T., Yuan, D., Hahn, S., Allgaier, N., & Garavan, H. (2021). Recalibrating expectations about effect size: A multi-method survey of effect sizes in the ABCD Study. *PLoS One*, *16*, (9), e0257535.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review*, *114*(1), 188–202.

Pedersen, M. L., Alnæs, D., van der Meer, D., Fernandez-Cabello, S., Berthet, P., Dahl, A., Kjelkenes, R., Schwarz, E., Thompson, W. K., Barch, D. M., Andreassen, O. A., & Westlye, L. T. (2023). Computational modeling of the N-Back task in the ABCD Study: Associations of drift diffusion model parameters to polygenic scores of mental disorders and cardiometabolic diseases. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*, (3), 290–299.

R Core Team. (2013). *R: A language and environment for statistical computing*.

Rahko, J. S., Vuontela, V. A., Carlson, S., Nikkinen, J., Hurtig, T. M., Kuusikko-Gauffin, S., Mattila, M.-L., Jussila, K. K., Remes, J. J., Jansson-Verkasalo, E. M., Aronen, E. T., Pauls, D. L., Ebeling, H. E., Tervonen, O., Moilanen, I. K., & Kiviniemi, V. J. (2016). Attention and working memory in adolescents with autism spectrum disorder: A functional MRI study. *Child Psychiatry & Human Development*, *47*, 503–517.

Rapport, M. D., Alderson, R. M., Kofler, M. J., Sarver, D. E., Bolden, J., & Sims, V. (2008). Working memory deficits in boys with attention-deficit/hyperactivity disorder (ADHD): The contribution of central executive and subsystem processes. *Journal of Abnormal Child Psychology*, *36*, 825–837.

Rapuano, K. M., Conley, M. I., Juliano, A. C., Conan, G. M., Maza, M. T., Woodman, K., Martinez, S. A., Earl, E., Perrone, A., Feczko, E., Fair, D. A., Watts, R., Casey, B. J., & Rosenberg, M. D. (2022). An open-access accelerated adult equivalent of the ABCD Study Neuroimaging Dataset (a-ABCD). *NeuroImage*, *255*, 119215.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*, (4), 260–281.

Rosenberg, M. D., Martinez, S. A., Rapuano, K. M., Conley, M. I., Cohen, A. O., Cornejo, M. D., Hagler, D. J., Meredith, W. J., Anderson, K. M., Wager, T. D., Feczko, E., Earl, E., Fair, D. A., Barch, D. M., Watts, R., & Casey, B. J. (2020). Behavioral and neural signatures of working memory in childhood. *Journal of Neuroscience*, *40*, (26), 5090–5104.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, (4), 1089.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*, (3), 414.

Schubert, A.-L., & Frischkorn, G. T. (2020). Neurocognitive psychometrics of intelligence: How measurement advancements unveiled the role of mental speed in intelligence differences. *Current Directions in Psychological Science*, *29*, (2), 140–146.

Schubert, A.-L., Frischkorn, G. T., Hagemann, D., & Voss, A. (2016). Trait characteristics of diffusion model parameters. *Journal of Intelligence*, *4*, (3), 7.

Skalaban, L. J., Cohen, A. O., Conley, M. I., Lin, Q., Schwartz, G. N., Ruiz-Huidobro, N. A., Cannonier, T., Martinez, S. A., & Casey, B. (2022). Adolescent-specific memory effects: Evidence from working memory, immediate and long-term recognition memory performance in 8–30 yr olds. *Learning & Memory*, *29*, (8), 223–233.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, (3), 161–168.

Somerville, L. H., & Casey, B. (2010). Developmental neurobiology of cognitive control and motivational systems. *Current Opinion in Neurobiology*, *20*, (2), 236–241.

Squeglia, L. M., & Gray, K. M. (2016). Alcohol and drug use and the developing brain. *Current Psychiatry Reports*, *18*, 1–10.

Sripada, C., & Weigard, A. (2021). Impaired evidence accumulation as a transdiagnostic vulnerability factor in psychopathology. *Frontiers in Psychiatry*, *12*, 627179.

Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language*, *70*, 36–52.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, *64*, (1–2), 1–34.

Stevens, L., Verdejo-García, A., Goudriaan, A. E., Roeyers, H., Dom, G., & Vanderplasschen, W. (2014). Impulsivity as a vulnerability factor for poor addiction treatment outcomes: A review of neurocognitive findings among individuals with substance use disorders. *Journal of Substance Abuse Treatment*, *47*, (1), 58–72.

Stevenson, N., Innes, R. J., Gronau, Q. F., Miletić, S., Heathcote, A., Forstmann, B. U., & Brown, S. D. (2024). Using group level factor models to resolve high dimensionality in model-based sampling. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000618

Stevenson, N., Innes, R. J., Boag, R. J., Miletić, S., Isherwood, S. J., Trutti, A. C., Heathcote, A., & Forstmann, B. U. (2024). Joint modelling of latent cognitive mechanisms shared across decision-making domains. *Computational Brain & Behavior*, *7*, (1), 1–22.

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, (1), 137.

Tervo-Clemmens, B., Calabro, F. J., Parr, A. C., Fedor, J., Foran, W., & Luna, B. (2023). A canonical trajectory of executive function maturation from adolescence to adulthood. *Nature Communications*, *14*, (1), 6922.

Townsend, J. T. (2008). Mathematical psychology: Prospects for the 21st century: A guest editorial. *Journal of Mathematical Psychology*, *52*, (5), 269–280.

Tulsky, D. S., Carlozzi, N. E., Chevalier, N., Espy, K. A., Beaumont, J. L., & Mungas, D. (2013). V. NIH Toolbox Cognition Battery (CB): Measuring working memory. *Monographs of the Society for Research in Child Development*, *78*, (4), 70–87.

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65–79.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*, (3), 368.

Verbruggen, F., McLaren, I. P., & Chambers, C. D. (2014). Banishing the control homunculi in studies of action control and behavior change. *Perspectives on Psychological Science*, *9*, (5), 497–524.

Wall, L., Gunawan, D., Brown, S. D., Tran, M.-N., Kohn, R., & Hawkins, G. E. (2021). Identifying relationships between cognitive processes across tasks, contexts, and time. *Behavior Research Methods*, *53*, (1), 78–95.

Weigard, A., Angstadt, M., Taxali, A., Heathcote, A., Heitzeg, M. M., & Sripada, C. (2024). Flexible adaptation of task-positive brain networks predicts efficiency of evidence accumulation. *Communications Biology*, *7*, (1), 801.

Weigard, A., Brislin, S. J., Cope, L. M., Hardee, J. E., Martz, M. E., Ly, A., Zucker, R. A., Sripada, C., & Heitzeg, M. M. (2021). Evidence accumulation and associated error-related brain activity as computationally-informed prospective predictors of substance use in emerging adulthood. *Psychopharmacology*, *238*, (9), 2629–2644.

Weigard, A., Clark, D. A., & Sripada, C. (2021). Cognitive efficiency beats top-down control as a reliable individual difference dimension relevant to self-control. *Cognition*, *215*, 104818.

Weigard, A., Huang-Pollock, C., Brown, S., & Heathcote, A. (2018). Testing formal predictions of neuroscientific theories of ADHD with a cognitive model–based approach. *Journal of Abnormal Psychology*, *127*, (5), 529.

Weigard, A., Matzke, D., Tanis, C., & Heathcote, A. (2023). A cognitive process modeling framework for the ABCD Study stop-signal task. *Developmental Cognitive Neuroscience*, *59*, 101191.

Weigard, A., McCurry, K. L., Shapiro, Z., Martz, M. E., Angstadt, M., Heitzeg, M. M., Dinov, I. D., & Sripada, C. (2023). Generalizable prediction of childhood ADHD symptoms from neurocognitive testing and youth characteristics. *Translational Psychiatry*, *13*, (1), 225.

Weigard, A., & Sripada, C. (2021). Task-general efficiency of evidence accumulation as a computationally defined neurocognitive trait: Implications for clinical neuroscience. *Biological Psychiatry Global Open Science*, *1*, (1), 5–15.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., & Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*, (11 Supplement 3), S54–S64.

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, (2), 385.

Wiker, T., Pedersen, M. L., Ferschmann, L., Beck, D., Norbom, L. B., Dahl, A., von Soest, T., Agartz, I., Andreassen, O. A., Moberget, T., Westlye, L. T., Huster, R. J., & Tamnes, C. K. (2023). *Assessing the longitudinal associations between decision-making processes and attention problems in early adolescence*.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, *57*, (11), 1336–1346.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*, (1), 152.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, (6), 1341.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, (3), 441–517.

Ziegler, S., Pedersen, M. L., Mowinckel, A. M., & Biele, G. (2016). Modelling ADHD: A review of ADHD theories through their predictions for computational models of decision-making and reinforcement learning. *Neuroscience & Biobehavioral Reviews*, *71*, 633–656.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Material

15516709, 2024, 11, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70019 by Uva Universiteitsbibliotheek, Wiley Online Library on [23/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License