Doctoral thesis submitted to
the Faculty of Behavioural and Cultural Studies
Heidelberg University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Dr. phil.)
in Psychology

Title of the publication-based thesis

Using the diffusion model to study individual differences

presented by
Michael von Krause

year of submission
2021

Dean: Prof. Dr. Dirk Hagemann
Advisor: Prof. Dr. Andreas Voß

Table of Contents

Acknowledgments

First, I would like to thank Prof. Dr. Andreas Voß for the great support and supervision during my dissertation project. Andreas was always there to help and answer my questions whenever I needed him, he was the one who excited me for the subject of my dissertation, and he was always ready for yet another project. I am also especially grateful to Prof. Dr. Jan Rummel, who immediately agreed to review this thesis, and my secondary and tertiary advisers from the SMiP group, Prof. Dr. Tanja Lischetzke and Prof. Dr. Rolf Ulrich.

The past three years would have been unthinkable without my amazing colleagues from the department of quantitative research methods. Dr. Veronika Lerche was especially important during the first part of my dissertation. She helped me a lot in finding my ways around the institute as a researcher, introduced me to sophisticated R loops, and made me feel very welcome at the department before I had even started. It was of course also a very pleasant surprise to find a fellow Starnberger See person among my colleagues. Stefan Radev became an increasingly important collaborator in what are now the first and the last papers of this dissertation. His wit and his wide range of interests from poetry and Tarkovsky movies to realms of mathematics far beyond my reach have been of the greatest importance for our projects. I am also very happy that Dr. Alica Mertens was my partner-in-crime for a number of great studies on the dark side of human nature – and together with Marie Wieschen part of the best office team ever. Annika Stump, Dr. Ulf Mertens and Lukas Schumacher also helped me with their ideas and the great times we have had together – especially during our shared lunches, that I hope will soon recommence in all their former glory. I would like to thank Prof. Dr. Cornelia Wrzus for her continued patience and support in the long-lasting project on longitudinal data. Lastly, the SMiP organizational team, most of all Dr. Anke Söllner, have been an incredible support for me as a doctoral student and father.

Outside of the university, I want to thank Max Geilke for teaching me to question everybody and everything and Maurizio Sicorello for convincing me to apply to this weird SMiP thing. To Prof. Dr. Christiane von Stutterheim I owe a great part of my conviction that trying to live life as a scientist is a worthwhile endeavor.

Throughout my entire life one group of people have been incredibly significant – my family. No matter what, I always know that you are there for me – thank you. Most importantly, none of this would have been remotely possible without my beloved wife Tine. Finally, my children Freya and Arvid bring into my life a kind of joy (and tiredness) that I could not even have imagined.

List of Scientific Publications of the Publication-Based Dissertation

Manuscript 1

von Krause, M., Radev, S. T., Voss, A., Quintus, M., Egloff, B., & Wrzus, C. (submitted). Stability and Change in Diffusion Model Parameters Over Two Years. *Journal of Intelligence*.

Manuscript 2

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*. Advance online publication.

* First authorship shared.

Manuscript 3

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2020). Age differences in diffusion model parameters: a meta-analysis. *Psychological Research*, 1-10.

Manuscript 4

von Krause, M., Lerche, V., Schubert, A. L., & Voss, A. (2020). Do Non-Decision Times Mediate the Association between Age and Intelligence across Different Content and Process Domains?. *Journal of Intelligence*, *8*(3), 33.

Manuscript 5

von Krause, M., Radev, S.T., & Voss, A. (submitted). Processing speed is high until age 60 - Insights from Bayesian modeling in a one million sample (with a little help of deep learning). *Proceedings of the National Academy of Sciences of the United States of America*.

1 Introduction

Already since the days of Gordon W. Allport (1937), psychological science has been partly divided. On the one hand, fields such as cognitive or experimental psychology have focused on principles and laws that supposedly generalize across people, possibly even all humans (Reisberg, 2013). On the other hand, fields such as personality psychology have focused on the individual differences between people (John et al., 2008). Though of course there has always been an exchange between these two “blocks” of psychology, the gap between them has more often than not proved hard to bridge. For example, there has been an abundance of research on individual differences in cognition, cognitive abilities like mental speed, and intelligence, dating back as far as Francis Galton (1908) and Alfred Binet (1904). Mostly separately from that, in the field of cognitive psychology, people have tried to understand the exact “hows” of cognitive processes in general, quite often by means of experimental methods and, in the past decades, mathematical or computational models of cognition (Busemeyer & Diederich, 2010). Through cognitive modeling, scientists try to map the distinct processes hypothesized to create certain behavioral data to parameters in a mathematical formulation (Farrell & Lewandowsky, 2018), ideally allowing for formalized testing of the theories underlying the models. As the hypothesized processes are usually not directly observable, modeling approaches also often have the advantage of providing estimates of these hidden or latent parameters, even on the level of an individual person. In this way, cognitive models provide a link between experimental psychology and research on individual differences.

One such cognitive model that has seen a huge rise in popularity over the past 40 years is the diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998; Voss et al., 2013). Roger Ratcliff originally developed the model in the 1970s, building on older work from Laming (1968) and Link and Heath (1975). It is a stochastic model of the decision process in simple binary decision tasks and part of the family of evidence accumulation models. Other models belonging to the same model family are the leaky competing accumulator model (LCA; Usher & McClelland, 2001) and the linear ballistic accumulator model (LBA; Brown & Heathcote, 2008). These latter models have several distinct properties when compared to the diffusion model, the most important ones probably being that they are applicable not only to binary decision tasks but also to multiple choice tasks, and try to (in case of the LCA) model the neural processes underlying decision making. Since I will only study binary decisions and this thesis will be concerned with the neural basis of decisions only to a minor extent, and because the diffusion model has been tried, tested, and validated by far most extensively in previous research, this model will also be the focus of my work. However, I will refer back to the LCA and LBA in the discussion.

The basic idea of the diffusion model is that people, for example when solving a recognition memory task (Ratcliff, 1978), continuously accumulate information until one of two thresholds is reached. These thresholds represent the two possible decision outcomes. The diffusion model uses response time distributions and accuracy rates to estimate different aspects of the decision process underlying the data obtained in an experimental setting. The main advantage of the model lies in its ability to disentangle the speed of information accumulation (called drift rate in the model) from speed-accuracy trade-offs (represented by the so-called boundary separation), decision biases, and the time needed for non-decisional processes like encoding and motor response execution (Voss et al., 2004). In this way, the model allows the specification of well-defined research questions, in contrast to, for example, the mere comparison of mean correct response times between different experimental conditions. For instance, the model can help explain whether the slower response times found in one experimental condition compared to another are due to higher average decision caution (maybe because of differently worded instructions) or differences in speed of information accumulation – both should map to different diffusion model parameters, namely boundary separation and drift rate.

There has been a great variety of studies employing the diffusion model, mostly in experimental or cognitive psychology, with a focus on research in memory (Arnold et al., 2015; Ball & Aschenbrenner, 2018; Boywitt & Rummel, 2012; Horn et al., 2013; McKoon & Ratcliff, 2012; Ratcliff, 1978; Spaniol et al., 2006; Voskuilen et al., 2018), perception (Dully et al., 2018; Kühn et al., 2010; McGovern et al., 2018; Ratcliff et al., 2003; Spaniol et al., 2011), language (Ratcliff, Gomez, et al., 2004; Ratcliff, Thapar, et al., 2004; Yap et al., 2012), and executive control (Madden et al., 2010). However, over the past two decades, more and more researchers have also started using the diffusion model to study individual differences in diffusion model parameter estimates.

When linking cognitive modeling and individual differences research in such a way, it is vital to first establish a clear definition of the diffusion model parameters as markers of individual differences. Do they represent trait-like entities? If this is the case, according to generally accepted definitions of personality and traits, they should exhibit consistency across tasks and time (John et al., 2008). Moreover, it is important that they are valid representations of the decision processes they are theorized to measure, and, in a correlational setting with other traits, show concurrent and discriminant validity. In the end, these issues lead to the question of how such parameters may help us tackle problems posed in individual differences research - a question I pursued in my dissertation project by employing the diffusion model to better understand the relations of the processes represented by the diffusion model parameters, intelligence, and age.

In doing so, I tried to bridge the gap between experimental psychology methodology and substantial individual differences research in two ways. On the one hand, I used the diffusion model, with its background in cognitive psychology, to obtain better-informed inferences regarding questions on individual differences in cognition than are achievable when solely relying on raw data. On the other hand, and this is the major novelty of my research program, I applied principles deemed important in individual differences research to diffusion model analyses, by using rich and diverse data to improve the reliability and scope of my results. Most previous diffusion model studies reporting on individual parameter values focused on very specific research questions grounded in experimental psychology and often not followed up on in a systematic way in subsequent research. In contrast to that, throughout the work that forms this thesis, my aim was to take the diffusion model parameters seriously primarily as constituents of individual differences and systematically probe their applicability and usefulness in such a framework throughout an entire research program.

My dissertation can roughly be divided in two parts, with the first, much shorter and more methodological part, setting the stage for the substantial research questions tackled in part two. Both parts are concerned with the question whether the diffusion model can help us study individual differences in cognition better and more precisely than by relying on raw data.

In the first section, I follow up on the previous literature on the question whether diffusion model parameters should be considered trait-like entities by studying a large sample of participants longitudinally over two years (*Manuscript 1*). There have been some attempts at systematically conceptualizing the study of individual differences with the diffusion model and establishing a framework for the interpretation of diffusion model parameters as person-specific measures of distinct cognitive processes (Lerche & Voss, 2017b; Ratcliff & Childers, 2015; Schubert et al., 2016). Lerche and Voss (2017b) reported that the main diffusion model parameters showed at least acceptable retest stability over a one week interval, while Schubert and colleagues (2016) used a latent state-trait model to establish their consistency (especially of drift rates) across two tasks as well as over an eight month interval. In our study, we go beyond the previous literature on temporal patterns in diffusion model parameters in several ways. We analyzed four different types of stability and change in diffusion model parameters in a large sample across four measurement occasions over two years.

In the second section, I use the estimates of individual differences in cognitive parameters provided by the model to better understand research questions on individual differences in cognition. An abundance of literature relies on mean response times as a measure of processing speed (Jensen, 2006; Sheppard & Vernon, 2008). As response times are composites of several processes, it is often hard to understand what exactly is measured in mean response times and how to interpret the corresponding literature that analyzes such data. Here, I focus on two “puzzles” in cognitive individual differences research that I hope to help better understand by means of disentangling the decision process components through diffusion modeling.

The first puzzle concerns the across-task structure of processing speed and its relationship to intelligence. Intelligence and processing speed were found to be positively related in numerous studies (Jensen, 2006; Sheppard & Vernon, 2008). Intelligence is also thought to be in parts content-specific, with the across-task structure showing both a general factor (*g*) and domain factors such as verbal or figural intelligence (Sternberg, 2000). Yet processing speed, as measured in mean response times, has in the past repeatedly been found to be largely unitary, that is, when analyzing the correlational pattern of processing speed across several tasks, their common variance seems best represented by a single general factor similar to *g* in intelligence research (Jensen, 2006; Schubert et al., 2017). Content-domain specific aspects, for example of processing speed specific to verbal or figural tasks, could not be found in previous research.

Taken together, these findings reported in the literature might seem somewhat puzzling: processing speed and intelligence are closely related, although the former does not exhibit the complex correlational patterns of the latter. The idea behind *Manuscript 2* is that this configuration of results can be attributed to the fact that response times are composite scores of several distinct processes. To get a theoretically pure measure of processing speed or speed of information accumulation, we focused on the drift rate diffusion model parameter in a study utilizing 18 tasks. We investigated the structure of processing speed and its relationship to intelligence across the figural, numerical, and verbal content domains. In this way, we could study the relative strength of domain-specific and domain-general aspects of drift rates and their correlational patterns to the respective intelligence components.

The second puzzle is already grounded in past diffusion model research. Older people often show longer response times across a great variety of cognitive tasks - this is a consistent finding in the literature on cognitive aging (Jensen, 2006). Repeatedly, this observation has been interpreted as representing a mental slow-down, that might even be the root of cognitive decline as a whole, including the lower intelligence scores found in older people (Salthouse, 1996). Yet studies employing the diffusion model, comparing college-aged persons with older adults aged 65 or more, have repeatedly shown that drift rates (as indices of processing speed) are unrelated to age – instead, older people tend to sample more information before taking a decision and need more time for encoding and motor response execution (e.g., Ball & Aschenbrenner, 2018; McKoon & Ratcliff, 2012, 2013; Ratcliff et al., 2003; Ratcliff, Thapar, et al., 2004; Spaniol et al., 2006; Thapar et al., 2003; Voskuilen et al., 2018). It must be noted that some studies did find drift rates to be negatively related to age, while others found a positive relation, making the overall picture quite unclear (Ratcliff, Thapar, et al., 2004; Voskuilen et al., 2018).

We used a three-step approach to better understand the inconsistent results regarding the effects of age on drift rates. First, we studied age differences in diffusion model parameters and especially drift rates in a meta-analysis to gain a quantitative description of the overall patterns reported in the literature (*Manuscript 3*). Second, we employed mediation analyses to study which of the diffusion model parameters mediate age-related differences in intelligence (*Manuscript 4)*, once more using the dataset of 18 different tasks also analyzed in *Manuscript 2*, to be able to scrutinize task-specificities within one sample. In this way, we could assess directly whether it was speed of information accumulation, boundary separation, or non-decision time that explained the slower response times found with increasing age, thus providing a test of the assumption that changes in processing speed are at the core of age-related cognitive decline. Finally, we studied age differences in mean response times and diffusion model parameters in a very large implicit association test (Greenwald et al., 1998, 2003; Nosek et al., 2007) dataset (*N* > 1,000,000; *Manuscript 5*). To obtain parameter estimates, we used a novel parameter estimation approach based on a deep neural network that makes handling such sample sizes feasible (Radev et al., 2020). The large number of participants allowed us to robustly analyze cross-sectional age differences on a year-specific level, yielding very interesting results on the age relationships of processing speed, decision caution, and non-decision time, almost over the entire lifespan (ages 10 to 80).

Together, the studies presented in this dissertation constitute an important step in the direction of a systematic use of the diffusion model in the study of individual differences. After introducing the five manuscripts on the following passages, I will then discuss their implications, possible limitations, and give some ideas on possible future research.

2 Stability and change in diffusion model parameters (Manuscript 1)[1]

Utilizing the full data resulting from binary decision tasks, the diffusion model allows researchers to obtain individual parameter estimates of processing speed (drift rates), decision caution (boundary separation), non-decision times, and response biases (Ratcliff & McKoon, 2008). These parameters have been validated both experimentally (Arnold et al., 2015; Voss et al., 2004) and neurophysiologically (McGovern et al., 2018; Ratcliff et al., 2007). However, in most studies using the diffusion model, the focus is on comparing differences in model parameter between experimental conditions. For example, one might study the question whether the IAT effect measured in implicit association tests (Greenwald et al., 1998, 2003; Nosek et al., 2007), that supposedly measures implicit bias, maps onto differences in drift rates, boundary separations, or non-decision times (Klauer et al., 2007). When employing the diffusion model in such a manner, that is, for studying group differences, the reliability and exact properties of each of the individual parameter estimates is of secondary importance to the general validity of the parameters. Contrarily, when researchers are interested in the diffusion model parameters as characterizing individuals, some new questions gain priority. Are the model parameters truly person specific? Is there reliable between-person variance? Are between-person differences in model parameters related across different paradigms? Are between-person differences stable across time? How do individual parameter estimates develop?

These questions relate to the concept of *traits* that is of central importance in theories of individual differences. Traits are often defined as characteristic patterns of thoughts, feelings and behaviors that show consistency across situations and stability across time (Allport, 1937; John et al., 2008). Whether the diffusion model parameters, for example, processing speed as measured by drift rates, can be interpreted as traits in the way they were just defined, is still unclear from past diffusion model studies. While a number of studies have started to employ the individual parameter estimates in correlational research, for example focusing on their relationships with intelligence (McKoon & Ratcliff, 2012; Ratcliff et al., 2010; Schmiedek et al., 2007), the underlying assumptions have rarely been tested systematically. The aspect that has received by far the most attention is whether the diffusion model parameters show consistency across tasks. Several studies have reported medium to high across-task correlations

[1] von Krause, M., Radev, S. T., Voss, A., Quintus, M., Egloff, B., & Wrzus, C. (submitted). Stability and Change in Diffusion Model Parameters Across Two Years. *Journal of Intelligence*.

for all of the core diffusion model parameters (e.g., Ratcliff et al., 2010; Schmiedek et al., 2007; Schubert et al., 2016). In contrast, their temporal stability has received far less attention.

The first study reporting test-retest correlations over a time period of at most a week found strong across-time correlation for drift rates, boundary separations, biases, and non-decision times (all correlations r >.70, Yap et al., 2012). In contrast, Lerche et al. (2017b) found far weaker across-time correlations for non-decision times (all *r*s <.50), although they did find strong stability for drift rates, boundary separations and biases (all *r*s >.70). Finally, Schubert et al. (2016) conducted a first systematic study of the trait characteristics of diffusion model parameters. Studying two different tasks over two measurement occasions eight months apart, the authors employed latent-state-trait structural equation models to separate the parts of diffusion model parameter variance specific to each task and each time point from trait variance. According to their analyses, drift rates show by far the greatest across-task and across-time stability, with boundary separation and non-decision times being less trait-like in their composition of variance (the authors did not study response bias).

While these studies were important first steps on the path of establishing the trait-like qualities of individual diffusion model parameter estimates in a temporal sense, they had several shortcomings. First, the time period studied was limited, with at most eight months separating the first form the second measurement occasion (Schubert et al., 2016). If diffusion model parameters should be considered trait-like entities, they might be expected to show stability over time periods of one year or even several years. Second, sample sizes were generally limited and so was the samples' heterogeneity – most participants were college-aged students. Third, the studies mostly focused on one aspect of temporal stability, namely test-retest correlations or the trait-factor in the structural equation model. Both these measures relate to the concept of rank-order stability, that is, the stability of the across-person relative positions of participants on the range of possible parameter values. Yet in the study of individual differences, a number of additional ways of studying stability and change has been proposed – not only rank-order stability, but also mean-level changes, inter-individual differences in change, and profile stability (Roberts, Brent et al., 2008). These aspects have been extensively studied for the Big Five personality traits (e.g., Roberts et al., 2001, 2006; Roberts & DelVecchio, 2000), but have received little attention in the literature on cognitive parameters and none in the diffusion model literature.

Our study that comprised Manuscript 1 seeks to address all three gaps just outlined. We studied diffusion model parameters in a personality IAT (Back et al., 2009; Schmukle et al., 2008) across four measurement occasions over two years, employing a hierarchical Bayesian parameter estimation approach (Wiecki et al., 2013). We used a diverse sample that included both college-aged people and old adults, and in both age groups included students and non-students. Finally, we studied four types of stability and change: rank-order stability, mean-level changes, inter-individual differences in change, and profile stability.

In short, we found all three diffusion model parameters studied (drift rates, boundary separations, non-decision times) to exhibit high rank-order stability over time, with drift rates over a time period of two years showing the lowest correlation (r = .64). Most across-time correlations of the three parameters each assessed at the four measurement occasions were even higher, in the range from r = .80 to r = .90. Regarding mean-level changes, the group-level drift rate parameters increased over time, while the boundary separations decreased. Non-decision times showed no changes. In terms of the rate of change, only drift rates exhibited credible individual differences. Finally, average profiles of the three core diffusion model parameters proved to be very stable across time.

All these results can be interpreted as supportive of the notion of individual diffusion model parameters as trait-like entities, at least regarding temporal aspects. Most importantly, the high rank-order stabilities, as well as the profile stabilities, make it clear that the individual relative expressions of processing speed, decision caution, and non-decision time most often remain stable even across longer time periods. Interestingly, rank-order stability was considerably higher than what has been suggested by previous studies (Lerche & Voss, 2017b; Schubert et al., 2016; Yap et al., 2012). Three main reasons for the higher stability found in our study might be a) the relatively high number of trials per person (600), b) the very robust hierarchical Bayesian modeling approach employed (both of which might have led to more reliable estimates), and c) the great heterogeneity in participant demographics in our sample, with the greater variance in diffusion model parameters possibly also leading to stronger covariances. The mean-level changes found for drift rates and boundary separations can be interpreted as practice effects – people sample information more efficiently and become less cautious over time. This is in line with previous studies on practice effects in diffusion model parameters, though it must be noted that these previous results (of within-session practice effects) also included decreases in non-decision time (Dutilh et al., 2009, 2011; Evans & Brown, 2017). Interestingly, in our study we could show the practice effects seem to persist over time periods of up to one year. Finally, we found that people differ in the extend they profit from the practice effect on processing speed.

All these results lead to the same conclusion: As the diffusion model parameters show considerable across-time rank-order and profile stability even over a period of two years and display interpretable mean-level changes, the notion of parameters-as-traits seems in this way justified. Our findings thus strengthen and expand the accounts presented in previous studies on individual differences in diffusion model parameters. In the following manuscript, we continued to test their applicability and usefulness for individual differences research, with the first application focusing on the relationship of the parameter drift rate, representing processing speed, with intelligence.

3 Diffusion modeling and intelligence (Manuscript 2)[2]

Cognitive processing speed is known to be related to general intelligence (*g*; Jensen, 2006; Sheppard & Vernon, 2008). Drawing on 172 studies, Sheppard and Vernon (2008) found small to medium correlations between mental speed as measured by mean response times (RTs) and intelligence across a variety of paradigms and heterogeneous types of sample; people with lower RTs tended to have higher intelligence (IQ) scores. Cognitive processing speed has also been hypothesized to contribute to age-related cognitive decline. Salthouse (1996) proposed the idea that a general slow-down of cognitive processes might be the reason for lower IQ scores found in older adults, highlighting the close relationship between processing speed and intelligence.

In the past two decades, a number of studies utilizing the diffusion model have started investigating the relationship between the model parameter drift rate and intelligence (McKoon & Ratcliff, 2012; Ratcliff et al., 2010; Schmiedek et al., 2007). The use of drift rates as measure of processing speed instead of mean RTs has several important advantages. First, by utilizing the full response time distributions of correct and error responses and also the accuracy rates, one can draw on a larger proportion of the available data. Second, by separating processing speed from decision caution and non-decision time, the diffusion model provides a theoretically pure measure of processing speed in its drift rate parameter, that should show more clearly interpretable correlational patterns to external criteria than mean RTs, which are a composite of several distinct processes. Schmiedek and colleagues (2007) found drift rates to strongly predict scores in reasoning, working memory, and psychometric speed. In a similar way, Ratcliff and colleagues found high positive correlations between drift rates and general

---

[2] 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*. Advance online publication.

intelligence across three different age groups (college-aged, 60-74 years old, 75-90 years old; Ratcliff et al., 2010). Both these studies used structural equation modeling to aggregate the drift rates from several tasks to a latent factor – eight in the case of Schmiedek and colleagues, while Ratcliff and colleagues used three. There is thus initial evidence that drift rates predict measures of intelligence in a similar way that raw mean RTs do.

When looking at the relationship between processing speed and intelligence more closely, a slightly puzzling finding becomes salient. Intelligence is typically assumed to have a hierarchical structure that contains both a strong general factor (*g*) and domain-specific abilities (Sternberg, 2000). Conversely, processing speed as measured by mean response times has been shown to be largely unitary, although it is linked to intelligence (Jensen, 2006). In order to better understand this issue, we analyzed the structure of processing speed (as measured by drift rates) and its relationship to intelligence in Manuscript 2.

We tested performance of 125 participants in a wide range of binary decision tasks. Six tasks stemmed from the verbal, figural and numerical content domains, respectively. In addition, we varied task complexity, with half of the tasks in each domain being simple tasks (mean RTs < 1 second), and the other half more complex tasks (mean RTs > 2 seconds). The rationale behind the latter distinction was as follows. Typically, the types of tasks analyzed with the diffusion model have been quick and simple tasks with very low response times (under one second). The reason is that basic assumptions of the diffusion model, namely within-trial stability of parameters and the idea that a single evidence accumulation process underlies the decision process, were thought to be violated in more complex tasks (Ratcliff & McKoon, 2008). However, in recent years it has been proposed that the diffusion model is also applicable to slower response time paradigms, with initial validation studies showing promising results (Lerche & Voss, 2017a). Following up on this research, we wanted to systematically include slow response time tasks from each content domain in our study, in order to better judge the applicability of the model to such tasks. Another reason was that in studies drawing on mean RTs, more complex tasks were shown to show stronger relationships to intelligence than simple, fast tasks (Sheppard & Vernon, 2008). We wanted to test whether the same holds true when using drift rates as the measure of processing speed.

As outcomes, we used the scores of a standard intelligence test (Jäger et al., 1997) for general intelligence and the verbal, figural and numerical content domains. Through structural equation modeling, we tested whether the latent general intelligence factor was related to a latent general factor of drift rates and a method factor representing the shared variance of drift rates in the slow response time tasks. In addition, we included content-domain specific factors for verbal, figural, and numerical intelligence, as well as for the respective drift rate content domains, and studied their cross-relationships.

Our results showed a very distinct pattern. General intelligence was related to the general drift rate factor (r = .45) and the factor encompassing the variance specific to drift rates in complex tasks (r = .68). Both drift rate factors jointly explained 67% of the variance in general intelligence. Regarding the latent content domain factors of drift rates, all of them showed strong correlations with the respective intelligence content domains (verbal: r = .50; figural: r = .90; numerical: r = .74), but not with the theoretically unrelated intelligence content domains. It should be noted that while both the verbal and numerical drift rate (residual) factor showed statistically significant variance, this was not the case for figural drift rates. Finally, non-decision times also showed strong relationships to intelligence, but here the latent structural equation models all failed to show satisfying fit.

Our results support the notion that processing speed, as measured by drift rates, is not unitary, but contains content-domain specific aspects. The fact that these were related to the respective intelligence content domains speaks in favor of the validity of the measurement of these aspects in our structural equation model. It might be that the domain-specificity of processing speed was hidden in previous studies utilizing mean correct response times due to composites of processes contributing to RTs. In our analyses, we did not find a latent measurement model of (raw or logarithmized) correct mean RTs incorporating all 18 tasks with acceptable model fit, no matter if we used a *g* factor only model or more complex models also representing content domains and specifics of the slow tasks. Mean RTs, possibly due to the entanglement of processing speed in speed-accuracy trade-offs and non-decision times, seem to be both less domain-specific and show stronger correlations between particular dyads of tasks, represented by implied residual covariances in a structural equation modeling framework.

We also found strong additional evidence for a positive relationship between drift rates and general intelligence. This was especially pronounced for the more complex, slower response time tasks. While this mirrors findings reported for mean RTs (Sheppard & Vernon, 2008), ours was the first study to analyze the implied moderation with drift rates.

Taken together, this attempt at utilizing individual estimates of diffusion model parameters, namely drift rates, as measures of individual differences, brought important insights into both the across-task structure of processing speed and the relationships of the obtained structural components to intelligence. By employing the diffusion model, a clear pattern of results emerged, that was in contrast to the state-of-the-art based on raw data. In this way, the model-based approach of obtaining individual decision process parameters to better understand cognition and mental abilities proved to be a promising avenue. We continued to probe its utility in the following manuscripts, which focused on an aspect of cognition that was only briefly touched in this thesis up to this point: the question of whether there are age differences in decision process parameters.

4 Age differences in diffusion model parameters – a meta-analysis (Manuscript 3) [3]

Older people show longer response times in elementary cognitive decision tasks. This finding has been replicated numerous times over the past decades and holds true across a variety of paradigms (Jensen, 2006). Already in young adulthood, increasing age is associated with longer mean RTs (Salthouse, 1996, 2004, 2010).

However, over the past twenty years, a number of studies utilizing the diffusion model have started to challenge the assumption that processing speed declines with age. When disentangling the decision process components contributing to empirical raw data, one finds that higher mean RTs can have at least three different (though possibly correlated) causes: lower processing speed (drift rates), higher decision caution (boundary separation), or slower encoding and motor response processes (non-decision times).

Several studies compared young college-aged adults to old adults aged at least 60 regarding their individual diffusion model parameter estimates (e.g., Ball & Aschenbrenner, 2018; McKoon & Ratcliff, 2012, 2013; Ratcliff et al., 2003; Ratcliff, Thapar, et al., 2004; Spaniol et al., 2006; Thapar et al., 2003; Voskuilen et al., 2018). Generally, older adults exhibited higher decision caution and slower non-decision times. For drift rates, findings were more complex and also differed across studies. While in most cases there were no differences in processing speed as measured by drift rates between young adults and old adults, in some cases drift rates were higher for the younger group (Voskuilen et al., 2018). Conversely, there are even reports of slightly higher drift rates in the older age group (Ratcliff, Thapar, et al., 2004). To address these issues more systematically, we conducted a meta-analysis, with the aim to study the age effects on drift rates (and the other two core diffusion model parameters) thoroughly and quantitatively.

Our multi-level meta-analysis comprised 25 samples with a total *N* of 1,503. In addition to the main effect of age group, we tested two potential moderators of this effect. One of them was

[3] Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2020). Age differences in diffusion model parameters: a meta-analysis. *Psychological Research*, 1-10.

the type of task – we categorized the studies as either using a perceptual task, a lexical decision task, or a memory task. The second moderator in our model was task difficulty as measured by across-person mean drift rates.

We found strong age effects for boundary separations and non-decision times: The older age group showed on average higher boundary separations and longer non-decision times. For these parameters, the inclusion of the moderators did not lead to a better model fit. For drift rates, the model including both moderators and their interaction showed a significantly better fit than the more parsimonious models. Results indicate that older adults have lower drift rates in perceptual tasks and memory tasks, but slightly higher drift rates in lexical decision tasks. Older adults also performed relatively better in more difficult tasks. Regarding the interaction between the moderators, we found that older adults showed higher drift rates in more difficult settings only for perceptual and lexical decision tasks, not for memory tasks. Finally, there was a large proportion of between-study variance in age effects sizes that was not explained by either moderator.

Our meta-analysis highlighted the importance of type of task and task difficulty in determining the difference in drift rates found between college-aged and old adults. At the same time, there were of course also many other factors potentially contributing to differences between the studies – most importantly, the studies were based on different samples. In Manuscript 4, we therefore studied age differences in diffusion model parameters across 18 different tasks within the same sample, utilizing the data we had already analyzed in Manuscript 2. This data set also had the advantage of incorporating people from a continuous age range (18 to 62 years) – including participants from middle adulthood, a period of life rarely analyzed in diffusion model studies so far. Finally, we also wanted to study the specific associations between age, the diffusion model parameters, and intelligence.

5 Relationships of age, intelligence, and diffusion model parameters (Manuscript 4)[4]

Increasing age is not only associated with longer mean response times, but also with decreases in a wide range of other cognitive abilities, including general intelligence (Hartshorne & Germine, 2015; Jensen, 2006; Salthouse, 2004; Verhaeghen & Salthouse, 1997). As was

[4] von Krause, M., Lerche, V., Schubert, A. L., & Voss, A. (2020). Do Non-Decision Times Mediate the Association between Age and Intelligence across Different Content and Process Domains?. *Journal of Intelligence*, *8*(3), 33.

already mentioned, the associated deterioration of mean response times and other cognitive abilities has given rise to the theory that cognitive slow-down might be part of the causal basis of general cognitive decline (Salthouse, 1996). Specifically, mean RTs were repeatedly found to mediate the association between age and intelligence, both in cross-sectional and (to a lesser degree) in longitudinal studies (Finkel et al., 2007; Salthouse, 1996; Zimprich & Martin, 2002).

Yet drift rates, as a theoretically process-pure measure of cognitive speed, fail to show clear-cut associations with age – instead, the relation to age depends strongly on the type and difficulty of the task. This gives rise to the question which of the decision process components are responsible for the age association found for mean RTs and ultimately for the mediation of age differences in intelligence through mean RTs. As both boundary separations and non-decisions were linked to age in our meta-analysis, both were potential candidate mediators.

Schubert and colleagues (2020) tested several mediation models of age, fluid intelligence, and the core diffusion model parameters in a sample of 223 participants, while also including the P3 event-related potential (ERP) latency from an electroencephalogram as a potential mediator. They found that while both drift rates and boundary separation failed to mediate the negative relationship between age and fluid intelligence, non-decision times and the P3 latency jointly fully mediated said association. The P3 latency is thought to be linked to anterior brain regions associated with response planning, as well as higher-order processing (Schubert et al., 2020). The authors offered two alternative explanations of this mediation via non-decision times and the P3 latency. As they did not find non-decision times to be related to ERPs associated with encoding processes (i.e., N1 and P1), they inferred that it should be the motor processes reflected in non-decision times that are the basis of the mediation. First, these motor response times might reflect age-related differences in anterior brain regions associated with response planning and response execution, as well as higher order processing, as they showed to be closely linked to the P3 latency. Second, the mediation might reflect the influence of motor response processes on the intelligence test scores. As the IQ test had strict time limits for each task and relied on hand-writing for recording the answers, motor response speed should have influenced the scores obtained.

In Manuscript 4, we followed up on these questions in several ways. First, we examined the associations between the diffusion model parameters and age across 18 different tasks from the verbal, figural, and numeric content domains to study the generalizability of previous results. Second, we estimated mediation models of age, the diffusion model parameters, and different aspects of intelligence, utilizing a broad range of outcomes. Schubert and colleagues (2020) had used a single outcome measure, fluid intelligence. We estimated mediation models for general (fluid) intelligence, for three different intelligence content domains (verbal, figural, numerical), as well as for three different intelligence process domains (processing capacity, memory, psychometric speed).

The differentiation of process domains allowed us to directly study the two different explanations offered by Schubert and colleagues (2020) for the mediation via non-decision times. If age differences in non-decision times reflected age-related differences in anterior brain regions associated with response planning, response execution, but also higher order processing, then the mediation should occur similarly across all intelligence process domains. Conversely, if the mediation of the relationship between age and fluid intelligence via non-decision times reflected non-decisional aspects influencing the intelligence test scores, then the mediation should be especially distinct for the psychometric speed tasks of the intelligence test, that relied extensively on quick handwriting; on the contrary, the mediation should be less pronounced for the processing capacity tasks, that were closest to a power test among the intelligence test tasks.

In our sample of 125 participants that covered an age range of 18 to 62, we found that boundary separations and non-decision times showed positive correlations with age. This generally held true across the 18 different tasks studied, although the magnitude of the correlations sometimes differed between tasks. Drift rates mostly did not show any linear age trends, although there were some tasks where older adults had lower, and there was even one task where they had higher drift rates. The distribution of drift rate correlations showed no interpretable pattern, neither when comparing the content domains, nor between simple/fast and more complex/slow tasks. In post-hoc analyses, we found that many of the task-specific drift rates showed non-linear age trends, in that they exhibited cross-sectional increases until age 30, and a slow decline thereafter. Given the exploratory nature of these results and the sparse sample size representing middle adulthood, these findings should be interpreted cautiously.

In our mediation models, we replicated the mediation of the association between age and intelligence via non-decision times. Neither boundary separation nor drift rates were a significant mediator in any of our models. However, non-decision time mediated the link for all outcomes except for figural intelligence and processing capacity. Regarding the two different explanations of the non-decision time mediation proposed by Schubert and colleagues (2020), we found that the mediation effect was strongest for psychometric speed, and (as was already mentioned) not at all found for processing capacity. In this way, our results speak in favor of the hypothesis that the mediation of age differences in intelligence test scores via non-decision times most likely reflects the fact that motor response execution influences the intelligence test scores via speed of hand-writing.

Taken together, our results shed some light on the complex relationship between age, the decision process components reflected in the core diffusion model parameters, and other cognitive abilities. Given the finding that it might partly be motor speed that gives rise to age differences in intelligence test scores, it would be interesting to follow up on these results with a study using, as measure of intelligence, a true power test without any time pressure. Even more interesting seems the still unclear pattern of results for the association between age and drift rates. While our meta-analysis suggested that age effects on drift rates depended on the type of task, in our study of 18 different response time tasks no clear pattern emerged between content domains and task complexities (that also moderated results in the meta-analysis as task difficulty). In addition, in our exploratory analyses, we found evidence for non-linear age trends in the majority of tasks, with an increase in drift rates up to about age 30. Unfortunately, our sample size was far too small to explore the age trends over middle adulthood in greater detail. It thus seemed imperative to study age differences in diffusion model parameters in a much greater sample, in order to gain a clear view of the relation of age and processing speed across the lifespan.

6 Age differences in diffusion model parameters in a large sample (Manuscript 5)[5]

In order to be able to study in depth the associations of age and the diffusion model parameters, especially drift rates, across the lifespan, we re-analyzed a very large dataset of response times and accuracy rates. As an example of a binary decision task, we used the race implicit association test (IAT; Greenwald et al., 1998, 2003; Greenwald & Farnham, 2000). Since the introduction of the race IAT at the end of the 1990s, a great number of people have completed the test at the websites provided by Project Implicit (Xu et al., 2014). In the race IAT, people are shown either words or images, which they have to classify as belonging to one of two categories, for example, “good” or “bad” for the words, and “White person” or “Black person” for the images. The answer categories are mapped to the same response buttons. In this way, across experimental conditions, “good” is in one block paired with “Black person”, but in

[5]von Krause, M., Radev, S.T., & Voss, A. (submitted). Processing speed is high until age 60 - Insights from Bayesian modeling in a one million sample (with a little help of deep learning). *Proceedings of the National Academy of Sciences of the United States of America.*

another block paired with “White person”. The (transformed) difference in mean response times is interpreted as reflecting the strength of implicit associations of the respective categories, and thus an implicit racial bias.

We were not interested in the race IAT as a measure of implicit cognition, but rather as an example of a binary decision task. IAT data have already been analyzed successfully with the diffusion model (Klauer et al., 2007), making the Project Implicit data a promising target for our analysis of age differences in diffusion model parameters in a large sample. We obtained the raw data from Project Implicit – summary statistics and demographics were already available at the Project’s OSF page (https://osf.io/y9hiq/). For our analyses, we used raw data collected between September 2016 and December 2018, adding up to a total of over 1,800,000 people – a sample size we deemed sufficient for fine-grained analyses of age differences.

When obtaining diffusion model parameters for such a large data set, standard estimation methods become computationally infeasible, especially when a Bayesian approach is employed. Thus, we used BayesFlow, a newly developed deep learning method based on invertible neural networks for extremely efficient parameter estimation (Radev et al., 2020). Utilizing the BayesFlow method, we obtained full individual posterior distributions of the three core diffusion parameters for our large sample on a standard laptop within a day. After data cleaning, our sample size was 1,185,898. Ages 10 to 80 were covered in sufficient depth for year-specific analyses, with oftentimes (tens of) thousands of participants for each year of age.

To be able to better compare our results with previous studies measuring processing speed with mean correct RTs (Finkel et al., 2007; Hartshorne & Germine, 2015; Jensen, 2006; Salthouse, 1996, 2010; Zimprich & Martin, 2002), we also analyzed the age relations with mean correct RTs in our sample. Specifically, we computed the across-person means of the individual mean correct RTs, and the individual posterior means of drift rates, boundary separations, and non-decision times, separately for each year of age.

Mean correct RTs showed, on average, decreases from age 10 to about the age of 20. They then exhibited a quasi-linear positive trend that continued until about age 60, after which this age-related tendency (i.e., cross-sectional slow-down) in mean RTs increased further, with the highest mean RTs found around the age of 80. This finding is in line with what was previously reported in the literature (see e.g., Hartshorne & Germine, 2015; Jensen, 2006; Salthouse, 2004).

Boundary separation, that is, decision caution, exhibited an age-related pattern that mirrored the one found for mean RTs, at least until about the age of 60. Decision caution displayed a decreasing trend over the teenage years, was lowest around age 20, and showed, on average, a quasi-linear increase thereafter up until the age of 80, with a steeper increasing trend found after age 60 for the incongruent experimental condition. Non-decision times, that is, the times needed for encoding and motor processes, exhibited a decreasing trend until ages 14 to 16, and a quasi-linear average increase starting thereafter, that continued over the entire age span studied.

Our most interesting results were found for drift rates, that is, processing speed. On average, the drift rates exhibited an increasing trend that lasted until about age 30. From ages 30 to 60, there were little age differences in processing speed as measured by drift rates, with a slow average decline starting at around age 50. From age 60 on, a clear and accelerating slow-down in average drift rates was present in our data, that continued until the age of 80 – the latter trend was more pronounced for the incongruent compared to the congruent experimental condition.

Also of interest was the fact that while mean RTs displayed an increase in across-person variance in old age, neither drift rates nor boundary separation exhibited a corresponding trend, but non-decision times noticeably did.

Taken together, our results help to explain many of the age patterns found in previous diffusion model studies, our meta-analysis (Manuscript 3) and in our 18 task study (Manuscript 4). Previous diffusion model studies most often compared drift rates found in college aged participants with those obtained from old adults, aged 65 or older. When taking the results of our large IAT analysis into account, it seems that the young age group in such studies might have not yet reached their maximum in processing speed, while the older age group might have already started age-related decline after a period of stability over middle adulthood. This might explain why no consistent differences in drift rates were found in these previous studies. Of course, given the cross-sectional nature of our data, such interpretations must remain cautious.

Referring back to Manuscript 4, our results from the large dataset replicate the finding that processing speed peaks around the age of 30 years. It seems that after the age of 60, trends in boundary separations, non-decision times and drift rates jointly contribute to an age-related slow-down, that is also mirrored in mean RTs. In this way, the results we found in the large IAT dataset complement the previous findings presented in this thesis on the relationships of age, decision process components as represented in the three core diffusion model parameters, and (other) cognitive abilities.

The results also shed a new light on our meta-analysis, where we found that differences in drift rates between young and old adults partly depended on the type of task studied. It seems plausible that the shifting point towards an accelerated decline in drift rates, that we found to be roughly at the age of 60 for the IAT, could be earlier or later in the lifespan for different types of task. For example, in lexical decision tasks, where old people might profit from their practice of language over a long period of time (Ratcliff, Gomez, et al., 2004; Ratcliff, Thapar, et al., 2004), processing speed might decline later than in simple perceptual tasks. In this way, the fact that most of the two group studies used simply one category for old adults (because of the low overall sample sizes) might have hidden these differential developmental patters by mapping them to a very simple three-point scale: that old adults generally either have similar, higher, or lower drift rates than young adults in a particular task.

7 Discussion

In this section, I will discuss the five manuscripts that are part of this thesis, uniting them in relation to the overall topic of this work: the use of diffusion model parameters in individual differences research. I will point out some of the limitations of this research, suggest some ideas for future projects, and finally give some concluding remarks.

7.1 Summary and General Discussion

My thesis can be divided in two main parts. In the first part, Manuscript 1, I tested the assumptions underlying the use of the diffusion model parameters as estimates of reliable individual differences. Specifically, I focused on an aspect that had been rarely studied in the diffusion model literature, that is, to what extent the model parameters exhibit stability and change across a longer time period. This question is essential to determine whether the diffusion model should be considered traits, as these are expected to be stable across situations and time. Because the stability across situations had already been studied in various research projects analyzing across-task correlations of the parameters (e.g., Ratcliff et al., 2010; Schmiedek et al., 2007; Schubert et al., 2016), my focus was on the temporal aspect.

We found that the main diffusion model parameters (drift rates/processing speed, boundary separation/decision caution, encoding and motor processes/non-decision times) showed great rank-order stability over time periods of up to two years. In addition, profiles of the relative (standardized) values of the parameters were also very stable for the majority of people. Finally, mean-level change and individual differences in change were easily interpretable as training effects leading to more effective information accumulation and lower decision caution. In this way our results support the assumption that the core diffusion model parameters can be considered trait-like, as they show great temporal stability and interpretable developmental patterns. Together with previous studies that also focused on the trans-situational aspect of stability (Lerche & Voss, 2017b; Ratcliff et al., 2010; Schubert et al., 2016;

Yap et al., 2012), Manuscript 1 thus provided a strong additional argument for the application of the diffusion model and the individual parameter estimates obtained from it in individual differences research. In Manuscripts 2 to 5, I shifted the focus from the question of how the diffusion model parameters should be interpreted towards the more applied question how these parameters can be helpful to conduct better research on individual differences in cognition.

I set out to seek answers to two questions arising out of the literature on individual differences in cognitive parameters: i) Is the structure of processing speed across tasks truly unitary, and how should its relationship to intelligence best be described? ii) What are the exact relationships between age, processing speed and other cognitive abilities? In both cases, previous studies had suffered from multiple shortcomings – they had, for example, either relied on mean RTs as a heuristic measure of processing speed, or had, in the case of diffusion model studies, used small numbers of tasks or tested only small and demographically homogeneous groups of participants. In our studies, we tried to address these limitations.

In Manuscript 2, we analyzed the across-task structure of processing speed and its relationship to intelligence. Utilizing 18 tasks from three different content domains, with half the tasks being simple and fast, and the other half being more complex, we found a distinct pattern of results. Processing speed, measured as drift rates, was best represented by a multi-faceted hierarchical structure, encompassing both a general factor, content-domain specific aspects, and a method factor representing the shared variance of the more complex tasks. This is in contrast to processing speed as measured in correct mean RTs, where we could not find a measurement model with adequate fit to our data. In addition, we found that the content domains of processing speed showed strong positive relations to the respective intelligence domains, while the factors representing general processing speed and the shared variance of the complex tasks predicted about 70% of the variance in general intelligence.

These results clearly spoke in favor of two interpretations: First, processing speed is not unitary, but multi-faceted and partly domain-specific; second, processing speed shows a robust and specific relationship to other cognitive abilities, most importantly general intelligence. The use of the drift rate parameter thus provided unique and novel insights that would not have been attainable had we relied on mean RTs as our estimates of processing speed.

The second main complex of substantial research I tackled in this thesis concerned the relationship between age, diffusion model parameters, and, once more, intelligence. In previous studies, decision caution and non-decision times were quite consistently higher in older people, while the pattern of results on the relation of age and drift rates remained ambiguous. This uncertainty regarding drift rates might have been caused by the fact that most of the studies reported only results based on a very low number of tasks from small samples, typically consisting of two age groups (young adults and old adults). To provide a clear picture, we studied age differences in three steps: i) we quantitatively analyzed the results reported in previous studies in a meta-analysis (Manuscript 3); ii) we studied age differences across 18 tasks within the same sample (Manuscript 4); iii) we provided fine-grained, year-specific age trend analyses based on a single very large sample (Manuscript 5).

Combined, our results underline the notion that boundary separations and non-decision times show higher across-person means with increasing age, even among young adults. For non-decision times, we even found that the lowest values were among teenagers aged 14 to 16 (see Manuscript 5). Conversely, drift rates seem to show increases over large parts of young adulthood (ages 20 to 30; see Manuscripts 4 and 5). Means in drift rates were roughly equal over middle adulthood, with an accelerated decrease in old adulthood in our analyses presented in Manuscript 5.

The task-specificities of the relationship between age and drift rates found in the meta-analysis (Manuscript 3) were only partly mirrored in our 18 task study (Manuscript 4). It must be noted that previous studies (that entered the meta-analysis) compared young adults and old people, often aged 65 and older. On the contrary, in our sample the oldest participants were 62 years old. Yet, we found a verbal task (though not the lexical decision task that formed a category in our meta-analysis) to be the only one to show a positive age trend.[6]

Finally, in Manuscript 4 we also studied which parts of a decision process might be responsible for the mediation of the relationship between age and intelligence via mean RTs reported in the literature (Salthouse, 1996; Zimprich & Martin, 2002). As it turned out, the most likely explanation of the mediation was via non-decision times – Schubert and colleagues (2020) had reported similar results. In our study, we replicated their findings in regard to general (fluid) intelligence, and expanded the results to other outcome measures, namely different intelligence content domains and intelligence process domains (which had not been done before).

We also provided evidence in favor of one of the two possible explanations of the mediation via non-decision times offered by Schubert and colleagues (2020). The differential results we found among the intelligence process domains, with the strongest mediation found for the psychometric speed intelligence tasks, and no mediation via non-decision times found

[6] In this task, participants had to judge whether a word shown on the screen was a noun or an adjective.

for the processing capacity tasks, support one specific interpretation. As the psychometric speed scores are in large part dependent on speed of handwriting, whereas the processing capacity scores are much closer to a power test, one might infer that it is precisely the motoric component inherent in intelligence test scores that at least partly drives the age differences found for them, and subsequently also their mediation via mean RTs. Of course, in the light of the non-linear age trends we found for drift rates in Manuscript 5, the (linear) mediation models estimated in Manuscript 4 also for drift rates should probably be reconsidered.

Bringing together the results reported in all five manuscripts that are part of this thesis, it seems that applying the diffusion model to obtain individual estimates of decision process components is both possible and fruitful. Diffusion model parameter estimates provide reliable, stable measures that show interpretable developmental patterns over a time period of up to two years and might therefore be at least in the temporal respect considered trait-like entities (Manuscript 1). When used to study substantial research questions, the parameters provide novel insights that would be impossible to obtain when relying on raw data. Manuscript 2 demonstrated this with regard to the structure of processing speed and its relationship to intelligence; we found content domain specific aspects of processing speed related to the respective intelligence components, that were not recoverable when analyzing mean RTs.

In Manuscripts 3 to 5, we scrutinized the relationship between age and the decision processes components represented by core diffusion model parameters. Once more, our results, especially regarding differences in processing speed across the lifespan, were in sharp contrast to what was previously inferred based on raw data. These combined findings are also a significant step forward from previous diffusion model analyses, given our strong data, with large numbers of tasks and participants and wide age ranges studied.

Another important keystone of our studies was the use of state-of-the-art parameter estimation methods. In Manuscript 1, we employed hierarchical Bayesian diffusion modeling (Wiecki et al., 2013), while in Manuscript 5, we utilized a novel deep learning approach for efficient Bayesian parameter estimation (Radev et al., 2020). These methods enabled us to reliably assess individual differences in diffusion model parameters, even in a large sample, and were thus an important prerequisite for our analyses.

7.2 Limitations and Ideas for Future Research

The research program described within this thesis has a number of unique features. Most importantly, we obtained robust, reliable and informative results by studying four different types of longitudinal development in diffusion model parameters over a long time period

(Manuscript 1), measuring the diffusion model parameters in 18 diverse tasks and studying their relations to a set of intelligence outcomes (Manuscript 2 and 4), systematically summarizing previous findings in a meta-analysis (Manuscript 3), and utilizing heterogeneous (Manuscripts 1, 2, 4, 5) and large (Manuscript 5) samples. These advantages were vital for obtaining the interesting findings of our studies. However, it must also be noted that this thesis also has some limitations.

First, the manuscripts concerned with the relationships of drift rates, non-decision times, and other cognitive abilities such as general intelligence, might have profited from incorporating a neurophysiological approach. The studies on the neural correlates of diffusion model parameters are plentiful (for overviews, see e.g. Dully et al., 2018; Schubert & Frischkorn, 2020). Yet specifically in the context of examining the structure of processing speed (Manuscript 2) it would have been interesting to note if differentiable patterns in processing speed map to differentiable patterns in brain activation.

Second, for all results on age differences in diffusion model parameters, it is important to note that we report purely cross-sectional data. Therefore, strictly speaking, statements about a longitudinal change are not possible. In order to get a better view of the true developmental patterns underlying the age-related mean differences we found in our studies, it would be vital to follow and test a group of participants over time, ideally for decades. A related aspect is that of cohort effects. We did not differentiate age effects and cohort effects in our analyses. In this way, it might for example be the case that the lower across-person means in drift rates we found for participants aged 60 and older in Manuscript 5 are partly explainable by the fact that these people had less experience in responding to computer tasks, independent of their age. Fortunately, the raw data published by Project Implicit (Xu et al., 2014) were collected between 2002 and 2020, and also include participant IDs. In this way, it should be possible to both study cohort effects (comparing, for example, people aged 60 in 2002 to people of the same age in 2020), and longitudinal developments in parameter values for people who participated several times over the years. Regarding cohort effects, it might also be interesting to study whether there is a Flynn effect (Flynn, 1987) in processing speed, possibly attributable to greater familiarity with computer-based assessments: Over the years of data collection, people might generally exhibit faster processing speed, regardless of age.

Third, in our analysis we did not take into account new developments in the diffusion model literature regarding the introduction of a possible additional model parameter, alpha, that describes the individual degree of heavy-tailedness in the noise distribution underlying the information accumulation process (Voss et al., 2019; Wieschen et al., 2020). If alpha is considerably lower than 2, this indicates a deviation from a standard diffusion process, and can model (random) jumps in the information sampling process, that might signify sudden insights. The literature on this topic is still in its infancy, but it seems worthwhile to study individual differences in this new parameter and its embedding in a nomological network of related constructs to be able to better interpret it.

Fourth, we put our focus strictly on the diffusion model and did not apply other types of evidence accumulation models that have been proposed in the literature, like the linear ballistic accumulator model (LBA; Brown & Heathcote, 2008) or leaky competing accumulator models (LCA; Usher & McClelland, 2001). The diffusion model as proposed by Roger Ratcliff (Ratcliff, 1978; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998) is but one member of the family of models trying to translate the underlying processes in simple decision making to a mathematical formulation. The competing models have a number of unique features, for example, the LCA has a stronger focus on mirroring the neural basis of information processing. In our studies, we only used the diffusion model. There were two main reasons for doing so. First and most importantly, the majority of the main model parameters are quite similar among the diffusion model, the LBA, and the LCA. This relates both to the implementation of the parameters and especially to their psychological interpretation. For example, Donkin et al. (2011) found that the diffusion model and the LBA agree on the mappings of the effects of experimental manipulations to the main model parameters (speed of information accumulation, decision caution, non-decision time), concluding that “inferences about psychological processes made from real data are unlikely to depend on the model that is used” (Donkin et al., 2011, p. 61; but see Goldfarb et al., 2014). Given that the diffusion model is the type of evidence accumulation model that has received by far the most attention in the literature and is also likely to continue being the most-researched approach (Voss et al., 2013), it made sense to probe the usefulness of this particular model. Second, as we only analyzed binary decision tasks, we had no need to employ one of the models that can also accommodate choices among multiple response options (Brown & Heathcote, 2008; Usher & McClelland, 2001).

After pointing out some limitations of the research project presented in this dissertation, I will now sketch a few ideas for additional analyses and possible future studies based on our results. Regarding the relation between drift rates and intelligence studied in Manuscript 2, it might be interesting to see whether drift rates show positive relationships to some of the real-life outcomes that intelligence is known to predict, for example, educational success (Sternberg, 2000). While in our 18 task study we did not assess educational background, I followed up on our results linking drift rates and intelligence after I had obtained diffusion model parameter estimates from the large IAT dataset described in Manuscript 5. This dataset also contains numerous additional measures, for example, detailed demographic questionnaires and personality items. I wanted to probe whether drift rates are higher for people with a stronger educational background in the large sample. The highest level of education attained was related to age, so I first regressed the drift rates (of trials from the incongruent condition, but results were similar for congruent trials) on age and age squared, to account for the non-linear relation of age to drift rates. I then analyzed the distributions of drift rates over levels of education. To my knowledge, no similar analyses have been published to this date.

Figure 1 shows the corresponding plot. The points indicate the group-specific means, with bars representing one standard deviation. As can be seen, within-group variance is high across all levels of education. Nevertheless, an increasing trend can be found with higher drift rates for people of a higher level of education, up until the point where people at least attended "some college". Please note that age and the quadratic effect of age were controlled for in these graphical analyses. When post-hoc dichotomizing the data in people with no college education vs. at least some college education, I found a corrected effect size of d = .307 for the difference in drift rates (in the incongruent condition). The humble effect size and large within-group standard deviations in the residualized analyses underline the scope of individual differences. Yet, it seems that drift rates are, on average, slightly higher among people with a higher education level, underlining their relationship to cognitive abilities such as intelligence.

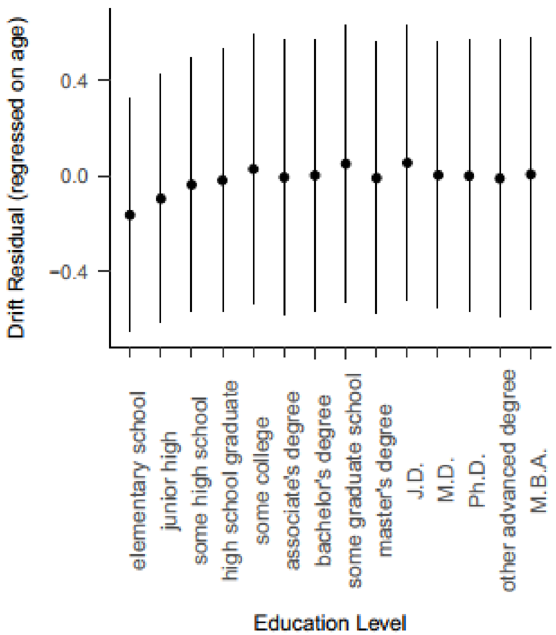

Figure 1. *Mean levels and standard deviations of drift rates (from the incongruent condition and after controlling for age and the quadratic effect of age) for each level of education in the dataset used in Manuscript 5 (N = 1,185,898). For the congruent condition, trends are similar.*

Another interesting line of research continuing on from our results would be to repeat the mediation analyses conducted for Manuscript 4 with participants from an age range that we would expect to show clear linear age trends in drift rates based on Manuscript 5, for example, people aged 50 to 80 – maybe also in combination with both a power test and a speeded task as intelligence outcomes. The precise differences in mean parameter values over the lifespan obtained in the very large dataset allow us to specifically define age ranges of interest. Similarly, the finding that non-decision times were lowest at around the age of 15 opens up an unexplored field of research, namely the developmental patterns of non-decisional processes in adolescence. It would also be intriguing to examine whether it is mostly encoding times or motor response execution times that drive the age differences in non-decision times.

Finally, given our results from the large sample (Manuscript 5), one possible approach for a future study assessing differences among types of response time tasks might be to employ a perceptual task, a memory task, and a lexical decision task (the categories we used in our meta-analysis) in a fairly large, representative sample covering large parts of adulthood (e.g., ages 18 to 80). Such an approach could be conducted with reasonable costs, if the number of trials per person was kept low. The results obtained from such a study might provide better insights in the differential (cross-sectional) temporal patterns between tasks. Based on the results found for the IAT data, one might expect the tasks to show quantitative differences regarding the age ranges where, on the one hand, the maximum in drift rates is observed, and, on the other hand, the trend towards a decline in drift rates becomes clear. The latter might start quite late for vocabulary-based tasks such as the lexical decision task, as high scores in verbal tasks were also found in less-speeded contexts for older adults (Hartshorne & Germine, 2015).

7.3 A measure of dark personality based on the diffusion model?

One possible avenue of extending the use of the diffusion model in individual differences research is to turn away from the interpretation of the parameters purely as cognitive process parameters as an end in itself, but rather to use them to calculate derived measures, for example, of implicit personality. As has already been noted, the diffusion model has successfully been applied to IAT data (Klauer et al., 2007, Manuscripts 1 and 5). The so-called IAT effect, that is, the difference in (adjusted) mean response times between the congruent and incongruent conditions in an IAT was shown to be closely mapped by the difference in drift rates between the conditions (Klauer et al., 2007). In this way, the drift rate difference scores constitute a measure of implicit association, or, in case of a personality IAT, of implicit personality.

Indirect measures such as the IAT try to capture implicit processes and associations and specifically aim at assessing socially aversive attitudes or traits, for example, implicit racial bias (Greenwald et al., 1998). Other examples of such constructs would be the so-called dark personality traits, with the most prominent being the Dark Triad of narcissism, Machiavellianism, and psychopathy (Furnham et al., 2013; Paulhus, 2014; Paulhus & Jones, 2015; Paulhus & Williams, 2002). In recent years, it has been proposed that the variety of dark trait constructs proposed in the literature shares a common core that is characterized by the "tendency to maximize one's individual utility - disregarding, accepting, or malevolently provoking disutility for other - accompanied by beliefs that serve as justifications" (Moshagen et al., 2018, p. 656).

As such traits should by their very definition be socially aversive (given the need for justifying beliefs), people might be inclined to present themselves incorrectly, either because of a conscious use of strategies aiming at making a positive impression, or unconsciously, because of insufficient insight in one's own trait expression on a particular dark trait (Back et al., 2009; Calanchini & Sherman, 2013; Greenwald & Farnham, 2000; Quintus et al., 2020). Thus, the development of an indirect assessment method relying on the use of drift rate difference scores seems a promising avenue for dark personality research to explore a novel type of assessment.

Over the course of two different studies (both *N*s > 300), we developed two different measures of an implicit "dark score". On the one hand, we tested an IAT, using the categories "me" and "other" (see Schmukle et al., 2008, for a similar approach aiming at the Big Five of personality), and adjectives associated with either the positive or negative pole of a hypothesized dark trait continuum (e.g., "spiteful", "good-hearted"). On the other hand, we tested a slightly different assessment method. Here, the binary decision options participants had to choose from were "that's me" and "that's not me" – in this way, there were no longer correct and wrong answers (as in the IAT). The stimuli were adjectives representing either "dark" personality or its opposite, based on a review of the literature on dark traits. As a measure of dark personality, we calculated the mean of drift rates between the answers for the positively- or negatively-coded dark trait stimuli.

We successfully fit the diffusion model to the data from both experimental paradigms tested. Figure 2 shows the exemplary scatterplot of the empirical response time quartiles and response choices for the indirect dark trait ("D") measure using the "that's me / that's not me" answer categories, plotted against the simulated data based on the diffusion model parameters.

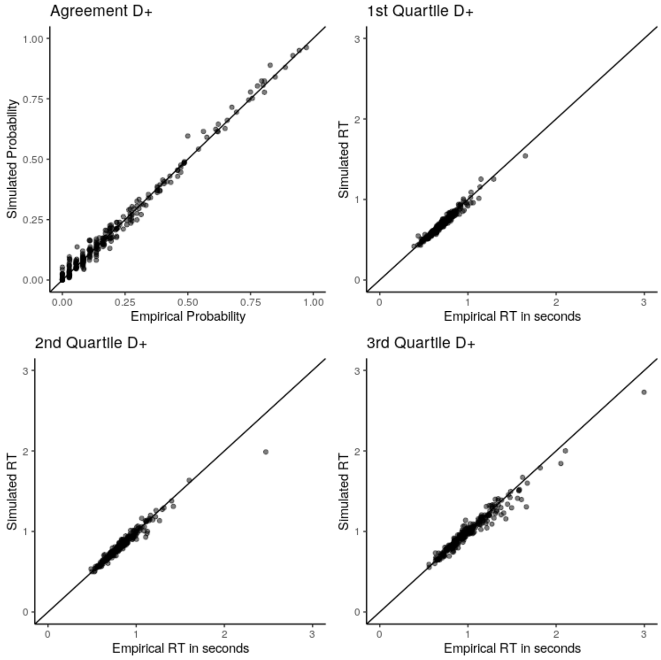

Figure 2. *Empirical response time quartiles and response choice probabilities plotted against corresponding simulated values based on the diffusion model parameter estimates. Participants either agreed ("that's me") or disagreed ("that's not me") to an adjective of the dark trait spectrum (e.g., "hateful"). Based on the data collected in Dark Trait Study 1.*

As an example, I show the answers to stimuli linked to the "dark" or socially aversive pole of "D" in Study 1. As can be seen, the recovered values closely mirror the empirical data, indicating reasonable generative performance of the diffusion model in the tasks employed.

We also obtained questionnaire data measuring several commonly studied dark trait constructs and different types of behavioral outcomes. To keep the results short, the measure of dark personality based on an IAT failed to show any considerable criterion-related correlations. The measure based on the "that's me" / "that's not me" distinction predicted actual behavior (sharing real money with other participants, lowering their payment, and cheating to avoid a tedious task) in a similar magnitude to dark trait questionnaire scales.

Data analysis of these studies is still ongoing and additional studies might be required to determine the usefulness of a measure of dark personality based on diffusion model estimates. In any case, opening the interpretation of parameter estimates and their differences to a context that is no longer based on ability testing, but seeks to obtain content-specific judgments on, for example, one's (dark) personality, seems a promising approach according to our initial results.

7.4 Conclusions

In this thesis, I presented a research program on the use of the diffusion model parameters as measures of individual differences. After studying the temporal patterns of individual parameter estimates, I applied the diffusion model to two different sets of substantive research questions from the field of individual differences in cognition. Employing the diffusion model to disentangle the different process components contributing to the raw data of response times and accuracy rates made it possible to gain novel insights in the across-task structure of processing speed, its relationships to other cognitive parameters, and its relationship to age. These findings could not have been obtained from raw data, and in many cases were in direct contrast to previous results based on mean RTs - the most important new finding probably being that in our large cross-sectional IAT dataset, processing speed was high throughout middle adulthood, although average mean RTs showed a positive age trend already from the beginning of young adulthood. In this way, we could show that individual differences research can profit from taking a model-based perspective on cognition.

As a member of the Research Training Group Statistical Modeling in Psychology (SMiP), I will at this point briefly point out the relation of my studies to the aims and conceptual framework of SMiP. One of the core elements of SMiP is the idea that there is a gap between psychological research focusing on developing statistical methods and substantive research. Novel statistical approaches are often largely ignored in applied studies (Sharpe, 2013) – a fact that might have detrimental consequences for scientific progress, and that SMiP is hoping to help overcome. My research focuses on how diffusion modeling, an elaborate statistical modeling technique, can be joined with individual differences research, and is thus in line with the core features and mission of SMiP.

The model-based study of decision process components to describe individual differences has in the past often adhered to research practices better suited for experimental psychology, for example, in the relatively low sample sizes used and the often-found loyalty to the comparison of parameter means between two groups as the main method of analysis. These same research practices are also assumed to form an important part of the so-called replication crisis still haunting large parts of psychology (Stanley et al., 2018). In this sense, implementing more robust research practices, for example by increasing statistical power, testing the generalizability of findings across a variety of paradigms, openly sharing data and also utilizing shared data for both replication studies and novel research should only bring fruitful results.

By following principles deemed important in individual differences research, for example, using longitudinal studies or improving reliability by employing numerous tasks and large and heterogeneous samples, the diffusion model parameters, originally stemming from a background in cognitive, experimental psychology, could successfully be transferred to a new context. In the end, we could show that experimental psychology can profit from incorporating ideas rooted in individual differences research, while scientists interested in the ways people differ from one another gain a powerful new tool by embracing mathematical modeling approaches. In this way, my thesis hopes to help bridge the gaps between these all-too-often separated fields of psychological research.

References

Allport, G. W. (1937). *Personality: A psychological interpretation*. H. Holt.

Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research*, *79*(5), 882–898. https://doi.org/10.1007/s00426-014-0608-y

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, *97*(3), 533–548. https://doi.org/10.1037/a0016229

Ball, B. H., & Aschenbrenner, A. J. (2018). The importance of age-related differences in prospective memory: Evidence from diffusion model analyses. *Psychonomic Bulletin & Review*, *25*(3), 1114–1122. https://doi.org/10.3758/s13423-017-1318-4

Binet, A., & Simon, Th. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'année psychologique*, *11*(1), 191–244. https://doi.org/10.3406/psy.1904.3675

Boywitt, C. D., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition*, *40*(1), 70–82. https://doi.org/10.3758/s13421-011-0128-6

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive Modeling*. SAGE.

Calanchini, J., & Sherman, J. W. (2013). Implicit Attitudes Reflect Associative, Non-associative, and Non-attitudinal Processes. *Social and Personality Psychology Compass*, *7*(9), 654–667. https://doi.org/10.1111/spc3.12053

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, *18*(1), 61–69. https://doi.org/10.3758/s13423-010-0022-4

Dully, J., McGovern, D. P., & O'Connell, R. G. (2018). The impact of natural aging on computational and neural indices of perceptual decision making: A review. *Behavioural Brain Research*, *355*, 48–55. https://doi.org/10.1016/j.bbr.2018.02.001

Dutilh, G., Krypotos, A.-M., & Wagenmakers, E.-J. (2011). Task-Related Versus Stimulus-Specific Practice: A Diffusion Model Account. *Experimental Psychology*, *58*(6), 434–442. https://doi.org/10.1027/1618-3169/a000111

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*(6), 1026–1036. https://doi.org/10.3758/16.6.1026

Evans, N. J., & Brown, S. D. (2017). People adopt optimal policies in simple decision-making, after practice and guidance. *Psychonomic Bulletin & Review*, *24*(2), 597–606. https://doi.org/10.3758/s13423-016-1135-1

Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press.

Finkel, D., Reynolds, C. A., McArdle, J. J., & Pedersen, N. L. (2007). Age changes in processing speed as a leading indicator of cognitive aging. *Psychology and Aging*, *22*(3), 558–568. https://doi.org/10.1037/0882-7974.22.3.558

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*(2), 171–191. https://doi.org/10.1037/0033-2909.101.2.171

Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of Personality: A 10 Year Review: Dark Triad of Personality. *Social and Personality Psychology Compass*, *7*(3), 199–216. https://doi.org/10.1111/spc3.12018

Galton, F. (1908). *Memories of my life*. Methuen & Company.

Goldfarb, S., Leonard, N. E., Simen, P., Caicedo-Núñez, C. H., & Holmes, P. (2014). A comparative study of drift diffusion and linear ballistic accumulator models in a reward maximization perceptual choice task. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00148

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*(6), 1022–1038. https://doi.org/10.1037/0022-3514.79.6.1022

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the lifespan. *Psychological Science*, *26*(4), 433–443. https://doi.org/10.1177/0956797614567339

Horn, S. S., Bayen, U. J., & Smith, R. E. (2013). Adult Age Differences in Interference From a Prospective-Memory Task: A Diffusion-Model Analysis. *Psychonomic Bulletin & Review*, *20*(6), 1266–1273. https://doi.org/10.3758/s13423-013-0451-y

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: BIS-Test*. Hogrefe.

Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences*. Elsevier.

John, Oliver P., Robins, R. W., & Pervin, L. A. (Eds.). (2008). *Handbook of personality: Theory and research* (3rd ed). Guilford Press.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology*, *93*(3), 353–368. https://doi.org/10.1037/0022-3514.93.3.353

Kühn, S., Schmiedek, F., Schott, B., Ratcliff, R., Heinze, H.-J., Düzel, E., Lindenberger, U., & Lövden, M. (2010). Brain Areas Consistently Linked to Individual Differences in Perceptual Decision-making in Younger as well as Older Adults before and after Training. *Journal of Cognitive Neuroscience*, *23*(9), 2147–2158. https://doi.org/10.1162/jocn.2010.21564

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. Academic Press.

Lerche, V., & Voss, A. (2017a). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research*. https://doi.org/10.1007/s00426-017-0945-8

Lerche, V., & Voss, A. (2017b). Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological Research*, *81*(3), 629–652. https://doi.org/10.1007/s00426-016-0770-5

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, *40*(1), 77–105. https://doi.org/10.1007/BF02291481

Madden, D. J., Costello, M. C., Dennis, N. A., Davis, S. W., Shepler, A. M., Spaniol, J., Bucur, B., & Cabeza, R. (2010). Adult Age Differences in Functional Connectivity during Executive Control. *NeuroImage*, *52*(2), 643–657. https://doi.org/10.1016/j.neuroimage.2010.04.249

McGovern, D. P., Hayes, A., Kelly, S. P., & O’Connell, R. G. (2018). Reconciling age-related changes in behavioural and neural indices of human perceptual decision-making. *Nature Human Behaviour*, *2*(12), 955–966. https://doi.org/10.1038/s41562-018-0465-6

McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in item recognition. *Journal of Memory and Language*, *66*(3), 416–437. https://doi.org/10.1016/j.jml.2011.12.001

McKoon, G., & Ratcliff, R. (2013). Aging and predicting inferences: A diffusion model analysis. *Journal of Memory and Language*, *68*(3), 240–254. https://doi.org/10.1016/j.jml.2012.11.002

Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. https://doi.org/10.1037/rev0000111

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at Age 7: A Methodological and Conceptual Review. In *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). Psychology Press.

Paulhus, D. L. (2014). Toward a Taxonomy of Dark Personalities. *Current Directions in Psychological Science*, *23*(6), 421–426. https://doi.org/10.1177/0963721414547737

Paulhus, D. L., & Jones, D. N. (2015). Measures of Dark Personalities. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 562–594). Academic Press. https://doi.org/10.1016/B978-0-12-386915-9.00020-6

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Quintus, M., Egloff, B., & Wrzus, C. (2020). Daily life processes predict long-term development in explicit and implicit representations of Big Five traits: Testing predictions from the TESSERA (Triggering situations, Expectancies, States and State Expressions, and ReActions) framework. *Journal of Personality and Social Psychology*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/pspp0000361

Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L., & Kothe, U. (2020). BayesFlow: Learning Complex Stochastic Models With Invertible Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15. https://doi.org/10.1109/TNNLS.2020.3042395

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, *2*(4), 237–279. https://doi.org/10.1037/dec0000030

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review*, *111*(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology*, *97*(2), 1756–1774. https://doi.org/10.1152/jn.00393.2006

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A Diffusion Model Analysis of the Effects of Aging in the Lexical-Decision Task. *Psychology and Aging*, *19*(2), 278–289. https://doi.org/10.1037/0882-7974.19.2.278

Ratcliff, R., Thapar, A., & Mckoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics*, *65*(4), 523–535. https://doi.org/10.3758/BF03194580

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*(3), 127–157. https://doi.org/10.1016/j.cogpsych.2009.09.001

Reisberg, D. (Ed.). (2013). *The Oxford handbook of cognitive psychology*. Oxford University Press.

Roberts, B. W., Caspi, A., & Moffitt, T. E. (2001). The kids are alright: Growth and stability in personality development from adolescence to adulthood. *Journal of Personality and Social Psychology*, *81*(4), 670–683. https://doi.org/10.1037/0022-3514.81.4.670

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3–25. https://doi.org/10.1037/0033-2909.126.1.3

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Roberts, Brent, Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 375–398). The Guilford Press.

Salthouse, T. A. (1996). The Processing-Speed Theory of Adult Age Differences in Cognition. *Psychological Review*, *103*(3), 403. https://doi.org/0033-295X/96/$3.00

Salthouse, T. A. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science*, *13*(4), 140–144. https://doi.org/10.1111/j.0963-7214.2004.00293.x

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society : JINS*, *16*(5), 754–760. https://doi.org/10.1017/S1355617710000706

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), 414–429. https://doi.org/10.1037/0096-3445.136.3.414

Schmukle, S. C., Back, M. D., & Egloff, B. (2008). Validity of the Five-Factor Model for the Implicit Self-Concept of Personality. *European Journal of Psychological Assessment*, *24*(4), 263–272. https://doi.org/10.1027/1015-5759.24.4.263

Schubert, A.-L., Frischkorn, G., Hagemann, D., & Voss, A. (2016). Trait Characteristics of Diffusion Model Parameters. *Journal of Intelligence*, *4*(3), 7. https://doi.org/10.3390/jintelligence4030007

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*, 140–146. https://doi.org/10.1177/0963721419896365

Schubert, A.-L., Hagemann, D., & Frischkorn, G. T. (2017). Is general intelligence little more than the speed of higher-order processing? *Journal of Experimental Psychology: General*, *146*(10), 1498–1512. https://doi.org/10.1037/xge0000325

Schubert, A.-L., Hagemann, D., Löffler, C., & Frischkorn, G. T. (2020). Disentangling the Effects of Processing Speed on the Association between Age Differences and Fluid Intelligence. *Journal of Intelligence*, *8*(1), 1. https://doi.org/10.3390/jintelligence8010001

Sharpe, D. (2013). Why the resistance to statistical innovations? Bridging the communication gap. *Psychological Methods*, *18*(4), 572–582. https://doi.org/10.1037/a0034177

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, *44*(3), 535–551. https://doi.org/10.1016/j.paid.2007.09.015

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 101–117. https://doi.org/10.1037/0278-7393.32.1.101

Spaniol, J., Voss, A., Bowen, H. J., & Grady, C. L. (2011). Motivational incentives modulate age differences in visual perception. *Psychology and Aging*, *26*(4), 932–939. https://doi.org/10.1037/a0023297

Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, *144*(12), 1325–1346. https://doi.org/10.1037/bul0000169

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence*. Cambridge University Press.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A Diffusion Model Analysis of the Effects of Aging on Letter Discrimination. *Psychology and Aging*, *18*(3), 415–429. https://doi.org/10.1037/0882-7974.18.3.415

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037/0033-295X.108.3.550

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, *122*(3), 231–249. https://doi.org/10.1037/0033-2909.122.3.231

Voskuilen, C., Ratcliff, R., & McKoon, G. (2018). Aging and confidence judgments in item recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 1–23. https://doi.org/10.1037/xlm0000425

Voss, A., Lerche, V., Mertens, U., & Voss, J. (2019). Sequential sampling models with variable boundaries and non-normal noise: A comparison of six models. *Psychonomic Bulletin & Review*, *26*(3), 813–832. https://doi.org/10.3758/s13423-018-1560-4

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology*, *60*(6), 385–402. https://doi.org/10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*(7), 1206–1220. https://doi.org/10.3758/BF03196893

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*. https://doi.org/10.3389/fninf.2013.00014

Wieschen, E. M., Voss, A., & Radev, S. (2020). Jumping to Conclusion? A Lévy Flight Model of Decision Making. *The Quantitative Methods for Psychology*, *16*(2), 120–132. https://doi.org/10.20982/tqmp.16.2.p120

Xu, K., Nosek, B., & Greenwald, A. (2014). Psychology data from the Race Implicit Association Test on the Project Implicit Demo website. *Journal of Open Psychology Data*, *2*(1), e3. https://doi.org/10.5334/jopd.ac

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 53–79. https://doi.org/10.1037/a0024177

Zimprich, D., & Martin, M. (2002). Can longitudinal changes in processing speed explain longitudinal age changes in fluid intelligence? *Psychology and Aging*, *17*(4), 690–695. https://doi.org/10.1037/0882-7974.17.4.690

List of Figures

Appendix A 1

Manuscript 1:

von Krause, M., Radev, S. T., Voss, A., Quintus, M., Egloff, B., & Wrzus, C. (submitted). Stability and Change in Diffusion Model Parameters Over Two Years. *Journal of Intelligence*.

Stability and Change in Diffusion Model Parameters Over Two Years

Mischa von Krause[1], Stefan T. Radev[1], Andreas Voss[1], Martin Quintus[2], Boris Egloff[2], & Cornelia Wrzus[1]

[1] Heidelberg University

[2] Mainz University

Author Note

This research was supported by grants from the German Research Foundation to Cornelia Wrzus (WR 160/1-1) and to the Graduate School SMiP (GRK 2277; Statistical Modeling in Psychology).

Correspondence concerning this article should be addressed to Mischa von Krause, Hauptstr. 47-51 69117 Heidelberg. E-mail: mischa.vonkrause@psychologie.uni-heidelberg.de

Abstract

In recent years, mathematical models of decision making, such as the diffusion model, have been endorsed in individual differences research. These models can disentangle different components of the decision process, like processing speed, speed-accuracy trade-offs, and duration of non-decisional processes. The diffusion model estimates individual parameters of cognitive process components, thus allowing the study of individual differences. These parameters are often assumed to show trait-like properties, that is, within-person stability across tasks and time. However, the assumption of temporal stability has so far been insufficiently investigated. With this work, we explore stability and change in diffusion model parameters by following over 270 participants across a time period of two years. We analysed four different aspects of stability and change: rank-order stability, mean-level change, individual differences in change, and profile stability. Diffusion model parameters showed strong rank-order stability and mean-level changes in processing speed and speed-accuracy trade-offs that could be attributed to practice effects. At the same time, people differed little in these developmental patterns across time. Also, profiles of individual diffusion model parameter proved to be stable over time. We discuss implications of these findings for the use of the diffusion model in individual differences research.

Keywords: diffusion model, cognitive modeling, individual differences, stability, longitudinal study

Stability and Change in Diffusion Model Parameters Over Two Years

Recently, the use of mathematical process models of cognition has seen an upsurge in research on individual differences in cognitive abilities and intelligence (Ratcliff & Childers, 2015; Ratcliff, Thapar, & McKoon, 2011; Schmiedek, Oberauer, Wilhelm, Süss, & Wittmann, 2007; Schubert & Frischkorn, 2020; Voss, Nagler, & Lerche, 2013). It has been proposed that our understanding of intelligence and cognition can profit from such modeling approaches, which disentangle different cognitive processes and components involved in solving cognitive tasks (Frischkorn & Schubert, 2018; Schubert & Frischkorn, 2020). One crucial aspect when employing mathematical models to estimate cognitive parameters to further our understanding of individual differences is whether these parameters have trait-like properties, that is, whether they measure processes which are stable and consistent across tasks and time.

Brief introduction of the diffusion model

One of the most prominent models of cognition is the diffusion model (Ratcliff, 1978). This model is a stochastic model for the analysis of response times and accuracy rates in binary decision tasks. It utilizes the full empirical response time distributions and accuracy rates simultaneously to estimate different parameters, which map onto specific components of the decision process. One of the main advantages of the diffusion model compared to the analysis of mean response times is that it can disentangle these different components. Most notably speed-accuracy trade-offs can be distinguished, that is, the fact that people sometimes show slower response times because they are more cautious. Among others, the model provides separate estimates of speed of information processing, decision caution (i.e., speed-accuracy trade off), and the time taken for encoding and motor response processes.

Figure 1 depicts the diffusion model and its core parameters. The decision process is modeled as a stochastic sampling of noisy information. The two possible responses are associated with the two decision boundaries named a and 0 in the graph. The drift rate (v) denotes the

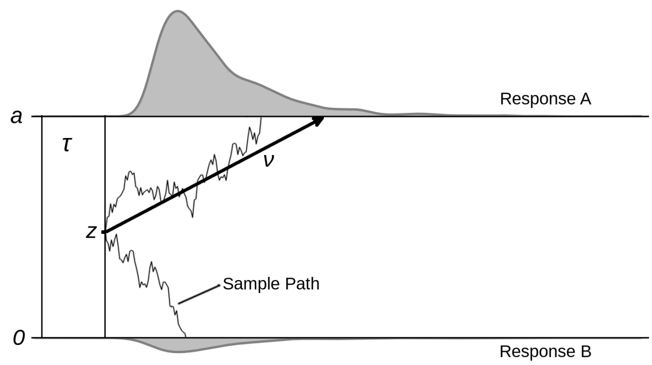

Figure 1. The diffusion model. The accumulation process starts at starting point z, moves with average slope ν, and terminates when one of the two thresholds (0 or a) has been reached. τ denotes the time taken for non-decisional processes, e.g., encoding and motoric response. On the top and the bottom of the figure, the two response time distributions are shown.

average speed of information accumulation towards one of the two boundaries. The separation between the two boundaries (a), determines how much information is sampled before a decision is taken - that is, when the noisy accumulation process reaches one of the two boundaries. Thus, a is a measure of decision conservatism or caution. The starting point, z, determines where the accumulation process starts, and maps a possible bias in the decision process in favour of one of the two responses. Finally, the non-decision time (τ) sums the duration of all non-decisional processes. On the top and the bottom of the graph are presented two example response time distributions generated by the model with a fixed parameter configuration. In addition to the parameters described above, the full diffusion model also contains parameters for the across-trial variability in drift rates, starting points, and non-decision times, that help explain certain special patterns found in empirical response time distributions, like quick or slow errors (Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998).

In the past decades, the diffusion model has been applied in various contexts, for instance, in studies on intelligence (Lerche et al., 2020; Ratcliff, Thapar, & McKoon, 2010; von Krause, Lerche, Schubert, & Voss, 2020) or aging studies (Ratcliff, 2008; Ratcliff et al., 2004a; Theisen, Lerche, von Krause, & Voss, 2020), and has found widespread use especially in the field of cognitive psychology (Ratcliff & McKoon, 2008; Voss et al., 2013). One particular question that crosses the boundaries of cognitive research towards the study of personality and individual differences is whether the diffusion model parameters constitute reliable measures of trait-like constructs that can be used to describe meaningful inter- and intra-individual differences between and within persons. A core aspect of traits as defined in the literature is their relative stability across time and measurement methods. While many studies have demonstrated that diffusion model parameters show substantial correlations across different experimental tasks (see e.g., Lerche et al., 2020; Ratcliff et al., 2010; Schubert, Frischkorn, Hagemann, & Voss, 2016), the question of temporal stability has received comparably little attention.

The first published results on the stability of diffusion model parameters were strong test-retest correlations of around $r = .70$ for all three main diffusion model parameters in a lexical decision task across a time interval of up to one week (Yap, Balota, Sibley, & Ratcliff, 2012). In another study across one week, medium to strong test-retest correlations were observed for the main diffusion model parameters (v, a, τ), with values ranging from $r > .70$ for drift rates and boundary separation and $r > .40$ for non-decision time (Lerche & Voss, 2017). Schubert et al. (2016) conducted a systematic study of the trait properties of diffusion model parameters over eight months, utilizing two different response time tasks and analysing them via latent state-trait structural equation models. The results showed stability across both tasks and time for all three main diffusion model parameters, with speed of information processing (drift rate) showing the highest stability and consistency: the latent trait factor generalizing over both time points and both tasks on average accounted for 44% of the manifest variance in drift rate. Task-specific across time correlations ranged from $r = .44$ to $r = .71$ for drift rates, from $r = .20$ to $r = .60$ for boundary separations, and from $r = .26$ to $r = .63$ for non-decision times (Schubert et al., 2016). These results suggest that some diffusion model parameters show considerable stability at least over the range of one week to eight months and might therefore in this regard be characterized as trait-like. However, findings warrant further research, because rank-order correlations across time are only one aspect of stability.

Different forms of stability and change in individual differences

While the notion of temporal stability remains a core feature of classical as well as contemporary definitions of personality traits (Allport, 1937; John, Robins, & Pervin, 2008), the idea that traits are essentially fixed at a certain point in life and remain stable thereafter, has come under more and more scrutiny in the past two decades (Wagner, Orth, Bleidorn, Hopwood, & Kandler, 2020). Thus it is now commonplace to study different forms of stability and change in personality traits to better understand their development over time.

One approach to studying stability and change that has found considerable echo in the literature was described by Roberts, Wood, and Caspi (2008). Mainly referring to the Big Five, they proposed to study four aspects of stability and change. First, rank-order stability (i.e., in most cases, test-retest correlations) refers to the stability of people's relative positions to others on the trait continuum. Second, mean-level change is the development of average (i.e., across person) levels in a certain trait over time. For example, people tend to become more agreeable and conscientious during young adulthood (Roberts, Walton, & Viechtbauer, 2006). Third, individual differences in change refer to the individual deviations in developmental patterns from the mean-level change in the sample. Finally, profile stability refers to the stability of the relative patterns of traits within a person across time: a person might stay more extraverted than she is agreeable, although both traits show changes in their absolute values. While the different forms of stability and change suggested by Roberts et al. (2008) have (to different degrees) been extensively studied for Big Five traits, the literature on diffusion model parameters has so far focused solely on rank-order stability over two time points.

In the present paper, we expand the scope of previous longitudinal studies of the diffusion model, and report findings on relative stability, mean-level change, individual differences in change, and profile stability in the main diffusion model parameters across four time points over two years.

We focus on a specific decision task that the diffusion model has repeatedly been applied to: the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; Greenwald & Farnham, 2000; Greenwald, Nosek, & Banaji, 2003; Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007). In the IAT, participants make binary decisions, typically classifying presented stimuli into one of two categories. In general, there are two different classification tasks (e.g., old vs. young, quick vs. slow) that are combined in some blocks of the experiment to form so-called congruent (e.g., old/slow) and incongruent (e.g., old/quick) combinations. The difference in mean response times between the congruent and incongruent block is then interpreted as a measure of the implicit association between the corresponding constructs (e.g., age and speed). The IAT has also been employed as a measure of implicit personality (Nosek, Greenwald, & Banaji, 2007). In this case, the classification categories are, for instance, "extraverted" vs. "introverted" on the one hand, and "me" vs. "other" on the other hand. The difference in response times between the blocks combining "me" and "extraverted" versus those combining "me" and "introverted" is then interpreted as a measure of implicit extraversion (Back, Schmukle, & Egloff, 2009).

When applying the diffusion model to the IAT, differences in performance can be decomposed into differences in speed of information processing (v), differences in decision caution (a), and differences in non-decision time (τ). Previous studies have shown that the IAT effect can mostly be attributed to differences in v that are strongly linked to the D scores usually employed to estimate the IAT effect (Klauer et al., 2007). At the same time, there were also differences in a and τ for the congruent and incongruent blocks (Klauer et al., 2007; van Ravenzwaaij, van der Maas, & Wagenmakers, 2011). Thus, the IAT could be an interesting example to study the stability and change in diffusion model parameters, as it can easily be analyzed with the diffusion model and such analyses improve the understanding of the underlying processes when working on the task. The focus of this paper is, however, not on the task-specific aspects and interpretation of the IAT, but on the longitudinal analysis of diffusion model parameter estimates as cognitive process parameters involved in the IAT. Namely, in our analyses we set aside the effects of the conditions (though we do include them in our model), and study the across-task and across-block estimates of the parameters. In this way, we account for the specific effects of each IAT condition and task, while keeping the results focused on the overall cognitive processes, and the number of analyses circumscribed.

The present study

In this paper, we analyze the stability of the diffusion model's measures for speed of information processing (drift rate), decision caution (boundary separation), and non-decision time using data from an implicit personality IAT across four time points over a period of two years. To our knowledge, this is the first study to assess the development of diffusion model parameters over more than two time points and over such an extended time period. We conducted analyses addressing the four forms of stability and change: rank-order stability, absolute mean-level change, individual differences in change, and profile stability, all with respect to drift rate (ν), boundary separation (a) and non-decision time (τ), to receive a comprehensive picture of stability and change in the cognitive parameters derived from the diffusion model.

Methods

Participants

The data used in this paper were collected in a large-scale longitudinal study that focused on temporal aspects of personality. This study included a wide range of measures of explicit and implicit personality traits, personality states, and cognitive abilities. Several papers drawing on these data have already been published (Lücke, Quintus, Egloff, & Wrzus, 2020; Quintus, Egloff, & Wrzus, 2017, 2020). These studies emphasized different aspects of personality processes and personality development. However, none of these papers focused on cognitive parameters or used the diffusion model in any of the analyses. The initial sample at the first time point (T1) comprised 382 participants (73% women, all with a similar educational background, the German Abitur). Of these, 255 were young adults ($M_{age} = 21.57$, $SD_{age} = 2.20$) and 127 were older adults ($M_{age} = 67.76$, $SD_{age} = 5.31$). The sample size was based on power analyses independent of the analyses reported in this paper. After six months (T2), 358 people from the original sample took part in the second time point. Both at T3 (one year after T1) and at T4 (two years after T1), 327 people participated. The sample consisted of five different subgroups: young people in their first year at university (Group 1, $n = 113$ at T1), young people in their second year at university (Group 2, $n = 109$), young non-students (Group 3, $n = 26$), older first-year students (Group 4, $n = 63$), and older non-students (Group 5, $n = 58$).

Procedure and Material

Laboratory data were collected in small age-homogeneous group sessions on a PC in a university setting. All participants provided informed consent. As was already mentioned, the study included a wide range of measures, most of which focused on personality traits and states. An overview of the instruments employed is available at https://osf.io/k9wsv/. In the following, we describe the Implicit Association Tests of the Big Five personality traits.

The Big Five IATs (Schmukle, Back, & Egloff, 2008) included five blocks of word classification tasks, with 20 trials in all training blocks and 60 trials in both the congruent and the incongruent test blocks, as is standard practice in IATs (Greenwald et al., 1998, 2003). Since we disregarded the practice trials in our analyses, this led to a total trial number of 600 per participant and time point (60 * 2 [congruent/incongruent] * 5 [Big Five traits]). For all Big Five traits, the same target categories (i.e., “me” and “others”) were used with a set of five different stimuli each (e.g., “I”, “they”). Attribute category labels were dependent on the specific Big Five traits (e.g., “conscientiousness” vs. “carelessness”) and also included five different stimuli for each of the traits (e.g., “helpful” for agreeableness or “reliable” for conscientiousness). In all blocks, stimuli were always presented in random order and then shuffled before the next presentation. In the test blocks, we alternated target and attribute stimuli. One specialty of the IAT data was the way error response times were recorded. The stimuli remained on screen until the correct response was given. In case of an error, only the response time of the later correct response was recorded. This coding is typical for IAT analyses but presents a particular challenge for diffusion model analysis. This is important for the modeling approach we used, since we tried to account for the differences in processes involved in creating the correct and error response times.

Data analysis

We used the programming language R (Version 4.0.3; R Core Team, 2020) and the R-packages *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2018), *blavaan* (Version 0.3.12; Merkle & Rosseel, 2018), *correlation* (Version 0.5.0; Makowski, Ben-Shachar, Patil, & Lüdecke, 2020), *papaja* (Version 0.1.0.9997; Aust & Barth, 2018), and *tidyverse* (Version 1.3.0; Wickham et al., 2019) for all statistical analyses. For all Bayesian analyses, the prior distributions used are available in the Appendix (A1). For the diffusion model parameters, we chose the default priors provided by the Python package HDDM (Wiecki, Sofer, & Frank, 2013), which are based on the recommendations by Matzke and Wagenmakers (2009).

Estimation of the diffusion model parameters. We used the hierarchical Bayesian method provided in HDDM (Wiecki et al., 2013) to estimate the diffusion model parameters. Prior to fitting the models, we removed trials that had not been recorded for technical reasons and also trials with latency below 300 ms or above 3000 ms, as these could be expected to qualitatively differ from the other trials regarding the processes involved in producing the answers. Separately for each time point, we also excluded all data from participants with low accuracy (across all five Big Five IATs). Low accuracy was defined as an accuracy rate lower than three interquartile ranges from the first quartile of accuracy rates across participants per time point (Tukey, 1977). Taken together, these pre-processing steps lead to the exclusion of 2.91% of the total number of trials. Finally, we excluded one warm-up trial per block per participant.

We fitted the same model separately for each time point. Using the Marcov chain Monte Carlo method implemented in HDDM, we obtained four chains with 6000 samples each from the posterior distribution per model. We discarded the first 1000 samples of each chain as a burn-in period. For all diffusion model parameters, we obtained posterior distributions both at group-level and at the person-level. We choose a parsimonious modelling approach, including only the core diffusion model parameters: drift rate, boundary separation, and non-decision time. The estimates of between-trial variability of the parameters are often unreliable and estimating them can actually have detrimental effects on the reliability of the main parameter estimates (Lerche & Voss, 2016). Thus, we fixed these parameters to zero, as they were also of no theoretical interest for our analyses. We also fixed the starting point to 0.5, as the decision boundaries were associated with correct and error responses and thus no implicit bias towards one of the alternatives could be expected.

To model the different experimental conditions (i.e., the five different Big Five traits, both in the congruent and the incongruent block), we used effect coding to estimate an intercept and effects per condition for both boundary separation and drift rates. Further, different non-decision times were estimated for correct and error responses. This was necessary, as the latency for the initial (erroneous) response was not recorded, but only the later, corrected response time. In our model, the time to correct the response is included in the error non-decision time. Figure 2 depicts our model formulation.

To ensure convergence of the Markov chains to the target posterior, we used several steps to inspect the group-level and individual parameters of drift rates, boundary separations and correct response non-decision times used in the further analyses (Kruschke, 2015). First, we visually inspected each chain via caterpillar plots. Second, we checked the $\hat{R}$ statistics and excluded estimates with a $\hat{R}$ value larger than 1.01 (Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2020). Third, we computed the effective sample sizes and excluded estimates with fewer than 400 effective samples (i.e., 100 per chain). To obtain full sets of the main diffusion model parameters for each participant at each time point, we excluded the individual parameter estimates of all of a, v and τ if signs of non-convergence were evident for any of these three parameters in a person (at a certain time point). Taken together, all preprocessing steps led to the exclusion of 7.44% of the

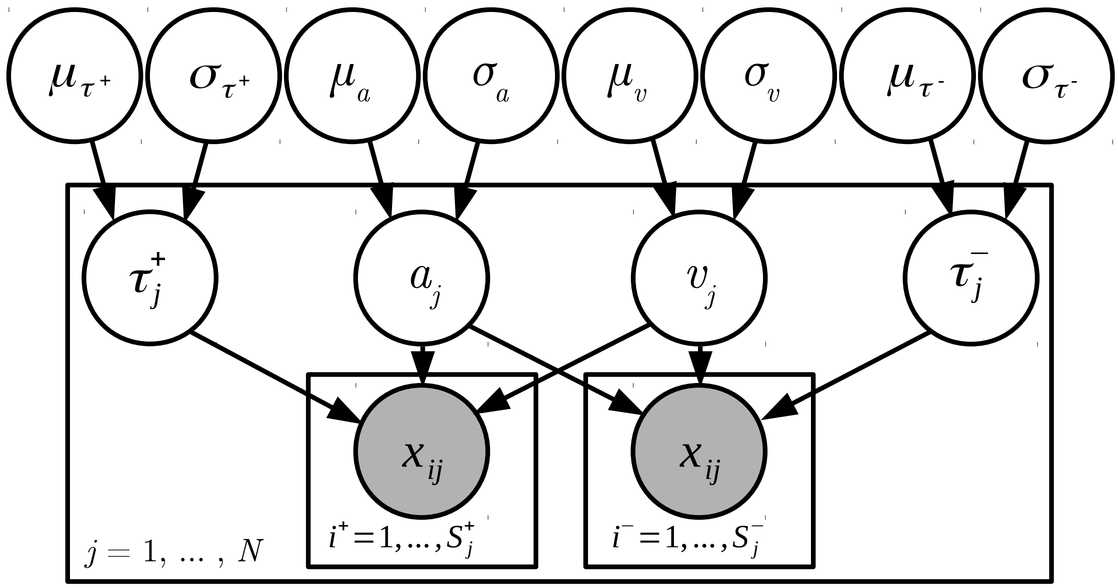

Figure 2. The hierarchical Bayesian model used for estimation of the diffusion model parameters. The inner plates relate to the trial level, the outer plate to the person level. On the outside are the group-level parameters. v = drift date, a = boundary separation, $\tau^{+/-}$ = non-decision time for correct and error responses, N = number of participants at a certain time point, $S^{+/-}$ = number of correct/error trials per person. x_{ij} denotes a single trial. The model does not show the effects on drift rate and boundary separation estimated on the group-level and person-level for the different experimental conditions and traits.

total individual parameter vectors. The corresponding statistics and plots can be found in the supplementary material.

To account for possible drop-out effects also due to non-converged chains only at later time points, we conducted Bayesian *t*-tests addressing whether the persons who had missing values at at least one of the later time points differed from the rest of the sample in any of the three diffusion model parameters. People with missing values had higher drift rates ($BF = 5.86$), higher boundary separation ($BF = 3.24$), and higher non-decision times ($BF = 195.03$). To account for this fact, we repeated all our analyses including the non-converged chains. No differences in the pattern of results emerged, notably also not for the pattern of mean-level changes across time. Also, when not excluding the non-converged chains, there were no more differences in means of diffusion model parameter for people dropping out (all $BFs < 1$).

To further assess model fit (generative performance), we conducted posterior predictive checks. For each time point, we randomly selected 500 samples from the joint posterior distribution of parameters and used each of these to generate person-specific simulated response times and response choices. As in the empirical data, 600 trials existed for each person at each time point (unless outlier trials had been removed as described above), we also obtained 600 trials per person for each of the 500 samples from the posterior distribution of diffusion model parameters (i.e., 60 for each of the trait/condition combinations with their specific effects). We then computed RT quartiles and error rates for each person and time point from both the empirical and simulated data. Figure 3 shows the resulting scatter plot for T1, the remaining plots can be found the the Appendix. As can be seen, the patterns found in the observed data are closely related to those found in the simulated data. Thus, the model fits the data quite well.

Following model evaluation, we extracted, for each time point, each person's individual posterior medians for the three diffusion model parameters. We used the intercept parameter estimates irrespective of condition and trait for a and v, and the non-decision times of correct responses. We did not further analyse error non-decision times because estimates were based on a low number of trials. We then utilized these posterior medians as summaries of the full posteriors in most of the further analyses. While it is true that such a two-step procedure makes no use of uncertainty estimates provided by Bayesian sampling procedures, it must be noted that our models already contained several thousands of parameters to be estimated for each time point and were thus very complex to estimate and converge.

Statistical Analyses of Stability and Change. To test the **rank-order stability** of the diffusion model parameters, we obtained Bayesian correlation estimates (between individual posterior medians). Hypothesis testing was performed with Bayes factors (instead of p values) using the *R* packages *correlation* and *BayesFactor*. As the sample contained different sub-groups

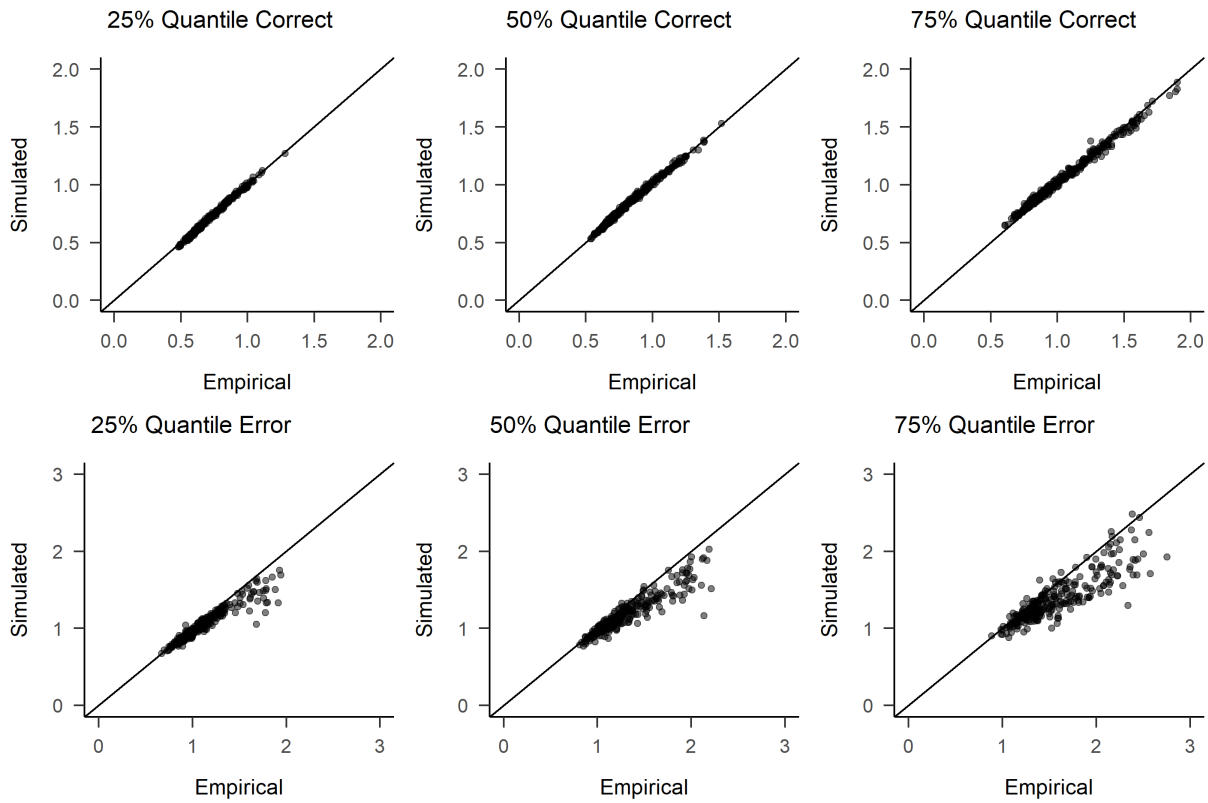

Figure 3. Posterior predictive check of RTs for T1. Error quantiles are based on far less data, with the median accuracy rate being 96 percent. Participants with 12 or less errors omitted from the error response time plots. See Appendix for posterior predictive checks for the other time points.

of participants (old/young, student/non-student, see above), we conducted separate analyses for each of the sub-groups to study whether the overall rank-order stability between participants might be due to the stability of differences between sub-groups. To analyse **mean-level change**, we compared the full posterior distributions of the group-level parameter estimates (i.e., across participants) across time points.

To study possible **individual differences** in stability and change in diffusion model parameters, we then estimated Bayesian growth curve models using the *blavaan* package (Merkle & Rosseel, 2018), separately for each parameter (v, a, and τ). The individual posterior medians at each time point served as observed variables in the model. We fixed all (unstandardized) loadings on the intercept factor to 1. For the slope factor (which reflects growth or change over time), we fixed the loading to 0 for T1 and to 1 for T2. We freely estimated the factor loadings for T3 and T4, as we did not have any hypotheses on the nature of change. Figure 4 shows a graphical representation of our growth curve models. For each of the models, we used three MCMC chains and obtained 10000 samples, discarding the first 5000 samples as burn-in (Merkle & Rosseel, 2018). To check the fit of the Bayesian growth curve models, we inspected the *bCFI* and *bGammaHat* metrics as advised by Garnier-Villarreal and Jorgensen (2020).

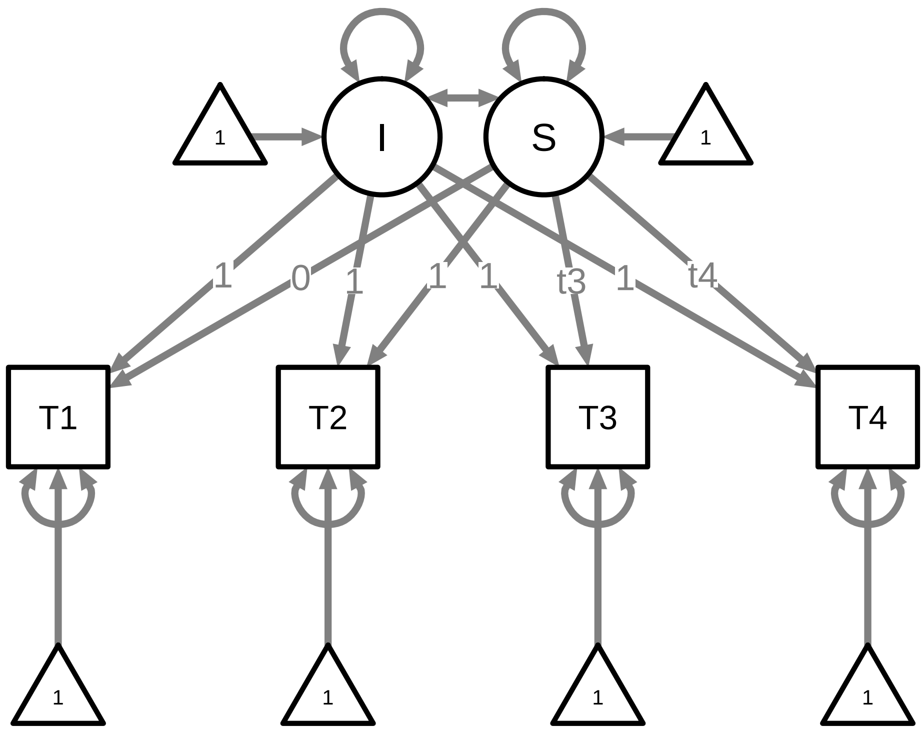

Figure 4. Growth curve model used for all three diffusion model parameters. T1 to T4 refer to the individual posterior medians of the respective diffusion model parameter at a certain time point. *I* = Intercept, *S* = Slope. The slope loadings *t3* and *t4* are treated as free parameters and thus estimated.

Finally, we calculated *q* correlations of individual posterior medians to study **profile stability** (Burt, 1937). In the *q* correlation framework, variables (i.e., ν, a, and τ) serve as cases which vary in relative strength and time points constitute the columns in separate datasets for each participant. In this way, it is possible to calculate the stability of the relative strength of the values (i.e., v, a, and τ), compared to one another. To this end, we first z-standardized the individual posterior medians, separately for each parameter, to make their relative strength comparable. We then calculated (frequentist) q correlations via the *multicon* package, separately for each participant, and created descriptive statistics and plots of correlations across participants. In order to reflect the exploratory nature of these calculations, we do not conduct inferential analyses of q correlations, but purely report the descriptive results.

Results

All data and analysis scripts can be found on the paper's OSF page (https://osf.io/cnr2a/). We report results on the rank-order stability, mean-level change and individual differences in change for each of the three main diffusion model model parameters (v, a, τ). For all these analyses, we used Bayesian methods to obtain our results. We also conducted all analyses using a frequentist, p-value based approach. This did not alter the interpretation of our findings. Finally, we report findings on the profile stability of the three parameters across time.

Table 1 shows the descriptive statistics of the individual posterior medians for the three diffusion model parameters for each of the four time points across the entire sample. Tables A2 to A6 in the Appendix contain the corresponding information, split up for each of the five sub-groups.

Rank-order stability

Table 2 shows the rank-order stability estimates of the diffusion model parameters for the entire sample. We report Bayesian correlation estimates, using a uniform prior for the correlation (see A1) and individual posterior medians as variables. Rank-order stability was high for drift rates (v; all $rs >= .64$) across the entire time span, with correlations getting slightly smaller for

Table 1

Summary statistics of the individual posterior medians of diffusion model parameters for each timepoint across all groups

Parameter	Symbol (Time Point)	*N*	*M*	*SD*	Minimum	Maximum
Drift Rate	ν (T1)	359	2.09	0.42	0.82	3.28
	ν (T2)	334	2.21	0.46	0.94	4.07
	ν (T3)	293	2.21	0.50	0.94	3.82
	ν (T4)	282	2.21	0.50	0.98	3.65
Boundary Separation	a (T1)	359	2.04	0.55	1.21	4.79
	a (T2)	334	1.89	0.51	1.03	3.98
	a (T3)	293	1.87	0.54	0.99	4.04
	a (T4)	282	1.85	0.56	0.97	4.39
Non-Decision Time	τ (T1)	359	0.43	0.08	0.29	0.72
	τ (T2)	334	0.42	0.08	0.28	0.78
	τ (T3)	293	0.44	0.09	0.25	0.72
	τ (T4)	282	0.43	0.09	0.27	0.75

Note. *M* = Mean. *SD* = Standard Deviation. Time 2 = Time 1 + 6 months. Time 3 = Time 1 + 12 months. Time 4 = Time 1 + 24 months.

Table 2

Correlation matrices of diffusion model parameters across four time points across all participants.

	Time 1	Time 2	Time 3
v Time 2	0.79 [0.76 - 0.82]		
v Time 3	0.73 [0.69 - 0.78]	0.78 [0.75 - 0.82]	
v Time 4	0.64 [0.59 - 0.70]	0.71 [0.66 - 0.76]	0.71 [0.65 - 0.76]
a Time 2	0.85 [0.82 - 0.87]		
a Time 3	0.83 [0.80 - 0.86]	0.90 [0.88 - 0.91]	
a Time 4	0.84 [0.82 - 0.87]	0.88 [0.86 - 0.91]	0.85 [0.82 - 0.88]
τ Time 2	0.88 [0.86 - 0.90]		
τ Time 3	0.87 [0.84 - 0.89]	0.90 [0.88 - 0.92]	
τ Time 4	0.80 [0.76 - 0.83]	0.86 [0.83 - 0.88]	0.84 [0.81 - 0.87]

Note. Means of Bayesian correlation estimates and 95 % credible interval reported. All Bayes factors > 999. Time 2 = Time 1 + 6 months. Time 3 = Time 1 + 12 months. Time 4 = Time 1 + 24 months.

larger time periods (e.g., $r = .79$ from T1 to T2, but only $r = .64$ from T1 to T4). We found the same pattern for boundary separation (a): Rank-order stability was high (all $rs >= .83$), with correlations getting slightly smaller across larger time periods (e.g., $r = .90$ from T2 to T3, but only $r = .83$ from T1 to T3). For non-decision times (τ), stability was again high (all $rs >= .80$) across the entire time span, with correlations once more getting smaller for larger time periods (e.g., $r = .90$ from T2 to T3, but only $r = .80$ from T1 to T4). All correlations showed Bayes factors > 999 when compared to a null-model.

Tables A7 to A9 show the estimates of rank-order stability separately for the three diffusion model parameters and split up across the five sub-groups studied. Generally, the interpretation of the pattern of results did not differ across groups, although within-group correlations often were slightly smaller than correlations for the total sample. Especially due to the smaller samples sizes, Bayes factor were also sometimes lower, for example, as low as $BF = 3.07$ for the correlation of drift rates at T2 to the ones at T4 in Group 3 ($n = 19$, $r = .46$).

Mean level change and individual differences in change

Figure 5 shows the group-level posterior distributions (i.e., across participants) for the three diffusion model parameters across the four time points. As can be seen, drift rates seem to rise after T1 (with the corresponding 95% Highest Density Interval (HDIs) showing no overlap with those of the other time points) and to a lesser degree also after T2 and T3. The pattern reverses for the boundary separation parameter, with a decline from T1 to the later time points. For non-decision times, no clear pattern of mean level change is evident. It should be noted that the group-level posterior distributions are not equivalent as the means of individual parameter posterior medians, due to the hierarchical modeling approach and due to the exclusion of individual parameter estimates with non-converged traces. However, the general pattern of results was the same for both group-level posteriors and means of individual posterior medians.

Table 3 shows the parameter estimates and fit indices for the Bayesian growth curve model of drift rates. The latent intercept and latent slope exhibited only a very weak estimated correlation, indicating that drift rates at T1 did not relate to the developmental patterns of drift rates. As the 95% CI of the covariance between intercept and slope included zero, we fixed this parameter to zero to help model convergence. All estimated parameters had effective sample sizes > 5000 and $\hat{R}$ values below 1.01, indicating that the chains had converged. Furthermore, model fit was good according to the mean Bayesian GammaHat estimate > 0.99 and the mean Bayesian CFI estimate > 0.99.

Latent slope loadings at t3 and t4 were estimated as 1.142 and 1.297. Both the mean level (intercept) of the latent intercept parameter and of the latent slope parameter were estimated as

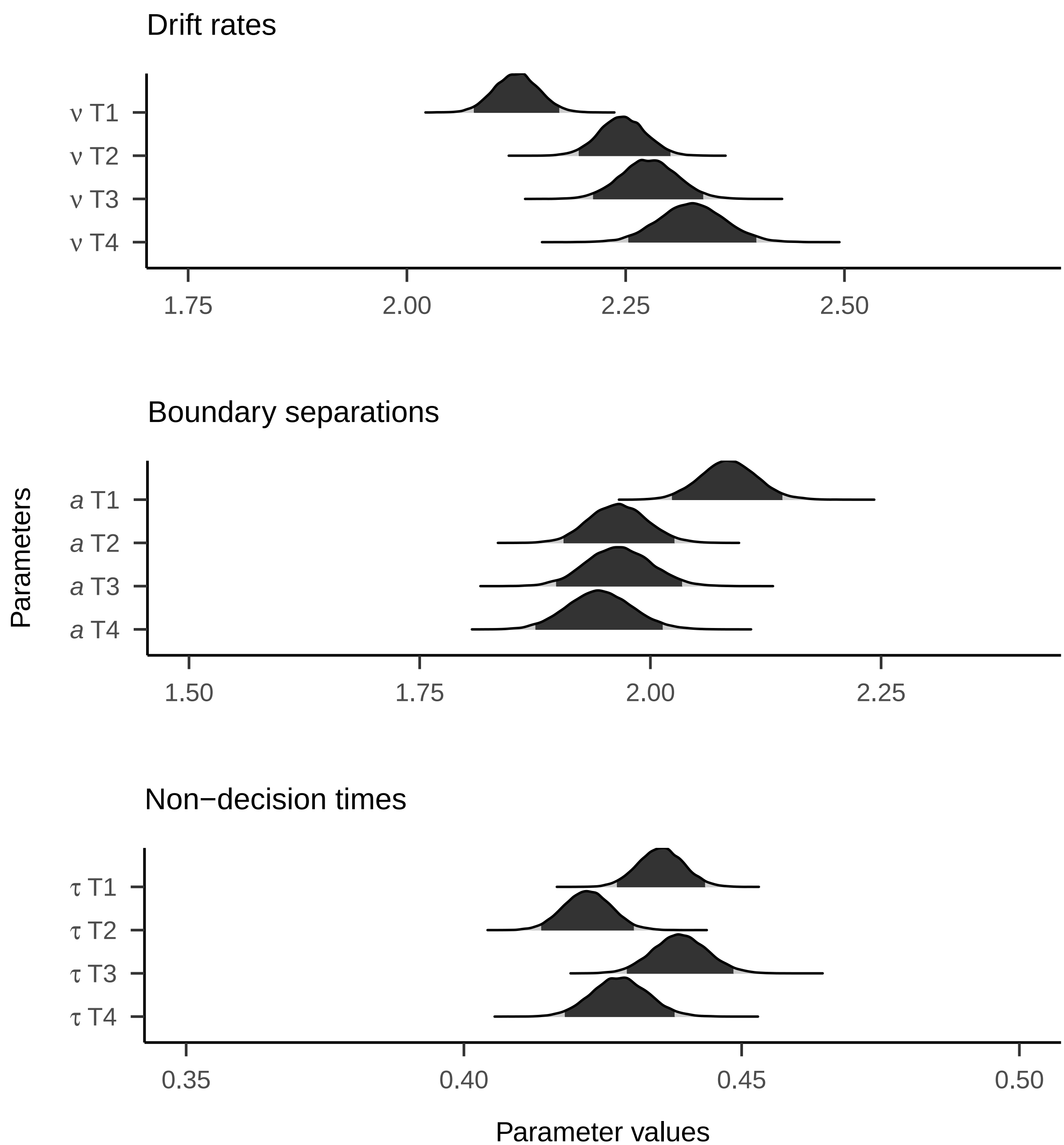

Figure 5. Group-level posterior plots of diffusion model parameters across time. 95% highest density intervals shown. T2 = T1 + 6 months. T3 = T1 + 12 months. T4 = T1 + 24 months.

positive and their 95% credibility intervals (CIs) did not include zero. This indicates that drift rates were generally positive at T1 (as would be expected) and tended to increase over time. The latent intercept showed considerable variance, indicating that people differed in their speed of information accumulation at T1. The latent slope parameter also indicated variance, meaning that people differed in their developmental patterns of drift rates across time - the 95% CI did not include zero.

Table 4 shows the parameter estimates and fit indices for the Bayesian growth curve model of boundary separations. The latent intercept and and latent slope exhibited only a very weak estimated correlation, indicating that boundary separation at T1 did not relate to the developmental patterns of boundary separation. As the 95% CI of the covariance between intercept and slope included zero, we fixed this parameter to zero to help model convergence. As the variance of the slope factor was also estimated to be zero and the model showed divergent transitions when estimating it, we also fixed this parameter. All estimated parameters had effective sample sizes > 5000 and $\hat{R}$ values below 1.01, indicating that the chains had converged. Model fit was good, with the mean Bayesian GammaHat estimate > 0.99 and the mean Bayesian CFI estimate > 0.99.

Latent slope loadings at t3 and t4 were estimated as 1.233 and 1.334. The mean level (intercept) of the latent intercept parameter was estimated as positive, while the mean level (intercept) of the latent slope parameter was estimated as negative. Both their 95% CIs did not include zero. This indicates that boundary separations were generally positive at T1 (as would be expected) and tended to decrease over time. The latent intercept showed considerable variance, indicating that people differed in their decision criteria at T1. As was already mentioned, the latent slope parameter was estimated and then fixed to be zero.

Table 5 shows the parameter estimates and fit indices for the Bayesian growth curve model of non-decision times. Latent intercept and and latent slope showed a very low estimated correlation, indicating that non-decision time at T1 did not relate to the developmental patterns of non-decision times. As the 95% CI of the covariance between intercept and slope included zero, we fixed this parameter to zero to help model convergence. As the variance of the slope factor was also estimated to be zero and the model showed divergent transitions when estimating it, we also fixed this parameter.

All estimated parameters had effective sample sizes > 5000 and $\hat{R}$ values below 1.01, indicating that the chains had converged. Model fit was good, with the mean Bayesian GammaHat estimate > 0.97 and the mean Bayesian CFI estimate > 0.98.

Latent slope loadings showed an unclear pattern, with loadings at t3 and t4 estimated as $-.358$ and .509. The mean level (intercept) of the latent intercept parameter was estimated as positive, while the mean level (intercept) of the latent slope parameter was estimated as negative. Both their 95% CIs did not include zero. This indicates that non-decision times were generally positive at T1 (as would be expected). Given the unclear pattern of loadings on the slope factor, no clear interpretation of the negative intercept of the latent slope factor emerged. The latent intercept showed considerable variance, indicating that people differed in their non-decision time at T1. As was already mentioned, the latent slope parameter was estimated and then fixed to be zero.

Profile Stability. We estimated *q* correlations of the *z*-standardized individual posterior medians for the three diffusion model parameters across all possible combinations of time points (T1 with T2/T3/T4, T2 with T3/T4, T3 with T4). Table 6 shows the means, standard deviations and medians across participants. Profile stability was generally high, with all median *q* correlations $> .85$. However, there was also considerable variance in correlations across participants (all *SD*s $> .42$), with lower mean correlations than median correlations. Figure 6 shows density plots of the individual *q* correlations for all six periods. As can be seen, a large part of the densities lies close to .95, but there are also much lower coefficients of stability and also participants showing negative *q* correlations.

Table 3

Parameter Estimates and Model Fit of the Drift Rate Growth Curve Model

	Variable	Estimate	Posterior SD	95 % CI	Std. Est.
Loadings Intercept	v (T1)	1.000		-	0.944
	v (T2)	1.000		-	0.852
	v (T3)	1.000		-	0.797
	v (T4)	1.000		-	0.751
Loadings Slope	v (T1)	0.000		-	0.000
	v (T2)	1.000		-	0.340
	v (T3)	1.142	0.143	0.875 - 1.439	0.364
	v (T4)	1.297	0.177	0.974 - 1.668	0.389
(Residual) Variances	v (T1)	0.020	0.007	0.008 - 0.033	0.110
	v (T2)	0.036	0.005	0.026 - 0.046	0.158
	v (T3)	0.060	0.007	0.046 - 0.075	0.233
	v (T4)	0.082	0.010	0.064 - 0.103	0.284
	I	0.164	0.014	0.139 - 0.191	1.000
	S	0.026	0.007	0.014 - 0.041	1.000
Covariance I & S		0.000		-	0.000
Intercepts	v (T1)	0.000		-	0.000
	v (T2)	0.000		-	0.000
	v (T3)	0.000		-	0.000
	v (T4)	0.000		-	0.000
	I	2.104	0.022	2.06 - 2.148	5.202
	S	0.112	0.015	0.081 - 0.142	0.691
bCFI = 0.998, bGammaHat = 0.997					

Note. Bayesian Parameter Estimates. Std. Est = Completely Standardized Solution. I = Latent Intercept. S = Latent Slope. CI = Credible Interval.

Table 4

Parameter Estimates and Model Fit of the Boundary Separation Growth Curve Model

	Variable	Estimate	Posterior SD	95 % CI	Std. Est.
Loadings Intercept	*a* (T1)	1.000		-	0.906
	a (T2)	1.000		-	0.966
	a (T3)	1.000		-	0.939
	a (T4)	1.000		-	0.927
Loadings Slope	*a* (T1)	0.000		-	0.000
	a (T2)	1.000		-	0.000
	a (T3)	1.233	0.127	1.008 - 1.505	0.000
	a (T4)	1.334	0.142	1.077 - 1.638	0.000
(Residual) Variances	*a* (T1)	0.060	0.006	0.049 - 0.072	0.180
	a (T2)	0.020	0.003	0.014 - 0.026	0.067
	a (T3)	0.036	0.004	0.029 - 0.046	0.118
	a (T4)	0.045	0.005	0.036 - 0.055	0.141
	I	0.274	0.021	0.235 - 0.318	1.000
	S	0.000		-	0.000
Covariance I & S		0.000		-	0.000
Intercepts	*a* (T1)	0.000		-	0.000
	a (T2)	0.000		-	0.000
	a (T3)	0.000		-	0.000
	a (T4)	0.000		-	0.000
	I	2.053	0.030	1.995 - 2.111	3.922
	S	-0.123	0.015	-0.153 - -0.093	-Inf
bCFI = 0.999, bGammaHat = 0.999					

Note. Bayesian Parameter Estimates. Std. Est = Completely Standardized Solution. I = Latent Intercept. S = Latent Slope. CI = Credible Interval.

Table 5

Parameter Estimates and Model Fit of the Non-Decision Time Growth Curve Model

	Variable	Estimate	Posterior SD	95 % CI	Std. Est.
Loadings Intercept	τ (T1)	1.000		-	0.932
	τ (T2)	1.000		-	0.967
	τ (T3)	1.000		-	0.931
	τ (T4)	1.000		-	0.894
Loadings Slope	τ (T1)	0.000		-	0.000
	τ (T2)	1.000		-	0.000
	τ (T3)	-0.358	0.354	-1.216 - 0.157	-0.000
	τ (T4)	0.509	0.291	-0.092 - 1.053	0.000
(Residual) Variances	τ (T1)	0.001	0.000	0.001 - 0.001	0.131
	τ (T2)	0.000	0.000	0 - 0.001	0.066
	τ (T3)	0.001	0.000	0.001 - 0.001	0.133
	τ (T4)	0.002	0.000	0.001 - 0.002	0.201
	I	0.006	0.000	0.005 - 0.007	1.000
	S	0.000		-	0.000
Covariance I & S		0.000		-	0.000
Intercepts	τ (T1)	0.000		-	0.000
	τ (T2)	0.000		-	0.000
	τ (T3)	0.000		-	0.000
	τ (T4)	0.000		-	0.000
	I	0.436	0.004	0.428 - 0.445	5.526
	S	-0.010	0.002	-0.014 - -0.006	-Inf
bCFI = 0.984, bGammaHat = 0.971					

Note. Bayesian Parameter Estimates. Std. Est = Completely Standardized Solution. I = Latent Intercept. S = Latent Slope. CI = Credible Interval.

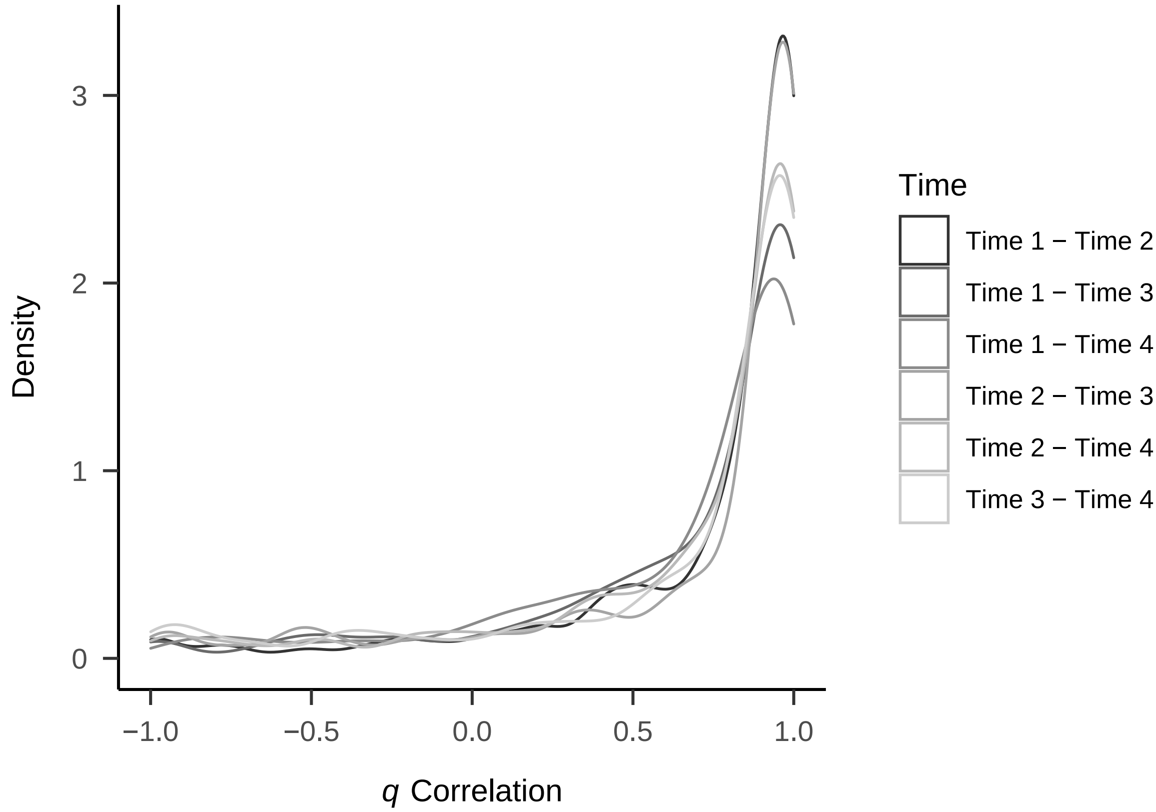

Figure 6. Density plots of q correlations.

Table 6

Descriptives of q correlations of main diffusion model parameters across time

Time	Mean	*SD*	Median	*N*
Time 1 - Time 2	0.73	0.43	0.91	318
Time 1 - Time 3	0.68	0.46	0.89	286
Time 1 - Time 4	0.65	0.47	0.86	275
Time 2 - Time 3	0.70	0.50	0.93	277
Time 2 - Time 4	0.68	0.48	0.91	268
Time 3 - Time 4	0.66	0.53	0.91	249

Discussion

In this article, we studied stability and change of cognitive processes as measured by the three main diffusion model parameters - processing speed (i.e., drift rates), decision caution (i.e., boundary separations), and speed of encoding and motor response (i.e., non-decision times) - using four different indices of stability and development. To our knowledge, this is the first study to analyse diffusion model parameters i) over such a long time period, ii) across more than two time points, and iii) in such a large, heterogeneous sample ($n = 353$ at Time 1). Moreover, our main statistical analyses relied on modern Bayesian estimation methods which offer multiple advantages compared to traditional methods. Overall, our analyses aimed to investigate whether the cognitive constructs encoded by diffusion model parameters exhibit a measurable trait-like nature. In the following, we briefly summarize the gist of our results.

Regarding rank-order stability, we found robust temporal stability of the main diffusion model parameters. Generally speaking, temporal correlations were high for all three parameters. This held true even when the entire period of the study (i.e., two years) was considered. The correlations we found were in many cases markedly higher than those previously reported in the literature (Lerche & Voss, 2017; Schubert et al., 2016; Yap et al., 2012). Especially for non-decision times, previous studies had sometimes found rank-order stability to be low ($r < .50$ across one week in Lerche and Voss (2017)). In contrast, our results indicate that non-decision times show even higher correlations across long time periods ($rs > .80$) than drift rates. This finding is worth discussing, since drift rates have so far been considered as the most "trait-like" parameters of the diffusion model (Schubert et al., 2016).

The latter difference might be attributable to several features of our study. First, in contrast to previous studies, we employed Bayesian hierarchical diffusion model estimation methods that in the past have been found to provide more robust results in correlational studies (Ratcliff & Childers, 2015; Wiecki et al., 2013). Bayesian methods incorporate prior knowledge on probable paramater values. Hierarchical Bayesian methods make use of shrinkage of the individual parameter estimates towards the group-level posteriors, balancing out extreme individual parameter estimates that might reflect noise in the data (Kruschke, 2015).

Second, we used a comparatively large number of response times for each participant at each time point (600 trials), which necessarily leads to more precise estimates. Finally, our sample included a large number of participants and exhibited a greater heterogeneity, especially in relation to age. The variance of parameter estimates might account for the higher correlations. However, it must be noted that correlations remained strong - though sometimes notably lower - even within sub-groups as small as around twenty participants (see Appendix). Thus, the present results cannot be attributed solely to sample size and sample heterogeneity. In the end, our estimates of (correct) non-decision times might be more reliable than the ones reported in previous studies, while boundary separation values might have already been estimated very reliably there. Conversely, drift rates might not show greater stability than in previous studies because of the specific content of the task: differences in drift rates also reflect differences in implicit personality, as their developmental patterns were the original focus of the study.

Regarding mean-level stability and change, we found evidence for systematic changes in both drift rates and boundary separations. Group-level drift rates increased from the first time point to the second time point six months later. The pattern of increase continued throughout the next two time points, but the posterior distributions showed much overlap there. The increase in drift rates might be interpreted as a practice effect. People tended to process the information needed to solve the IAT tasks more efficiently after they had completed the first time point. Conversely, group-level boundary separations decreased from the first to the second time point and to a lesser degree (once more marked by overlap in the posteriors) thereafter. That is, people tended to apply more liberal decision criteria and gathered less information until they made their decisions in the second to fourth time points. We suppose that participants reduce their decision caution at later time points mainly in response to the increased drift rate: that is, participants notice that they may lower their response criteria without deteriorating accuracy. Additionally, a decrease in accuracy motivation over time might also contribute to the reduction of decision caution.

In the literature on the diffusion model, practice effects in the form of increasing drift rates and decreasing boundary separations (but sometimes also non-decision times and shifting starting points) have repeatedly been reported (Dutilh, Krypotos, & Wagenmakers, 2011; Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009; Evans & Brown, 2017; Lerche & Voss, 2017; Petrov, Van Horn, & Ratcliff, 2011). However, none of these previous studies focused on training effects across such long time periods as in our study, but investigated primarily within-session training effects. It is interesting to note that training effects seem to be stable over months. Evans and Brown (2017) found that people often first adopt non-optimal decision criteria when working on a new task, that is, they are overly cautious and try to avoid mistakes, as is mirrored in high boundary separation in the diffusion model. Having practiced the task many times, people then adapt more lenient decision criteria that are closer to the optimum. Thus, a possible interpretation of our results states that people tend to keep the more lenient decision criterion when returning to the task months or even a year later.

Finally, we did not find systematic changes in non-decision times. Group-level posterior distributions remained roughly the same across the two year time period studied. This is in contrast to the results found in earlier studies on training effects that sometimes found decreasing non-decision times (Dutilh et al., 2011, 2009). Task-specific aspects of the IAT might be responsible for our findings. For instance, Dutilh et al. (2011) found that the effects on non-decision times were partly task-specific as well as item-specific.

Regarding inter-individual differences in intra-individual change, our growth curve models indicate that inter-individual differences are mainly based on across-time intercepts: We found substantial variance in the latent intercepts of all three diffusion model parameters. For boundary separation and non-decision times, people varied in their intercepts (which contribute equally to all time points) but not in their slope parameters, which reflect the rate of change across time. The slope parameter for boundary separation showed a negative trend; this means that the decrease in boundary separation, that is, the use of more liberal decision criteria, is close to universal in our data. As the estimated slope factor loadings in the non-decision time model mirror the unclear and mostly stable group-level trends found for this parameter, the slope factor is hard to interpret. In any case, its variance was estimated to be zero. The slope factor in the drift rate growth curve model was the only slope factor to show substantial inter-individual differences. Thus, people seem to differ in the ways they profit from training effects in terms of task-related information processing. In post-hoc analyses, we regressed the slope factor on age and found a clear and strong positive correlation. This means that older people tended to increase their drift rates more than their younger counterparts. As older adults did not show lower mean level drift rates (Ratcliff et al., 2004b; Schubert, Hagemann, Loeffler, & Frischkorn, 2019; von Krause et al., 2020) this implies that they generally profited more from practice. Of course, these post-hoc analyses must be interpreted cautiously and warrant further developmental research. To sum up, people tended to show great inter-individual differences in their overall levels of drift rates, boundary separations, and non-decisions time, but differed little in their developmental patterns, with the exception of drift rates. It would be interesting to follow up on these results in a longitudinal study with a stronger focus on training effects, as these were only of periphery interest here.

Regarding profile stability, the estimated *q* correlations were strongly positive across time in the majority of cases, but not in all. We also found a considerable across-participant variance in correlations, with some people showing *q* values close to zero or even negative. Correlations tended to get lower across larger periods of time. The profiles comprising the relative strengths of drift rate, boundary separation and non-decision might be seen a configuration of process components that together lead to certain empirical response time distributions and accuracy rates. For example, the same accuracy data could be the results of high drift rates and low boundary separation, and vice-versa. In a similar way, some people might show low boundary separation in combination with high drift rates, others in combination with low drift rates. It seems that, for most participants in the study, this parameter configuration remained very much the same across time.

All in all, we found that the three main diffusion model parameters are broadly consistent across time, thus fulfilling a central prerequisite of being identified as traits. This is particularly interesting as the diffusion model can be applied to a large range of binary decision tasks (not just from the cognitive domain). Our results reveal positive change in drift rates and negative change in boundary separation, but little individual differences in change, with the exception of drift rates. Profiles of the three parameters were also quite stable.

Limitations

While our study has a number of unique features, for instance, the distinction between the four forms of stability and change, the four time points over a period of two years, and the relatively large sample size, it also has some limitations. First, the variety of tasks was rather restricted. While we used five different IATs and combined them to obtain task-general parameter estimates, we did not use any other tasks. It is known that diffusion model parameters obtained in different tasks sometimes show only weak correlations among each another (Lerche et al., 2020; Ratcliff et al., 2010; Schubert et al., 2016). Thus, some of the results presented here might be specific to the tasks studied.

Second, it must be noted that the posterior predictive checks did not perfectly recover the error response time distributions. Several different factors might contribute to this. First of all, due to the small number of errors, the empirical quantiles are numerically unstable and thus may not be a good representation of the actual (latent) distribution. Also due to the low number of error responses per person, the group-level parameter of error non-decision times greatly influenced the estimates of individual error non-decision times (because of hierarchical shrinkage). This means that individual deviations in error non-decision times might sometimes have been underestimated. In turn, this might have led to a situation where our approach of modeling error response times with a separate non-decision time parameter was less successful among the very slow errors. Nevertheless, as the focus of this paper is on the psychometric properties and developmental patterns of diffusion model parameters, the relative misfit of this small proportion of trials is of secondary importance. Finally, there are alternative plausible ways to analyze the present data within a purely Bayesian framework. Intuitively, the most straightforward way to approach the question would have been to formulate and fit a full hierarchical model with time included as an additional level. However, despite being intuitive from a Bayesian lens, such an approach involves an enormous computational cost due to the large number of posteriors that need to be estimated simultaneously.

In fact, estimating the full hierarchical model turned out to be practically infeasible using the available computational software. Thus, our two-step approach using posterior medians as summary statistics might underestimate the epistemic uncertainty around parameter estimates. However, we deem our approach a reasonable trade-off, since it incorporates more information than frequentist approaches used in most of the diffusion model literature. Further, it also utilizes hierarchical shrinkage within each time point, thereby rendering point and uncertainty estimates more robust than a non-hierarchical approach.

Conclusions

We examined four different forms of stability and change in the three main diffusion model parameters: drift rate, boundary separation, and non-decision time. Our main aim was to study whether and in which way the assumption of temporal stability that is inherent in the interpretation of model-parameters-as-traits holds. Across a time period of up to two years, all three diffusion model parameters showed strong rank-order stability. Group-level drift rates tended to increase, whereas group-level boundary separations decreased, and group-level non-decision times exhibited no clear change. These findings could be interpreted as practice effects, which is remarkable given the long time intervals between the sessions (up to one year). People differed from one another in their base rates of all three main diffusion model parameters (intercepts in the growth curve models), but only drift rates showed inter-individual differences in change across time (slopes). Profiles of the three parameters mostly stayed stable across time, but some participants showed strong deviations from this pattern. We believe our study makes a strong case for the - with regard to temporal aspects - trait-like qualities of the three core diffusion model parameters. In the light of our results, the use of diffusion model parameters in individual differences research seems warranted and promising.

References

Allport, G. W. (1937). *Personality: A psychological interpretation*. H. Holt. Retrieved from https://books.google.de/books?id=AOV9AAAAMAAJ

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, *97*(3), 533–548. https://doi.org/10.1037/a0016229

Burt, C. (1937). Correlations between Persons. *British Journal of Psychology. General Section*, *28*(1), 59–96. https://doi.org/10.1111/j.2044-8295.1937.tb00862.x

Dutilh, G., Krypotos, A.-M., & Wagenmakers, E.-J. (2011). Task-Related Versus Stimulus-Specific Practice: A Diffusion Model Account. *Experimental Psychology*, *58*(6), 434–442. https://doi.org/10.1027/1618-3169/a000111

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*(6), 1026–1036. https://doi.org/10.3758/16.6.1026

Evans, N. J., & Brown, S. D. (2017). People adopt optimal policies in simple decision-making, after practice and guidance. *Psychonomic Bulletin & Review*, *24*(2), 597–606. https://doi.org/10.3758/s13423-016-1135-1

Frischkorn, G., & Schubert, A.-L. (2018). Cognitive Models in Intelligence Research: Advantages and Recommendations for Their Application. *Journal of Intelligence*, *6*(3), 34. https://doi.org/10.3390/jintelligence6030034

Garnier-Villarreal, M., & Jorgensen, T. D. (2020). Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihood. *Psychological Methods*, *25*(1), 46–70. https://doi.org/10.1037/met0000224

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*(6), 1022–1038. https://doi.org/10.1037/0022-3514.79.6.1022

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

John, O. P., Robins, R. W., & Pervin, L. A. (Eds.). (2008). *Handbook of personality: Theory and research* (3rd ed). New York: Guilford Press.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology*, *93*(3), 353–368. https://doi.org/10.1037/0022-3514.93.3.353

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (Edition 2). Boston: Academic Press.

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*. https://doi.org/10.1037/xge0000774

Lerche, V., & Voss, A. (2016). Model Complexity in Diffusion Modeling: Benefits of Making the Model More Parsimonious. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01324

Lerche, V., & Voss, A. (2017). Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological Research*, *81*(3), 629–652. https://doi.org/10.1007/s00426-016-0770-5

Lücke, A. J., Quintus, M., Egloff, B., & Wrzus, C. (2020). You can't always get what you want: The role of change goal importance, goal feasibility and momentary experiences for volitional personality development. *European Journal of Personality*, 0890207020962332. https://doi.org/10.1177/0890207020962332

Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2020). Methods for correlation analysis. *CRAN*. Retrieved from https://github.com/easystats/correlation

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*(5), 798–817. https://doi.org/10.3758/PBR.16.5.798

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, *85*(4), 1–30. https://doi.org/10.18637/jss.v085.i04

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at Age 7: A Methodological and Conceptual Review. In *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). New York, NY, US: Psychology Press.

Petrov, A. A., Van Horn, N. M., & Ratcliff, R. (2011). Dissociable perceptual-learning mechanisms revealed by diffusion-model analysis. *Psychonomic Bulletin & Review*, *18*(3), 490–497. https://doi.org/10.3758/s13423-011-0079-8

Quintus, M., Egloff, B., & Wrzus, C. (2017). Predictors of volitional personality change in younger and older adults: Response surface analyses signify the complementary perspectives of the self and knowledgeable others. *Journal of Research in Personality*, *70*, 214–228. https://doi.org/10.1016/j.jrp.2017.08.001

Quintus, M., Egloff, B., & Wrzus, C. (2020). Daily life processes predict long-term development in explicit and implicit representations of Big Five traits: Testing predictions from the TESSERA (Triggering situations, Expectancies, States and State Expressions, and ReActions) framework. *Journal of Personality and Social Psychology*, Advance online publication. https://doi.org/10.1037/pspp0000361

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging*, *23*(4), 900–916. https://doi.org/10.1037/a0013930

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, *2*(4), 237–279. https://doi.org/10.1037/dec0000030

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004a). A Diffusion Model Analysis of the Effects of Aging in the Lexical-Decision Task. *Psychology and Aging*, *19*(2), 278–289. https://doi.org/10.1037/0882-7974.19.2.278

Ratcliff, R., Thapar, A., & McKoon, G. (2004b). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*(4), 408–424. https://doi.org/10.1016/j.jml.2003.11.002

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*(3), 127–157.

https://doi.org/10.1016/j.cogpsych.2009.09.001

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, *140*(3), 464–487. https://doi.org/10.1037/a0023810

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roberts, B., Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. p. 375–398). The Guilford Press.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

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), 414–429. https://doi.org/10.1037/0096-3445.136.3.414

Schmukle, S. C., Back, M. D., & Egloff, B. (2008). Validity of the Five-Factor Model for the Implicit Self-Concept of Personality. *European Journal of Psychological Assessment*, *24*(4), 263–272. https://doi.org/10.1027/1015-5759.24.4.263

Schubert, A.-L., Frischkorn, G., Hagemann, D., & Voss, A. (2016). Trait Characteristics of Diffusion Model Parameters. *Journal of Intelligence*, *4*(3), 7. https://doi.org/10.3390/jintelligence4030007

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*, 096372141989636.

https://doi.org/10.1177/0963721419896365

Schubert, A.-L., Hagemann, D., Loeffler, C., & Frischkorn, G. T. (2019). Disentangling the Effects of Processing Speed on the Association between Age Differences and Fluid Intelligence. *Journal of Intelligence*, *8*(1), 1. https://doi.org/10.3390/jintelligence8010001

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2020). Age differences in diffusion model parameters: A meta-analysis. *Psychological Research*. https://doi.org/10.1007/s00426-020-01371-8

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, Mass: Addison-Wesley Pub. Co.

van Ravenzwaaij, D., van der Maas, H. L. J., & Wagenmakers, E.-J. (2011). Does the Name-Race Implicit Association Test Measure Racial Prejudice? *Experimental Psychology*, *58*(4), 271–277. https://doi.org/10.1027/1618-3169/a000093

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2020). Rank-normalization, folding, and localization: An improved RHat for assessing convergence of MCMC. *Bayesian Analysis*. https://doi.org/10.1214/20-BA1221

von Krause, M., Lerche, V., Schubert, A.-L., & Voss, A. (2020). Do Non-Decision Times Mediate the Association between Age and Intelligence across Different Content and Process Domains? *Journal of Intelligence*, *8*(3), 33. https://doi.org/10.3390/jintelligence8030033

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology*, *60*(6), 385–402. https://doi.org/10.1027/1618-3169/a000218

Wagner, J., Orth, U., Bleidorn, W., Hopwood, C. J., & Kandler, C. (2020). *Towards an Integrative Model of Sources of Personality Stability and Change* (Preprint). PsyArXiv. https://doi.org/10.31234/osf.io/qzef8

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

https://doi.org/10.21105/joss.01686

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*. https://doi.org/10.3389/fninf.2013.00014

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(1), 53–79. https://doi.org/10.1037/a0024177

Appendix

Table A1

Prior distributions used in all analyses

Parameter	Prior
Diffusion model parameters	
μ_a	$Gamma(1.5, 0.75)$
σ_a	$Half-normal(0.1)$
a_j	$Gamma(\mu_a, \sigma_a^2)$
μ_v	$Normal(2, 3)$
σ_v	$Half-normal(2)$
v_j	$Normal(\mu_v, \sigma_v^2)$
μ_τ	$Gamma(0.4, 0.2)$
σ_τ	$Half-normal(1)$
τ_j	$Normal(\mu_\tau, \sigma_\tau^2)$
Growth curve model	
Factor loading	$Normal(0, 10)$
Latent variable covariance	$LKJcorrelation(1)$
Latent Intercept	$Normal(0, 10)$
Latent SD	$Gamma(1, 0.5)$
All correlations	$Beta(1, 1)$

Note. The diffusion model parameters are HDDM standards based on the suggestions by Matzke and Wagenmakers, 2009. The index j refers to individual participants (at a certain time point).

Table A2

Summary statistics of the individual diffusion model parameter estimates for each timepoint for Group 1

Parameter	Symbol (Time Point)	*N*	*M*	*SD*	Minimum	Maximum
Drift Rate	v (T1)	112	2.12	0.45	1.14	3.28
	v (T2)	102	2.21	0.46	1.24	3.38
	v (T3)	93	2.16	0.49	1.31	3.63
	v (T4)	89	2.16	0.47	1.05	3.46
Boundary Separation	a (T1)	112	1.76	0.32	1.23	2.91
	a (T2)	102	1.64	0.28	1.03	2.48
	a (T3)	93	1.58	0.26	0.99	2.22
	a (T4)	89	1.54	0.26	0.97	2.26
Non-Decision Time	τ (T1)	112	0.39	0.04	0.29	0.48
	τ (T2)	102	0.38	0.03	0.30	0.46
	τ (T3)	93	0.39	0.04	0.29	0.49
	τ (T4)	89	0.38	0.04	0.28	0.49

Note. *M* = Mean. *SD* = Standard Deviation. Individual posterior medians used.

Table A3

Summary statistics of the individual diffusion model parameter estimates for each timepoint for Group 2

Parameter	Symbol (Time Point)	*N*	*M*	*SD*	Minimum	Maximum
Drift Rate	v (T1)	103	2.01	0.39	1.08	2.88
	v (T2)	104	2.15	0.45	1.24	3.29
	v (T3)	85	2.08	0.42	1.12	3.14
	v (T4)	82	2.11	0.46	1.20	3.36
Boundary Separation	a (T1)	103	1.78	0.34	1.21	3.60
	a (T2)	104	1.65	0.35	1.08	3.40
	a (T3)	85	1.60	0.27	1.15	2.20
	a (T4)	82	1.58	0.29	1.03	2.24
Non-Decision Time	τ (T1)	103	0.40	0.05	0.30	0.51
	τ (T2)	104	0.39	0.05	0.28	0.52
	τ (T3)	85	0.39	0.05	0.25	0.51
	τ (T4)	82	0.38	0.04	0.27	0.55

Note. *M* = Mean. *SD* = Standard Deviation. Individual posterior medians used.

Table A4

Summary statistics of the individual diffusion model parameter estimates for each timepoint for Group 3

Parameter	Symbol (Time Point)	*N*	*M*	*SD*	Minimum	Maximum
Drift Rate	v (T1)	26	1.93	0.43	1.28	3.15
	v (T2)	23	2.02	0.49	1.30	3.13
	v (T3)	18	1.93	0.58	0.99	3.50
	v (T4)	20	1.87	0.44	1.08	2.68
Boundary Separation	a (T1)	26	1.88	0.36	1.38	2.98
	a (T2)	23	1.72	0.27	1.23	2.24
	a (T3)	18	1.74	0.27	1.22	2.14
	a (T4)	20	1.72	0.36	1.27	2.62
Non-Decision Time	τ (T1)	26	0.40	0.06	0.30	0.52
	τ (T2)	23	0.39	0.05	0.30	0.48
	τ (T3)	18	0.38	0.05	0.31	0.47
	τ (T4)	20	0.40	0.06	0.30	0.53

Note. *M* = Mean. *SD* = Standard Deviation. Individual posterior medians used.

Table A5

Summary statistics of the individual diffusion model parameter estimates for each timepoint for Group 4

Parameter	Symbol (Time Point)	*N*	*M*	*SD*	Minimum	Maximum
Drift Rate	v (T1)	58	2.23	0.45	0.82	3.24
	v (T2)	55	2.30	0.47	0.94	4.07
	v (T3)	44	2.45	0.57	0.94	3.82
	v (T4)	44	2.39	0.52	0.98	3.65
Boundary Separation	a (T1)	58	2.59	0.64	1.79	4.79
	a (T2)	55	2.44	0.51	1.61	3.88
	a (T3)	44	2.42	0.54	1.56	4.04
	a (T4)	44	2.44	0.57	1.69	4.39
Non-Decision Time	τ (T1)	58	0.52	0.07	0.36	0.72
	τ (T2)	55	0.52	0.08	0.33	0.78
	τ (T3)	44	0.55	0.08	0.36	0.72
	τ (T4)	44	0.53	0.07	0.37	0.71

Note. M = Mean. *SD* = Standard Deviation. Individual posterior medians used.

Table A6

Summary statistics of the individual diffusion model parameter estimates for each timepoint for Group 5

Parameter	Symbol (Time Point)	*N*	*M*	*SD*	Minimum	Maximum
Drift Rate	ν (T1)	53	2.12	0.38	1.25	2.90
	ν (T2)	48	2.36	0.41	1.73	3.65
	ν (T3)	51	2.42	0.39	1.77	3.27
	ν (T4)	44	2.45	0.47	1.12	3.36
Boundary Separation	a (T1)	53	2.56	0.44	1.75	3.56
	a (T2)	48	2.40	0.41	1.74	3.98
	a (T3)	51	2.44	0.52	1.66	3.93
	a (T4)	44	2.46	0.50	1.71	4.07
Non-Decision Time	τ (T1)	53	0.51	0.07	0.38	0.66
	τ (T2)	48	0.50	0.07	0.36	0.62
	τ (T3)	51	0.52	0.07	0.37	0.64
	τ (T4)	44	0.52	0.09	0.28	0.75

Note. *M* = Mean. *SD* = Standard Deviation. Individual posterior medians used.

Table A7

Correlation matrix of drift rates across four time points split by groups.

Time Point	Group	Time 1	Time 2	Time 3
v Time 2	Group 1	0.77 [0.70 - 0.83]		
v Time 3		0.70 [0.61 - 0.78]	0.71 [0.63 - 0.79]	
v Time 4		0.63 [0.53 - 0.73]	0.66 [0.57 - 0.76]	0.62 [0.51 - 0.73]
v Time 2	Group 2	0.80 [0.74 - 0.85]		
v Time 3		0.69 [0.60 - 0.79]	0.79 [0.72 - 0.85]	
v Time 4		0.66 [0.57 - 0.77]	0.76 [0.68 - 0.83]	0.66 [0.55 - 0.76]
v Time 2	Group 3	0.72 [0.55 - 0.87]		
v Time 3		0.76 [0.59 - 0.91]	0.86 [0.76 - 0.95]	
v Time 4		0.47 [0.20 - 0.74]	0.46 [0.17 - 0.72]	0.82 [0.68 - 0.95]
v Time 2	Group 4	0.80 [0.72 - 0.88]		
v Time 3		0.77 [0.68 - 0.87]	0.86 [0.79 - 0.92]	
v Time 4		0.70 [0.57 - 0.81]	0.85 [0.76 - 0.91]	0.76 [0.64 - 0.87]
v Time 2	Group 5	0.80 [0.70 - 0.88]		
v Time 3		0.77 [0.67 - 0.86]	0.80 [0.70 - 0.87]	
v Time 4		0.55 [0.38 - 0.73]	0.55 [0.37 - 0.73]	0.57 [0.40 - 0.74]

Note. Means of Bayesian correlation estimates and 95 % credible interval reported.

Table A8

Correlation matrix of boundary separation across four time points split by groups.

Time Point	Group	Time 1	Time 2	Time 3
a Time 2	Group 1	0.69 [0.60 - 0.77]		
a Time 3		0.72 [0.64 - 0.80]	0.81 [0.75 - 0.86]	
a Time 4		0.65 [0.55 - 0.75]	0.71 [0.63 - 0.80]	0.77 [0.69 - 0.83]
a Time 2	Group 2	0.85 [0.80 - 0.89]		
a Time 3		0.67 [0.58 - 0.77]	0.84 [0.79 - 0.89]	
a Time 4		0.69 [0.59 - 0.78]	0.77 [0.70 - 0.84]	0.83 [0.77 - 0.89]
a Time 2	Group 3	0.70 [0.51 - 0.86]		
a Time 3		0.69 [0.46 - 0.87]	0.87 [0.77 - 0.96]	
a Time 4		0.79 [0.63 - 0.91]	0.90 [0.81 - 0.96]	0.79 [0.59 - 0.92]
a Time 2	Group 4	0.69 [0.57 - 0.81]		
a Time 3		0.61 [0.44 - 0.76]	0.76 [0.66 - 0.87]	
a Time 4		0.73 [0.61 - 0.84]	0.81 [0.71 - 0.88]	0.60 [0.43 - 0.78]
a Time 2	Group 5	0.60 [0.44 - 0.74]		
a Time 3		0.63 [0.49 - 0.77]	0.72 [0.60 - 0.84]	
a Time 4		0.58 [0.40 - 0.74]	0.62 [0.45 - 0.77]	0.58 [0.39 - 0.72]

Note. Means of Bayesian correlation estimates and 95 % credible interval reported.

Table A9

Correlation matrix of non-decision times across four time points split by groups.

Time Point	Group	Time 1	Time 2	Time 3
τ Time 2	Group 1	0.68 [0.60 - 0.76]		
τ Time 3		0.63 [0.53 - 0.73]	0.61 [0.50 - 0.71]	
τ Time 4		0.55 [0.43 - 0.66]	0.57 [0.45 - 0.68]	0.62 [0.50 - 0.72]
τ Time 2	Group 2	0.72 [0.64 - 0.80]		
τ Time 3		0.59 [0.47 - 0.70]	0.77 [0.70 - 0.84]	
τ Time 4		0.49 [0.35 - 0.62]	0.65 [0.54 - 0.74]	0.66 [0.56 - 0.76]
τ Time 2	Group 3	0.78 [0.65 - 0.91]		
τ Time 3		0.73 [0.53 - 0.89]	0.64 [0.41 - 0.84]	
τ Time 4		0.68 [0.47 - 0.86]	0.73 [0.54 - 0.88]	0.60 [0.31 - 0.82]
τ Time 2	Group 4	0.71 [0.60 - 0.82]		
τ Time 3		0.68 [0.54 - 0.81]	0.75 [0.62 - 0.84]	
τ Time 4		0.50 [0.33 - 0.69]	0.70 [0.56 - 0.82]	0.51 [0.31 - 0.71]
τ Time 2	Group 5	0.71 [0.58 - 0.82]		
τ Time 3		0.73 [0.62 - 0.84]	0.83 [0.75 - 0.90]	
τ Time 4		0.58 [0.43 - 0.75]	0.70 [0.56 - 0.83]	0.67 [0.51 - 0.79]

Note. Means of Bayesian correlation estimates and 95 % credible interval reported.

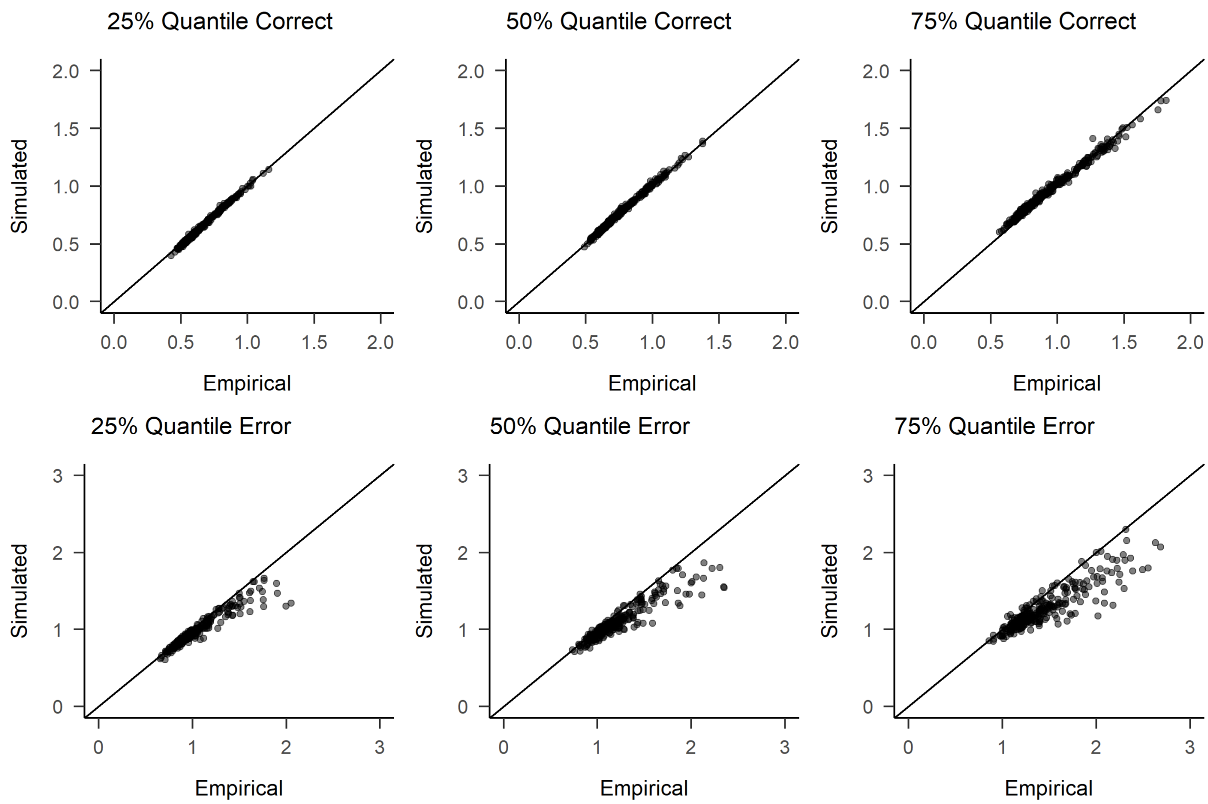

Figure A1. Posterior predictive check of RTs for T2. Participants with 10 or less errors omitted from the error response time plots.

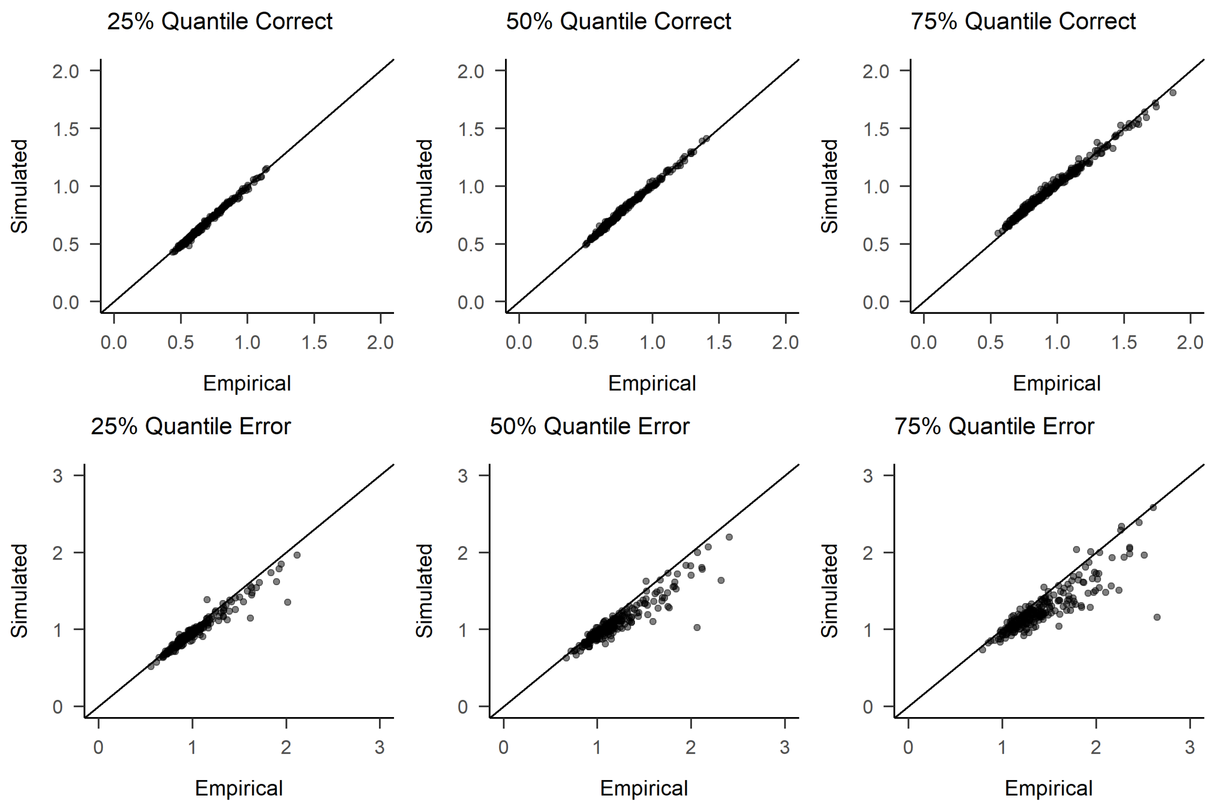

Figure A2. Posterior predictive check of RTs for T3. Participants with 10 or less errors omitted from the error response time plots.

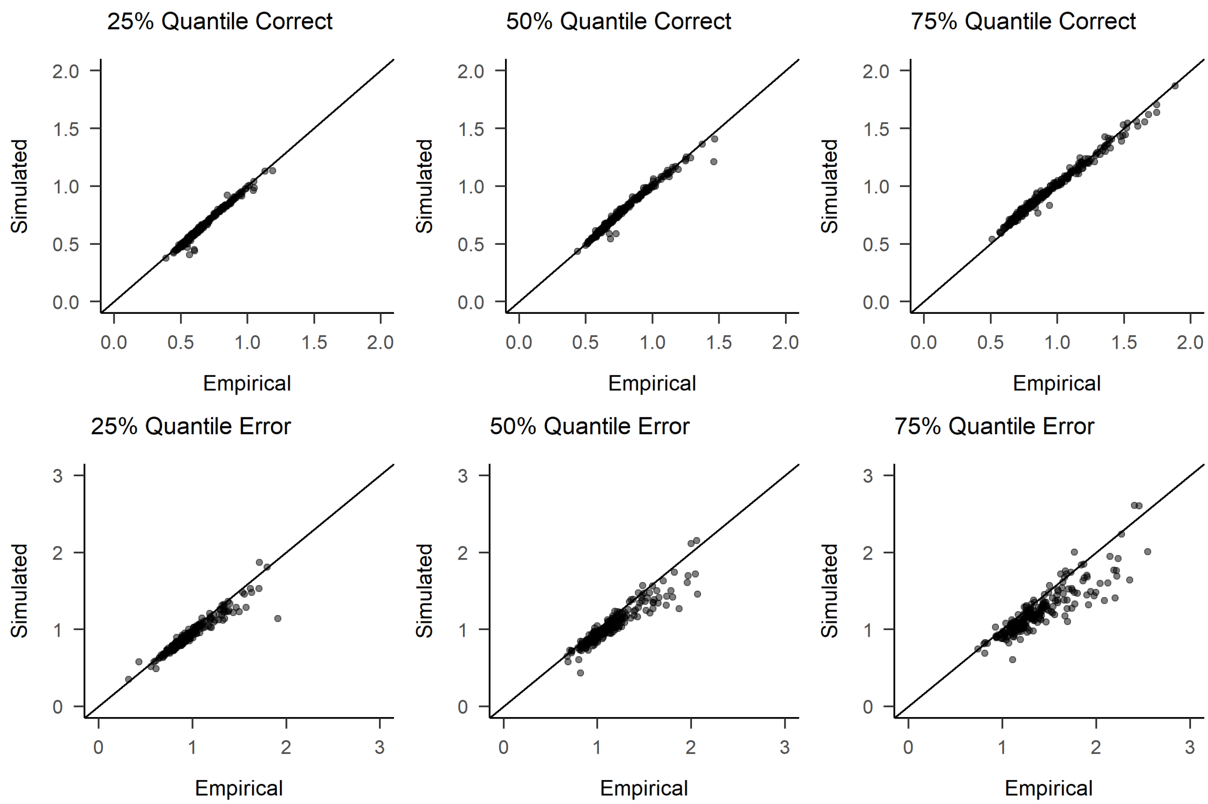

Figure A3. Posterior predictive check of RTs for T4. Participants with 10 or less errors omitted from the error response time plots.

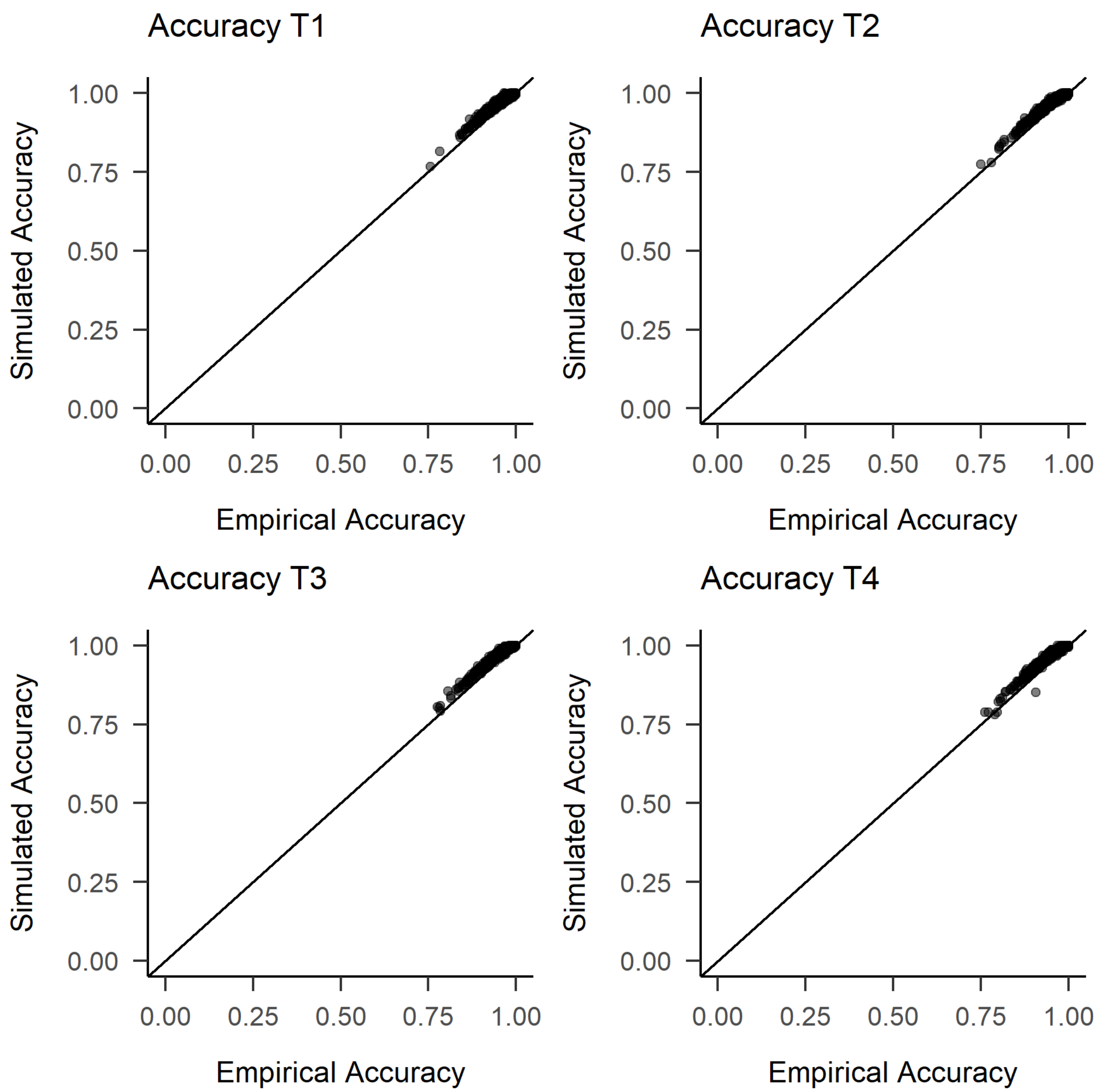

Figure A4. Posterior predictive checks of accuracy rates for all time points.

Appendix A 2

Manuscript 2:

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*. Advance online publication.

* First authorship shared.

 doi: 10.1037/xge0000774

Diffusion Modeling and Intelligence:

Drift rates show both domain-general and domain-specific relations with intelligence

Veronika Lerche*, Mischa von Krause*, Andreas Voss, Gidon T. Frischkorn, Anna-Lena Schubert & Dirk Hagemann

*Veronika Lerche and Mischa von Krause contributed equally to this project.

Word Count: 10,413

Author Note

Veronika Lerche, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany; Mischa von Krause, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany; Andreas Voss, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany; Gidon Frischkorn, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany; Anna-Lena Schubert, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany; Dirk Hagemann, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany

Acknowledgments

This research was supported by grants from the German Research Foundation to Andreas Voss (Grant Nr. VO1288/2-2) and to the Graduate School SMiP (GRK 2277; Statistical Modeling in Psychology). We further thank our students Anne Bülow, Sarah Hladik, Henrike Jungeblut, Julia Karl, Sontje Nordholt, Yannick Roos, and Moritz Spielberger for their help during data acquisition.

Correspondence concerning this article should be addressed to Veronika Lerche, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, email: veronika.lerche@psychologie.uni-heidelberg.de, telephone: +49-6221-54-7322 or to Mischa von Krause, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, email: mischa.vonkrause@psychologie.uni-heidelberg.de, telephone: +49-6221-54-7747.

Abstract

Several previous studies reported relationships between speed of information processing as measured with the drift parameter of the diffusion model (Ratcliff, 1978) and general intelligence. Most of these studies utilized only few tasks and none of them used more complex tasks. In contrast, our study (*N* = 125) was based on a large battery of 18 different response time tasks that varied both in content (numeric, figural, and verbal) and complexity (fast tasks with mean RTs of ca. 600 ms vs. more complex tasks with mean RTs of ca. 3000 ms). Structural equation models indicated a strong relationship between a domain-general drift factor and general intelligence. Beyond that, domain-specific speed of information processing factors were closely related to the respective domain scores of the intelligence test. Furthermore, speed of information processing in the more complex tasks explained additional variance in general intelligence. In addition to these theoretically relevant findings, our study also makes methodological contributions showing that there are meaningful interindividual differences in content specific drift rates and that not only fast tasks, but also more complex tasks can be modeled with the diffusion model.

Keywords: intelligence, diffusion model, mathematical models, reaction time methods, fast-dm

Diffusion Modeling and Intelligence:

Drift rates show both domain-general and domain-specific relations with intelligence

One of the processes that has often been discussed as basis of individual differences in intelligence is speed of information processing (Jensen, 2006). This notion is supported by consistent empirical results showing moderate relationships between general intelligence[1] and response times (RTs) from a broad range of cognitive tasks (Sheppard & Vernon, 2008). Regarding these relationships between intelligence and RTs, (at least) two important observations have been made in the last decades: (1) The relationship between RT and intelligence does not seem to be specific to content domains (verbal, figural, numeric; Levine, Preddy, & Thorndike, 1987; Neubauer & Bucik, 1996). (2) The slower responses within one task are more highly related to intelligence than the faster responses, resulting in the formulation of the *worst performance rule* (Larson & Alderton, 1990; for a review, see Coyle, 2003; for methodological considerations, see Frischkorn, Schubert, Neubauer, & Hagemann, 2016; for a meta-analysis, see Schubert, 2019). Thus, in brief, the relationship between intelligence and speed of information processing seems to depend on the speed of trials, but not or only to a small degree on the specific task content.

However, there are some methodological limitations of previous studies that examined the relationship between intelligence and speed of information processing. One of these limitations has been pointed out by Schmiedek, Oberauer, Wilhelm, Süß, and Wittmann (2007): Regarding the worst performance rule, they noted that previous studies employed different RT bands resulting in only restricted numbers of trials per band, thereby limiting the reliability of estimates. Instead of employing RT bands, Schmiedek et al. (2007) used a mathematical model that takes into account information about RT distributions, and thus has

[1] In this paper, we use the term *general intelligence* to denote a general factor that statistically emerges in intelligence tests (in the sense of sampling theories, e.g., Kovacs & Conway, 2016). Our use of the term general intelligence does not imply that we assume this factor to be a causal factor. In fact, our study does not have the aim of providing any inferences regarding the question of causality.

a considerably higher information usage—the diffusion model (Ratcliff, 1978; see Voss, Nagler, & Lerche, 2013, for a review).

The diffusion model is a stochastic model that is applicable to binary response time tasks and allows the separation of different, otherwise confounded, processes. One parameter of this model—drift rate—is supposed to provide a pure measure of speed of information processing, with other processes (such as speed of motoric response execution, or speed-accuracy settings) "partialled out". It is a known property of the diffusion model that changes in drift rate have a larger influence on the tail than on the leading edge of RT distributions. More specifically, Ratcliff and McKoon (2008) report that changes in the .9 quantile of RT distributions are typically four times as large as changes in the .1 quantile. Changes in other parameters of the diffusion model—which measure processes such as speed-accuracy settings (threshold separation parameter) or the duration of encoding and motoric processes (non-decision time parameter)—on the other hand, do not have this asymmetric influence on fast vs. slow RTs. In line with this reasoning, Schmiedek et al. (2007) found the drift rate (but not other diffusion model parameters) to be related to intelligence. In the following years, other studies also supported the notion that intelligence as measured by classical intelligence tests is associated with the drift rate (e.g., Ratcliff, Thapar, & McKoon, 2011; Schmiedek et al., 2007; Schmitz & Wilhelm, 2016; Schubert, Hagemann, Voss, Schankin, & Bergmann, 2015).

In contrast to drift rate, mean RTs are influenced by a number of different processes (e.g., how cautious individuals are and how fast they execute the motoric response). In fact, for these other processes, for which the diffusion model provides distinct measures, no consistent correlations with intelligence have been found. The only relationship that has been reported several times is a small negative correlation of intelligence with non-decision time, indicating that more intelligent people are faster in non-decisional processes, that is, in encoding and/or motoric processes (McKoon & Ratcliff, 2012; Schubert et al., 2015; Schulz-

Zhecheva, Voelkle, Beauducel, Biscaldi, & Klein, 2016). In several other studies, however, this relationship between intelligence and non-decision time has not been found (e.g., Schmiedek et al., 2007; Schmitz & Wilhelm, 2016). Critically, previous studies that examined relationships between diffusion model parameters and intelligence are based on only limited numbers of tasks and they used different estimation approaches, which might account for inconsistencies in the findings.

To sum up, according to the literature distinct effects of speed of information processing on RT distributions account for the worst performance rule. Furthermore, whereas drift rate seems to be consistently related to intelligence, for the other diffusion model parameters the current state of research is inconsistent. We will now come back to the question of domain-specificity of mental speed. The diffusion model, which has proved useful for the examination of the worst performance rule, might also help to gain further insights into this finding.

Interestingly, previous studies did not find clear support for a three-factor structure (numeric, figural, verbal) in RT tasks, suggesting that there are no substantial domain-specific factors of speeds of information processing (Levine et al., 1987; Neubauer & Bucik, 1996). This observation is in contrast to findings from intelligence tests that assume a hierarchical structure of intelligence with both a general factor and domain-specific factors (e.g., verbal, numeric, figural; Jäger, Süß, & Beauducel, 1997). However, it might be difficult to draw definite conclusions from the mental speed studies by Levine et al. (1987) and Neubauer and Bucik (1996) as they did not explicitly disentangle processing speed from other processes. The mental speed measures used in these studies might, thus, have been distorted and may therefore have been no valid indicators of actual speed of information processing. Notably, the studies did find a tendency for domain-specific correlations (i.e., higher correlations between intelligence and mental speed in the respective domains) although their data did not contain compelling evidence for a hierarchical factor structure of mental speed. Moreover, effects were not consistent and very small. Thus, we hypothesize that the measures of processing speed used might not have been pure enough to find clear support for domain-specificity. Using drift rate as a purer measure of cognitive speed provides a more powerful and fairer test for the question, whether cognitive speed has stable domain-specific components. The diffusion model literature, though, so far only reports one general drift rate factor, and Schmiedek et al. (2007) see their results as suggesting that “underlying mechanisms could be relatively task-independent” (p. 425). Notably, however, previous diffusion model studies only used a very restricted number of tasks per domain. Accordingly, the existing literature does not allow to draw clear inferences as to whether there is only one common speed of information processing or whether there are domain-specific speeds. It is further unclear whether domain-specific processing speeds (if they exist) are related to the respective intelligence test scores or just to general intelligence.

To sum up, we see two important research gaps that have not been addressed by previous studies analyzing the association of cognitive speed and intelligence with the diffusion model framework. These gaps originate from restrictions in the number and breadth of the employed tasks. First, whereas previous studies found clear evidence for an association of drift rate and general intelligence, results regarding the other diffusion model parameters are less clear-cut. Second, previous diffusion model studies did not vary task content systematically, so it remains an open question whether there are also domain-specific factors of cognitive speed, and whether such domain-specific speeds are related to the respective intelligence test scores.

Another perspective on the research aims listed above relates to the diffusion model as a diagnostic tool: Whereas, in the past, the diffusion model was mainly employed for the analysis of differences between groups or conditions, in recent years it has been proposed to use this methodology also for the analysis of *interindividual differences* in cognitive processes (e.g., Frischkorn & Schubert, 2018; Ratcliff & Childers, 2015; White, Curl, & Sloane, 2016). Our study allows for an examination of whether there are in fact meaningful content-domain specific interindividual differences in the processing of information.

One further important goal of the present study is the comparison of easy (perceptual) tasks vs. complex tasks (requiring more complex mental operations). In the past, it was often recommended to apply the diffusion model only to tasks with mean trial RTs of up to 1.5 seconds (e.g., Ratcliff & Frank, 2012; Ratcliff & McKoon, 2008; Ratcliff, Thapar, Gomez, & McKoon, 2004). Following this rule of thumb, the previous studies that examined links between intelligence and drift rate used easy tasks that required no complex mental operations and thus allowed for very rapid responding. Interestingly, first studies indicate that the diffusion model might also be applicable to more complex tasks, requiring several seconds for response selection (Aschenbrenner, Balota, Gordon, Ratcliff, & Morris, 2016; Lerche, Christmann, & Voss, 2018; Lerche & Voss, 2017a). These studies, however, only examined single tasks (e.g., a complex figural task in the studies by Lerche & Voss, 2017a) and did not compare easy with more complex tasks. In the present study, we use a large number of both easy and more complex tasks and examine whether the goodness-of-fit of the diffusion model differs between data from easy vs. complex tasks.

Furthermore, we test the criterion validity of drift rate in the more complex tasks, analyzing whether drift rate is related to intelligence not only in the fast, but also in the more complex tasks. In fact, for more complex conditions stronger associations of intelligence and mental speed have been reported (Sheppard & Vernon, 2008; see also Coyle, 2017; Marshalek, Lohman, & Snow, 1983). More precisely, the relationship between intelligence and mental speed increases from very simple tasks (RTs of about 300 ms) to moderately complex tasks (RTs around 500-900 ms), but decreases again if tasks get even more complex

(RTs of more than 1200 ms; Jensen, 2005; see also Lindley, Wilson, Smith, & Bathurst, 1995). Thus, there seems to be an inverted-U-shaped relationship between task complexity and the correlation between intelligence and mental speed. In our study, we examine “easy” tasks (around 600 ms; i.e., moderately complex tasks according to the definition by Jensen) and “complex” tasks (around 3000 ms). Jensen states the hypothesis that one reason for the decrease from moderately complex to complex tasks is that individual differences in performance strategies play a more important role in complex tasks. Furthermore, Lindley et al. (1995) point out that in their complex task participants had to repeatedly scan between different task elements resulting in supplemental motor time so that RT became a less accurate measure of processing speed. Notably, drift rate is a more specific measure of processing speed with some strategies (different speed-accuracy settings) or the duration of encoding processes partialled out. Jensen also mentions that complex tasks show more task-specific factors that can weaken the correlation between RT and *g*. As we use a large number of tasks, we can use a structural equation modeling (SEM) approach, which helps us to control for task specificities. Thus, the use of diffusion modeling and SEM provides us with more specific measures of mental speed and the relationship between mental speed and intelligence. Accordingly, in our study we assume a substantial relationship between drift rate and intelligence also for the more complex tasks.

In the following paragraphs, we first give a brief introduction to the diffusion model (for more detailed information, see Ratcliff, Smith, Brown, & McKoon, 2016; Voss, Nagler, et al., 2013; Wagenmakers, 2009). Next, we present a review of previous studies that examined relationships between intelligence and diffusion model parameters. In the subsequent section, we present theoretical underpinnings of the relationship between drift rate and intelligence. After that, we examine the question of whether the diffusion model is also applicable to more complex RT tasks. Finally, we present the method and results of our study.

Introduction to the Diffusion Model

The diffusion model (Ratcliff, 1978) is a mathematical model that is applicable to decision tasks with two response options. When a participant works on a trial of such a binary task (e.g., color discrimination task, see Voss, Rothermund, & Voss, 2004) she is assumed to accumulate information continuously until she reaches one of two thresholds (see Figure 1). The two thresholds represent either the two response options (response coding) or the response accuracy (accuracy coding; e.g., Figure 1). The distance between the thresholds, the so-called threshold separation (a) reflects how much information needs to be accumulated to reach a decision. If individuals are more cautious, they will accumulate more information before they decide for one option. In this case, a larger threshold separation will cause longer RTs and—at the same time—higher accuracy because the decision processes will terminate at the wrong threshold more rarely.

Speed of information processing is denoted as drift (v) and is illustrated by the arrows in Figure 1, with steeper arrows indicating faster accumulation of information. During information sampling, Gaussian noise is added constantly to the drift, reflecting random fluctuations in the decision process. Due to this noise, the accumulation process does not terminate after the same time and not always at the same threshold, even if the available information (i.e., the stimulus) is identical. The two panels of Figure 1 illustrate the influence of differences in drift on the RT distributions. It can be seen that if the drift is higher (Panel B) fewer errors are made resulting in a smaller distribution at the error threshold and a larger distribution at the correct response threshold. In addition, RT distributions for lower drift rates (Panel A) are more spread out than those for higher drift rates. Another diffusion model parameter is non-decision time (t_0) which subsumes the duration of all non-decision processes, such as encoding of information (preceding the decision process) and motoric response execution (succeeding the decision process). The last parameter of the basic diffusion model is starting point, which maps whether a decision is biased for one of the two response options.

Next to these four main model parameters, often three more parameters mapping intertrial variability of drift s_v, starting point s_{zr} (Ratcliff & Rouder, 1998) and of non-decision time s_{t0} (Ratcliff & Tuerlinckx, 2002) are estimated. However, the intertrial variability of drift and starting point cannot be estimated reliably and fixation of these parameters to zero can improve estimation of the main diffusion model parameters (Lerche & Voss, 2016; see also van Ravenzwaaij, Donkin, & Vandekerckhove, 2017).

Intelligence and Diffusion Modeling

It is well-known that intelligence shows a high stability over long time periods (e.g., Carroll, 1993; Larsen, Hartmann, & Nyborg, 2008). Accordingly, the rank-order stability of a diffusion model parameter is a prerequisite for it to be related to intelligence. Test-retest studies by Lerche and Voss (2017b) provide first evidence that drift rates are rather time stable. More specifically, in Study 1, a lexical decision task and a recognition memory task were completed at two sessions, separated by a one-week interval. In a second study, participants worked on an associative priming task (again with a test-retest interval of one week). In all three tasks, drift showed acceptable test-retest correlations. The authors further conducted simulation studies based on the parameters estimated for the empirical data. Specifically, they simulated two data sets (reflecting the two sessions) based on identical parameter values. Interestingly, test-retest correlations of drift rates estimated from the real data were very similar to correlations based on simulated data. This suggests that the speed of information processing was very stable across measurements, and situation influences on drift rate are rather small.

A study by Schubert, Frischkorn, Hagemann, and Voss (2016) corroborates this idea. The authors conducted a test-retest study with a time interval of eight months. They then used latent state-trait analyses to disentangle trait influences and situation influences. The most important finding was that drift rates had the highest consistencies, indicating that they were the most trait-like parameters. Accordingly, drift rate might be a good candidate for associations with intelligence, which is characterized by high temporal stability and great consistency (Danner, Hagemann, Schankin, Hager, & Funke, 2011).

In support of this hypothesis, in several studies relationships between general intelligence and drift rate have been reported (McKoon & Ratcliff, 2012; Ratcliff, Thapar, & McKoon, 2010; Ratcliff et al., 2011; Schmiedek et al., 2007; Schmitz & Wilhelm, 2016; Schubert et al., 2015; Schulz-Zhecheva et al., 2016). These studies measured drift rates from performance in different types of binary tasks. For example, Ratcliff et al. (2010) used a numerosity discrimination task, a recognition memory task, and a lexical decision task. Intelligence was assessed by means of the Vocabulary and Matrix Reasoning subtests of the Wechsler Adult Intelligence Scale. The authors observed substantial correlations between IQ (mean over the two scales) and drift rate as measured in the lexical decision (r = .53) and recognition memory task (r = .55). The correlation was smaller for the numerosity task (r = .24). As also alluded to by the authors this is not astonishing, as the subscales of the intelligence test that were administered did not address the numeric domain, but the verbal (vocabulary subtest) and figural domain (matrix reasoning subtest). Only small-to-moderate values were observed for the correlation of intelligence with threshold separation and non-decision time ($|r|_{max}$ = .33).

In a subsequent paper, Ratcliff et al. (2011) reported correlations between IQ and diffusion model parameters from an item recognition memory task and an associative recognition memory task. Again, there were substantial correlations between the IQ scales and drift rate with *r* = .36-.68 for college age participants and *r* = .47-.67 for participants aged 60-74 years. For the oldest group (75-90 years old), correlations were smaller (*r* = .18-.34), which was seen as partly attributable to floor effects and lower reliability of the vocabulary subtest. For threshold separation and non-decision time, an inconsistent pattern of mostly small correlations with IQ emerged across tasks and age groups. McKoon and Ratcliff (2012), who assessed participants of the same three age groups with the same two subtests of the Wechsler Intelligence Scale, also found IQ to be correlated with drift rates for associative recognition (*r*s between .24 and .68) and item recognition (*r*s between .49 and .68). In addition, non-decision times were negatively related to IQ, suggesting faster encoding and/or response execution of more intelligent participants.

Schubert et al. (2015) report results from three elementary cognitive tasks (Hick task, Sternberg memory scanning task, and Posner letter matching task). Intelligence was assessed in this study with Raven’s Advanced Progressive Matrices and with a shortened version of the knowledge test of the German Intelligenz-Struktur-Test 2000-R. In line with the results of the previously reported studies, the authors observed a correlation of *r* = .50 between the component score of drift rates from the different tasks (extracted from principal component analyses) and general intelligence. In addition, like in the study by McKoon and Ratcliff (2012), a negative relationship between intelligence and non-decision time emerged (*r* = -.42). Thus, the more intelligent individuals not only showed higher drift rates but also shorter non-decision times.

Schmiedek et al. (2007) used a larger number of different tasks: two lexical tasks, two numeric tasks, and four spatial tasks. For the assessment of intelligence, the authors employed tasks of the Berlin Structure of Intelligence Test (BIS; Jäger et al., 1997). More specifically, three numeric, figural, and verbal tasks from the reasoning and psychometric speed operation scales were used. Based on structural equation modeling (SEM), the authors found that the latent factor of psychometric speed correlated highest with latent drift rate (r = .59), whereas the correlations were smaller for threshold separation (r = -.42) and non-decision time (r = -.04). Similarly, for reasoning the highest correlation emerged for drift rate (r = .79; threshold separation: r = -.48; non-decision time: r = .25).

Schmitz and Wilhelm (2016) also reported relationships of drift with intelligence. Using two different cognitive tasks and also employing SEM to link the drift rates to a measure of fluid intelligence (a figural sequence reasoning test from the BEFKI; Wilhelm, Schroeders, & Schipolowski, 2014) they found correlations with drift of r = .15 (non-significant) for visual search and of r = .29 for visual comparison. The authors did not report any significant correlations between fluid intelligence and the other diffusion model parameters.

Schulz-Zhecheva et al. (2016) tested a sample of participants aged 8 to 18 years with Cattell's Culture Fair Intelligence Test (CFT 20-R; Cattell & Cattell, 1960; Weiss, 2006) of fluid intelligence and measured diffusion model drift rates across four simple decision tasks. The latter consisted of deciding whether a number was odd or even, whether a number was smaller or larger than 50, whether an arrow pointed upward or downward and whether a line was shown in the upper or lower half of the screen. Once more, drift rate was by far the strongest correlate of fluid intelligence (*gf*; r =.41; non-decision time: r = -.20; threshold separation: r = -.13). The total *gf* factor variance explained by the diffusion model parameters was 19%.

In sum, drift rate seems to have a trait-like characteristic, showing moderate consistency across different tasks and temporal stability. Moreover, robust relationships between drift rates and intelligence have been reported across different studies and experimental tasks. In contrast, correlations of the other diffusion model parameters with intelligence are smaller and the pattern is less consistent. Apart from the relationship with drift rate, the finding that has been most often reported is a negative correlation between intelligence and non-decision time. However, this relationship only showed up in some of the studies.

From the previous diffusion model literature, no clear conclusions can be drawn regarding the existence of domain-specific drift rates. Whereas the findings by Schmiedek et al. (2007) speak in favor of task-independence of speed of information processing, other studies lend first support to the hypothesis that speed of information processing might differ between domains. For example, Ratcliff et al. (2010) who measured intelligence with a verbal and a figural test found a smaller correlation of intelligence with drift in a numeric task than in a verbal or a figural task. Furthermore, in the study by Schubert et al. (2016) drift rates showed smaller consistencies than typically observed in intelligence tests, suggesting that individual differences in drift rates also reflect task- and content-specific properties to a substantial degree. Importantly, a study that combines domain-specific intelligence assessment with a battery of various RT tasks that tackle these domains is still missing. It is an open question whether a domain-specific structure of speed of information processing can be found and if so, if such domain-specific drift rates correlate with the respective domain scores of an established intelligence measure. To address these questions, in our study, we put together a battery of 18 different binary RT tasks that address the three different domains of intelligence.

Relationship between Drift Rate and Intelligence: Theoretical Considerations

As we described in the last section, empirical findings support the view that speed of information processing as measured by the drift rate of the diffusion model is related to intelligence. Next, we will outline why this relationship is theoretically plausible and why we assume that in more complex tasks relationships between drift rate and intelligence might be even stronger than in less complex tasks.

For illustration, let us consider the two mechanisms proposed by Salthouse (1996) to describe the assumed effect of age-related slowing on cognition, the *limited time mechanism* and the *simultaneity mechanism*. The *limited time mechanism* is supposed to be in effect when the time for solving a problem is limited and only little time is available for the higher-order integration of information, because earlier stages of information processing occupied too much time. The *simultaneity mechanism* assumes that, over time, information becomes less available in working memory. If older individuals need more time to process information, a greater amount of information will then be lost or at least fragmented by the time they start to integrate all processed information. Accordingly, we assume that individuals who have a reduced speed of information processing (i.e., a smaller drift rate) will suffer more from time constraints, as they have less time available for higher-order processing. Furthermore, for these individuals (in contrast to individuals with higher drift rates) more information will get lost during the accumulation process. The importance of temporal aspects in information-processing has also been stressed, for example, by the Time-Based Resource-Sharing (TBRS) model (Barrouillet, Bernardin, & Camos, 2004; Camos & Barrouillet, 2014). The model supports the view of a time-related decay of memory traces and regards the number of necessary memory retrievals and the time given to perform them as important factors influencing performance. More complex tasks will often require more memory retrievals than simple RT tasks (e.g., perceptual or recognition memory tasks), with time pressure kept constant between task types. Accordingly, more complex RT tasks might be more vulnerable to deficits in speed of accumulation of information. In other words, task-related differences in working memory demands might underlie higher relationships between more complex tasks and intelligence.

A similar idea is part of the *process overlap theory* (Conway & Kovacs, 2015; Kovacs & Conway, 2016, see also Kan, van der Maas, & Kievit, 2016), a recently proposed intelligence theory. According to this theory “executive/attentional processes” play an important role, underlying—amongst other—both the worst performance rule and the finding of higher relationships with intelligence for more complex tasks. Process overlap theory is considered a modern version of Thomson`s sampling theory (Thomson, 1916). According to Thomson (1916), each mental test addresses a number of what has later often been called “bonds” (see Deary, Lawn, & Bartholomew, 2008, for a historical analysis). This account explains correlations of performance across tasks by an overlap of required psychological processes (in the intelligence literature also often referred to as *positive manifold*). Rather than assuming a causal general factor of intelligence, process overlap theory regards the *g* factor—that undoubtedly shows up in any factor analysis of cognitive ability test data—as an “emergent property” (p. 162, Kovacs & Conway, 2016).

In contrast to Thomson`s theory, process overlap theory does not postulate an additive overlap of processes but assumes a bottleneck in form of multiplicatively linked “executive/attentional processes” (Kovacs & Conway, 2016; see Schubert & Rey-Mermet, 2019, for a critical discussion of the empirical testability of this hypothesis). Kovacs and Conway (2016) state that “*g* loadings depend on the involvement of executive processes seated primarily in the prefrontal cortex rather than on the number of processes measured” (p. 170) and define *complexity* as “the extent to which a test taps executive/attentional processes” (p. 164). Accordingly, they suppose the relationship between more complex tasks and intelligence is driven by the engagement of executive processes. Similarly, it is assumed that the slower trials in a task are more highly related to intelligence because they are indicators of failures in executive processes. We support this view of a common explanation of both these empirical observations. More specifically, we assume that the drift rate of the diffusion model might provide a methodological account for both observations. It has already been demonstrated that the drift rate provides an explanation for the worst performance rule (e.g., Schmiedek et al., 2007). So far, however, no study has examined relationships between intelligence and drift rate in more complex tasks. In our study, we examine complex tasks with RTs of about 3000 ms, thus tasks for which according to Jensen (2005) relationships between mental speed and intelligence should be small because of higher influences of individual differences in strategies. As the diffusion model provides a more specific measure of mental speed (e.g., partialling out speed-accuracy settings), we assume that also for more complex tasks there should be a substantial relationship between mental speed (measured by means of the drift rate) and intelligence. This relationship might even be larger than for less complex tasks because of higher memory demands.

In short, we suppose that a higher speed of information processing helps to counteract time-related decay of memory. This might be particularly relevant for tasks with higher memory demands. In our study, we examine both fast tasks with little memory demands and more complex tasks with higher memory demands. As we will outline in the next section, we assume that the diffusion model is also applicable to such more complex tasks.

Diffusion Modeling for Fast vs. More Complex Tasks

In the past, the diffusion model has almost exclusively been applied to *fast* tasks. By this term, we refer here to tasks with a mean trial duration of below 1.5 seconds. The claim that the diffusion model is only applicable to such fast tasks has been repeatedly put forth (e.g., Ratcliff & Frank, 2012; Ratcliff & McKoon, 2008; Ratcliff, Thapar, et al., 2004) and has strongly influenced the choice of tasks for diffusion modeling for a long time. The reasoning underlying this restriction is that tasks with longer RTs were seen as more likely to violate basic assumptions of the diffusion model (such as the assumption that decisions are based on a single processing stage and that parameters remain constant over time within one trial). However, we question the idea that data from more complex tasks are more likely to violate assumptions of the diffusion model.

Let us first consider response time tasks that fulfill the 1.5 second rule, that is, typical RT tasks to which the diffusion model has been applied frequently, such as a color discrimination task. In this task, participants have to decide whether, for example, the color orange or blue prevails in a square filled with pixels of these two colors (e.g., Germar, Schlemmer, Krug, Voss, & Mojzisch, 2014; Voss et al., 2004). Participants are assumed to sample evidence from the perceptual dimension (here, color). In such perceptual tasks, it is plausible that participants continuously sample information (i.e., perceptions of color), until they are reasonable sure that one color prevails. However, the diffusion model has also often been applied to tasks in which a continuous sampling of information is less plausible. Imagine, for example, the lexical decision task (Ratcliff, Gomez, & McKoon, 2004). Here it is unclear, whether—during decision making—information of "wordiness" of a stimulus is accumulated with constant drift. Rather, different pre-lexical (e.g., bigram frequencies) and post-lexical (e.g., similarity to existing words) processes could inform the decision with different impact, thus resulting in separate decision stages with different drift rates.

Since there is no way to assess the assumptions of the diffusion model analytically, the model has to be validated empirically, both regarding its general ability to fit empirical data and regarding the external validity of all model parameters. Such validation studies are essential for any cognitive model and any new type of task. One important tool in this regard are so-called selective influence studies that demonstrate that specific experimental manipulations with high face validity take impact on specific model parameters in a specific way. Importantly, such selective influence studies have shown comparably good validity of the diffusion model parameters for color discrimination (Voss et al., 2004) and recognition memory (Arnold, Bröder, & Bayen, 2015). Accordingly, even in the recognition memory task the model assumptions are apparently not seriously violated.

Imagine now a more complex task, for example, the complex figural task used in our study (see Figure 2, for an example stimulus). In each trial of this task, participants see several rectangles. Half of the rectangles are surrounded by a blue border and half of them by a red border. Participants have to estimate the total area of the blue-bordered rectangles and compare it to the total area of the red-bordered rectangles in order to assess which of these summed areas is larger. In studies by Lerche and Voss (2017a), the variant of the complex figural task employed led to mean RTs of about 7 seconds per trial. Answers of participants to an open-framed question about their use of strategies revealed that a typical strategy is to sequentially pick pairs of rectangles and compare the two rectangles within one pair to each other (i.e., one red- and one blue-bordered rectangle). Apart from the high perceptual and spatial affordances (e.g., considering color of borders, and both width and height of rectangles at different positions on the screen), also memory processes are relevant. Participants need to remember which of the rectangles they have already compared and how large the differences were. Thus, this task can be partitioned into several sub-tasks. For example, each pair of rectangles could be seen as one sub-task (with each of these sub-tasks consisting of further sub-tasks). Each sub-task might be conceived of as having its own speed of information processing. Following the concept of the law of large numbers, with an increase in the number of sub-tasks, extreme values of drift rate in single sub-tasks might become less influential, allowing for an even better measurement of overall mental speed. Thus, we assume that the data of tasks such as the complex figural task can be modelled adequately by a constant drift (i.e., on average, information accrues towards the correct boundary) with Gaussian noise (reflecting non-systematic influences).

Importantly, in selective influence studies based on the complex figural task, convergent and discriminant validity of the diffusion model parameters were comparable to what has been observed in the validation studies based on faster tasks (Lerche & Voss, 2017a). Furthermore, in another study, data from a complex *verbal* task were entered into a diffusion model analysis (Lerche et al., 2018). In this task, participants had to assess the meaningfulness of sentences, which took 2.2 seconds on average. Results again demonstrated an excellent fit of the diffusion model. Thus, these first empirical findings support our claim that the diffusion model can also be applied to tasks with mean response times above 1.5 seconds. In the present study, we build upon these promising results and employ both fast and more complex tasks. We compare the model fit between these two types of tasks and examine the external validity (analyzing the relationship of drift rate with intelligence).

The Present Study

In the present study, an intelligence test battery and a battery of 18 binary RT tasks were administered to a sample of 125 participants. The RT tasks included both simple and complex tasks addressing three content domains (numeric, figural, and verbal). With our study, we pursued three main objectives: First, we aimed to replicate findings from previous studies showing that general intelligence correlates with drift rate measured across a variety of different tasks. That is, we expected a substantial relationship between general intelligence and the drift rates across tasks. Second, we wanted to examine whether there are domain-specific aspects of cognitive speed as measured by drift rates and—if so—whether these are related to the respective numeric, verbal, and figural aspects of intelligence, as measured by an intelligence test. Third, we aimed at further investigating the applicability of the diffusion model to more complex RT tasks, which require more time for response selection. Specifically, we compare model fit from nine fast and nine more complex tasks. We also examine how drift rates estimated from the more complex tasks specifically predict general intelligence.

Method

Participants

We determined the required minimum sample size for structural equation analyses with a power analysis following the procedure described by Kim (2005). According to this procedure, the proposed minimum sample size for a test of close model fit according to the Root Mean Squared Error of Approximation (RMSEA) is 113 (df = 350, α =.05, β = .05). We recruited 125 participants for the study to ensure adequate power.[2]

We used different recruitment methods. The largest part of participants was recruited via a newspaper article. Others were hired via the participants' pool of the Psychological Institute of Heidelberg University in Germany using the software hroot (Bock, Baetge, & Nicklisch, 2014) or by means of fliers that were distributed at public places. We obtained informed consent from all participants. Participants were remunerated with 35€ after data collection was completed. In addition, all participants received feedback about their performance. Participants were between 18 and 65 years old (M = 36.0, SD = 14.3). Sixty-three percent were females. The percentage of students amounted to 50%.

Design and Procedure

The study consisted of three sessions. In the first session, participants had to work on an intelligence test[3]. In the second and third session, all RT tasks were administered (with nine of these tasks in each session). The order of tasks was identical for all participants and is provided in Table 1. Tasks of the three different domains and fast and slow tasks were presented alternatingly. After the third and the sixth task within each session, participants took a break of three minutes.

Each of the 18 tasks started with four practice trials. In these trials, participants

[2] Following suggestions of our reviewers, we kept the structural equation models simpler than in our original analysis plan. Most importantly, for the intelligence data, we used scale means rather than the single task scores, leading to a lower number of *df*s in our models.
[3] N = 11 participants had already participated in a previous study in which the same intelligence test was administered. These participants, therefore, only took part in the two PC assessments and received 25€.

received feedback about the correctness of their response (green checkmark vs. red cross for correct vs. erroneous responses, respectively; presentation duration: 1500 ms). After the practice trials, 100 test trials (preceded by one warm-up trial) were administered. All tasks had a binary response format, with both responses correct in half of the trials. Simulation studies have shown that the diffusion model can provide reliable parameter estimates for about 100 or even fewer trials (Lerche, Voss, & Nagler, 2017). The practice and warm-up trials were discarded from subsequent analyses. The order of trials was determined randomly and was held constant for all participants. In each trial, participants had to press one of two keys (“A” or “L”). The key assignment was identical for all participants. Each trial started with the presentation of a fixation cross for 500 ms. Subsequently, the target was shown and remained on the screen until the participant responded. Participants were instructed always to respond as fast and accurately as possible. The next trial started after an inter-trial-interval of 500 ms.

The fast tasks took between 528 and 810 ms on average per trial (M = 655 ms) and the slow tasks took between 2469 and 4314 ms (M = 3319 ms). The mean duration of assessment sessions was 71 minutes for session 2 and 69 minutes for session 3.

Intelligence Assessment

For the assessment of intelligence we used the Berlin Intelligence Structure Test (BIS; Jäger et al., 1997) which relies on the bimodal Berlin intelligence structure model (Jäger, 1982). This model comprises operation-related and content-related components of general intelligence. Of interest to our study were the content-related components (numeric, figural, and verbal). The intelligence assessment was run in sessions of six participants at maximum and took on average 50 minutes.

Whereas Schmiedek et al. (2007) selected only nine tasks that were all taken from the reasoning and psychometric speed operations, we also used the memory tasks of the short scale BIS (BIS; Jäger et al., 1997), which resulted in a total of 12 tasks originating from three of the four operations tapped in the test (reasoning, psychometric speed, memory, and idea fluency). We excluded the tasks on idea fluency because they are more related to creativity than to the construct of intelligence (cf. Schmitz & Wilhelm, 2016). Consequently, verbal, numeric, and figural domains were represented by four tasks each. To keep the structural equation models as simple as possible, we used scale means as manifest variables for each of the three content domains.

Response-time Tasks

The study consisted of 3 (domain: numeric vs. verbal vs. figural) × 2 (speed: fast vs. slow) × 3 (number of tasks) = 18 different RT tasks (Table 1). In the following, we briefly describe the different tasks and materials.

Numeric Tasks

The *fast numeric tasks* were the number discrimination task, the odd-even task, and the simple inequation task. In the *number discrimination task*, participants saw a number in each trial and had to assess whether this number was smaller or larger than 500. The numbers were randomly drawn from a uniform distribution ranging from 100 to 900 (excluding 500), with the restriction that half of the numbers were larger than 500 and that the mean deviation from 500 was identical for the numbers smaller and the numbers larger than 500. In the *odd-even task*, participants had to assess whether a presented number was odd or even. The numbers were randomly drawn from a uniform distribution ranging from 100 to 899 (i.e., a vector including 400 odd and 400 even numbers). In the *simple inequation task*, participants had to decide which of two numbers displayed left and right of the center of the screen was larger. The two simultaneously presented numbers were randomly drawn from a uniform distribution ranging from 1 to 20, with the restrictions that numbers were never identical and that the difference between the numbers did not exceed 3.

The *slow numeric tasks* were the mean value computation task, the equation task and the complex inequation task. In the *mean value computation task*, 16 numbers were presented on the screen. Participants had to assess whether the mean of these numbers was smaller or larger than 500. The mean of the 16 simultaneously presented numbers of each trial was either 400 or 600, and the numbers were presented at random positions on the screen (overlapping of numbers was prevented). In the *equation task*, in each trial an equation was shown and participants had to assess whether the equation was correct or wrong. In half of the trials, a multiplication or division had to be performed, respectively. The erroneous equations were generated using several different principles. Specifically, for erroneous equations either the tens digit or the ones digit of the solution were set to incorrect values (e.g., $5 \cdot 7 = 25$ or $4 \cdot 12 = 40$, respectively), the operator was wrong (e.g., $11/3 = 33$), or the order of numerator and denominator was reversed (e.g., $8/64 = 8$). In the *complex inequation task*, participants had to decide which solution of two equations displayed on the left and right side of the screen was larger. The equations were sums and differences of two numbers (e.g., “9 – 6” vs. “19 – 17”). The two numbers were drawn randomly from a uniform distribution between 1 and 20, and the solutions of the sums and differences were in that range as well. The operations for the two equations were randomly determined and could be the same or different for the two equations. Furthermore, the difference between the solutions of the two equations was restricted to a maximum of 3.

Verbal Tasks

The *fast verbal tasks* were the word category task, the lexical decision task, and the animacy task. In the *word category task*, in each trial a word was presented and participants had to assess whether the word was an adjective or a noun. All words comprised of six letters and had one or two syllables. The words had frequency classes of 12 or above (according to the online dictionary project of the university of Leipzig, retrieved in May 2017, see http://wortschatz.uni-leipzig.de/de), which indicates that the German word “der” (“the”) is used at least 2^{12} times as often as the selected stimuli. The mean frequency class of adjectives and nouns was identical (M = 15). Thus, all words had a low frequency in German language. In the *lexical decision task*, letter combinations were presented and participants had to assess whether or not these were German words. The stimuli were selected from a lexical decision study by Lerche and Voss (2017b). The words were nouns consisting of one or two syllables and four to six letters. The words had a frequency class of 14 or 15 (retrieved in November 2014). The non-words had been generated by replacement of vowels from valid word. Thus, all non-words were pronounceable and had plausible bigram frequencies. In the *animacy task*, nouns were presented and participants had to classify these as living vs. nonliving. The "living" stimuli could refer to humans, animals or plants. Two of the authors and two further independent raters classified the words unambiguously as living vs. nonliving. The words consisted of one to three syllables, four to eight letters, and had frequency classes between 11 and 16 (retrieved in June 2017). The mean frequency class was identical for words classified as living or nonliving (M = 13).

The *slow verbal tasks* were the grammar task, the statement task, and the semantic category task. In the *grammar task*, participants read German sentences with grammatical errors and had to indicate whether the error was located in the possessive pronoun or in the noun. All sentences consisted of five words and had a very similar structure: They always started with a personal pronoun and further contained a predicate and an object with a possessive pronoun (e.g., “Er widerspricht seine Chef oft.” = “He often contradicts his boss.”; the error in the German statement is in the possessive pronoun that should read “seine*m*” instead of “seine”). In each trial, by changing one word—either the possessive pronoun or the object—the sentence could be corrected. The errors were generated using the wrong case (e.g., accusative instead of dative), the wrong gender, the wrong declension, or the wrong number.

In the *statement task*, four to six words were presented at different positions of the screen. The participants had to assess whether or not it was possible to create a true statement using all of the presented words. The words were distributed randomly across the screen. From each set of words one grammatically correct sentence could be composed. An example for a true statement is “ein Lastwagen ist sehr schwer” (“A truck is very heavy”) and for a wrong statement is “reiche Menschen haben kein Geld” (“Rich people have no money”).

In the *semantic category task*, five nouns were presented one above the other. There was one superordinate category to which most of the words (that is, three or four words) belonged. Either one or two words did not belong to this category. Participants had to indicate whether one or two words did not belong to this superordinate category. The selected words were members of the superordinate categories planets, seating furniture, fruit, tools, baking ingredients, medical specialists, geometric figures, grain, craftsmen, or organs reported by Scheithe and Bäuml (1995). Either three or four words belonged to the same category and one or two belonged to another superordinate category. For example, in one trial the words “Stuhl” (= chair), “Sonne” (= sun), “Sessel” (= armchair), “Sofa” (= sofa), and “Bank” (= bench) were shown. Here, the correct response was 1 because all words except one ("sun") belong to the same superordinate category “seating furniture”. In another example, “Weizen” (= wheat), “Mond” (= Moon), “Jupiter” (= Jupiter), “Merkur” (= Mercury), and “Hirse” (= sorghum) were presented. In this case, the correct response was 2, because two nouns (“wheat” and “sorghum”) do not belong to the dominant category (planets). There are 10 different possibilities for the positioning of two minority category members among the five words and five possibilities for the positioning of one minority category member. Each possible positioning was used equally often.

Figural Tasks

Example illustrations of the figural tasks are depicted in Figure 2. The *fast figural tasks* were the dot-rectangle task, the simple area task, and the polygon task. In the *dot-rectangle task,* a rectangle and a dot were shown. Participants had to indicate whether the dot was located within or outside of the rectangle. The rectangles varied in size while the dot was always of the same size. The form of the rectangle and the exact positioning of the dot were determined randomly. In the *simple area task*, two rectangles were shown side by side. Participants had to assess which of the two rectangles was larger. The edge lengths of the rectangles were determined randomly, with the area of the smaller rectangle always comprising 70% of the area of the larger rectangle. In the *polygon task*, polygons were shown and participants had to indicate whether the stimulus was a triangle or a quadrangle. The shapes of polygons were generated randomly.

The *slow figural tasks* were the maze task, the complex area task, and the pie task. In the *maze task*, mazes were presented with a dot positioned inside the maze. Participants had to assess whether or not it was possible to leave the labyrinth (starting from the position of the dot). The mazes were drawn manually with a graphics program. In the *complex area task* (cf. Lerche & Voss, 2017a), in each trial six rectangles were shown. Three of them had a red border and three of them had a blue border. Participants had to compare the total area of all red-bordered rectangles with the total area of all blue-bordered rectangles and decide which area was larger. The larger area was always 1.3 times larger than the smaller area. The rectangles were generated randomly based on some restrictions (most importantly, the largest or smallest area was not indicative of the correct answer so that participants really had to assess the total area, see Lerche & Voss, 2017a, for details). In the *pie task*, three pie slices were shown in each trial. Participants had to judge whether the three slices—if put together—add up to more or less than a full circle. Between trials, the slices summed up to either 95%

or 105%, and each slice comprised between 5% and 95% of a full circle each. The combinations of slices were generated randomly with the restriction that from the summing of only two slices it was not possible to derive a correct answer.

Data preparation

For all RT tasks, we discarded all responses faster than 300 ms. Furthermore, for each task, trials lying more than three interquartile ranges beneath the first or above the third quartile of the intra-individual logarithmized RT distributions were excluded (see also Tukey, 1977). The percentage of excluded trials was on average 1.3% per task and participant.

One participant interrupted accidentally the experimental program at the beginning of the penultimate task of the session, so that data from two tasks (mean value computation task and dot-rectangle task) are missing for this participant. Furthermore, separately for the different RT tasks, we removed the diffusion model parameter estimates of participants with inadequate model fit (i.e., fit < 1% quantile of the simulated data, see below for details on the assessment of model fit; this resulted in an exclusion of 0.93% of the diffusion model parameter estimates). Next, we also excluded the diffusion model parameter estimates, mean RT and accuracy for a specific person and task if the accuracy rate or mean RT for this specific task and person exceeded the Tukey criterion (i.e., distance from first or third quartile larger than three times the interquartile range; Tukey, 1977)[4]. Finally, based on the estimated diffusion model parameters (v, a, t_0), accuracy rates, mean RTs and intelligence scale scores, we computed the Mahalanobis distances to detect multivariate outliers. Two of our participants exceeded the critical value of $\chi^2 = 140.89$ ($df = 93$, $p = .001$) and thus had to be excluded.

[4] To test the robustness of our main findings, in additional analyses we excluded univariate outliers in the diffusion model parameters (because we had obtained some extreme estimates, e.g., $t_0 \approx 0$, $a \approx 10$, $v > 10$). The pattern of results remained unchanged when we excluded these values.

Parameter Estimation

We estimated the diffusion model parameters using the maximum likelihood optimization criterion implemented in *fast-dm-30* (Voss & Voss, 2007; Voss & Voss, 2008; Voss, Voss, & Lerche, 2015). Parameters were estimated separately for each participant and each task. Thresholds were associated with correct (upper threshold) and erroneous (lower threshold) responses. Accordingly, the starting point was centered between thresholds ($z_r = 0.5$). In addition, we fixed the intertrial variabilities of drift rate and starting point to zero. These two parameters cannot be estimated reliably from low trial numbers and the fixation of these parameters can even improve the estimation of the other model parameters (Lerche & Voss, 2016; see also van Ravenzwaaij, Donkin, & Vandekerckhove, 2016). In sum, for each participant and each task we obtained estimates for threshold separation, drift rate, non-decision time, and the intertrial variability of non-decision time.

In order to examine the robustness of our results, we also conducted three additional types of parameter estimation. In the first, we associated the thresholds with the two response categories of the respective task (instead of correct and erroneous responses) and freely estimated the starting point. This way, we could check if accounting for a possible bias in starting point alters our results. With this estimation approach, we obtained two different drift rate estimates per task, one for each response category, and—after multiplying the drift rate for the category associated with the lower threshold by -1—computed the mean of the two drift rates as an overall estimate of drift per task. In our second additional estimation procedure, we examined whether practice effects might influence our pattern of results. Therefore, prior to parameter estimation, we excluded not only the four practice trials and the warm-up trial of each task, but also the subsequent 20 trials. Finally, we combined the two alternative estimation approaches obtaining parameter estimates with a freely estimated starting point while also excluding the 20 additional practice trials.

Some of the tasks employed in our study were similar to tasks that have already been used for diffusion model analyses: Specifically, lexical decision tasks (e.g., Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008; Yap, Balota, Sibley, & Ratcliff, 2012), number discrimination (Ratcliff, 2014; Ratcliff, Thompson, & McKoon, 2015), odd-even tasks (Schmiedek et al., 2007; Schmitz & Voss, 2012), animacy discrimination tasks (Aschenbrenner et al., 2016; Spaniol, Madden, & Voss, 2006; Voss, Rothermund, Gast, & Wentura, 2013), and the complex area task (Lerche & Voss, 2017a) have been analyzed with the diffusion model before. However, most tasks, in particular the slow RT tasks (with the exception of the complex area task), have not yet been examined by means of diffusion modeling. Thus, we were particularly interested in whether the model can fit data from all tasks (and especially from the slow tasks) reasonably well. Accordingly, we examined the model fit for all tasks (our procedure is reported in the Results section).

Structural Equation Modeling

Our structural equation modeling approach consisted of two main steps. First, we established a measurement model for drift rates and a model of the intelligence test scales, separately. Then, we combined these two models into one complete model. We used the *R* package *lavaan* (Rosseel, 2012) for the structural equation analyses. To deal with missing data we employed the full information maximum likelihood (FIML) estimator included in *lavaan*, which utilizes all available information.

We standardized all observed variables before they were entered into the structural equations to avoid estimation problems resulting from differing variances between the drift rates and the intelligence scale scores. As we were not interested in absolute values, fixing all means to zero is unproblematic. However, the analysis of correlations instead of covariances can lead to biased standard errors and fit indices (Cudeck, 1989). We accounted for this by fixing the model implied indicator variances to one, equal to the manifest indicator variances, as proposed by Cudeck. For examination of model fit we used several fit indices: the χ^2 statistic, the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Tucker-Lewis Index (TLI). We used the cut-off criteria proposed by Hu and Bentler (1999) for evaluation of fit. Please note that due to the use of the FIML estimator, a mean structure was also estimated. We fixed all estimated indicator means to zero (as the variables were standardized), a fact that informs the degrees of freedom for all reported models.

We compared four different measurement models of drift rate. Because it was essential to keep the models as parsimonious as possible, we assumed parallel measurement of all factors by fixing all factor loadings to one and setting all residual variances of items loading onto the same factor equal (see Lord & Novick, 1968, Equations 3.3.1a and 3.3.1b, for the outline of a model of parallel measurement). The four models are shown in Figure 3. The first model (Model 1) assumed a general (*g*) factor of drift rate. This equals the assumption that the common variance in speed of information processing can be explained by a single, general factor contributing to all tasks. Model 2 did not include a *g* factor, but three uncorrelated domain factors. The idea behind this model is that there are different types of speed of information processing for figural, verbal and numeric tasks, and that these are unrelated to one another. In Model 3, we assumed a hierarchical structure of the factors: *g* was modeled as a higher-order factor and the domain factors as lower part of the factor hierarchy. The general factor is here interpreted as the common variance of the domain factors, which—in contrast to Model 2—are thought to be correlated. Thus, Model 3 assumes that speed of information processing has both a general component and domain-specific components[5]. Finally, in Model 4, we fit an extended version of Model 3 adding a factor that captures the specific variance of the slow tasks (M-1 approach; Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). Here, the idea is that speed of information processing in the slower, more complex tasks shares specific common variance. This way, the interpretation of the *g* factor changes: It now comprises the domain-general shared variance of speed of information processing except for the variance solely shared by the slow tasks. As not all of the models are nested, we compare model fit based on AIC and BIC values.

For the BIS intelligence scales, we used a hierarchical model of domains and a superordinate *g* factor (Intelligence Model, see Figure 4). We employed scale means (instead of single item values) as single indicators for each domain (figural, numerical, verbal) to keep the model as simple as possible, fixing residual indicator (not: domain) variances to zero.[6] Domain factor variances were set equal for the three domains. We also fixed the unstandardized loadings of the indicators on *g* and on the domain factors to 1. While this assumption of perfect measurement and parallel structure is certainly an oversimplification, we made this decision because the BIS is an established instrument and the focus of this study is less on the structure of intelligence, but on the structure of speed of information processing and its relationship to intelligence. In the last step, we combined the best fitting model of drift rates and the BIS model (Combined Drift-Intelligence Model).

Although the focus of this work is on drift rate, we also fit the same model structures (Models 1 to 4, see Figure 3) to estimates of threshold separation (a), non-decision time (t_0)

[5] In the literature on the structure of mental abilities, there is an ongoing debate on how hierarchical models compare to so-called bifactor models (see, e.g., Morgan, Hodge, Wells, & Watkins, 2015). The latter assume a structure of both uncorrelated domain factors and a *g* factor, also orthogonal to the other factors. Thus, bifactor models do not make the presumption that the common variance shared by all tasks is due to the variance shared between the domain factors. Empirically, bifactor models often tend to fit better, while at the same time being less understood from a substantive, theoretical perspective (Kan, van der Maas, & Levine, 2019). Bifactor models fit better because with all loadings estimated freely hierarchical models are more constrained: The hierarchical models assume that the proportions of indicator variance accounted for by the domain (residual) factors and the proportions accounted for by *g* are the same for all indicators within a domain (Gignac, 2016). In our modeling approach, we fixed all factor loadings to be equal within each factor, which leads to a case were hierarchical and bifactor models are mathematically equivalent, yielding identical fit indices and estimates of the corresponding variances. We decided to use a hierarchical model instead of a bifactor model because it can be interpreted more intuitively and because it is also the more common model of cognitive abilities found in the literature.

[6] Fixing the indicator variances to zero and using the domain factors as de-facto residuals was necessary to estimate the covariances between the drift domain residuals and the respective intelligence test components.

and mean logarithmized response times of correct responses. If a measurement model with acceptable fit emerged, we further tested the combined model (i.e., including the intelligence model). In the tables and plots, models are labeled accordingly (e.g., Drift Model 1 or RT Model 1). The data of our study is available on the Open Science Framework project page: https://osf.io/xpbwe/.

Results

Tables A1 to A6 in the Appendix report descriptive statistics of response times, accuracy rates, drift rates, threshold separations, non-decision times, and intelligence scores. Figures A1 (fast tasks) and A2 (slow tasks) in the Appendix show boxplots of the response times for all 18 tasks.

Fit of the diffusion model

Our analyses of model fit comprise two different approaches: First, we examined the fit values of the maximum likelihood optimization. For better interpretation of these values, we conducted simulation studies based on the estimated parameters to infer a criterion for the assessment of model fit (Voss, Nagler, et al., 2013). Second, we analyzed model fit by means of graphical illustrations comparing observed and estimated descriptive statistics.

In the maximum likelihood approach, parameter estimation is based on the maximization of the sum of logarithmized densities over all responses. Boxplots illustrating log-likelihood values for all tasks are given in Figure B1 (fast tasks) and Figure B2 (slow tasks) in the Appendix. Higher likelihood values indicate a better fit of data to the model. One problem with the interpretation of the log-likelihood values is that they depend on the parameter ranges of the specific task. For example, the RT distributions of slower tasks are more spread so that the sum of logarithmized densities is smaller (for an example illustration, see Fig. 4 in Lerche & Voss, 2017a). This makes it difficult to compare the performance of tasks with different parameter ranges.

To account for this, we conducted simulation studies. More specifically, for each task, we generated 1,000 random parameter sets from multivariate normal distributions, with means, variances, and covariances based on the distribution of estimated parameters. Thus, simulated parameter sets were similar to observed parameters. From each parameter set, we simulated one random data set (using *construct-samples,* which is part of the program *fast-dm*). Therefore, simulated data reflects the assumption that data is based on a diffusion process. Next, we re-estimated parameters from simulated data using the same *fast-dm* settings as for the analyses of observed data (i.e., same number of estimated and fixed parameters, same optimization criterion). If the fit values for the real data are worse than those of the simulated data, the observed data probably do not result from a diffusion process only, and consequently, results from the diffusion model analyses might be invalid. Importantly, the distributions of log-likelihood values did not differ systematically between observed data and simulated data, suggesting an excellent model fit (see Figures B1 and B2).

We further defined a criterion to quantify the percentage of observed data sets with poor fit. Specifically, we computed the 1% quantile of the distribution of fit values from simulated data. Maximum likelihood values below this criterion are assumed to indicate poor model fit. This criterion is depicted as horizontal line in each plot. In addition, the plots give the percentage of data sets with fit values below this criterion. The percentages of suspicious fits are very low (at maximum 3.2%) and they are equal for the slow and fast tasks (M = 1.1%). This suggests that the diffusion model fits equally well for the fast and slow RT tasks of our study.

We also examined the model fit graphically, in terms of the precision of predictions for accuracy rates and RT quartiles. Specifically, we constructed scatter plots for each type of task (domain × speed) that show the correspondence of different statistics (RT quantiles and accuracy rates) of observed data (x-axis) with the respective values predicted from the diffusion model results (y-axis; see Figures B3 and B4 in the Appendix for the fast and slow tasks, respectively). In these figures, each point represents one participant in one task. The figures illustrate that the diffusion model fit the data very well as for all tasks the points are close to the diagonals (all correlations between the empiric and the respective estimated quartiles were larger than .97). Interestingly, the model fits at least as well for slow as for fast RT tasks. Thus, the graphical fit analyses are in accordance with the simulation-based analyses of maximum likelihood values.

The simulation studies and graphical analyses of model fit for the three alternative types of estimation (including estimates of starting point, excluding additional practice trials, and doing both) yielded similar results. The according plots are in the supplementary online material.

Structural Equation Modeling[7]

We started by fitting the measurement models described above (Models 1 to 4, see Figure 3) to the drift rate estimates: Model 1, a *g* factor model; Model 2, a model of uncorrelated domains; Model 3, a hierarchical model of domains and a *g* factor; and Model 4, a model that further added a method factor for all slow decision tasks. Table 2 shows the fit indices for all drift rate models. Figures C1 to C4 in the Appendix show the results for Drift Models 1 to 4 and Tables C1 to C4 in the Appendix report the parameter estimates for each of the four structural equation models, including the unstandardized solution, the corresponding standard errors and *p* values, and completely standardized estimates.

Model 4, the model containing a hierarchical structure of three content domain factors, a superordinate *g* factor, and a method factor for the slow tasks had the best fit in terms of AIC and BIC values (see Table 2) and also regarding the measures of absolute model fit (χ^2 [df = 184] = 254.40, CFI = .88; TLI = 0.90; RMSEA = 0.06). Accordingly, we

[7] All the structural equation modeling analyses can be examined and replicated by executing the R Markdown file that we provide on the OSF project page.

decided to retain this model. It should be noted that the estimated residual variance of the figural drift factor did not differ significantly from zero und should therefore be interpreted accordingly. We kept it in the model in order to a) refrain from post hoc model adjustments and b) make possible replications easier to compare.

The Intelligence Model is illustrated in Figure C5 in the Appendix, Table 2 shows the fit, and Table C5 in the Appendix the parameter estimates. As the fit was good (χ^2 [*df* = 8] = 0.18, CFI = 1.00; TLI = 1.03; RMSEA = 0.00), we used this model for the combined analyses.

Finally, we combined the best measurement model of drift rates (i.e., Model 4) and the Intelligence Model into a Combined Drift-Intelligence Model. We allowed freely estimated covariances between residual figural drift rate and residual figural BIS intelligence, residual numeric drift rate and residual numeric BIS intelligence, residual verbal drift rate and residual verbal BIS intelligence, and the superordinate *g* factor for drift rate and *g* BIS intelligence.[8] In addition, the covariance between the slow decision task factor and the *g* BIS intelligence factor was freely estimated, reflecting our hypothesis that speed of information processing in slow tasks might be especially closely related to general intelligence. Figure 5 shows the resulting model. Model fit was acceptable (χ^2 [*df* = 241] = 406.49; CFI = .82; TLI = 0.84; RMSEA = 0.07; see Table 2). Table 3 shows the parameter estimates. All latent factors except the figural drift factor had variances significantly different from zero; the same was true for the covariances between them. The relative parts of the variances of the manifest indicators explained by the latent factors are reported in Table 4. Across all tasks, 20% of the variance of drift rates could be attributed to the *g* Drift factor, while 3-16% were based on the

[8] We also fitted a Combined Drift-Intelligence Model freely estimating the covariances between all domain residuals. Only the theoretically implied covariances (Figural Drift <-> Figural IQ, Numeric Drift <-> Numeric IQ, Verbal Drift <-> Verbal IQ) reached statistical significance, except for a negative correlation between verbal drift and figural intelligence (*r* = -.34, *p* = .048).

domain-specific factors. For the complex tasks, an additional 10% of the variance was explained by the slow factor. Overall, the mean task specific and error variance was 63%.

The estimated correlation between figural intelligence and figural drift rate was .90. However, this value should not be over-interpreted because of the very low residual variance of figural drift rate, which did not differ significantly from zero. Numeric intelligence and numeric drift rate correlated with .74. The correlation between verbal intelligence and verbal drift rate was .50, while the correlation between domain general drift rate and general intelligence as measured by the BIS was .45. Finally, the method factor for slow decision tasks and the BIS *g* factor were also strongly correlated (r = .68). If the links of the *g* drift and slow drift factors to *g* BIS intelligence were modeled as a regression, the R^2 value of *g* BIS was .67. Thus, the domain general drift factor and the slow drift factor jointly explained two thirds of the variance in general intelligence.

We conducted several robustness checks to ensure our main findings would hold. First, we fit models with completely freely estimated factor loadings and residual indicator variances for both the best measurement model (Drift Model 4, freely estimated, see Figure C6 and Table C6 in the Appendix; see Table 2 for fit indices) and the Combined Drift-Intelligence Model (freely estimated, see Figure C7 and Table C7 in the Appendix; see Table 2 for fit indices). In terms of AIC and BIC values, the constrained Drift Model 4 was preferred to the freely estimated version. For the Combined Drift-Intelligence Model, AIC was lower for the free model, but the constrained model had the lower BIC value (i.e., better fit). Please note that the number of estimated parameters in the freely estimated models is very large for our sample size and the results should thus be interpreted with caution. In addition, estimation of the Combined Drift-Intelligence Model (freely estimated) yielded a non-positive definite estimated covariance matrix.[9] Still, while the estimated unstandardized factor loadings in the freed models sometimes differed widely from unity and standard errors were much higher than in the constrained model, leading to statistically insignificant estimates, the main resulting covariances remained much the same. Namely, the estimated correlations between the factors in the freely estimated Combined Drift-Intelligence Model (compared to the constrained Combined Drift-Intelligence Model) were: .56 (.90) for the figural, .90 (.74) for the numeric, and .52 (.50) for verbal drift residual factors and their respective intelligence counter-parts. A correlation of .42 (.45) was now found for the relation of *g* Drift and *g* BIS and a correlation of .74 (.68) for the association of the slow factor and *g* BIS.

Further evidence for the robustness of our results was provided by additional analyses based on different specifications of the diffusion models: Similar results emerged for the structural equation models when drift was estimated using the alternative diffusion model architectures that a) also estimated the starting point, b) excluded 20 additional practice trials, or c) did both. Fit indices and parameter estimates for these models are given in the supplementary online material.

Table 5 shows the fit values for the measurement models of threshold separation, non-decision time, and mean logarithmized response times. Parameter estimates for all these models can be found in the supplementary online material. Of all the measurement models, only t_0 Models 1, 3, and 4 showed somewhat acceptable model fit (RMSEA < 0.08, CFI and TLI at least > 0.82), with Model 4 showing the lowest values in AIC and BIC. Thus, for non-decision time, a hierarchical model of domain factors, a superordinate g t_0 factor and a method factor for slow tasks provided the best fit. However, the residual variances for the

[9] This problem could be overcome by fixing the residual variance of the Figural Drift factor, that did not differ significantly from zero, to zero.

figural and numerical domain factors did not reach statistical significance. Table C8 shows the complete parameter estimates for this model. We also fit a combined model of non-decision time and the BIS intelligence scales (Combined t_0-Intelligence Model, see Table 5 for the fit measures). The model structure was identical to the Combined Drift-Intelligence Model. Table 6 shows the resulting estimates. The non-decision time domain factors were negatively correlated to the respective intelligence factor residuals, as were the g_{t0} factor and the slow_{t0} factor to general intelligence.

Notably, none of our predefined models showed acceptable fit to the mean logarithmized response times. However, the relationship between response times and intelligence is of particular theoretical interest because response times are the measures of mental speed used in most previous studies. Therefore, we additionally conducted an exploratory principal components analysis to explore the covariance structure of response times in our sample. A parallel analysis (Horn, 1965) suggested the extraction of one general component that explained 58 % of variance in response time variables. When added to the Intelligence (structural equation) Model as a manifest variable, the component scores explained 65 % of the variance in *g*IQ ($\beta = .80$, $p < .001$; RMSEA = 0.00, CFI and TLI ≥ 1.00 for this model).

Discussion

Our study focused on the relationship between intelligence and drift rate—a measure of speed of information processing estimated in diffusion model analyses (Ratcliff, 1978). In contrast to previous studies that examined such relationships (e.g., Ratcliff et al., 2011; Schmiedek et al., 2007; Schmitz & Wilhelm, 2016; Schubert et al., 2015), we used a much larger set of RT tasks, and these tasks systematically addressed three content domains (verbal, numeric, and figural). More specifically, we employed six tasks for each of the three domains with half of the tasks of each domain being typical fast diffusion model tasks (mean

RT of 660 ms), and the other half being more complex, slower tasks (mean RT of 3320 ms). Thereby, our study is the first diffusion model study on intelligence that includes not only fast but also more complex RT tasks and uses a large number of tasks per content domain. This allowed us to examine three main substantial questions: First, we tested whether we can replicate the relationship between general intelligence and drift rate that has been found in previous diffusion model studies (e.g., Ratcliff et al., 2011; Schmiedek et al., 2007; Schmitz & Wilhelm, 2016; Schubert et al., 2015). Additionally, we also examined relationships of intelligence with mean RT and other diffusion model parameters. Second, we analyzed whether there are domain-specific aspects of speed of information processing and—if so—whether these domain-specific drift rate factors are related to the respective domains of the intelligence test BIS (Berlin Intelligence Structure Test; Jäger et al., 1997).

In addition to these substantial questions, our study also allows the examination of two methodological issues. First, in the last years it has been proposed to use the diffusion model not only for the analysis of differences between groups or conditions (the typical application in most previous studies), but also for the examination of interindividual differences (e.g., Frischkorn & Schubert, 2018; Ratcliff & Childers, 2015; White et al., 2016). Our study is the first to allow a profound analysis of whether there are meaningful interindividual differences in the content-domain specific aspects of drift rates. Second, in the past, the diffusion model was typically only applied to fast RT tasks. Our study allows inferences about whether the diffusion model fits slower, more complex RT tasks similarly well as typical fast RT tasks. Furthermore, we could examine the external validity of drift rate in more complex tasks, analyzing the relationship with intelligence.

Summary of Results

The presented structural equation analyses replicated findings of previous diffusion model studies in that we found a strong relationship between a general drift rate factor and general intelligence as measured by the BIS. As the general latent factor of drift rates in our study captured the shared variance of 18 different tasks, this provides strong support for the hypothesis that speed of information processing is closely linked to general intelligence. Furthermore, for two out of three content domains (verbal and numeric), we found significant domain-specific drift factors, indicating that there are domain-specific interindividual differences in mental speed that can be assessed with a diffusion model analysis. Strikingly, the three domain-specific latent factors accounted for roughly one third of the shared variance between tasks. Moreover, the domain-specific drift factors were closely related to the respective components of the standard intelligence test. Finally, fit of diffusion models was equally good for fast and more complex RT tasks and speed of information processing in the more complex tasks explained additional variance in general intelligence.

Domain-specific speeds of information processing

Our study is the first to reveal domain-specific drift factors, which we further found to be related to the respective domain scores of the intelligence test. The variance proportions explained by the domain-specific drift factors for numeric and verbal drift are substantial (15% and 16%), challenging the view of only one general mental speed factor. Thereby, our study helps to reconcile research on mental speed with the literature that is based on standard intelligence testing. In the latter, a hierarchical structure with both a *g* factor and domain-specific factors is a very common assumption. Previous mental speed studies might have failed to reveal domain-specific factors due to measurement issues. Specifically, studies that did not employ the diffusion model might have examined a measure of mental speed that is confounded by other processes such as encoding speed, motoric speed, or speed-accuracy settings. The diffusion model has the great advantage of providing a more process-pure measure of mental speed. Furthermore, previous studies employing the diffusion model might have failed to find domain-specific drift rates because the number of tasks that had been used for each domain might have been too low.

Diffusion modeling for slower, more complex RT tasks?

In the past, it was assumed that the diffusion model is only applicable to fast RT tasks with mean trial RTs below 1.5 seconds (e.g., Ratcliff, Thapar, et al., 2004). However, first studies support the notion that the model might also be utilized for more complex tasks. Lerche and Voss (2017a) conducted experimental validation studies (also often called “selective influence studies”) based on a complex figural RT task, and Lerche et al. (2018) examined model fit of a complex verbal task. The present study offers a unique possibility to compare model fit between easy and more complex tasks, because participants completed both nine complex tasks and nine fast tasks, which were—beside the differences in cognitive demands—very similar. Thus, we could compare model fit (in statistical terms and graphically) between fast and slow tasks and examine correlations with intelligence. Interestingly, the fit of the diffusion model was as good for the more complex as for the simpler tasks.

Furthermore, in our structural equation modeling analyses, a model that included an additional “slow drift factor” (i.e., a factor on which the drift rates of all slow tasks loaded) fitted data better than models without this factor. Furthermore, this slow drift factor was closely linked to general intelligence (r = .68). The explained variance (R^2) for drift rates from slow tasks was slightly higher than for drift from fast tasks, due to the latent slow factor that explained 10% of their variance. Thus, drift rates in the more complex tasks are closely related to intelligence, which provides evidence for a good criterion validity of drift rates in this kind of tasks.

The complex tasks that we employed in our study apparently differed in their demands in terms of, for example, memory (e.g., high demands in the “complex area task”) or reasoning (e.g., high demands in the “word category task”). We did not manipulate or measure the specific demands in our study. However, it is notable that the diffusion model fit all of our complex tasks very well, thus, fit was independent of the specific task demands. In line with this finding are other recent studies that successfully applied sequential sampling models to tasks with high demands on memory or reasoning. One of them applied the diffusion model to a difficult recognition memory task (Aschenbrenner et al., 2016) and another one applied the linear ballistic accumulator model (Brown & Heathcote, 2008) to an inductive reasoning task (Hawkins, Hayes, & Heit, 2016).

Advantages of the diffusion model

Notably, the slow drift factor and the general drift factor together accounted for an impressive 67% of the variance of general intelligence assessed by the BIS. It is striking that drift rate has such a close relation to intelligence in the present study. In our view, this strong relation—and the advantage of drift rate over mean RT—can be explained by two advantages of the diffusion model.

First, unlike mean RT, the drift provides a common metric that combines both RT and accuracy (Spaniol et al., 2006). Thus, when effects of cognitive ability spread over response latencies and accuracy (i.e., higher ability is negatively related to RT and positively related to accuracy of a task), a common metric is required that captures both effects. This is of special importance, when the main impact of cognitive ability is for one group of participants on speed and for others on accuracy.

Second, the diffusion model makes it possible to disentangle different processes of information processing. Most important, different—and conceptually independent—parameters map speed of information processing, speed-accuracy settings, and non-decision times. For example, participants might be faster or slower, because they are less or more cautious (i.e., error avoiding), respectively. Participants might also differ in the time needed for encoding or motoric responses (i.e., non-decision time parameter). For example, it has been consistently found that older participants are more cautious (i.e., higher threshold separations) and that they have higher non-decision times than younger participants (see Theisen, Lerche, von Krause, & Voss, 2019, for a meta-analysis). This example shows that the validity of pure RT as a measure for mental speed might be problematic (see Coyle, 2017, for a similar argument). In diffusion modeling, the response style (threshold separation) and non-decision time are removed analytically from the index for mental speed (drift). Therefore, drift rate is a more process-pure measure of mental speed than is mean RT, and is thus a better predictor for intelligence.

Are relationships with intelligence specific for drift rate?

Importantly, in our structural equation analyses drift rates showed a clear pattern of correlations with intelligence, distinguishing between domain-general and domain-specific aspects, whereas the structural equation models of mean RT did not have a satisfactory fit. Similarly, previous studies that used chronometric tasks and varied the type of material (numeric, verbal, figural) failed to find clear support for domain-specific factors (Levine et al., 1987; Neubauer & Bucik, 1996). These studies examined behavioral variables which—as outlined in more detail in the previous section—are confounded with other processes involved in task execution such as speed-accuracy settings.

Apart from drift rate, for non-decision time, we also observed relationships with intelligence (fitting the same models as for drift rate resulted in a worse, but still acceptable, model fit). Higher scores in the intelligence test were associated with shorter non-decision times. Also in some previous studies, negative relationships between non-decision time and intelligence have been reported (McKoon & Ratcliff, 2012; Schubert et al., 2015; Schulz-Zhecheva et al., 2016), whereas in other studies no such relationship was found (e.g., Schmiedek et al., 2007; Schmitz & Wilhelm, 2016). Our study—which is based on a large number of RT tasks and might thus allow more solid inferences than previous studies—supports the view that there is also a relationship between non-decision time and intelligence (even though this relationship is smaller than for drift rate).

What does this relationship between intelligence and non-decision time indicate? It suggests that “intelligence” as measured by classical paper-and-pencil based intelligence tests is more than speed of information processing. In fact, as already mentioned previously, not only mean RTs in response time tasks, but also performance in paper-and-pencil-based intelligence tests like the BIS can be influenced by different processes. In intelligence tests, it is difficult to distinguish between the different processes that are involved in task completion, such as decision settings (i.e., whether individuals prefer speed or accuracy), motoric elements (e.g., how fast individuals write down their answers), encoding processes, and speed of information processing.[10] Thus, we suppose that non-decision time is related to the BIS because also the paper-and-pencil-based test measures to a certain extent non-decisional components. The non-decision time parameter of the diffusion model includes time needed for encoding and motoric processes. We hypothesize that the correlations with intelligence are probably mainly based on encoding processes rather than on motoric processes. It seems implausible that for motoric components a model with not only a general factor, but also domain-specific factors and a complex task factor emerges. In line with this argument, when the *Jensen box* is used—which allows a separation of the time needed for decision making (termed RT) from the time needed for finger movement (movement time)—RTs clearly increase with increasing task complexity, whereas movement times do not (Jensen, 1987; 2006; see also the Differential–Developmental Model by Coyle, 2017). It is, however, highly plausible that encoding processes differ between domains. Furthermore, the complex task

[10] One notable exception is the explanatory model for performance in the Raven matrices by Carpenter, Just, and Shell (1990), in which different processes (incremental encoding, rule induction, goal management) were identified that contributed to the solution of the matrices. However, its application remains limited and its focus on Raven matrices forbids the generalization to other types of intelligence tests.

factor could be attributed to the fact that the stimuli in the more complex tasks consisted of more elements than the stimuli in the fast tasks (e.g., several numbers distributed over the screen in the mean value computation task in contrast to a single number presented in the center of the screen in the number discrimination task). Accordingly, more complex tasks pose higher demands on encoding than easier tasks. Importantly, by means of diffusion modeling, we get a purer measure of speed of information processing with the time needed for encoding and motoric components partialled out.

Limitations and directions for future research

We want to make clear that we do not claim that mental speed is causally related to intelligence. In fact, a recent study based on an experimental approach did not find support for a causal link between mental speed (as measured by the drift rate of the diffusion model) and intelligence (Schubert, Hagemann, Frischkorn, & Herpertz, 2018). Rather, the authors suggest that structural properties of the brain may give rise to the association between mental speed and intelligence. The aim of our project was not to make any inferences regarding the question of causality.

Diffusion modeling allows for an examination of interesting research questions surrounding the *g* factor and other intelligence-related phenomena. One of these questions, which we addressed in our study, is the examination of whether there are domain-specific mental speeds. However, there are certainly further interesting research questions that could be examined by means of diffusion modeling in the future, for example the factor differentiation finding (e.g., Detterman & Daniel, 1989), which is regarded as one main feature of *g* (Kovacs & Conway, 2016).

Apart from the examination of further intelligence-related phenomena, it would also be important to explore relationships between drift rate and external criteria (e.g., grades at school/university, or job performance). Presently, we have no data on the predictive validity of drift rates for success in life; however, we think that future studies investigating this issue are important. Because our analyses revealed that in particular drift rate in more complex RT tasks showed strong relationships with intelligence, future research might focus on these more complex tasks.

In future studies, one might also examine whether the results that we observed in our study are moderated by the number of trials used in the RT tasks. Several diffusion model studies found that drift rate grows over time (Dutilh et al., 2009; Lerche & Voss, 2017b; Petrov, Van Horn, & Ratcliff, 2011). Possibly, the 100 trials per task used in our study still give room for learning effects and relationships with intelligence might be even stronger or possibly smaller if higher trial numbers were employed, so that more trials could be discarded as practice trials.[11] A higher trial number would also increase reliability of estimates for drift (Lerche & Voss, 2017b; Lerche et al., 2017). Further, in future studies we advise to use higher numbers of participants. The sample size of our study was relatively small for the application of structural equation modeling, leading to the use of very parsimonious parallel measurement models to ensure model convergence.

One aspect that is common to both the assessment of intelligence with the BIS and our computerized RT tasks (both “fast” and “slow” tasks) is the focus on speed. Chuderski (2013) showed that this focus on speed can have an important impact. He found that working memory capacity and fluid intelligence are isomorphic constructs when both are measured under time pressure. If, on the other hand, fluid intelligence is measured with no real time pressure, the relationship with working memory capacity decreases. The findings from the study by Chuderski (2013) suggest that relationships between drift rate in speeded RT tasks and intelligence measured under unspeeded conditions will probably be lower than the

[11] Notably, our additional analyses in which we estimated parameters after exclusion of a larger number of practice trials did not result in a different pattern of results.

relationships we observed in our study which focused on speed. However, the difference in relationships between drift rate and speeded vs. unspeeded intelligence tests would possibly be smaller than the differences between working memory capacity and speeded vs. unspeeded fluid intelligence as measured by Chuderski, because the isomorphic relation between working memory and fluid intelligence both assessed under speeded conditions might be partly attributable to non-decision time (e.g., speed of encoding). If the diffusion model is used, such influences can be “partialled out” so that we expect more similar relationships between speeded vs. unspeeded intelligence testing and our performance measure (drift rate). It would be interesting to examine the size of the relationship between drift rate and unspeeded vs. speeded intelligence testing in future research and compare it to the effect sizes found by Chuderski.

One final aspect that we want to point out is that our findings do not lend support to an application of the diffusion model to all kinds of more complex, slower RT task. In tasks that require significantly more time than the approximately three seconds observed in our study, it becomes more likely that central assumptions of the diffusion model are seriously violated. In future studies it would be interesting to analyze tasks with substantially longer RTs (e.g., a matrices task with a mean RT of more than a minute; Partchev & De Boeck, 2012). Probably more important than the mean RT of a task are characteristics of the specific task. Even fast tasks can be poor candidates for diffusion modeling (e.g., because no continuous information uptake takes place). At the same time, even highly complex tasks that consist of many sub-tasks might be compatible with the diffusion model. In our study, the diffusion model provided a good fit for all employed tasks, and the relationships with intelligence speak in favor of the validity of the parameter drift rate. These tasks are interesting candidates for future diffusion model studies. If, however, researchers are interested in applying the diffusion model to any new tasks, these tasks (whether fast or slow) need to be carefully tested in terms of model fit and—even better—additionally with validation studies.

Conclusions

Prior research revealed relationships between general intelligence and the drift parameter of the diffusion model. This pattern proved to be robust in our structural equation modeling of a set of 18 binary RT tasks. Additionally, we expanded this research showing that there are content-domain specific (verbal, numeric, figural) aspects of cognitive speed, which are related to the respective components of a standard intelligence test. Moreover, slower, more complex tasks also proved to be closely linked to intelligence. Finally, we supply several more complex binary RT tasks that were fit well by the diffusion model and could thus be employed in future research projects.

Context of the Research

This research project is a cooperation of researchers from the departments of Quantitative Research Methods (VL, MVK, and AV) and Personality Research (GTF, ALS, and DH) of the Psychological Institute of Ruprecht-Karls-Universität Heidelberg. In this project, we could nicely combine the main expertise of the two labs, that is, diffusion modeling and intelligence research. In the preceding years, VL and AV have been contacted repeatedly by researchers who asked whether they could use the diffusion model also for more complex RT tasks. VL and AV conducted studies that provide first support for an extension to more complex tasks. Thereby arose the idea for a larger project, which includes numerous both fast and more complex RT tasks. GTF, ALS, and DH were always wondering whether there are domain-specific speeds of information processing but—because they usually additionally collect EEG data—they so far had refrained from running a study with such a large number of different RT tasks (N = 18). MVK is a PhD student who joined the team at the beginning of the recruitment for the study and has taken over an important role in the running of the study and the data analyses. He is currently examining the data further, focusing on age effects. One future research project will be the examination of relationships between drift rate in more complex tasks and external measures of performance (e.g., job performance).

References

Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research, 79*(5), 882-898. doi: 10.1007/s00426-014-0608-y

Aschenbrenner, A. J., Balota, D. A., Gordon, B. A., Ratcliff, R., & Morris, J. C. (2016). A diffusion model analysis of episodic recognition in preclinical individuals with a family history for Alzheimer's disease: The adult children study. *Neuropsychology, 30*(2), 225-238. doi: 10.1037/neu0000222

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: General, 133*(1), 83-100. doi: 10.1037/0096-3445.133.1.83

Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review, 71*, 117-120.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153-178. doi: 10.1016/j.cogpsych.2007.12.002

Camos, V., & Barrouillet, P. (2014). Attentional and non-attentional systems in the maintenance of verbal information in working memory: The executive and phonological loops. *Frontiers in Human Neuroscience, 8*. doi: 10.3389/fnhum.2014.00900

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*(3), 404-431. doi: 10.1037/0033-295X.97.3.404

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY, US: Cambridge University Press.

Cattell, R. B., & Cattell, A. (1960). *Measuring intelligence with the culture fair tests*. Champaign, IL, USA: Institute for Personality and Ability Testing.

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence, 41*(4), 244-262. doi: 10.1016/j.intell.2013.04.003

Conway, A. R. A., & Kovacs, K. (2015). New and emerging models of human intelligence. *WIREs Cognitive Science, 6*(5), 419-426. doi: 10.1002/wcs.1356

Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence, 31*(6), 567-587. doi: 10.1016/S0160-2896(03)00054-0

Coyle, T. R. (2017). A Differential–Developmental Model (DDM): Mental Speed, Attention Lapses, and General Intelligence (g). *Journal of Intelligence*, *5*(2), 25.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological bulletin, 105*(2), 317-327. doi: 10.1037/0033-2909.105.2.317

Deary, I. J., Lawn, M., & Bartholomew, D. J. (2008). A conversation between Charles Spearman, Godfrey Thomson, and Edward L Thorndike: The International Examinations Inquiry Meetings 1931-1938. *History of Psychology, 11*(2), 122-142. doi: 10.1037/1093-4510.11.2.122

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, *13*(4), 349-359. doi: https://doi.org/10.1016/S0160-2896(89)80007-8

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review, 16*(6), 1026-1036. doi: 10.3758/16.6.1026

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods, 8*(1), 38-60. doi: 10.1037/1082-989X.8.1.38

Frischkorn, G. T., & Schubert, A.-L. (2018). Cognitive Models in Intelligence Research: Advantages and Recommendations for Their Application. *Journal of Intelligence, 6*(3), 34.

Frischkorn, G. T., Schubert, A.-L., Neubauer, A. B., & Hagemann, D. (2016). The Worst Performance Rule as Moderation: New Methods for Worst Performance Analysis. *Journal of Intelligence, 4*(3), 9.

Germar, M., Schlemmer, A., Krug, K., Voss, A., & Mojzisch, A. (2014). Social influence and perceptual decision making: A diffusion model analysis. *Personality and Social Psychology Bulletin, 40*(2), 217-231. doi: 10.1177/0146167213508985

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55*, 57-68. doi: 10.1016/j.intell.2016.01.006

Hawkins, G. E., Hayes, B. K., & Heit, E. (2016). A dynamic model of reasoning and memory. *Journal of Experimental Psychology: General, 145*(2), 155-180. doi: 10.1037/xge0000113

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *psychometrika, 30*(2), 179-185. doi: 10.1007/bf02289447

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. doi: 10.1080/10705519909540118

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterenwicklung eines deskriptiven Intelligenzstrukturmodells. = Multimodal classification of intelligence achievement: Experimentally controlled, further development of a descriptive intelligence structure model. *Diagnostica, 28*(3), 195-225.

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test : BIS-Test, Form 4*. Göttingen [u.a.]: Hogrefe.

Jensen, A. R. (1987). Process differences and individual differences in some cognitive tasks. *Intelligence, 11*(2), 107-136. doi: 10.1016/0160-2896(87)90001-8

Jensen, A. R. (2005). Mental chronometry and the unification of differential psychology. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence: Identifying the mechanisms of the mind.* (pp. 26-50). New York, NY: Cambridge University Press.

Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences*. Amsterdam: Elsevier.

Kan, K.-J., van der Maas, H. L. J., & Kievit, R. A. (2016). Process overlap theory: Strengths, limitations, and challenges. *Psychological inquiry, 27*(3), 220-228. doi: 10.1080/1047840X.2016.1182000

Kan, K.-J., van der Maas, H. L. J., & Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against g in favor of mutualism? *Intelligence, 73*, 52-62. doi: 10.1016/j.intell.2018.12.004

Kim, K. H. (2005). The Relation Among Fit Indexes, Power, and Sample Size in Structural Equation Modeling. *Structural Equation Modeling, 12*(3), 368-390. doi: 10.1207/s15328007sem1203_2

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological inquiry, 27*(3), 151-177. doi: 10.1080/1047840X.2016.1153946

Larsen, L., Hartmann, P., & Nyborg, H. (2008). The stability of general intelligence from early adulthood to middle-age. *Intelligence, 36*(1), 29-34. doi: 10.1016/j.intell.2007.01.001

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence*, *14*(3), 309-325. doi: https://doi.org/10.1016/0160-2896(90)90021-K

Lerche, V., Christmann, U., & Voss, A. (2018). Impact of context information on metaphor elaboration: A diffusion model study. *Experimental Psychology*, *65*(6), 370-384. doi: 10.1027/1618-3169/a000422

Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology*, *7*(1324). doi: 10.3389/fpsyg.2016.01324

Lerche, V., & Voss, A. (2017a). Experimental Validation of the Diffusion Model based on a Slow Response Time Paradigm. *Psychological Research*. doi: 10.1007/s00426-017-0945-8

Lerche, V., & Voss, A. (2017b). Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological Research, 81*(3), 629-652. doi: 10.1007/s00426-016-0770-5

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods, 49*(2), 513-537. doi: 10.3758/s13428-016-0740-2

Levine, G., Preddy, D., & Thorndike, R. L. (1987). Speed of information processing and level of cognitive ability. *Personality and individual differences, 8*(5), 599-607. doi: 10.1016/0191-8869(87)90057-2

Lindley, R. H., Wilson, S. M., Smith, W. R., & Bathurst, K. (1995). Reaction time (RT) and IQ: Shape of the task complexity function. *Personality and individual differences, 18*(3), 339-345. doi: 10.1016/0191-8869(94)00154-K

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*(2), 107-127. doi: 10.1016/0160-2896(83)90023-5

McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in item recognition. *Journal of Memory and Language, 66*(3), 416-437. doi: 10.1016/j.jml.2011.12.001

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are Fit Indices Biased in Favor of Bi-Factor Models in Cognitive Ability Research?: A Comparison of Fit in Correlated Factors, Higher-Order, and Bi-Factor Models via Monte Carlo Simulations. *Journal of Intelligence, 3*(1), 2-20.

Neubauer, A. C., & Bucik, V. (1996). The mental speed—IQ relationship: unitary or modular? *Intelligence, 22*(1), 23-48. doi: https://doi.org/10.1016/S0160-2896(96)90019-7

Partchev, I., & De Boeck, P. (2012). Can fast and slow intelligence be differentiated? *Intelligence, 40*(1), 23-32. doi: 10.1016/j.intell.2011.11.002

Petrov, A. A., Van Horn, N. M., & Ratcliff, R. (2011). Dissociable perceptual-learning mechanisms revealed by diffusion-model analysis. *Psychonomic Bulletin & Review, 18*(3), 490-497. doi: 10.3758/s13423-011-0079-8

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59-108. doi: 10.1037/0033-295x.85.2.59

Ratcliff, R. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 870-888. doi: 10.1037/a0034954

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision, 2*(4), 237-279. doi: 10.1037/dec0000030

Ratcliff, R., & Frank, M. J. (2012). Reinforcement-Based Decision Making in Corticostriatal Circuits: Mutual Constraints by Neurocomputational and Diffusion Models. *Neural Computation, 24*(5), 1186-1229. doi: 10.1162/NECO_a_00270

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review, 111*(1), 159-182. doi: 10.1037/0033-295x.111.1.159

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873-922. doi: 10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347-356. doi: 10.1111/1467-9280.00067

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. doi: 10.1016/j.tics.2016.01.007

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging, 19*(2), 278. doi: 10.1037/0882-7974.19.2.278

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology, 60*(3), 127-157. doi: 10.1016/j.cogpsych.2009.09.001

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General, 140*(3), 464-487. doi: 10.1037/a0023810

Ratcliff, R., Thompson, C. A., & McKoon, G. (2015). Modeling individual differences in response time and accuracy in numeracy. *Cognition, 137*, 115-136. doi: 10.1016/j.cognition.2014.12.004

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*(3), 438-481. doi: 10.3758/bf03196302

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *2012, 48*(2), 36. doi: 10.18637/jss.v048.i02

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428. doi: 10.1037/0033-295X.103.3.403

Scheithe, K., & Bäuml, K.-H. (1995). Deutschsprachige Normen für Vertreter von 48 Kategorien. = German-language norms for representatives of 48 conceptual categories. *Sprache & Kognition, 14*(1), 39-43.

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), 414-429. doi: 10.1037/0096-3445.136.3.414

Schmitz, F., & Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 222-250. doi: 10.1037/a0026003

Schmitz, F., & Wilhelm, O. (2016). Modeling Mental Speed: Decomposing Response Time Distributions in Elementary Cognitive Tasks and Correlations with Working Memory Capacity and Fluid Intelligence. *Journal of Intelligence, 4*(4), 13.

Schubert, A.-L. (2019). A meta-analysis of the worst performance rule. *Intelligence, 73*, 88-100. doi: 10.1016/j.intell.2019.02.003

Schubert, A.-L., Frischkorn, G. T., Hagemann, D., & Voss, A. (2016). Trait Characteristics of Diffusion Model Parameters. *Journal of Intelligence, 4*(3), 7. doi: 10.3390/jintelligence4030007

Schubert, A.-L., Hagemann, D., Frischkorn, G. T., & Herpertz, S. C. (2018). Faster, but not smarter: An experimental analysis of the relationship between mental speed and mental abilities. *Intelligence, 71*, 66-75. doi: https://doi.org/10.1016/j.intell.2018.10.005

Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A., & Bergmann, K. (2015). Decomposing the relationship between mental speed and mental abilities. *Intelligence, 51*, 28-46. doi: 10.1016/j.intell.2015.05.002

Schubert, A.-L., & Rey-Mermet, A. (2019). Does Process Overlap Theory Replace the Issues of General Intelligence with the Issues of Attentional Control? *Journal of Applied Research in Memory and Cognition, 8*(3), 277-283. doi: https://doi.org/10.1016/j.jarmac.2019.06.004

Schulz-Zhecheva, Y., Voelkle, M., Beauducel, A., Biscaldi, M., & Klein, C. (2016). Predicting Fluid Intelligence by Components of Reaction Time Distributions from

Simple Choice Reaction Time Tasks. *Journal of Intelligence, 4*(3), 8. doi: 10.3390/jintelligence4030008

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and individual differences, 44*(3), 535-551. doi: 10.1016/j.paid.2007.09.015

Spaniol, J., Madden, D. J., & Voss, A. (2006). A Diffusion Model Analysis of Adult Age Differences in Episodic and Semantic Long-Term Memory Retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 101-117. doi: 10.1037/0278-7393.32.1.101

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2019). Age differences in diffusion model parameters: A meta-analysis. *Manuscript submitted for publication*.

Thomson, G. H. (1916). A Hierarchy without a General Factor. *British Journal of Psychology, 8*(3), 271.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading: Addison-Wesley.

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2016). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review*, 1-10. doi: 10.3758/s13423-016-1081-y

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review, 24*(2), 547-556. doi: 10.3758/s13423-016-1081-y

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology, 60*(6), 385-402. doi: 10.1027/1618-3169/a000218

Voss, A., Rothermund, K., Gast, A., & Wentura, D. (2013). Cognitive processes in associative and categorical priming: a diffusion model analysis. *J Exp Psychol Gen, 142*(2), 536-559. doi: 10.1037/a0029459

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32*(7), 1206-1220. doi: 10.3758/BF03196893

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods, 39*(4), 767-775. doi: 10.3758/bf03192967

Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology, 52*(1), 1-9. doi: 10.1016/j.jmp.2007.09.005

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: a tutorial based on fast-dm-30. *Frontiers in Psychology, 6*(336). doi: 10.3389/fpsyg.2015.00336

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21*(5), 641-671. doi: 10.1080/09541440802205067

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58*(1), 140-159. doi: 10.1016/j.jml.2007.04.006

Weiss, R. H. (2006). *Grundintelligenztest Skala 2-Revidierte Fassung (CFT 20-R)[GeneralIntelligence Test Scale 2-Revised]*. Goettingen: Hogrefe.

White, C. N., Curl, R. A., & Sloane, J. F. (2016). Using Decision Models to Enhance Investigations of Individual Differences in Cognitive Neuroscience. *Frontiers in Psychology, 7*(81). doi: 10.3389/fpsyg.2016.00081

Wilhelm, O., Schroeders, U., & Schipolowski, S. (2014). *Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 8. bis 10. Jahrgangsstufe (BEFKI 8-10)*. Göttingen: Hogrefe.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 53-79. doi: 10.1037/a0024177

Table 1

Overview of the 3 (domain: numeric vs. verbal vs. figural) × 2 (speed: fast vs. slow) × 3 (number of tasks) = 18 RT tasks

	fast	**Slow**
numeric	– FN1: number discrimination task (2.2) number is greater vs. smaller than 500	– SN1: mean value computation task (1.8) 16 numbers with mean greater vs. smaller than 500
	– FN2: odd-even task (1.5) number is odd vs. even	– SN2: equation task (2.5) equation is correct vs. wrong
	– FN3: simple inequation task (2.8) inequation is correct vs. wrong	– SN3: complex inequation task (1.2) equation on left or right side is larger
verbal	– FV1: word category task (2.6) word is adjective vs. noun	– SV1: grammar task (1.4) sentence with grammatical error in possessive pronoun vs. noun
	– FV2: lexical decision task (1.1) letter combination is word vs. non-word	– SV2: statement task (2.3) statement is correct vs. wrong
	– FV3: animacy task (1.7) noun is living vs. nonliving	– SV3: semantic category task (2.9) several nouns with one vs. two nouns not belonging to the superordinate category
figural	– FF1: dot-rectangle task (1.9) dot within vs. outside of rectangle	– SF1: maze task (2.1) maze solvable vs. insolvable
	– FF2: simple area task (2.4) rectangles with larger area on the left vs. right side	– SF2: complex area task (1.6) six rectangles with larger total area of red vs. blue bordered rectangles
	– FF3: polygon task (1.3) polygon is triangle vs. rectangle	– SF3: pie task (2.7) three pie slices making more vs. less of a total pie

Note. The first letter indicates the task complexity (F = fast, S = slow); the second letter denotes the domain (N = numeric, V = verbal, F = figural). The numbers in parentheses indicate the time point of assessment (session and number in sequence).

Table 2

Fit indices of Drift Rate Models, Intelligence Model, and Combined Drift-Intelligence Model

Model	AIC	BIC	χ^2	df	CFI	TLI	RMSEA
Drift Model 1	5,773.69	5,776.50	350.71	188	.73	0.78	0.08
Drift Model 2	5,795.32	5,803.75	368.34	186	.69	0.75	0.09
Drift Model 3	5,711.05	5,722.30	282.07	185	.84	0.86	0.07
Drift Model 4	**5,685.38**	**5,699.44**	254.40	184	**.88**	**0.90**	**0.06**
Drift Model 4, freely estimated	5,688.59	5,772.96	207.61	159	.92	0.92	0.05
Intelligence Model	945.39	948.21	0.18	8	1.00	1.03	0.00
Combined Drift-Intelligence Model	6,507.19	6,538.12	406.49	241	.82	0.84	0.07
Combined Drift-Intelligence Model, freely estimated	6,496.67	6,603.53	341.97	214	.86	0.86	0.07

Note. Model 1: *g* factor model; Model 2: model of uncorrelated domains; Model 3: hierarchical model of domains and a *g* factor; Model 4: Model 3 with additional method factor for all slow decision tasks. AIC = Akaike's Information Criterion. BIC = Bayesian Information Criterion. CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. RMSEA = Root Mean Squared Error Of Approximation. The best-fitting drift rate model among the four alternative models (Models 1 to 4) is highlighted. In the freely estimated models, all loadings and residual variances were unconstrained.

Table 3
Combined Drift-Intelligence Model

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
			Loadings		
Fν on ν (each figural task)	1	0			0.487
Nν on ν (each numeric task)	1	0			0.603
Vν on ν (each verbal task)	1	0			0.591
sν on ν (each slow task)	1	0			0.322
gν on Fν	1	0			0.919
gν on Nν	1	0			0.742
gν on Vν	1	0			0.758
gIQ on F_Mean/on N_Mean/V_Mean	1	0			0.734
FIQ on F_Mean/NIQ on N_Mean/VIQ on V_Mean	1	0			0.679
			Covariances		
gν with gIQ	0.148	0.035	[0.080; 0.216]	<.001	0.450
sν with gIQ	0.162	0.030	[0.102; 0.222]	<.001	0.684
Fν with FIQ	0.117	0.031	[0.057; 0.177]	<.001	0.899
Nν with NIQ	0.202	0.035	[0.134; 0.269]	<.001	0.736
Vν with VIQ	0.130	0.034	[0.063; 0.197]	<.001	0.497
			Latent (Residual) Variances		
gν	0.200	0.025	[0.152; 0.248]	<.001	1
gIQ	0.539	0.039	[0.462; 0.617]	<.001	1
sν	0.104	0.023	[0.059; 0.149]	<.001	1
Fν	0.037	0.028	[-0.017; 0.091]	.182	0.156
Nν	0.163	0.032	[0.100; 0.227]	<.001	0.449
Vν	0.149	0.031	[0.089; 0.209]	<.001	0.426
FIQ/NIQ/VIQ	0.461	0.039	[0.383; 0.538]	<.001	0.461
			Residual Indicator Variances		
ν (each fast figural task)	0.763	0.033	[0.698; 0.827]	<.001	0.763
ν (each fast numeric task)	0.637	0.031	[0.576; 0.697]	<.001	0.637
ν (each fast verbal task)	0.651	0.032	[0.589; 0.713]	<.001	0.651
ν (each slow figural task)	0.659	0.034	[0.593; 0.725]	<.001	0.659
ν (each slow numeric task)	0.533	0.034	[0.467; 0.599]	<.001	0.533
ν (each slow verbal task)	0.547	0.032	[0.486; 0.609]	<.001	0.547

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized.

Table 4

Percentage of variance explained by latent variables in manifest indicators in Combined Drift-Intelligence Model

Task type	*g* Factor	Slow factor	Domain Factor	Residual
Fast Figural	20.03		3.70	76.27
Slow Figural	20.03	10.37	3.70	65.90
Fast Numeric	20.03		16.30	63.67
Slow Numeric	20.03	10.37	16.30	53.29
Fast Verbal	20.03		14.85	65.12
Slow Verbal	20.03	10.37	14.85	54.75

Table 5

Fit indices of threshold separation (a), non-decision time (t_0) and RT models

Model	AIC	BIC	χ^2	*df*	CFI	TLI	RMSEA
a Model 1	5,594.45	5,597.26	485.09	188	.67	0.73	0.11
a Model 2	5,813.55	5,821.99	700.20	186	.43	0.53	0.15
a Model 3	5,597.19	5,608.44	481.84	185	.67	0.73	0.11
a Model 4	**5,502.78**	**5,516.84**	385.42	184	**.78**	**0.82**	**0.09**
t_0 Model 1	5,610.96	5,613.77	316.75	188	.82	0.86	**0.07**
t_0 Model 2	5,791.36	5,799.80	493.15	186	.58	0.65	0.12
t_0 Model 3	5,607.52	5,618.77	307.31	185	.83	0.86	**0.07**
t_0 Model 4	**5,587.65**	**5,601.71**	285.44	184	**.86**	**0.88**	**0.07**
Combined t_0-Intelligence Model	6,457.09	6,488.03	390.73	241	.84	0.86	0.07
RT Model 1	4,887.05	4,889.87	801.89	188	.70	0.75	0.16
RT Model 2	5,076.96	5,085.40	987.80	186	.60	0.67	0.19
RT Model 3	4,794.67	4,805.92	703.50	185	.74	0.79	**0.15**
RT Model 4	**4,760.91**	**4,774.97**	667.75	184	**.76**	**0.80**	**0.15**

Note. Model 1: *g* factor model; Model 2: model of uncorrelated domains; Model 3: hierarchical model of domains and a *g* factor; Model 4: Model 3 with additional method factor for all slow decision tasks. AIC = Akaike's Information Criterion. BIC = Bayesian Information Criterion. CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. RMSEA = Root Mean Squared Error Of Approximation. The best-fitting model among the four alternative models (Models 1 to 4) is always highlighted.

Table 6

Combined t_0-Intelligence Model

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
			Loadings		
Ft_0 on t_0 (each figural task)	1	0			0.540
Nt_0 on t_0 (each numeric task)	1	0			0.579
Vt_0 on t_0 (each verbal task)	1	0			0.614
st_0 on t_0 (each slow task)	1	0			0.275
gt_0 on Ft_0	1	0			1.016
gt_0 on Nt_0	1	0			0.948
gt_0 on Vt_0	1	0			0.894
gIQ on F_Mean/N_Mean/V_Mean	1	0			0.731
VIQ on V_Mean/NIQ on N_Mean/FIQ on F_Mean	1	0			0.682
			Covariances		
gt_0 with gIQ	-0.266	0.031	[-0.327; -0.206]	<.001	-0.663
st_0 with gIQ	-0.023	0.025	[-0.071; 0.026]	.358	-0.112
Ft_0 with FIQ	-0.047	0.027	[-0.101; 0.007]	.086	-0.709
Nt_0 with NIQ	-0.103	0.030	[-0.161; -0.045]	.001	-0.819
Vt_0 with VIQ	-0.113	0.032	[-0.176; -0.051]	<.001	-0.604
			Latent (Residual) Variances		
gt_0	0.301	0.021	[0.260; 0.343]	<.001	1
gIQ	0.535	0.041	[0.455; 0.615]	<.001	1
st_0	0.076	0.019	[0.039; 0.113]	<.001	1
Ft_0	-0.010	0.022	[-0.052; 0.033]	.657	-0.033
Nt_0	0.034	0.023	[-0.012; 0.080]	.146	0.101
Vt_0	0.076	0.026	[0.025; 0.127]	.003	0.201
FIQ/NIQ/VIQ	0.465	0.041	[0.385; 0.545]	<.001	1
			Residual Indicator Variances		
t_0 (each fast figural task)	0.708	0.029	[0.651; 0.765]	<.001	0.708
t_0 (each fast numeric task)	0.665	0.029	[0.609; 0.721]	<.001	0.665
t_0 (each fast verbal task)	0.623	0.028	[0.567; 0.678]	<.001	0.623
t_0 (each slow figural task)	0.633	0.030	[0.574; 0.691]	<.001	0.633
t_0 (each slow numeric task)	0.589	0.030	[0.529; 0.649]	<.001	0.589
t_0 (each slow verbal task)	0.547	0.031	[0.486; 0.608]	<.001	0.547
F_Mean/N_Mean/V_Mean	0	0			

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized.

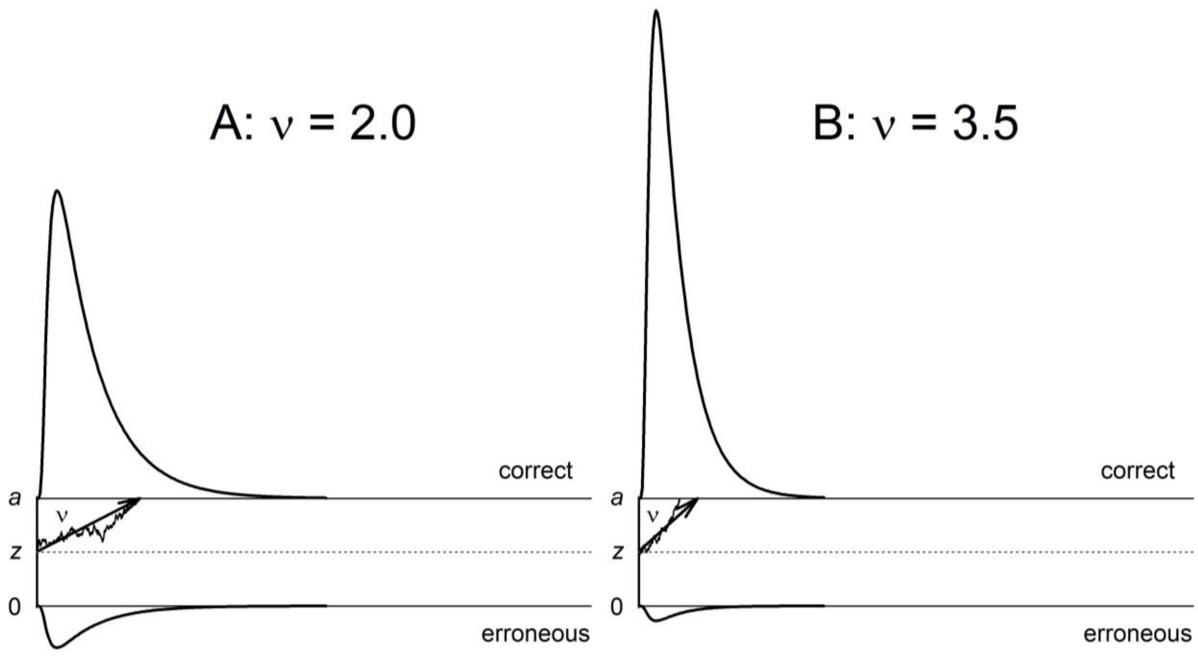

Figure 1. Illustration of the diffusion model. The most important model parameters are threshold separation (*a*), starting point *z* (here situated at the center between the two thresholds), non-decision time (t_0, not depicted in the figure) and drift rate *v*. In Panel B, drift (v = 3.5) is higher than in Panel A (v = 2.0), which results in more accurate and faster responses.

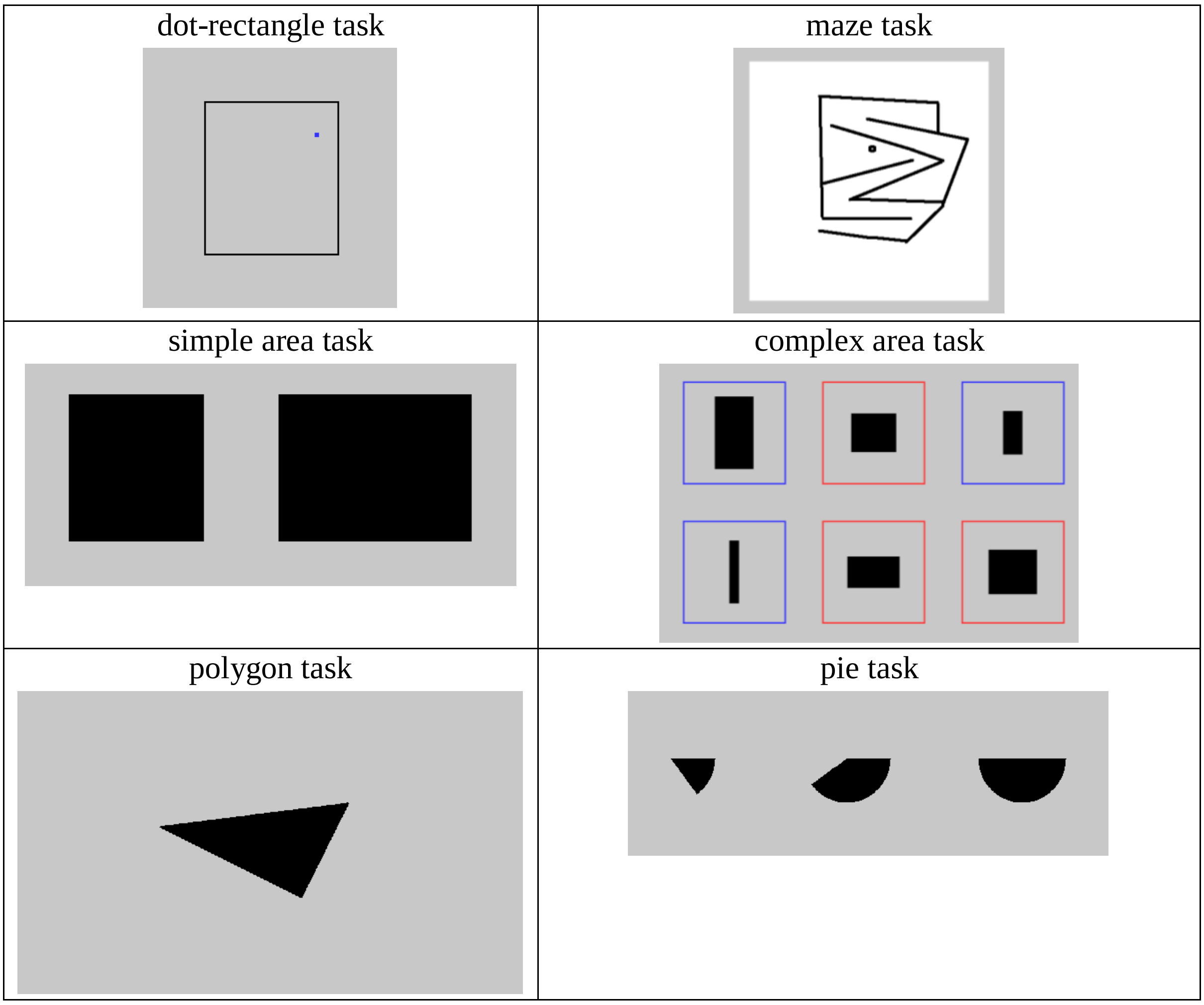

Figure 2. Example for stimuli from the fast figural tasks (left) and the slow figural tasks (right).

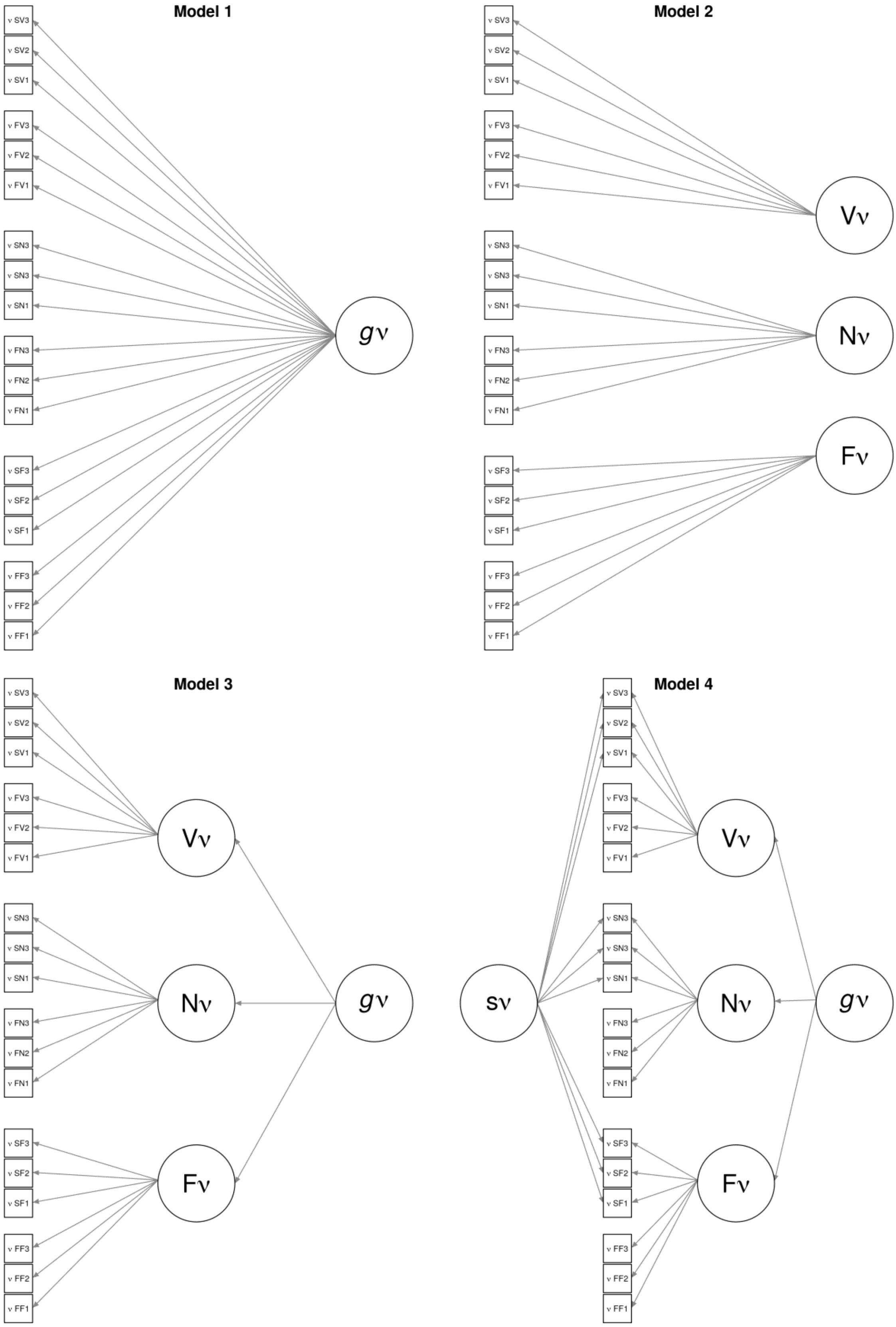

Figure 3. Drift Rate Models 1 to 4. The first letter of the task indices denotes the type of task (F = fast, S = slow); the second letter indicates the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. *g*ν = general drift rate factor, Vν = verbal drift rate factor, Nν = numeric drift rate factor, Fν = figural drift rate factor, sν = method factor for drift rate in slow tasks. All unstandardized factor loadings are fixed to 1. Residuals are omitted from the plot for simplicity. We used the same model structures also for threshold separation, non-decision time and mean logarithmized response times.

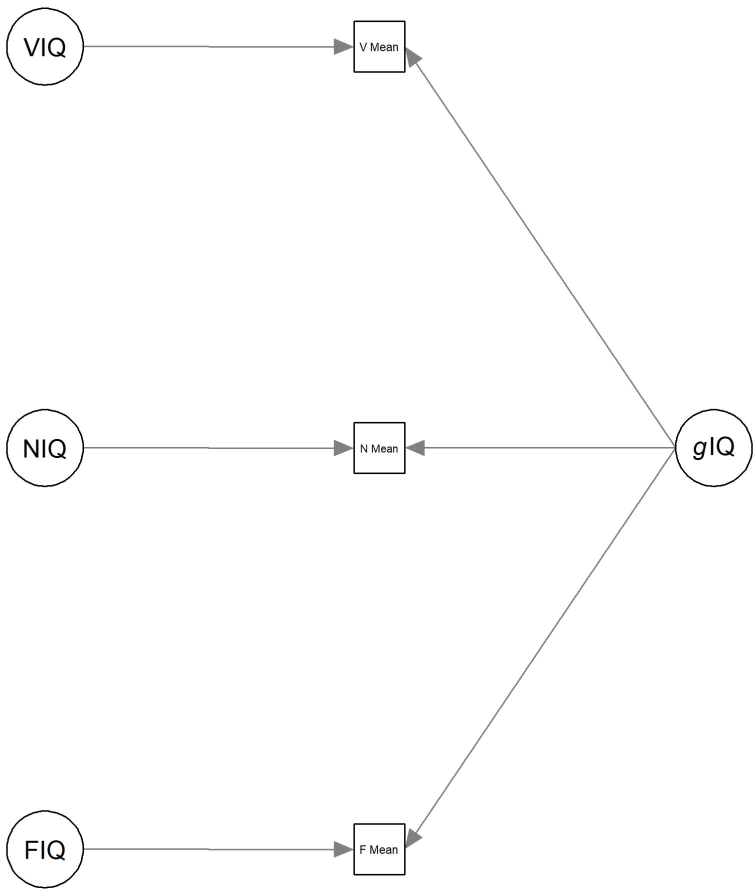

Figure 4. Intelligence Model. Scale means are used as indicators for verbal (VIQ), numeric (NIQ) and figural (FIQ) intelligence. *g*IQ = general intelligence. Indicator residuals are fixed to zero, domain factors serve as quasi-residuals, see Methods.

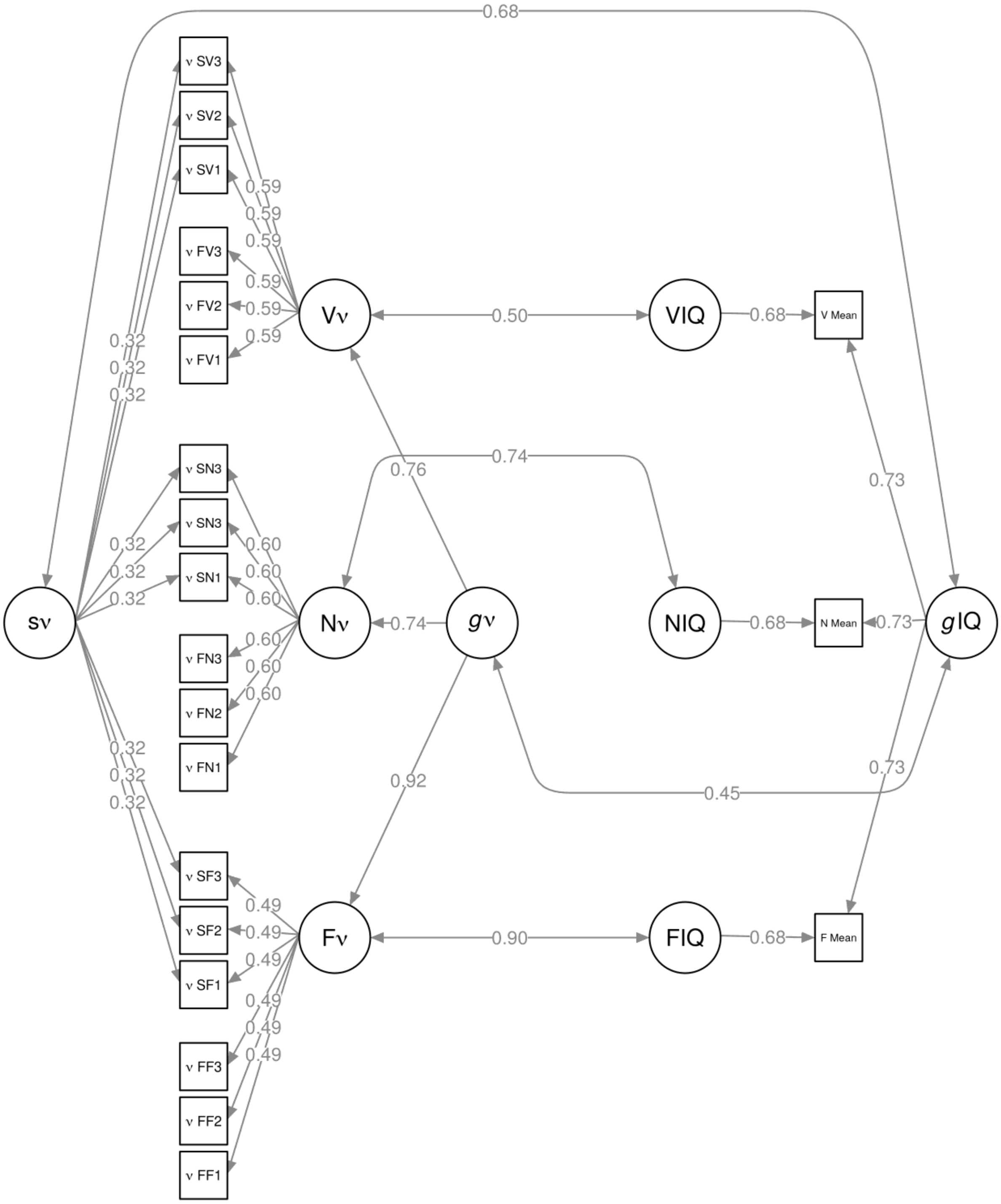

Figure 5. Combined Drift-Intelligence Model. The first letter of the task indices denotes the type of task (F = fast, S = slow); the second letter indicates the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. Completely standardized loadings are reported. Residuals are omitted from the plot for simplicity. The latent correlations between the drift domains and intelligence domains are between the drift domain residuals and the (quasi-residual) intelligence domain factors (see Methods). *g*ν = general drift rate factor. Vν = verbal drift rate factor. Nν = numeric drift rate factor. Fν = figural drift rate factor. sν = method factor for drift rate in slow tasks. Scale means are used as indicators for verbal (VIQ), numeric (NIQ) and figural (FIQ) intelligence. *g*IQ = general intelligence. As the loadings of the drift domain factors are standardized on the different freely estimated variances of the domain factors, their standardized values differ although the unstandardized loadings are all fixed to 1.

Appendix A: Task Descriptives

Table A1

Descriptives of RT (in ms)

Task	Mean	*SD*	Minimum	Maximum
FF1	560	96	398	846
FF2	620	176	372	1,278
FF3	551	96	393	877
FN1	527	78	395	758
FN2	590	107	409	947
FN3	670	135	467	1,168
FV1	792	164	542	1,350
FV2	781	162	513	1,397
FV3	737	124	530	1,161
SF1	3,234	1,091	1,517	7,354
SF2	4,189	2,009	1,355	10,366
SF3	2,856	906	1,021	5,171
SN1	4,168	1,904	1,004	11,074
SN2	2,761	1,098	1,014	6,670
SN3	2,805	885	1,571	5,780
SV1	2,380	709	1,145	4,516
SV2	3,030	1,002	1,654	6,599
SV3	3,600	895	1,935	6,808

Note. The first letter indicates the task complexity (F = fast, S = slow); the second letter denotes the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. *SD* = standard deviation.

Table A2

Descriptives of Accuracy Rate (in %)

Task	Mean	*SD*	Minimum	Maximum
FF1	93.65	2.88	84.54	97.00
FF2	98.68	1.60	93.00	100.00
FF3	97.71	1.90	91.58	100.00
FN1	98.03	2.26	89.00	100.00
FN2	97.68	2.03	91.00	100.00
FN3	97.17	2.74	88.00	100.00
FV1	96.22	3.76	79.55	100.00
FV2	95.11	3.97	78.35	100.00
FV3	97.18	2.41	87.00	100.00
SF1	95.53	2.91	87.00	100.00
SF2	86.69	6.50	69.00	100.00
SF3	80.47	9.10	53.06	97.00
SN1	90.76	8.11	61.00	100.00
SN2	91.16	5.48	72.00	98.00
SN3	93.51	3.71	82.00	100.00
SV1	96.36	2.39	88.00	100.00
SV2	95.11	2.61	85.86	99.00
SV3	94.24	4.77	80.21	100.00

Note. The first letter indicates the task complexity (F = fast, S = slow); the second letter denotes the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. *SD* = standard deviation.

Table A3

Descriptives of drift rate

Task	Mean	*SD*	Minimum	Maximum
FF1	3.16	0.73	1.79	6.42
FF2	3.26	1.02	1.43	7.16
FF3	4.27	0.96	2.38	8.01
FN1	4.97	1.82	2.41	16.50
FN2	3.95	0.97	2.12	8.52
FN3	3.97	1.39	2.00	12.23
FV1	2.81	0.88	1.37	6.25
FV2	2.68	0.78	1.12	4.83
FV3	3.21	0.89	1.54	6.61
SF1	0.94	0.20	0.52	1.61
SF2	0.58	0.17	0.17	0.97
SF3	0.50	0.18	0.09	1.02
SN1	0.70	0.22	0.15	1.30
SN2	0.80	0.25	0.39	1.48
SN3	1.08	0.33	0.57	2.15
SV1	1.17	0.20	0.64	1.79
SV2	1.03	0.29	0.54	1.99
SV3	0.90	0.23	0.39	1.63

Note. The first letter indicates the task complexity (F = fast, S = slow); the second letter denotes the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. *SD* = standard deviation.

Table A4

Descriptives of threshold separation

Task	Mean	*SD*	Minimum	Maximum
FF1	0.91	0.21	0.46	1.71
FF2	1.53	0.53	0.66	3.61
FF3	1.16	0.61	0.63	5.52
FN1	1.47	1.31	0.44	10.00
FN2	1.20	0.51	0.62	3.90
FN3	1.36	1.03	0.50	10.00
FV1	1.52	0.73	0.53	5.76
FV2	1.33	0.44	0.55	2.62
FV3	1.35	0.55	0.66	5.61
SF1	3.75	1.44	1.73	10.00
SF2	3.71	1.37	1.45	8.05
SF3	3.06	0.81	1.36	5.10
SN1	4.00	1.53	1.21	10.00
SN2	3.25	0.92	1.13	6.35
SN3	2.85	0.92	1.52	6.79
SV1	3.08	0.84	1.71	7.07
SV2	3.19	0.87	1.35	5.14
SV3	3.69	1.23	1.75	10.00

Note. The first letter indicates the task complexity (F = fast, S = slow); the second letter denotes the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. *SD* = standard deviation.

Table A5

Descriptives of non-decision time (in ms)

Task	Mean	*SD*	Minimum	Maximum
FF1	423	65	273	587
FF2	359	66	242	592
FF3	411	56	236	555
FN1	388	67	135	539
FN2	427	57	313	678
FN3	499	96	192	789
FV1	513	76	226	850
FV2	527	74	367	749
FV3	520	65	333	732
SF1	1,286	495	137	2,969
SF2	1,480	918	63	5,874
SF3	913	397	230	2,657
SN1	1,628	1,207	0	5,794
SN2	844	309	36	2,097
SN3	1,501	422	628	2,983
SV1	1,092	348	366	2,525
SV2	1,448	420	910	3,746
SV3	1,635	413	68	3,280

Note. The first letter indicates the task complexity (F = fast, S = slow); the second letter denotes the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. *SD* = standard deviation.

Table A6

Descriptives of BIS domain scale scores

	Mean	*SD*	Minimum	Maximum
F_Mean	96.35	7.74	76.50	114.25
N_Mean	99.94	8.38	80.50	120.75
V_Mean	102.78	7.83	79.75	121.50

Note. V = Verbal, N = Numeric, F = Figural. *SD* = standard deviation.

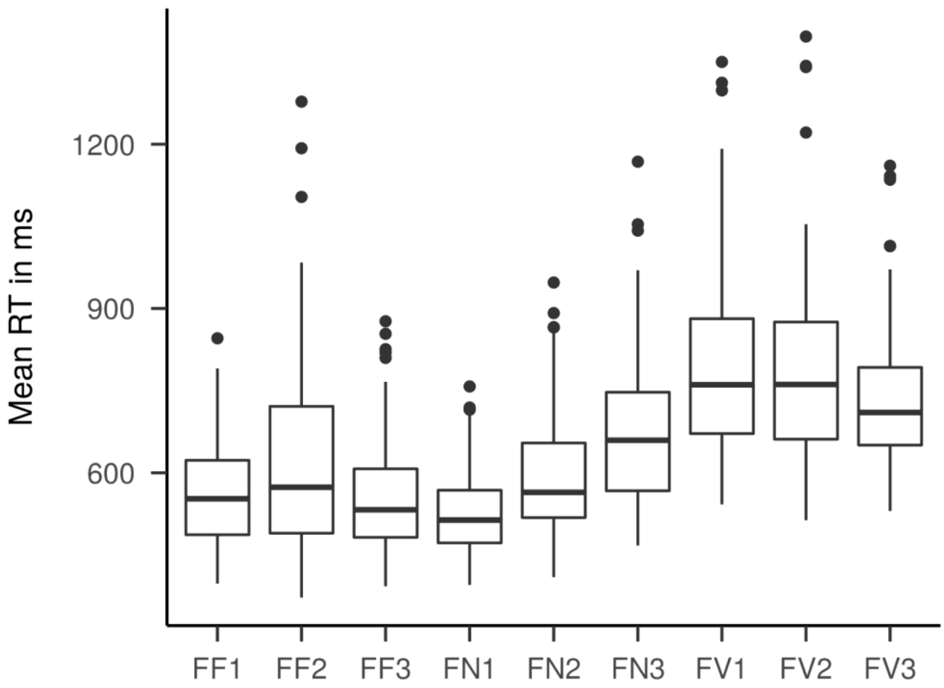

Figure A1. Boxplots of mean response times for all *fast* tasks. The first letter indicates the task complexity (F = fast); the second letter denotes the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. The boxplots display the first, second and third quartile. Outliers are values greater than 1.5 times the interquartile range from either end of the box.

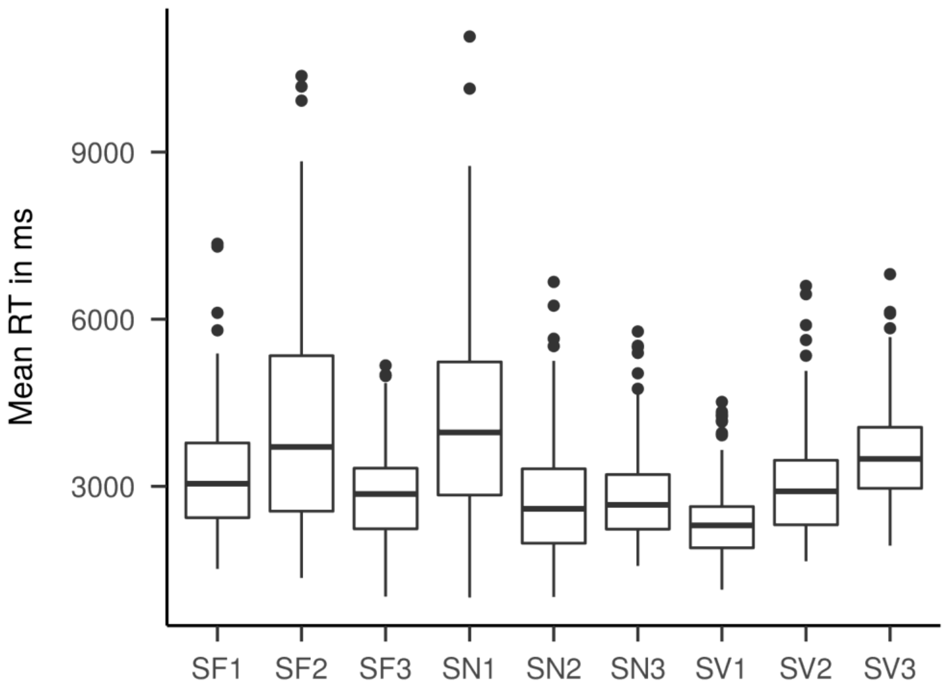

Figure A2. Boxplots of mean response times for all *slow* tasks. The first letter indicates the task complexity (S = slow); the second letter denotes the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. The boxplots display the first, second and third quartile. Outliers are values greater than 1.5 times the interquartile range from either end of the box.

Appendix B: Diffusion Model Fit

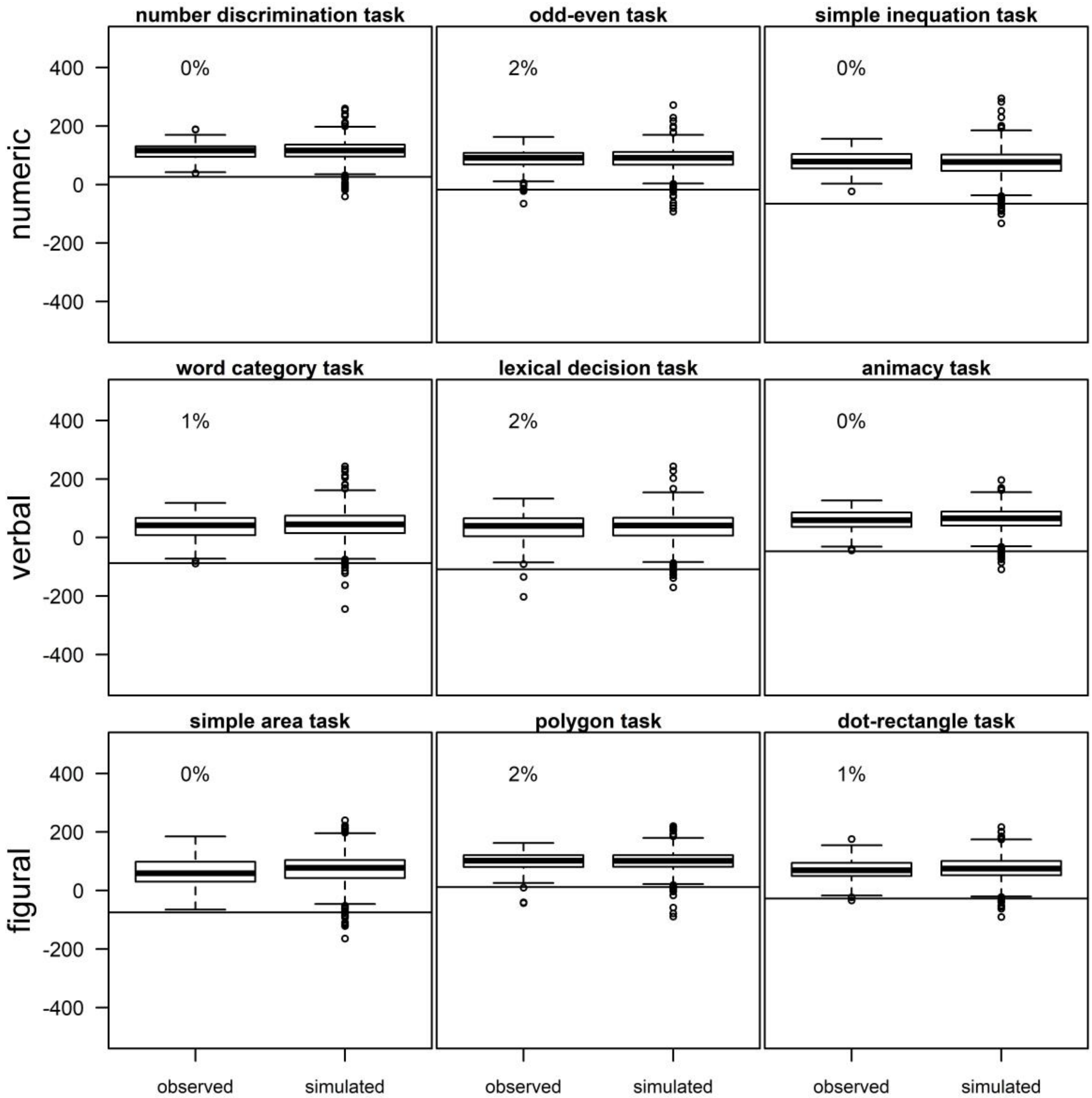

Figure B1. Model fit of all *fast* RT tasks. The boxplots show the maximum likelihood statistic (sum of logarithmized densities). Lower values indicate worse model fit. The horizontal line is the 1% percentile of fit values from 1000 simulated data sets. For observed data, the percentage of fits that are worse than this critical value is also given.

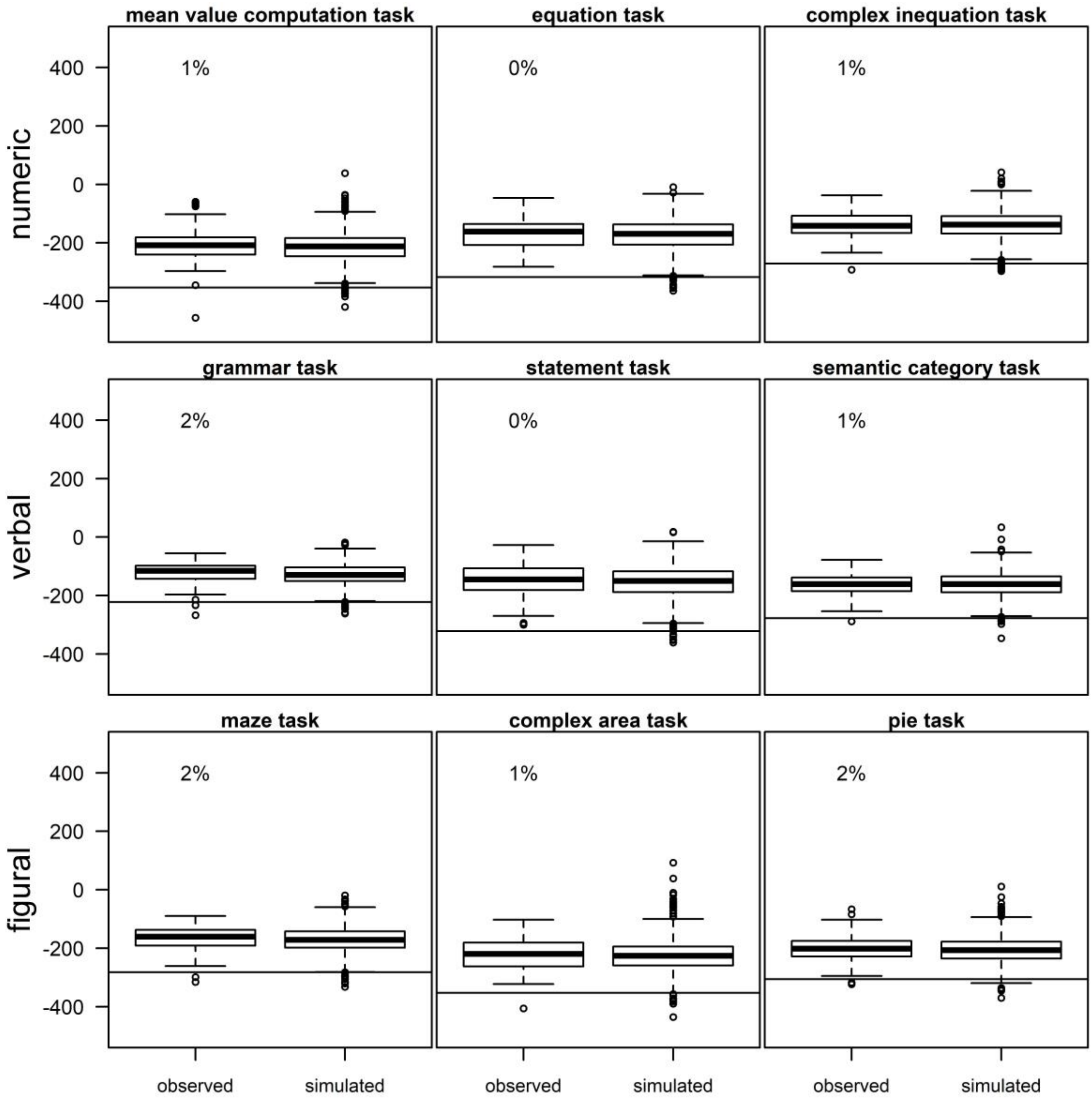

Figure B2. Model fit of all *slow* RT tasks. The boxplots show the maximum likelihood statistic (sum of logarithmized densities). Lower values indicate worse model fit. The horizontal line is the 1% percentile of fit values from 1000 simulated data sets. For observed data, the percentage of fits that are worse than this critical value is also given.

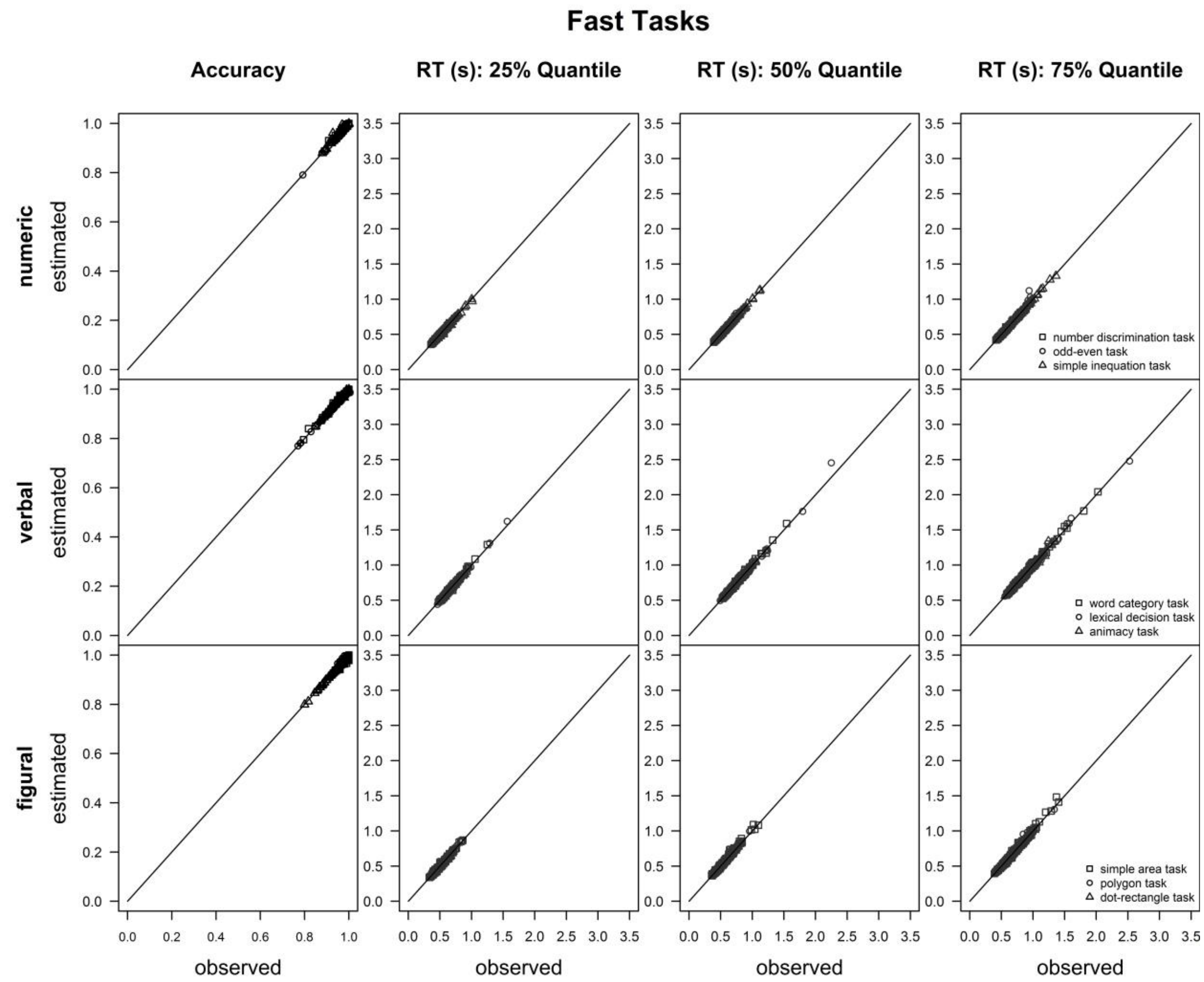

Figure B3. Model fit of the *fast* RT tasks based on the comparison of statistics (accuracy rate, first, second and third RT quartile) of observed data and models' predictions. Each point represents one participant in one task. The diagonals indicate perfect model fit. One data point exceeding the scales of the third RT quartile plot was omitted.

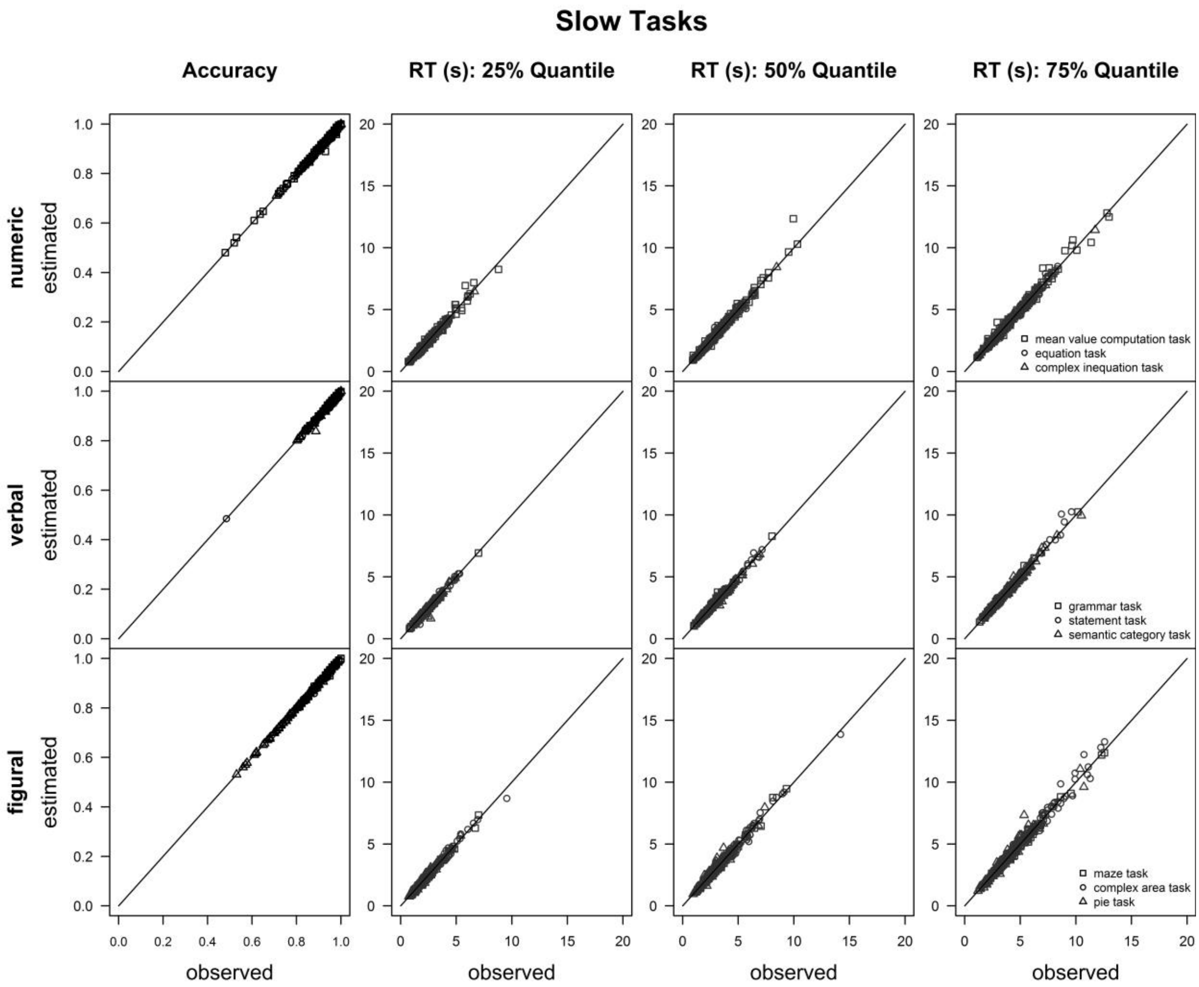

Figure B4. Model fit of the *slow* RT tasks based on the comparison of statistics (accuracy rate, first, second and third RT quartile) of observed data and models' predictions. Each point represents one participant in one task. The diagonals indicate perfect model fit. Two data points exceeding the scales of the third RT quartile plot were omitted.

Appendix C: Structural Equation Models

Table C1

Drift Model 1 (g factor)

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
			Loadings		
*g*ν on ν (each task)	1	0			0.509
			Latent (Residual) Variances		
*g*ν	0.259	0.020	[0.219; 0.298]	<.001	1
			Residual Indicator Variances		
ν (each task)	0.741	0.020	[0.702; 0.781]	<.001	0.741

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized.

Table C2

Drift Model 2 (uncorrelated domains)

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
		Loadings			
Fν on ν (each figural task)	1	0			0.506
Nν on ν (each numeric task)	1	0			0.610
Vν on ν (each verbal task)	1	0			0.615
		Latent (Residual) Variances			
Fν	0.256	0.035	[0.188; 0.325]	<.001	1
Nν	0.371	0.033	[0.308; 0.435]	<.001	1
Vν	0.378	0.033	[0.314; 0.442]	<.001	1
		Residual Indicator Variances			
ν (each figural task)	0.744	0.035	[0.675; 0.812]	<.001	0.744
ν (each numeric task)	0.629	0.033	[0.565; 0.692]	<.001	0.629
ν (each verbal task)	0.622	0.033	[0.558; 0.686]	<.001	0.622

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized.

Table C3

Drift Model 3 (hierarchical model of domains & g factor)

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
			Loadings		
Fν on ν (each figural task)	1	0			0.514
Nν on ν (each numeric task)	1	0			0.605
Vν on ν (each verbal task)	1	0			0.617
*g*ν on Fν	1	0			0.922
*g*ν on Nν	1	0			0.784
*g*ν on Vν	1	0			0.769
			Latent (Residual) Variances		
*g*ν	0.225	0.024	[0.178; 0.271]	<.001	1
Fν	0.039	0.029	[-0.017; 0.096]	.171	0.149
Nν	0.141	0.033	[0.077; 0.206]	<.001	0.386
Vν	0.156	0.032	[0.092; 0.219]	<.001	0.409
			Residual Indicator Variances		
ν (each figural task)	0.736	0.032	[0.672; 0.800]	<.001	0.736
ν (each numeric task)	0.634	0.031	[0.573; 0.696]	<.001	0.634
ν (each verbal task)	0.620	0.031	[0.559; 0.680]	<.001	0.620

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized.

Table C4

Drift Model 4 (hierarchical model of domains & g factor & slow method factor)

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
			Loadings		
sν on ν (each slow task)	1	0			0.308
Fν on ν (each figural task)	1	0			0.486
Nν on ν (each numeric task)	1	0			0.600
Vν on ν (each verbal task)	1	0			0.598
*g*ν on Fν	1	0			0.926
*g*ν on Nν	1	0			0.750
*g*ν on Vν	1	0			0.751
			Latent (Residual) Variances		
*g*ν	0.202	0.025	[0.154; 0.251]	<.001	1
sν	0.095	0.022	[0.051; 0.138]	<.001	1
Fν	0.034	0.028	[-0.022; 0.089]	.235	0.142
Nν	0.158	0.033	[0.094; 0.222]	<.001	0.438
Vν	0.156	0.031	[0.095; 0.217]	<.001	0.435
			Residual Indicator Variances		
ν (each fast figural task)	0.764	0.034	[0.698; 0.830]	<.001	0.764
ν (each fast numeric task)	0.640	0.031	[0.579; 0.701]	<.001	0.640
ν (each fast verbal task)	0.642	0.032	[0.580; 0.704]	<.001	0.642
ν (each slow figural task)	0.670	0.034	[0.602; 0.737]	<.001	0.670
ν (each slow numeric task)	0.545	0.034	[0.479; 0.612]	<.001	0.545
ν (each slow verbal task)	0.547	0.032	[0.485; 0.610]	<.001	0.547

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized.

Table C5

Intelligence Model

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
		Loadings			
*g*IQ on F_Mean/N_Mean/V_Mean	1	0			0.736
VIQ on V_Mean/NIQ on N_Mean/FIQ on F_Mean	1	0			0.677
		Latent (Residual) Variances			
*g*IQ	0.542	0.040	[0.465; 0.620]	<.001	1
FIQ/NIQ/VIQ	0.458	0.040	[0.380; 0.535]	<.001	1
V_Mean/N_Mean/F_Mean	0	0			

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized.

Table C6

Drift Model 4 (hierarchical model of domains & g factor & slow method factor), freely estimated

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
			Loadings		
Fν on ν.FF1	1	0			0.365
on ν.FF2	1.213	0.685	[-0.128; 2.555]	.076	0.443
on ν.FF3	1.996	1.266	[-0.486; 4.477]	.115	0.729
on ν.SF1	0.793	0.624	[-0.430; 2.017]	.204	0.290
on ν.SF2	0.974	0.532	[-0.067; 2.016]	.067	0.356
on ν.SF3	1.364	0.802	[-0.207; 2.935]	.089	0.498
Nν on ν.FN1	1	0			0.610
on ν.FN2	1.035	0.144	[0.753; 1.318]	<.001	0.632
on ν.FN3	0.802	0.158	[0.492; 1.112]	<.001	0.489
on ν.SN1	0.673	0.188	[0.304; 1.042]	<.001	0.411
on ν.SN2	1.172	0.203	[0.774; 1.570]	<.001	0.715
on ν.SN3	1.206	0.217	[0.780; 1.632]	<.001	0.736
Vν on ν.FV1	1	0			0.690
on ν.FV2	1.045	0.126	[0.799; 1.291]	<.001	0.721
on ν.FV3	0.942	0.135	[0.678; 1.207]	<.001	0.650
on ν.SV1	0.828	0.123	[0.586; 1.070]	<.001	0.571
on ν.SV2	0.628	0.130	[0.372; 0.883]	<.001	0.433
on ν.SV3	0.741	0.136	[0.474; 1.008]	<.001	0.511
sν on ν.SF1	1	0			0.378
on ν.SF2	1.339	1.182	[-0.978; 3.656]	.257	0.507
on ν.SF3	1.080	0.997	[-0.875; 3.034]	.279	0.408
on ν.SN1	1.543	1.299	[-1.002; 4.088]	.235	0.584
on ν.SN2	0.587	0.673	[-0.733; 1.907]	.383	0.222
on ν.SN3	0.579	0.744	[-0.879; 2.038]	.436	0.219
on ν.SV1	0.749	0.501	[-0.233; 1.731]	.135	0.283
on ν.SV2	0.895	0.653	[-0.385; 2.175]	.170	0.339
on ν.SV3	1.099	0.654	[-0.182; 2.381]	.093	0.416
gν on Fν	1	0			0.748
*g*ν on Nν	1.860	1.370	[-0.825; 4.545]	.175	0.833
*g*ν on Vν	1.768	1.188	[-0.560; 4.096]	.137	0.700
			Latent (Residual) Variances		
*g*ν	0.075	0.100	[-0.121; 0.270]	.455	1
sν	0.143	0.214	[-0.276; 0.562]	.503	1
Fν	0.059	0.050	[-0.038; 0.156]	.235	0.441
Nν	0.114	0.071	[-0.026; 0.254]	.110	0.307
Vν	0.243	0.082	[0.081; 0.404]	.003	0.510
			Residual Indicator Variances		
ν.FF1	0.867	0.142	[0.589; 1.144]	<.001	0.867
ν.FF2	0.804	0.085	[0.637; 0.970]	<.001	0.804
ν.FF3	0.469	0.170	[0.136; 0.802]	.006	0.469
ν.FN1	0.628	0.090	[0.451; 0.804]	<.001	0.628
ν.FN2	0.601	0.094	[0.418; 0.784]	<.001	0.601
ν.FN3	0.760	0.074	[0.615; 0.906]	<.001	0.760

ν.FV1	0.524	0.083	[0.361; 0.687]	<.001	0.524
ν.FV2	0.480	0.086	[0.312; 0.648]	<.001	0.480
ν.FV3	0.577	0.082	[0.416; 0.738]	<.001	0.577
ν.SF1	0.773	0.158	[0.463; 1.083]	<.001	0.773
ν.SF2	0.617	0.096	[0.428; 0.806]	<.001	0.617
ν.SF3	0.585	0.090	[0.408; 0.762]	<.001	0.585
ν.SN1	0.491	0.098	[0.298; 0.684]	<.001	0.491
ν.SN2	0.439	0.071	[0.300; 0.578]	<.001	0.439
ν.SN3	0.411	0.073	[0.268; 0.553]	<.001	0.411
ν.SV1	0.594	0.079	[0.440; 0.748]	<.001	0.594
ν.SV2	0.698	0.082	[0.538; 0.858]	<.001	0.698
ν.SV3	0.566	0.094	[0.381; 0.750]	<.001	0.566

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized.

Table C7

Combined Drift-Intelligence Model, freely estimated

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
		Loadings			
Fv on v.FF1	1	0			0.392
on v.FF2	1.180				0.463
on v.FF3	1.630				0.639
on v.SF1	0.758				0.297
on v.SF2	1.215				0.477
on v.SF3	1.554				0.610
Nv on v.FN1	1	0			0.526
on v.FN2	1.011	0.187	[0.645; 1.377]	<.001	0.532
on v.FN3	0.756	0.181	[0.401; 1.112]	<.001	0.398
on v.SN1	0.860	0.202	[0.464; 1.257]	<.001	0.453
on v.SN2	1.472	0.261	[0.960; 1.985]	<.001	0.775
on v.SN3	1.572	0.252	[1.078; 2.066]	<.001	0.827
Vv on v.FV1	1	0			0.679
on v.FV2	1.043	0.123	[0.803; 1.284]	<.001	0.709
on v.FV3	0.970	0.131	[0.714; 1.226]	<.001	0.659
on v.SV1	0.846	0.118	[0.615; 1.076]	<.001	0.575
on v.SV2	0.679	0.117	[0.450; 0.907]	<.001	0.461
on v.SV3	0.740	0.120	[0.505; 0.976]	<.001	0.503
sv on v.SF1	1	0			0.564
on v.SF2	0.537	0.230	[0.087; 0.988]	.019	0.303
on v.SF3	0.399	0.191	[0.025; 0.773]	.036	0.225
on v.SN1	0.641	0.219	[0.212; 1.070]	.003	0.362
on v.SN2	0.469	0.236	[0.008; 0.931]	.046	0.265
on v.SN3	0.151	0.188	[-0.218; 0.520]	.421	0.085
on v.SV1	0.392	0.168	[0.063; 0.721]	.020	0.221
on v.SV2	0.717	0.214	[0.297; 1.137]	.001	0.404
on v.SV3	0.747	0.182	[0.391; 1.104]	<.001	0.421
*g*v on Fv	1	0			0.885
*g*v on Nv	1.091				0.720
*g*v on Vv	1.191				0.608
*g*IQ on F_Mean	1	0			0.808
*g*IQ on N_Mean	0.858	0.033	[0.794; 0.923]	<.001	0.693
*g*IQ on V_Mean	0.833				0.673
FIQ on F_Mean	1	0			0.590
NIQ on N_Mean	1	0			0.721
VIQ on V_Mean	1	0			0.740
		Covariances			
*g*v with gIQ	0.117				0.418
sv with gIQ	0.336	0.062	[0.214; 0.458]	<.001	0.739
Fv with FIQ	0.060	0.035	[-0.008; 0.128]	.082	0.561
Nv with NIQ	0.237	0.038	[0.162; 0.312]	<.001	0.899
Vv with VIQ	0.208	0.046	[0.119; 0.298]	<.001	0.522

	Latent (Residual) Variances				
*g*ν	0.121				1
*g*IQ	0.652	0.038	[0.578; 0.727]	<.001	1
sν	0.318	0.127	[0.068; 0.567]	.013	1
Fν	0.033	0.017	[0.000; 0.067]	.053	0.217
Nν	0.134	0.036	[0.000; 0.067]	<.001	0.482
Vν	0.291	0.080	[0.134; 0.448]	<.001	0.630
FIQ	0.348	0.038	[0.273; 0.422]	<.001	1
NIQ	0.519	0.059	[0.404; 0.634]	<.001	1
VIQ	0.548	0.052	[0.446; 0.649]	<.001	1
	Residual Indicator Variances				
ν.FF1	0.846				0.846
ν.FF2	0.786	0.067	[0.655; 0.916]	<.001	0.786
ν.FF3	0.591	0.097	[0.402; 0.780]	<.001	0.591
ν.FN1	0.723	0.085	[0.557; 0.890]	<.001	0.723
ν.FN2	0.717	0.075	[0.571; 0.863]	<.001	0.717
ν.FN3	0.842	0.064	[0.716; 0.967]	<.001	0.842
ν.FV1	0.538	0.080	[0.382; 0.695]	<.001	0.538
ν.FV2	0.497	0.077	[0.346; 0.649]	<.001	0.497
ν.FV3	0.566	0.079	[0.412; 0.720]	<.001	0.566
ν.SF1	0.594	0.102	[0.393; 0.794]	<.001	0.594
ν.SF2	0.681	0.076	[0.531; 0.831]	<.001	0.681
ν.SF3	0.578	0.061	[0.458; 0.697]	<.001	0.578
ν.SN1	0.664	0.078	[0.512; 0.817]	<.001	0.664
ν.SN2	0.330	0.054	[0.225; 0.435]	<.001	0.330
ν.SN3	0.309	0.051	[0.209; 0.409]	<.001	0.309
ν.SV1	0.621	0.076	[0.471; 0.771]	<.001	0.621
ν.SV2	0.624	0.080	[0.466; 0.782]	<.001	0.624
ν.SV3	0.570	0.082	[0.409; 0.731]	<.001	0.570
F_Mean/N_Mean/V_Mean	0	0			

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized. Caveat: unreliable estimates with some missing standard errors.

Table C8

Non-Decision Time Model 4 (hierarchical model of domains & g factor & slow method factor)

Parameter	Estimate	*SE*	95% *CI*	*p*	Std. Est.
			Loadings		
Ft_0 on t_0 (each figural task)	1	0			0.539
Nt_0 on t_0 (each numeric task)	1	0			0.582
Vt_0 on t_0 (each verbal task)	1	0			0.613
st_0 on t_0 (each slow task)	1	0			0.273
gt_0 on Ft_0	1	0			1.020
gt_0 on Nt_0	1	0			0.944
gt_0 on Vt_0	1	0			0.897
			Latent (Residual) Variances		
gt_0	0.302	0.021	[0.261; 0.344]	<.001	1
st_0	0.075	0.019	[0.038; 0.112]	<.001	1
Ft_0	-0.012	0.021	[-0.054; 0.031]	.592	-0.040
Nt_0	0.037	0.023	[-0.009; 0.083]	.117	0.108
Vt_0	0.074	0.026	[0.023; 0.124]	.004	0.196
			Residual Indicator Variances		
t_0 (each fast figural task)	0.709	0.029	[0.652; 0.767]	<.001	0.709
t_0 (each fast numeric task)	0.661	0.029	[0.605; 0.717]	<.001	0.661
t_0 (each fast verbal task)	0.624	0.028	[0.568; 0.680]	<.001	0.624
t_0 (each slow figural task)	0.635	0.030	[0.575; 0.694]	<.001	0.635
t_0 (each slow numeric task)	0.587	0.030	[0.527; 0.646]	<.001	0.587
t_0 (each slow verbal task)	0.550	0.031	[0.488; 0.611]	<.001	0.550

Note. Missing *p* values indicate fixed parameters. The standardized solution is completely standardized.

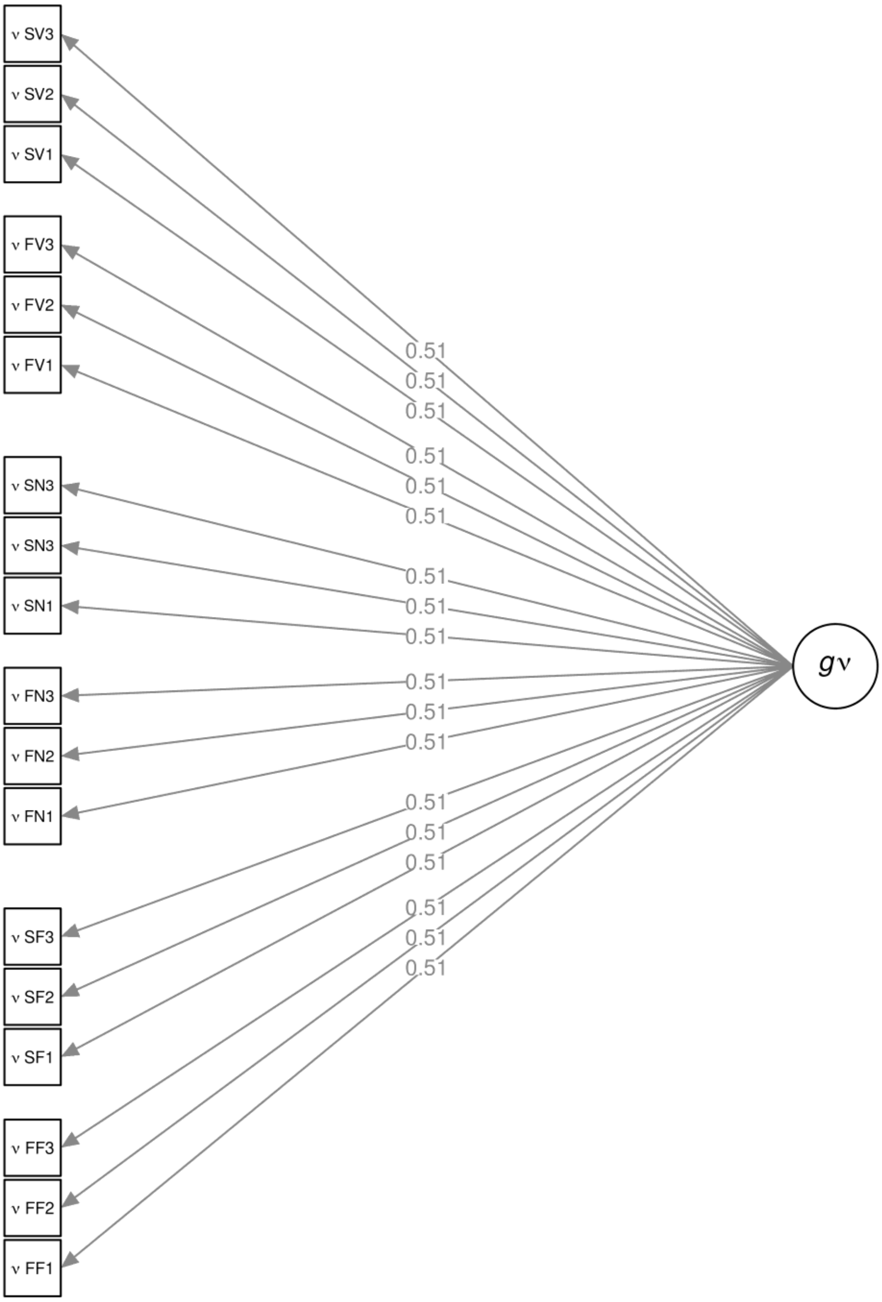

Figure C1. Drift Model 1. The first letter of the task indices denotes the type of task (F = fast, S = slow); the second letter indicates the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. Standardized loadings reported. Residuals are omitted from the plot for simplicity. *g*ν = general drift rate factor.

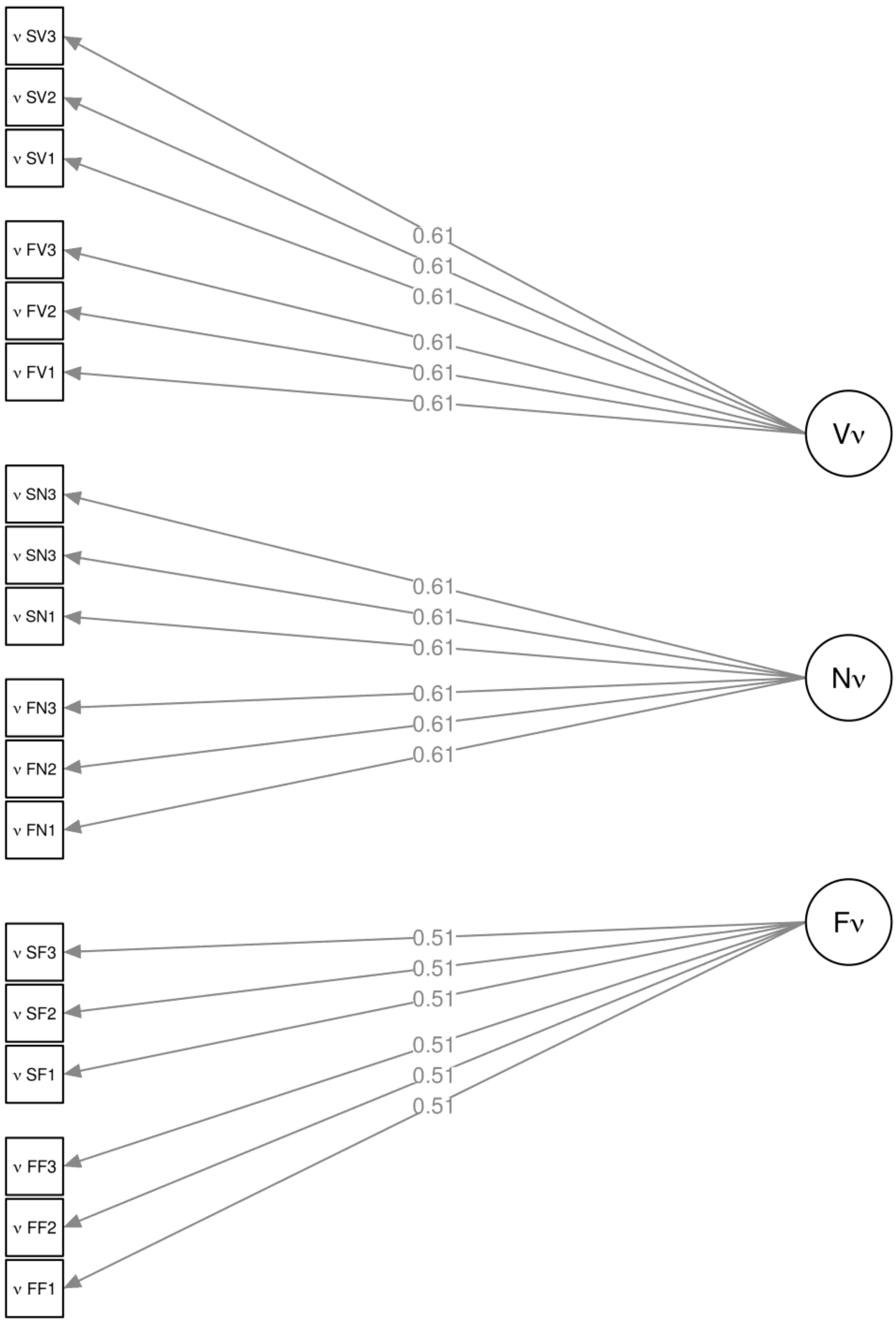

Figure C2. Drift Model 2. The first letter of the task indices denotes the type of task (F = fast, S = slow); the second letter indicates the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. Standardized loadings reported. Residuals are omitted from the plot for simplicity. Vν = verbal drift rate factor. Nν = numeric drift rate factor. Fν = figural drift rate factor.

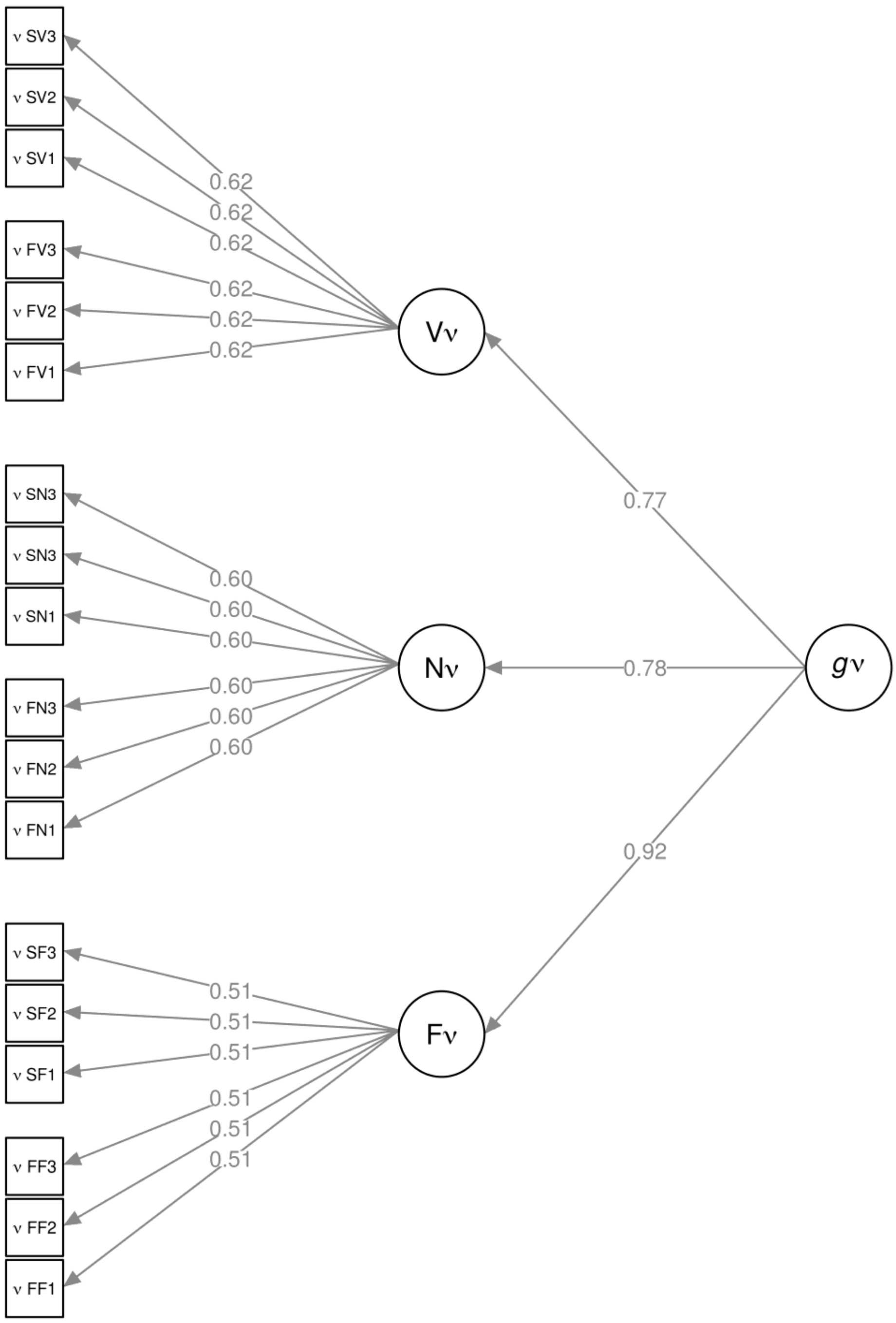

Figure C3. Drift Model 3. The first letter of the task indices denotes the type of task (F = fast, S = slow); the second letter indicates the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. Standardized loadings reported. Residuals are omitted from the plot for simplicity. gν = general drift rate factor. Vν = verbal drift rate factor. Nν = numeric drift rate factor. Fν = figural drift rate factor. As the loadings of the drift domain factors are standardized on the different freely estimated variances of the domain factors, their standardized values differ although the unstandardized loadings are all fixed to 1.

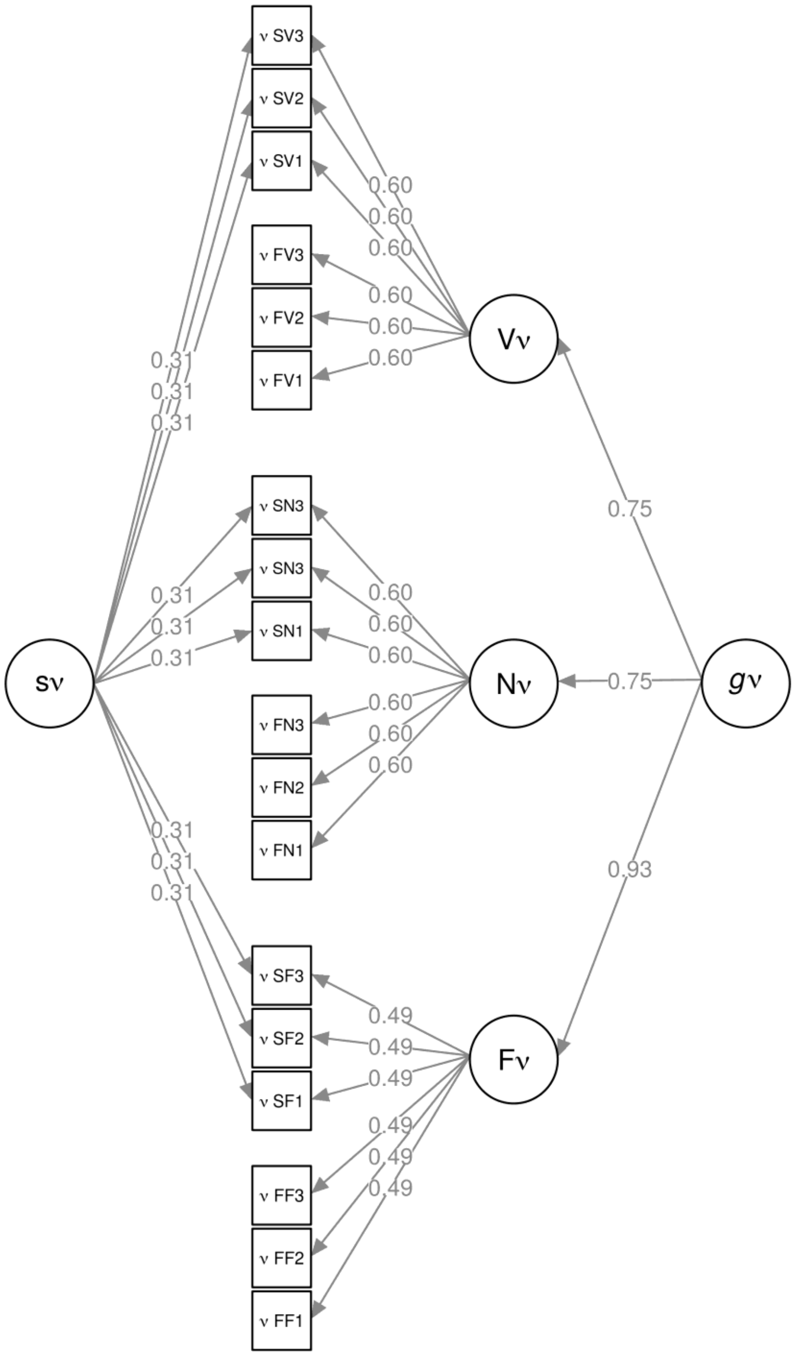

Figure C4. Drift Model 4. The first letter of the task indices denotes the type of task (F = fast, S = slow); the second letter indicates the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. Standardized loadings reported. Residuals are omitted from the plot for simplicity. gν = general drift rate factor. Vν = verbal drift rate factor. Nν = numeric drift rate factor. Fν = figural drift rate factor. sν = method factor for drift rate in slow tasks.

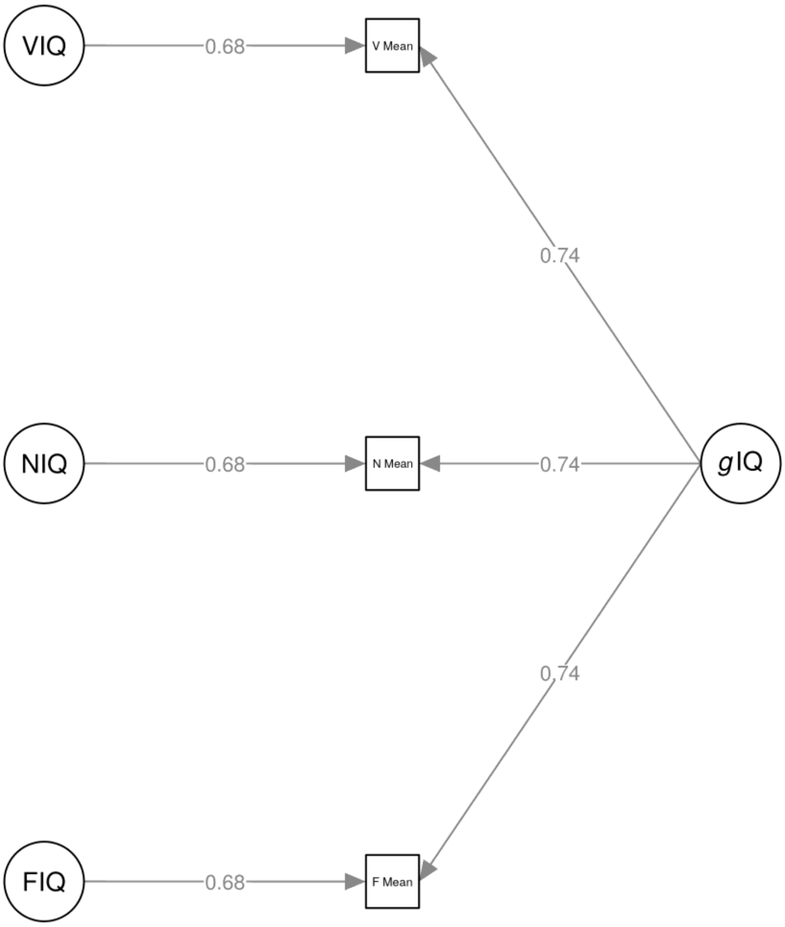

Figure C5. Intelligence Model. Scale means are used as indicators for verbal (VIQ), numeric (NIQ) and figural (FIQ) intelligence. gIQ = general intelligence. Completely standardized loadings are reported. Indicator residuals are fixed to zero, domain factors serve as quasi-residuals, see Methods.

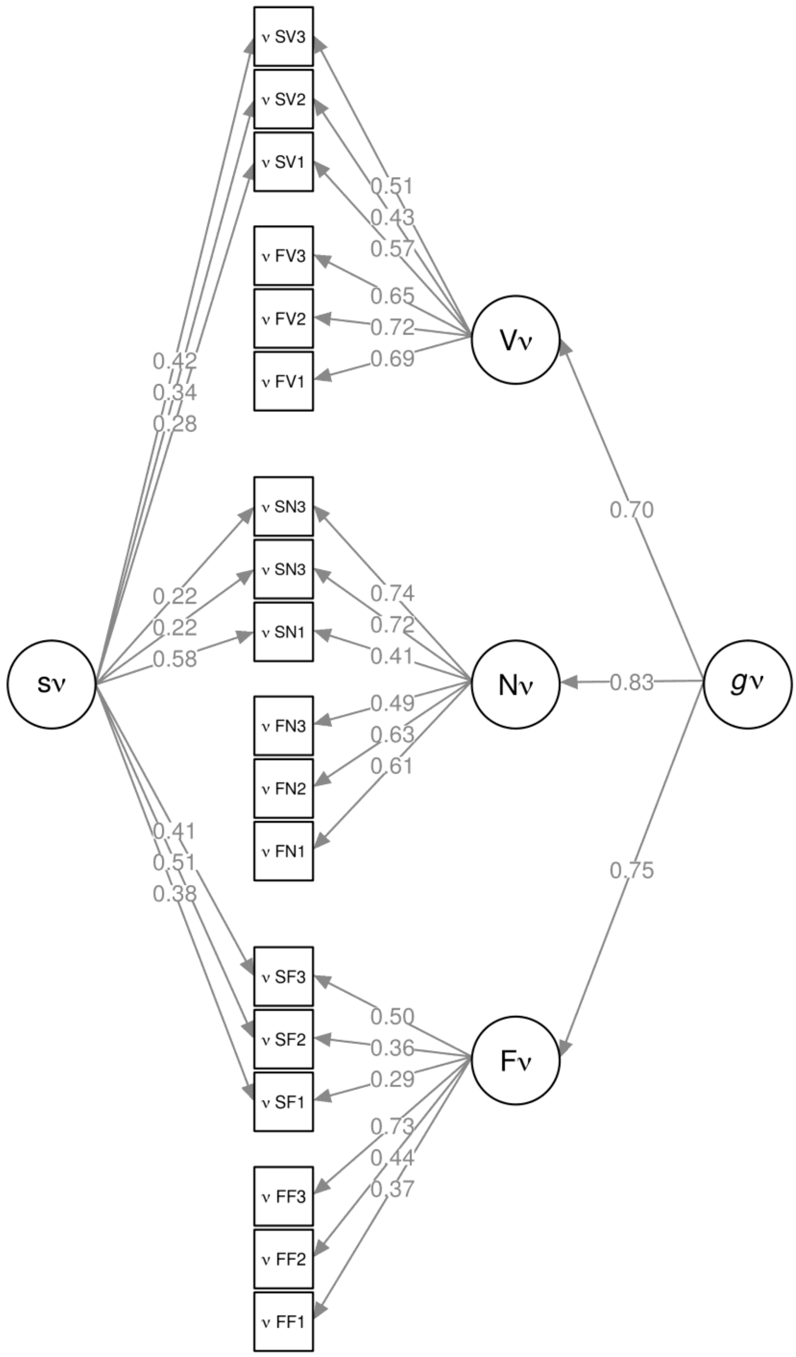

Figure C6. Drift Model 4 (freely estimated). The first letter of the task indices denotes the type of task (F = fast, S = slow); the second letter indicates the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. Standardized loadings reported. Residuals are omitted from the plot for simplicity. gν = general drift rate factor. Vν = verbal drift rate factor. Nν = numeric drift rate factor. Fν = figural drift rate factor. sν = method factor for drift rate in slow tasks.

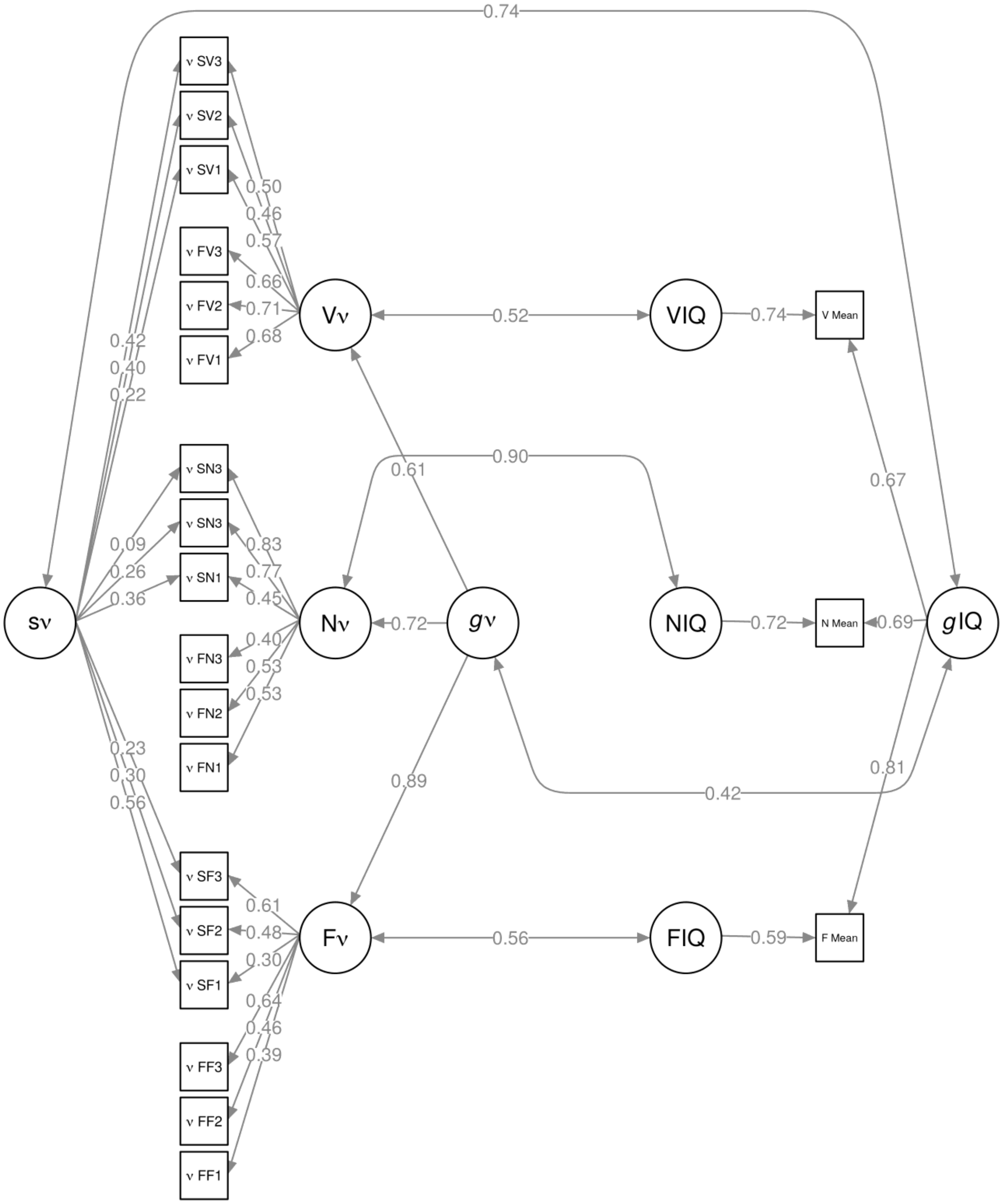

Figure C7. Combined Drift-Intelligence Model (freely estimated). The first letter of the task indices denotes the type of task (F = fast, S = slow); the second letter indicates the domain (N = numeric, V = verbal, F = figural). See Table 1 for a brief description of all tasks. Standardized loadings reported. Residuals are omitted from the plot for simplicity. The latent correlations between the drift domains and intelligence domains are between the drift domain residuals and the (quasi-residual) intelligence domain factors (see Methods).gν = general drift rate factor. Vν = verbal drift rate factor. Nν = numeric drift rate factor. Fν = figural drift rate factor. sν = method factor for drift rate in slow tasks. Scale means are used as single indicators for verbal (VIQ), numeric (NIQ) and figural (FIQ) intelligence). gIQ = general intelligence.

Appendix A 3

Manuscript 3:

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2020). Age differences in diffusion model parameters: a meta-analysis. *Psychological Research*, 1-10.

 doi: 10.1007/s00426-020-01371-8

Age differences in diffusion model parameters: A meta-analysis

Maximilian Theisen, Veronika Lerche, Mischa von Krause, & Andreas Voss

Heidelberg University

Authors‘ Note

Maximilian Theisen, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany; Veronika Lerche, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany; Mischa von Krause, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany; Andreas Voss, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany;

Correspondence concerning this article should be addressed to Maximilian Theisen, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, email: maximilian.theisen@psychologie.uni-heidelberg.de.

Abstract

Older adults typically show slower response times in basic cognitive tasks than younger adults. A diffusion model analysis allows the clarification of why older adults react more slowly by estimating parameters that map distinct cognitive components of decision making. The main components of the diffusion model are the speed of information uptake (drift rate), the degree of conservatism regarding the decision criterion (boundary separation), and the time taken up by non-decisional processes (i.e., encoding and motoric response execution; non-decision time). While the literature shows consistent results regarding higher boundary separation and longer non-decision time for older adults, results are more complex when it comes to age differences in drift rates. We conducted a multi-level meta-analysis to identify possible sources of this variance. As possible moderators, we included task difficulty and task type. We found that age differences in drift rate are moderated both by task type and task difficulty. Older adults were inferior in drift rate in perceptual and memory tasks, but information accumulation was even increased in lexical decision tasks for the older participants. Additionally, in perceptual and lexical decision tasks, older individuals benefitted from high task difficulty. In the memory tasks, task difficulty did not moderate the negative impact of age on drift. The finding of higher boundary separation and longer non-decision time in older than younger adults generalized over task type and task difficulty. The results of our meta-analysis are consistent with recent findings of a more pronounced age-related decline in memory than in vocabulary performance.

Keywords: Age differences; Aging; Decision Making; Diffusion Model; Response Time Models

Age differences in diffusion model parameters: A meta-analysis

It is a common finding from the literature on cognitive aging that older people show larger response times (RTs) in basic cognitive tasks than younger adults (Jensen, 2006). In the last decades, the mechanisms underlying this age-related slowing have become a subject of debate. On the one hand, the higher RTs of the older adults might be the result of a general decline in cognitive processing speed due to increased neural noise (Salthouse, 1996). On the other hand, however, it is also possible that the slow responses are based on encoding problems (e.g., due to impaired vision), reduced motoric speed, or more cautious response criteria. Such different accounts can be differentiated by means of diffusion model analyses (Voss, Nagler, & Lerche, 2013). The diffusion model (Ratcliff, 1978) is a stochastic model used to analyze response time distributions and error rates in binary decision tasks. It thus utilizes a more complete representation of decision outcomes than just mean RTs. The model aims to disentangle three main components of the decision process: the speed of information uptake (*drift rate*), the degree of conservatism regarding the decision criterion (i.e., speed-accuracy trade-offs; *boundary separation*) and the time taken up by non-decisional processes such as encoding and motoric response execution (*non-decision time*).

Several diffusion model studies have challenged the view that age differences in RT are indicative of a general decline in cognitive speed (e.g., Spaniol, Madden, & Voss, 2006). Quite often, age differences in RT were not due to differences in the mean speed of information uptake, but due to the fact that older people tended to be more cautious (i.e., they favored accurate over fast responses) and that they took longer in terms of the non-decisional components of the response time (e.g., Ratcliff, Thapar, & McKoon, 2001). However, in some studies, older people additionally showed a lower speed of information uptake (Voskuilen, Ratcliff, & McKoon, 2018), consistent with the notion that processing speed declines with age.

So far, to our knowledge no attempt has been made to bring together the inconsistent results regarding drift rates in a quantitative way. It is an open question whether the discrepancies are simply due to random sample differences or can be explained by specific study attributes. As Dully, McGovern, and O'Connell (2018) note in their literature review, there are "task-specific differences in evidence accumulation rates" (p. 3). However, these task-specific differences have not yet been examined quantitatively.

In this paper, we present the results of a meta-analysis regarding age differences in diffusion model parameters. The focus is on drift rates because of the variability in findings for this parameter. We were interested in whether characteristics of the task (specifically, content and difficulty of task) might explain the inconsistent findings in the literature. We also analyzed the parameters boundary separation and non-decision time. In terms of these parameters, we expected that age differences generalize across tasks. In the next chapter, we briefly introduce the diffusion model (for further introductory information, see e.g., Ratcliff & McKoon, 2008; Voss et al., 2013; Wagenmakers, 2009).

Introduction to Diffusion Modeling

The diffusion model is a mathematical model that can be applied to examine the processes underlying RT tasks with two response options. It has most frequently been used with three main task types (Voss et al., 2013). The first group of tasks comprise *memory tasks*. Here, participants usually have to decide whether a stimulus has been presented to them before or not (recognition memory tasks, e.g., Spaniol et al., 2006; Yap, Sibley, Balota, Ratcliff, & Rueckl, 2015). Second, there are *perceptual tasks* in which participants have to discriminate, for example, between two levels of brightness (bright vs. dark, e.g., Ratcliff, 2002; Ratcliff, Thapar, & McKoon, 2003), between two different letters (e.g., F vs. Q, Thapar, Ratcliff, & McKoon, 2003) or between two different quantities of stimuli (small vs. large, e.g., Ratcliff, Thompson, & McKoon, 2015; Ratcliff & Van Dongen, 2009). The third category of task types includes *lexical decision tasks*. In these tasks, participants have to assess whether a presented letter string is a word or not (e.g., Ratcliff, Gomez, & McKoon, 2004; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008).

For these three categories of tasks, it is assumed that the four central assumptions of the diffusion model are met: (1) Information about the two response options is accumulated continuously, (2) decisions are based on single-stage processing, (3) parameters are constant over time, and (4) the tasks are fast response time tasks with mean RTs below 1.5 seconds. Note, however, that this latter criterion has recently been questioned. Studies demonstrated that also for RT tasks that take up to several seconds per trial, the diffusion model fits well and provides valid parameter estimates (Aschenbrenner, Balota, Gordon, Ratcliff, & Morris, 2016; Lerche, Christmann, & Voss, 2018; Lerche & Voss, 2017). In fact, for such slower tasks, the standard diffusion model that is based on random Gaussian noise might even fit better than for very fast response time tasks (Voss, Lerche, Mertens, & Voss, 2019).

Four main parameters affect the position and shape of response time distributions in the diffusion model framework. These parameters are also visualized in

Figure 1. First, there is the distance between the two boundaries that are associated with correct (upper boundary) and error responses (lower boundary) in the example figure. This *boundary separation* (*a*) defines the quantity of information that needs to be accumulated before a decision is made. Under accuracy (speed) instructions, participants typically adopt more distant (more close) boundaries (e.g., Ratcliff & Rouder, 1998; Voss, Rothermund, & Voss, 2004).

Second, there is the speed of information accumulation, the so-called *drift rate* (v). Drift rate is higher in easier compared to more difficult tasks (e.g., Arnold, Bröder, & Bayen, 2015; Lerche & Voss, 2017) and drift is also positively related to cognitive abilities (e.g., Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Schubert, Hagemann, Voss, Schankin, & Bergmann, 2015).

Third, the *starting point* (z) of the accumulation process is modelled. In many tasks, decision processes start from the center between the two boundaries. However, if one of the two response options has a higher expected value (e.g., the response is correct in more trials or higher reward is associated with this response), participants shift the starting point towards the favored option (e.g., Leite & Ratcliff, 2011; Voss et al., 2004).

Finally, *non-decision time* (t_0) subsumes the total duration of all the non-decisional processes (e.g., encoding of information and motor response). Moreover, inter-trial variabilities are often included in the model. However, these variability parameters (in particular, inter-trial variability of drift rate and starting point) often cannot be estimated very reliably (Boehm et al., 2018; Lerche & Voss, 2016; van Ravenzwaaij, Donkin, & Vandekerckhove, 2017) and thus are not very useful to assess inter-individual differences in decision making.

Method

As the main focus of this meta-analysis is on examining inconsistent findings concerning age differences in drift rate, the literature search concentrated on studies comparing mean drift rates between two age groups. For these studies, we additionally coded effect sizes for boundary separation and non-decision time. Below, we report our procedure in detail.

Inclusion Criteria and Literature Search

For our literature search, we used the following two inclusion criteria:

1. All studies had to refer to the original publication introducing the diffusion model in psychology (Ratcliff, 1978) and they had to report results from a diffusion model analysis. Articles applying the EZ-diffusion model (Wagenmakers, van der Maas, & Grasman, 2007) were included. However, we excluded all studies in which parameter estimation was based on fitting the Ex-Gaussian or the shifted Wald distributions due to concerns about the interpretability of their parameters (Matzke & Wagenmakers, 2009).
2. The second required inclusion criterion regards an obligatory comparison between younger adults (college age) and healthy older adults (youngest participant older than 55 or mean age > 60). We excluded studies reporting continuous age analyses if no categorical age data could be extracted from the reported results (e.g., the relation between age and the corresponding parameter of the diffusion model was only provided as a correlation, without raw data being available). We included studies reporting results from more than two age groups if college-aged adults and older adults were among these groups. In case of two higher age groups, we used only the younger one of them to enhance comparability between studies.

We used Google Scholar's search engine to collect studies, as it allows to combine a descendant approach with the use of specific keywords (for the comparability of Google Scholar with established scientific databases, see Anders & Evans, 2010; Gehanno, Rollin, & Darmoni, 2013; Shultz, 2007). In a first step, we identified all the papers citing Ratcliff's seminal work on the diffusion model (1978) (*k* = 3341). The next step consisted in searching these studies using age-related terms[1], resulting in *k* = 561 publications. The search was finished on January 16, 2019.

We conducted a full-text scan of these papers searching for studies that fulfilled the inclusion criteria, resulting in *k* = 48 articles. After removal of duplicates, *k* = 46 articles remained. Several articles did not report sufficient data to calculate effect sizes on drift rate, resulting in a final dataset of 21 papers. Some papers reported data from more than one sample (e.g., if more than one experiment conducted on different participants is reported in the same publication) and/or more than one effect size per sample (e.g., if different tasks were reported for the same participants). We retrieved effect sizes from 25 samples. For boundary separation and non-decision time, we had to exclude one sample, respectively, as the reported data was not sufficient. In total, we retrieved 146 effect sizes for drift rate, 47 effect sizes for boundary separation, and 40 effect sizes for non-decision time.

Calculation of Effect Sizes

As effect size measure we used Hedges' *g* (Hedges, 1981). We computed effect sizes using the *compute.es* package (version 0.2-4; Del Re, 2013) of the *R* open-source software environment (version 3.5.1; R Core Team, 2018). If a paper did not report means or standard deviations, we used inferential statistics to determine effect sizes. Positive effect sizes indicate higher values for higher age.

Coding of Moderator Variables

For each study, we coded the *type of task* using the categories described in Voss, Nagler, and Lerche (2013). Following this classification, most binary decision tasks analyzed with the

[1] search string: ("age differences" OR "old adults" OR "old participants" OR "older adult" OR "older adults" OR "older participants" OR "higher age" OR "older group" OR "old group" OR "age-related" OR "effects of aging" OR "effects of age" OR "aging effects" OR "age effects")

diffusion model are either perceptual, lexical decision or memory tasks. Some experimental tasks did not fit this classification scheme. We omitted the according effect sizes from the analyses (11 effect sizes for drift rate, 2 effect sizes for boundary separation and 4 effect sizes for non-decision time). See Table 1 for the articles included in this final dataset.

A second moderator variable in our analyses was *task difficulty*. We used drift rate as measure of task difficulty as the literature suggests that more difficult tasks go along with lower drift rates (e.g., Arnold, Bröder, & Bayen, 2015; Voss, Rothermund, & Voss, 2004). In several studies, task difficulty varied between conditions. Here we computed a mean drift rate across the different difficulty levels and age groups (weighting by the number of participants per group). Next, we z-transformed and inverted the mean drift so that higher values of the variable indicate enhanced task difficulty.

Statistical Analyses

As several effect sizes are based on the same samples, we assumed effect sizes to be dependent. We accounted for this dependent structure by conducting multilevel meta-analyses using the *metafor* package (version 2.0-0; Viechtbauer, 2010) in *R*. We specified the levels as effect size nested in sample with task type as an inner grouping factor. This means that effects stemming from different samples are assumed to be independent, while effects of the same task type within a sample share correlated random effects. The variance structure of the inner factor was assumed to be a heteroscedastic compound symmetric structure.[2]

We used the maximum likelihood estimation procedure included in the function rma.mv( ) and analyzed the three outcome variables in independent sets of analyses (i.e., drift

[2] As recommended by Wolfgang Viechtbauer (Personal communication, April 2018).

rate, boundary separation, and non-decision time). In a first step, we ran multilevel meta-analyses without any moderators (Model 1). Then, in a second step, we included task type and task difficulty as moderators (Model 2). As we were also interested in a possible interaction between task type and task difficulty, we further added the interactions in a third step (Model 3).

The validity of meta-analytical models can suffer because of influential outliers. To date, the development of tools for outlier and influence diagnostics for multilevel meta-analyses is still in progress (Viechtbauer & Cheung, 2010). We followed the procedure of Habeck and Schultz (2015), removing any influential outliers, defined by effect sizes with both hat values greater than two times the average hat value and standardized residual values exceeding 3.29.

We tested for publication bias using Egger's regression test (Egger, Smith, Schneider, & Minder, 1997; Sterne & Egger, 2005) by modifying Model 1 to include the variance of the effect size as moderator (Moreno et al., 2009). If the intercept of this model significantly deviates from zero, the relationship between variance and effect size can be assumed to be asymmetrical, indicating a bias. Because of the low power of this test for publication bias, we set the alpha-level to $\alpha = .10$ (Egger et al., 1997).

Furthermore, we assessed heterogeneity among effect sizes using Cochran's Q statistic and the I^2 statistic. Large Q values indicate that differences among effect sizes can be attributed to differences among the true effects and do not solely result from sampling errors. If the Q test is significant, the integrated effect size is not an estimator of the true effect but rather an estimator of the mean of the distribution of different true effect sizes (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Results

Study Characteristics

The included studies stem from the period of 2003 to 2018. In total, we analyzed the data of 1,503 participants (M = 62.63 per sample, SD = 34.90). The mean age of the young groups was 21.15 (SD = 1.75), the mean age of the older groups was 69.77 (SD = 2.17). Table 2 shows the distribution of task types over the three diffusion model parameters (see Table S1 in the Supplementary Material for a detailed description of the respective task and condition for each included effect size).

Diagnostics

For drift rate, there were two cases with standard residual values greater than 3.29. However, their hat values did not exceed 2 and, therefore, we did not omit them from the analyses. For boundary separation and non-decision time, we found no outliers. The intercepts of Egger's regression models indicated publication bias for all three diffusion model parameters (drift rate: $\beta_0 = 1.049$, $p < .001$; non-decision time: $\beta_0 = 0.988$, $p < .001$; boundary separation: $\beta_0 = -0.637$, $p = .063$).

Meta-analysis

Drift rate

The meta-analytical model with task type and task difficulty as moderators (Model 2; AICc = 376.2) had a better fit than the model without moderators (Model 1; AICc = 379.1, $p = .022$). Including the interaction between task type and difficulty improved the model fit even further (Model 3; AICc = 374.1, $p = .034$). Thus, our final meta-analytical model contained task type, task difficulty, and their interaction as moderator variables (see Figure S1 in the Supplementary Material for a forest plot of the final model). For the final model, the Q test was highly significant, $Q(129) = 1016.331$, $p < .001$. The estimated standard deviations of true effects per task type were $\tau = 0.848$ (perceptual tasks), $\tau = 1.153$ (lexical decision tasks), and $\tau = 0.549$ (memory tasks). The I^2 values for the three levels of task type were 91.61% (perceptual tasks), 95.28% (lexical decision tasks), and 82.07% (memory tasks)[3].

The mean effect sizes per task type were $g = -0.608$, 95% CI [-1.032, -0.184], $p = .005$ for perceptual tasks, $g = 0.620$, 95% CI [0.037, 1.203], $p = .037$ for lexical decision tasks and $g = -0.326$, 95% CI [-0.587, -0.065], $p = .014$ for memory tasks (see Figure 2 for a graphical representation). This indicates reduced drift rates for older adults in perceptual and memory tasks but increased drift rates of older adults for lexical decision tasks. Furthermore, in more difficult tasks older adults performed relatively better compared to younger adults ($\beta = 0.181$, $p = .010$).

To examine the task type by task difficulty interaction, we additionally conducted separate analyses for each type of task, with and without task difficulty as moderator. We then compared the fit of the model with and without difficulty to test if this moderator explains variance within a task type (Table 3). For perceptual and lexical decision tasks, the model with difficulty as moderator performed significantly better than the model without moderator. More specifically, task difficulty significantly predicted effect sizes for perceptual ($\beta = 0.203$, $p = .004$) and lexical decision tasks ($\beta = 0.719$, $p = .022$): Older adults profited from high task difficulty. For memory tasks, on the other hand, task difficulty did not predict effect sizes, $\beta = 0.016$, $p = .816$. In the supplementary materials, we provide a full forest plot showing all drift rate effect sizes analyzed.

[3]To compute the I^2 on task type level we used the approach for multivariate models as described in Viechtbauer (2018, December 8). To compute the I^2 for boundary separation and non-decision time, we used Higgins and Thompson's (2002) formula.

Boundary separation

The meta-analytical model without any moderators (AICc = 143.8) showed a better fit to the data than the model with task type and task difficulty as moderators (AICc = 149.1, $p = .372$). Therefore, we kept the model without any moderators. The mean effect size of age on boundary separation was $g = 0.731$, 95% CI [0.472, 0.989], $p < .001$. Results indicate that older adults generally adopt higher boundary separations (i.e., a more conservative response criterion) than young adults. The Q test was highly significant, $Q(44) = 669.203$, $p < .001$; I^2 for the whole model was 93.13%.

Non-decision time

The meta-analytical model without any moderators (AICc = 89.9) showed a better fit than the model with task type and task difficulty (AICc = 96.1, $p = .379$). Therefore, we kept the model without any moderators. The mean effect size of age on non-decision time was $g = 1.673$, 95% CI [1.404, 1.942], $p < .001$. Our results suggest that older people show a longer non-decision time than younger people. The Q test was highly significant, $Q(37) = 388.946$, $p < .001$; I^2 for the whole model was 90.487%.

Discussion

In the last decades, the diffusion model (Ratcliff, 1978) has become a popular approach for the analysis of age differences in response time tasks. The findings from the diffusion model analyses have challenged the view that cognitive processing speed generally declines with age. Rather, the studies revealed a more complex picture, which we wanted to examine further in our meta-analysis. Most importantly, we were interested in the drift rate, which is a measure of speed of information accumulation that is closely related to intelligence (e.g., Ratcliff, Thapar, & McKoon, 2011; Schmitz & Wilhelm, 2016). Regarding age effects on drift, previous studies provided inconsistent findings. Whereas some studies report reduced drift rates for older adults (e.g., Thapar et al., 2003), other studies do not find differences in this model parameter (e.g., Ratcliff et al., 2001), or even higher drift rates for older adults (e.g., Ratcliff, Thapar, & McKoon, 2010). With the present meta-analysis, we aim to identify reasons for this heterogeneity. To this aim, we assembled studies that report drift rate differences between a young (college age) and an old age group (> 55 years). Then, we examined the influence of task difficulty and task type (perceptual, lexical decision, and memory) on age effects in diffusion model parameters. Thus, we could uncover possible important moderators that might explain (part of) the inconsistent findings in the literature.

Boundary Separation and Non-decision Time

Results provided two most clear-cut findings: First, older adults are slower than young adults in non-decisional processes (such as encoding of information and motoric response execution). The corresponding effect size was large ($g = 1.673$). Second, older adults generally use more conservative response criteria (i.e., larger boundary separations) than young adults. Even if the effect size is smaller than for non-decision time, it is still substantial ($g = 0.731$). Thus, older individuals are more cautious in their decisions. These effects did not depend on either task type or task difficulty. Note that both boundary separation and non-decision time influence RT (e.g., Ratcliff & McKoon, 2008). Thus, the common finding of higher RTs of older adults seems to be highly attributable to these two parameters.

Drift Rate

Whereas age differences in boundary separation and non-decision time generalized across different task types and difficulties, we found moderator effects for speed of information accumulation (drift rate). In perceptual tasks and memory tasks, older adults had lower drift rates than younger adults. However, the older groups were superior in speed of information accumulation compared to their younger counterparts in lexical decision tasks. Furthermore, task difficulty also influenced age effects on drift: In terms of perceptual and lexical decision tasks, older participants profited from more difficult tasks. In the memory tasks, task difficulty did not moderate the effect of age on drift.

Thus, our study shows that the pattern of results is clearly more complex for drift rate than for boundary separation and non-decision time and that it seems to be important to consider the specific cognitive processes required by different experimental paradigms. In line with this finding are the results from a recent diffusion model study that is based on a set of 18 different RT tasks (Lerche et al., 2020). The study revealed domain-specific drift factors (numeric, verbal, figural) that were further related to the respective components of an intelligence test. Thus, speed of information accumulation seems to be dependent on the task content. Furthermore, also neurophysiological studies found that aging effects depend on the task (Dully et al., 2018).

Our meta-analysis suggests that older adults outperform young adults in lexical decision tasks, whereas they perform worse in memory tasks. This is in line with the findings from studies that date back to the 1920s and 30s (e.g., Conrad, Jones, & Hsiao, 1933; Foster & Taylor, 1920; Willoughby, 1929). The results of these studies suggest that age differences are more pronounced in measures of memory than vocabulary. Also, recent studies generally confirm this observation. For example, Salthouse (2004), aggregating across several studies from 1998 till 2003, reports a substantial, linear age decline in performance in a memory test. On the other hand, performance improved with age in a vocabulary test, at least until about the mid-50s. After that, it remained stable or somewhat declined (confer also Spaniol et al., 2006). Our meta-analysis showed that such task-specificities are captured in the drift rate of the diffusion model. Furthermore, our analysis revealed that not only the type of task, but also task difficulty needs to be considered. Older adults profited from the more difficult tasks.

Limitations and Future Research

Even if the overall number of effect sizes for drift rate used for the meta-analysis was substantial ($N = 135$), analyses of the moderator task type were based on smaller case numbers. Here, the distribution was not balanced with clearly fewer lexical decision effect sizes ($n = 16$) than effect sizes from perceptual ($n = 30$) or memory tasks ($n = 89$). In future research, one might try to replicate our findings in large-scale studies that are explicitly designed to measure the influence of task type (and difficulty) on age differences. Further note that despite consideration of two moderator variables, there was still substantial unexplained variance in our meta-analysis. Accordingly, in future studies, one might try to identify further possible moderators.

The focus of our meta-analysis was on drift rate because findings in the literature seemed to be more inconsistent for this parameter. Therefore, our search strategy was based on finding all studies that report age differences in drift rate. With this strategy we do not identify studies that report age differences only in boundary separation or non-decision time, but not in drift rate. Accordingly, the superiority of the model without moderators might also be partly attributable to the small cell numbers (between 6 and 23 for the different task types). Thus, if one would like to examine moderator influences for non-decision time and boundary separation in more detail, separate meta-analyses should be conducted.

Finally, it would be highly interesting to examine age effects more systematically also for other sequential sampling models, e.g., the popular linear ballistic accumulator model (LBA; Brown & Heathcote, 2008). So far, the literature on age effects in LBA model parameters is more limited than the respective diffusion model literature. The previous LBA findings seem to be generally in line with the results from our meta-analysis: In comparison with younger adults, older adults have been found to have higher threshold separations (Forstmann et al., 2011; Garton, Reynolds, Hinder, & Heathcote, 2019), and longer non-decision times (Ben-David, Eidels, & Donkin, 2014; Garton et al., 2019), whereas the results for drift rate are less clear-cut. Further, previous research suggests that the diffusion model parameters and the LBA model parameters have very similar meanings (Donkin, Brown, Heathcote, & Wagenmakers, 2011). However, to note, in a recent multi-lab project one systematic difference between the two models emerged (Dutilh et al., 2019). More specifically, for instructions that focused either on accuracy or speed the teams that used the diffusion model often found an effect in non-decision time (in addition to an effect in threshold separation), whereas the LBA teams often detected an effect in drift rate. The reasons for this pattern of results will need to be investigated further in future research (for a recent discussion of this topic, see Evans, 2020; Lerche & Voss, 2018). Based on these varying findings, we hypothesize that somehow different age effects might emerge if older and younger adults are compared based on different sequential sampling models. For example, effect sizes for age effects in non-decision time might be larger for the diffusion model than for the LBA.

Compliance with Ethical Standards

Funding: This research was supported by grants from the German Research Foundation to Andreas Voss (Grant Nr. VO1288/2-2) and to the Graduate School SMiP (GRK 2277; Statistical Modeling in Psychology).

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

Allen, P. A., Lien, M.-C., Ruthruff, E., & Voss, A. (2014). Multitasking and aging: Do older adults benefit from performing a highly practiced task? *Experimental Aging Research, 40*(3), 280-307. doi: 10.1007/s00426-014-0608-y

Anders, M. E., & Evans, D. P. (2010). Comparison of PubMed and Google Scholar Literature Searches. *Respiratory Care, 55*(5), 578-583.

Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research, 79*(5), 882-898. doi: 10.1007/s00426-014-0608-y

Aschenbrenner, A. J., Balota, D. A., Gordon, B. A., Ratcliff, R., & Morris, J. C. (2016). A diffusion model analysis of episodic recognition in preclinical individuals with a family history for Alzheimer's disease: The adult children study. *Neuropsychology, 30*(2), 225-238. doi: 10.1037/neu0000222

Ball, B. H., & Aschenbrenner, A. J. (2018). The importance of age-related differences in prospective memory: Evidence from diffusion model analyses. *Psychonomic Bulletin & Review, 25*(3), 1114-1122. doi: 10.3758/s13423-017-1318-4

Ben-David, B. M., Eidels, A., & Donkin, C. (2014). Effects of Aging and Distractors on Detection of Redundant Visual Targets and Capacity: Do Older Adults Integrate Visual Targets Differently than Younger Adults? *PLoS ONE, 9*(12), 1-29. doi: 10.1371/journal.pone.0113551

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., . . . Wagenmakers, E.-J. (2018). Estimating across-trial variability parameters of the Diffusion Decision Model: Expert advice and recommendations. *Journal of Mathematical Psychology, 87*, 46-75. doi: 10.1016/j.jmp.2018.09.004

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153-178. doi: 10.1016/j.cogpsych.2007.12.002

Conrad, H. S., Jones, H. E., & Hsiao, H. H. (1933). Sex differences in mental growth and decline. *Journal of Educational Psychology, 24*(3), 161-169. doi: 10.1037/h0073913

Del Re, A. C. (2013). compute.es: Compute Effect Sizes. R package version 0.2-2. Retrieved from http://cran.r-project.org/web/packages/compute.es

Dirk, J., Kratzsch, G. K., Prindle, J. P., Kröhne, U., Goldhammer, F., & Schmiedek, F. (2017). Paper-based assessment of the effects of aging on response time: A diffusion model analysis. *Journal of Intelligence, 5*(12). doi: 10.3390/jintelligence5020012

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review, 18*(1), 61-69. doi: 10.3758/s13423-010-0022-4

Dully, J., McGovern, D. P., & O'Connell, R. G. (2018). The impact of natural aging on computational and neural indices of perceptual decision making: A review. *Behavioural Brain Research, 355*, 48-55. doi: 10.1016/j.bbr.2018.02.001

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P. P. P., . . . Donkin, C. (2019). The Quality of Response Time Data Inference: A Blinded, Collaborative

Assessment of the Validity of Cognitive Models. *Psychonomic Bulletin & Review, 26*(4), 1051-1069. doi: 10.3758/s13423-017-1417-2

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj-British Medical Journal, 315*(7109), 629-634. doi: 10.1136/bmj.315.7109.629

Evans, N. J. (2020). Think fast! The implications of emphasizing urgency in decision-making. . *PsyArXiv*. doi: https://doi.org/10.31234/osf.io/pfrb4

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *Journal of Neuroscience, 31*(47), 17242-17249. doi: 10.1523/JNEUROSCI.0309-11.2011

Foster, J. C., & Taylor, G. A. (1920). The applicability of mental tests to persons over fifty years of age. *Journal of Applied Psychology, 4*(1), 39-58. doi: 10.1037/h0071260

Garton, R., Reynolds, A., Hinder, M. R., & Heathcote, A. (2019). Equally flexible and optimal response bias in older compared to younger adults. *Psychology and Aging, 34*(6), 821-835. doi: 10.1037/pag0000339

Gehanno, J. F., Rollin, L., & Darmoni, S. (2013). Is the coverage of google scholar enough to be used alone for systematic reviews. *Bmc Medical Informatics and Decision Making, 13*. doi: 10.1186/1472-6947-13-7

Habeck, C. W., & Schultz, A. K. (2015). Community-level impacts of white-tailed deer on understorey plants in North American forests: a meta-analysis. *Aob Plants, 7*. doi: 10.1093/aobpla/plv119

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128. doi: 10.2307/1164588

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-1558. doi: 10.1002/sim.1186

Huff, M. J., & Aschenbrenner, A. J. (2018). Item-specific processing reduces false recognition in older and younger adults: Separating encoding and retrieval using signal detection and the diffusion model. *Memory & Cognition, 46*(8), 1287-1301. doi: 10.3758/s13421-018-0837-1

Kapucu, A. (2010). *Emotional recognition memory for younger and older adults: Combining ROC analysis and the diffusion model (Dissertation).* University of Massachusetts Amherst, Amherst, MA.

Kordella, B. C. (2009). *A diffusion model analysis of the effects of aging on sentence memory (Master thesis).* Ohio State University, Athens, OH.

Kühn, S., Schmiedek, F., Schott, B., Ratcliff, R., Heinze, H.-J., Düzel, E., . . . Lövden, M. (2011). Brain areas consistently linked to individual differences in perceptual decision-making in younger as well as older adults before and after training. *Journal of Cognitive Neuroscience, 23*(9), 2147-2158. doi: 10.1162/jocn.2010.21564

Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making, 6*(7), 651-687.

Lerche, V., Christmann, U., & Voss, A. (2018). Impact of Context Information on Metaphor Elaboration: A Diffusion Model Study. *Experimental Psychology, 65*(6), 370-384. doi: 10.1027/1618-3169/a000422

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*. doi: 10.1037/xge0000774

Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology, 7*(1324). doi: 10.3389/fpsyg.2016.01324

Lerche, V., & Voss, A. (2017). Experimental Validation of the Diffusion Model based on a Slow Response Time Paradigm. *Psychological Research*. doi: 10.1007/s00426-017-0945-8

Lerche, V., & Voss, A. (2018). Speed–accuracy manipulations and diffusion modeling: Lack of discriminant validity of the manipulation or of the parameter estimates? *Behavior Research Methods, 50*(6), 2568-2585. doi: 10.3758/s13428-018-1034-7

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16*(5), 798-817. doi: 10.3758/PBR.16.5.798

McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in item recognition. *Journal of Memory and Language, 66*(3), 416-437. doi: 10.1016/j.jml.2011.12.001

McKoon, G., & Ratcliff, R. (2013). Aging and predicting inferences: A diffusion model analysis. *Journal of Memory and Language, 68*(3), 240-254. doi: 10.1016/j.jml.2012.11.002

Moreno, S. G., Sutton, A. J., Ades, A. E., Stanley, T. D., Abrams, K. R., Peters, J. L., & Cooper, N. J. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *Bmc Medical Research Methodology, 9*. doi: 10.1186/1471-2288-9-2

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59-108.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: fitting real data and failing to fit fake but plausible data. *Psychon Bull Rev, 9*(2), 278-291. doi: 10.3758/bf03196283

Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging, 23*(4), 900-916. doi: 10.1037/a0013930

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review, 111*(1), 159-182. doi: 10.1037/0033-295x.111.1.159

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput, 20*(4), 873-922. doi: 10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347-356. doi: Doi 10.1111/1467-9280.00067

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging, 19*(2), 278. doi: 10.1037/0882-7974.19.2.278

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychol Aging, 16*(2), 323-341.

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Percept Psychophys, 65*(4), 523-535. doi: 10.3758/bf03194580

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*(4), 408-424. doi: 10.1016/j.jml.2003.11.002

Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review, 13*(4), 626-635. doi: 10.3758/BF03193973

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology, 60*(3), 127-157. doi: 10.1016/j.cogpsych.2009.09.001

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General, 140*(3), 464-487. doi: 10.1037/a0023810

Ratcliff, R., Thompson, C. A., & McKoon, G. (2015). Modeling individual differences in response time and accuracy in numeracy. *Cognition, 137*(0), 115-136. doi: 10.1016/j.cognition.2014.12.004

Ratcliff, R., & Van Dongen, H. P. A. (2009). Sleep deprivation affects multiple distinct cognitive processes. *Psychonomic Bulletin & Review, 16*(4), 742-751. doi: 10.3758/PBR.16.4.742

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428. doi: 10.1037/0033-295X.103.3.403

Salthouse, T. A. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science, 13*(4), 140-144. doi: 10.1111/j.0963-7214.2004.00293.x

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), 414-429. doi: 10.1037/0096-3445.136.3.414

Schmitz, F., & Wilhelm, O. (2016). Modeling Mental Speed: Decomposing Response Time Distributions in Elementary Cognitive Tasks and Correlations with Working Memory Capacity and Fluid Intelligence. *Journal of Intelligence, 4*(4), 13.

Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A., & Bergmann, K. (2015). Decomposing the relationship between mental speed and mental abilities. *Intelligence, 51*, 28-46. doi: 10.1016/j.intell.2015.05.002

Shultz, M. (2007). Comparing test searches in PubMed and google scholar. *Journal of the Medical Library Association, 95*(4), 442-445. doi: 10.3163/1536-5050.95.4.442

Spaniol, J., Madden, D. J., & Voss, A. (2006). A Diffusion Model Analysis of Adult Age Differences in Episodic and Semantic Long-Term Memory Retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 101-117. doi: 10.1037/0278-7393.32.1.101

Spaniol, J., Voss, A., Bowen, H. J., & Grady, C. L. (2011). Motivational incentives modulate age differences in visual perception. *Psychology and Aging, 26*(4), 932-939. doi: 10.1037/a0023297

Spaniol, J., Voss, A., & Grady, C. L. (2008). Aging and emotional memory: Cognitive mechanisms underlying the positivity effect. *Psychology and Aging, 23*(4), 859-872. doi: 10.1037/a0014218

Sterne, J. A., & Egger, M. (2005). Regression Methods to Detect Publication and Other Bias in Meta-Analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication Bias in Meta-Analysis* (pp. 99–110). Chichester, UK: John Wiley & Sons, Ltd.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging, 18*(3), 415-429. doi: 10.1037/0882-7974.18.3.415

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review, 24*(2), 547-556. doi: 10.3758/s13423-016-1081-y

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software, 36*(3), 1-48.

Viechtbauer, W. (2018, December 8). I² for Multilevel and Multivariate Models. Retrieved from: http://www.metafor-project.org/doku.php/tips:i2_multilevel_multivariate/

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112-125. doi: 10.1002/jrsm.11

Voskuilen, C., Ratcliff, R., & McKoon, G. (2018). Aging and confidence judgments in item recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(1), 1-23. doi: 10.1037/xlm0000425

Voss, A., Lerche, V., Mertens, U., & Voss, J. (2019). Sequential sampling models with variable boundaries and non-normal noise: A comparison of six models. *Psychonomic Bulletin & Review, 26*(3), 813-832. doi: 10.3758/s13423-018-1560-4

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: a practical introduction. *Exp Psychol, 60*(6), 385-402. doi: 10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: an empirical validation. *Mem Cognit, 32*(7), 1206-1220.

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21*(5), 641-671. doi: 10.1080/09541440802205067

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58*(1), 140-159. doi: 10.1016/j.jml.2007.04.006

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3-22. doi: 10.3758/bf03194023

Willoughby, R. R. (1929). Incidental learning. *Journal of Educational Psychology, 20*(9), 671-682. doi: 10.1037/h0071404

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *J Exp Psychol Learn Mem Cogn, 41*(3), 597-613. doi: 10.1037/xlm0000064

Table 1

Samples included in the final dataset of the meta-analysis

Articles	*n*	*n* young	*n* old	Age range young	Age range old	Mean age young	Mean age old
Allen, Lien, Ruthruff, and Voss (2014)	21	11	10	18-24	64-80	21.7	71.8
Ball and Aschenbrenner (2018)	125	67	58	18-21	60-90	18.9	75.0
Dirk et al. (2017)	40	20	20	18-36	64-75	25.7	68.1
Huff and Aschenbrenner (2018)	163	85	78			21	74.6
Kapucu (2010)	56	30	26			19.8	71.9
Kordella (2009)							
Experiment 2	41	22	19	18-24	61-74	20.2	68.9
Experiment 3	38	22	16	18-25	60-74	20.1	68.3
Kühn et al. (2011)	39	24	15	20-31	65-80	25.2	70.2
McKoon and Ratcliff (2012)	78	39	39	18-25	60-74	20.6	68.4
McKoon and Ratcliff (2013)	67	30	37		60-74	20.8	69.7
Ratcliff (2008)	38	19	19		60-75	20.8	69.2
Ratcliff, Thapar, Gomez, and McKoon (2004)							
Experiment 1	98	54	44			19.8	68.5
Experiment 2	94	54	40			20.2	67.2
Ratcliff, Thapar, and McKoon (2004)	80	39	41		60-74	19.6	70
Ratcliff, Thapar, and McKoon (2006)	20	10	10		60-74		
Ratcliff, Thapar, and McKoon (2010)	88	45	43	18-25	60-74	20.8	68.6
Ratcliff, Thapar, and McKoon (2011)	91	46	45		60-74	20.4	68.3
Spaniol, Voss, and Grady (2008)							
Experiment 1	43	22	21	19-28	60-75	22.5	67.5
Experiment 2	47	24	23	18-32	61-85	22.3	71.8
Spaniol, Voss, Bowen, and Grady (2011)	53	26	27	18-32	61-85	23.0	71.5
Thapar, Ratcliff, and McKoon (2003)	78	40	38		60-75	19.8	69.1
Voskuilen et al. (2018)	23	11	12		60-80		

Table 2

Number of available effect sizes for each diffusion model parameter depending on the task type

Parameter	Perceptual tasks	Lexical decision tasks	Memory tasks
Drift rate	30	16	89
Boundary separation	16	6	23
Non-decision time	14	6	16

Table 3

Drift rate: Comparisons between models with and without task difficulty as moderator for each task type subset

Statistic	Perceptual tasks	Lexical decision tasks	Memory tasks
$AICc_{\text{with task difficulty}}$	69.710	63.887	244.189
$AICc_{\text{without task difficulty}}$	75.465	65.395	242.100
p	.004	.032	.817

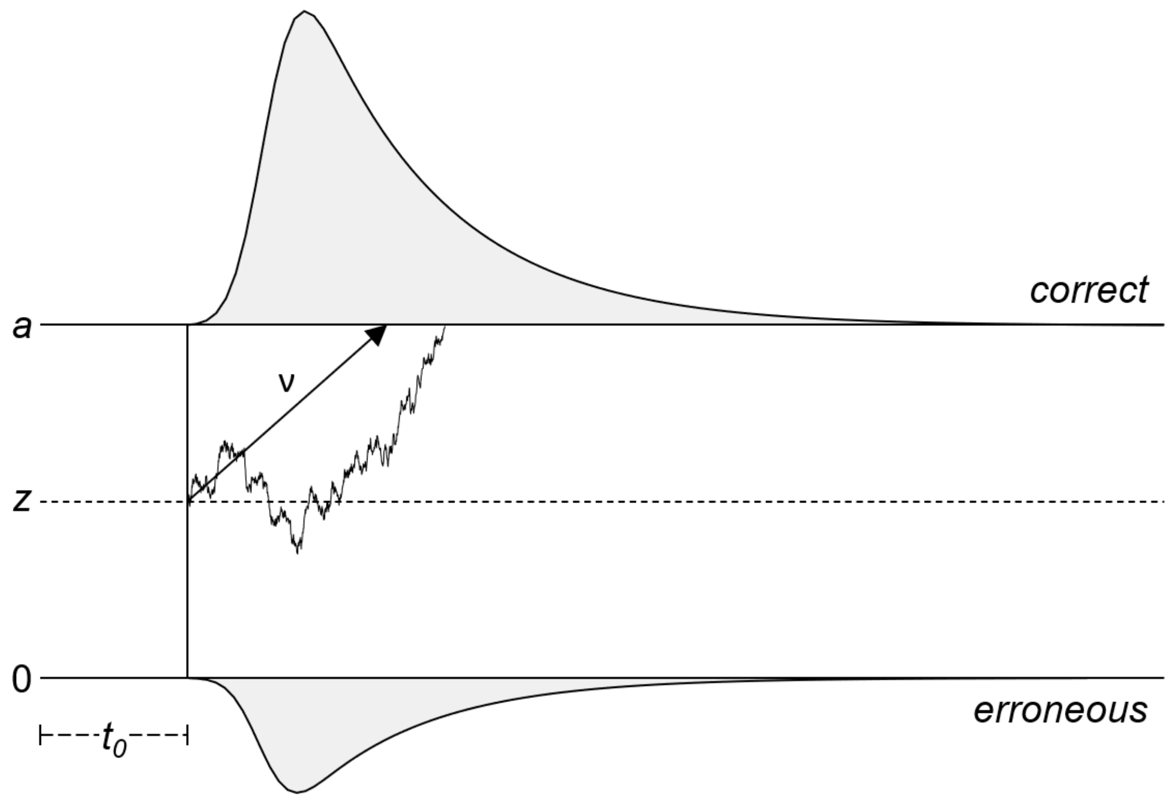

Figure 1. Diffusion model with its four main parameters. The boundaries are associated with correct and erroneous responses here. One exemplary trial is illustrated. In this trial, the accumulation process starts at starting point z, which is here right in the center between the two boundaries (0 and a). The process moves with speed v toward the upper boundary. To this straight process adds random Gaussian noise. For convenience, parameter t_0 is illustrated at the left from the decision process. Note that it also includes processes succeeding the decisional process (the motoric response).

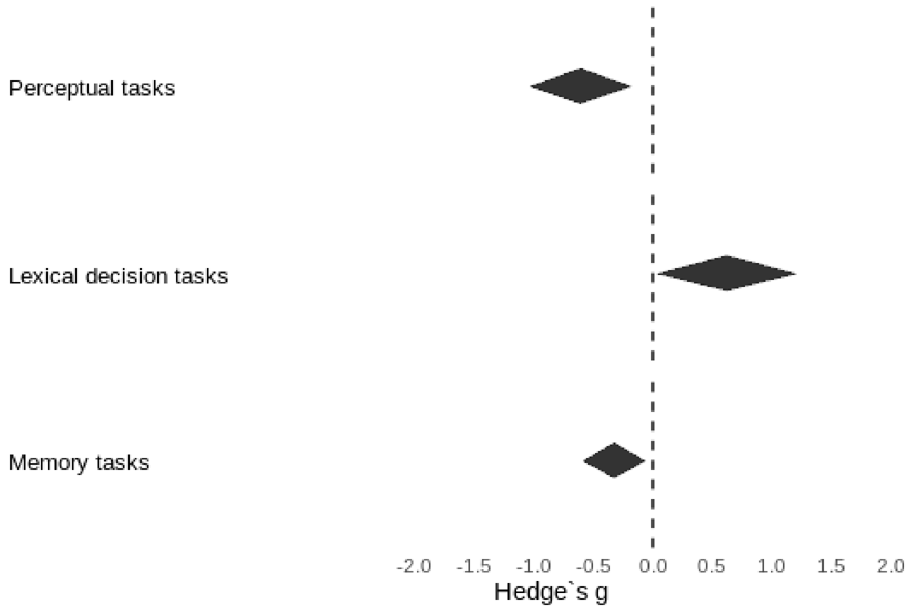

Figure 2. Mean age effects in drift rate for each of the task types analyzed. 95% confidence intervals indicated by the width of the polygons.

Appendix A 4

Manuscript 4:

von Krause, M., Lerche, V., Schubert, A. L., & Voss, A. (2020). Do Non-Decision Times Mediate the Association between Age and Intelligence across Different Content and Process Domains?. *Journal of Intelligence*, *8*(3), 33.

 doi: 10.3390/jintelligence8030033

Do Non-Decision Times Mediate the Association Between Age and Intelligence Across Different Content and Process Domains?

Mischa von Krause[1], Veronika Lerche[1], Anna-Lena Schubert[1], and & Andreas Voss[1]

[1] Heidelberg University

Author Note

This research was supported by grants from the German Research Foundation to Andreas Voss (Grant Nr. VO1288/2-2) and to the Graduate School SMiP (GRK 2277; Statistical Modeling in Psychology). We thank Stefan Radev for his valueable comments on an earlier draft of this paper. We further thank our students Anne Bülow, Sarah Hladik, Henrike Jungeblut, Julia Karl, Sontje Nordholt, Yannick Roos, and Moritz Spielberger for their help during data acquisition.

Correspondence concerning this article should be addressed to Mischa von Krause, Hauptstraße 47-51, 69115 Heidelberg. E-mail: mischa.vonkrause@psychologie.uni-heidelberg.de

Abstract

In comparison to young adults, middle-aged and old people show lower scores in intelligence tests and slower response times in elementary cognitive tasks. Whether these well-documented findings can both be attributed to a general cognitive slow-down across the life-span has become subject to debate in the last years. The drift diffusion model can disentangle three main process components of binary decisions, namely the speed of information processing, the conservatism of the decision criterion and the non-decision time (i.e., time needed for processes such as encoding and motor response execution). All three components provide possible explanations for the association between response times and age. We present data from a broad study using 18 different response time tasks from three different content domains (figural, numeric, verbal). Our sample included people between 18 to 62 years of age, thus allowing us to study age differences across young-adulthood and mid-adulthood. Older adults generally showed longer non-decision times and more conservative decision criteria. For speed of information processing, we found a more complex pattern that differed between tasks. We estimated mediation models to investigate whether age differences in diffusion model parameters account for the negative relation between age and intelligence, across different intelligence process domains (processing capacity, memory, psychometric speed) and different intelligence content domains (figural, numeric, verbal). In most cases, age differences in intelligence were accounted for by age differences in non-decision time. Content domain-general, but not content domain-specific aspects of non-decision time were related to age. We discuss the implications of these findings on how cognitive decline and age differences in mental speed might be related.

Keywords: diffusion modeling, cognitive aging, response time, intelligence

Do Non-Decision Times Mediate the Association Between Age and Intelligence Across Different Content and Process Domains?

Most cognitive abilities decline across the life-span (Hartshorne & Germine, 2015; Timothy A. Salthouse, 2004; Salthouse, 2010). This well-established finding holds true across a variety of process domains (e.g., general intelligence, reasoning, memory) and across different content domains (e.g., figural, numeric, verbal). Only for so-called crystallized abilities (Cattell, 1963), which are largely knowledge-based, ability scores increase until people are in their sixties (Horn & Cattell, 1967). One clear-cut result, found in cross-sectional as well as longitudinal data, is that older adults show slower response times than younger people in elementary cognitive tasks - a pattern that already starts in mid-adulthood (Hartshorne & Germine, 2015; Jensen, 2006; Salthouse, 2010; Schaie, 2005). As response times are linked to intelligence (Sheppard & Vernon, 2008), it has been proposed that these age differences in response times might form the basis for the decline of a wide range of cognitive abilities - cognitive decline in general might be based on a slow-down of basic cognitive processes (Finkel et al., 2007; Salthouse, 1996; Verhaeghen & Salthouse, 1997). In fact, response times have been found to mediate the relationship between age and intelligence, lending support to the assumption that differences in cognitive speed might be responsible for age differences in intelligence (although findings in longitudinal studies show that the correlation between age-related change in processing speed and age-related change in intelligence is lower than the cross-sectional data suggest, see Lindenberger et al., 2011; Zimprich & Martin, 2002).

Salthouse (1996) proposed that an age-related slow-down might affect cognition in two ways. First, because the time available for problem solving is typically limited, less time is available for higher-order information integration if the basic processes in early stages of information processing take too long. Second, based on the idea that information stored in working memory deteriorates over time, a slow-down in early processes might lead to greater losses of information before integration starts. Both accounts assume that processing speed reflects a general component of information processing that generalizes across content domains and task types. Thus, the same base might be responsible for all kinds of cognitive decline, across process domains and content domains. If that is the case, different aspects of cognitive ability should show correlated change. Findings in support of this notion of a general decline have been reported in the literature, although there is also evidence for domain-specific and task-specific aspects (T. Salthouse, 2004; Sliwinski & Hall, 1998; Tucker-Drob, 2011).

Response times measured in elementary cognitive tasks are a widely used and long-established instrument to assess cognitive speed (Jensen, 2006). However, the use of response times as a single indicator leads to at least two problems, both of which are related to the fact that response times are not a process-pure measure of cognitive speed. First, there can be a trade-off between speed and accuracy: Some people might try to respond as quickly as possible at the risk of making more mistakes, whereas others might be more inclined to be as accurate as possible, even if this leads to slower responses. Second, the time needed for sensory encoding and for motoric response execution is intermingled with the time needed for information processing. Thus, mean response times (and response time variances, too) are influenced by several processes that might not actually reflect processing speed.

To gain more process-pure measures, the diffusion model (Ratcliff, 1978; R. Ratcliff, Schmiedek, et al., 2008; Voss et al., 2013) can be applied – a stochastic model that takes into account both accuracy rates and response time distributions from binary decision tasks. Figure 1 shows a graphical representation of the model. The decision process is described as an evidence accumulation process with constant drift and random noise, starting at the point z between two decision boundaries. A decision is made, and motor response execution starts as soon as the evidence accumulation process reaches one of the boundaries. One main advantage of the diffusion model is that it allows to disentangle (a) speed-accuracy trade-offs, (b) the speed of information processing, and (c) non-decisional components of response times. Among others, the model yields estimates of three parameters, that reflect (a) the conservatism of the decision criterion (i.e., boundary separation a), (b) the speed of information processing or the quality of the evidence entering the decision process (i.e., drift rate v), and (c) the time needed for encoding and motor response execution (i.e., non-decision time t_0). Experimental studies have demonstrated that these diffusion model parameters are valid measures of the respective components of the decision process (Arnold et al., 2015; Lerche & Voss, 2017; Voss et al., 2004).

The diffusion model thus provides parameter estimates that allow a model-based scrutiny of why older people's response are slower. Are elder persons more careful in selecting the correct answer, focusing less on speed? Are they slower in their speed of information accumulation? Or do they take longer for encoding and motor response execution? Or does age-related slowing reflect a combination of these processes? The answers to these questions hint at different interpretations of what underlies the correlation between age differences in response times and age differences in cognitive abilities.

There is a growing number of studies on age differences in diffusion model parameters (Ball & Aschenbrenner, 2018; Janczyk et al., 2018; Madden et al., 2010; McGovern et al., 2018; McKoon & Ratcliff, 2012, 2013; Ratcliff et al., 2004, 2010, 2001; R. Ratcliff, Thapar, et al., 2006a; R. Ratcliff, 2008; Spaniol et al., 2006, 2011; Thapar et al., 2003; Voskuilen et al., 2018). Dully et al. (2018) gave an overview of the state of the literature in their systematic review. They found consistent and robust age effects for boundary separation a and non-decision time t_0. This suggests that elder people respond slower because they employ more conservative decision criteria and need more time for extra-decisional processes. In contrast to these clear-cut findings, age differences in drift rate vary notably across studies. This finding is surprising as it implies that there might be no general age-related slow-down in information processing. Age differences in response times might arise primarily or even exclusively due to the fact that older people are more careful and take longer for encoding and motor processes. In a recent meta-analysis of age differences in diffusion model parameters summarizing 25 samples, Theisen et al. (2020) studied task type as potential moderator of the link between age and drift rate. The authors found small negative age effects on drift rate in memory and simple perception tasks, but small positive age effects for drift rate in lexical decision tasks. The latter might be explained by the fact that performance in lexical decision tasks is partly based on vocabulary knowledge, an aspect of cognition that has been found to peak later in life than most other cognitive abilities, showing increases at least until the age of 50 (Hartshorne & Germine, 2015; Horn & Cattell, 1967; Timothy A. Salthouse, 2004). Theisen et al. (2020) further examined task difficulty as a potential moderator. The meta-analysis suggested that in perceptual and lexical decision tasks older adults profited from increased task difficulty. However, for the moderator analyses the number of available data sets was rather low so that these results should be interpreted with caution. Nevertheless, the findings from this meta-analysis suggest that age effects in drift rate might be highly dependent on the type of task (e.g., its domain and difficulty). An important limitation of most previous studies on age differences in diffusion model parameters is that they used only very few different tasks, typically only one (Spieler, 2001). Thus, it remains an open question whether the effects found in different studies for different tasks are comparable.

Extending this argument, it should be noted that the studies examined in the meta-analysis all employed tasks with relatively short latencies and thus a restricted variance in complexity. In the past, most tasks analyzed with the diffusion model had mean response times of less than 1.5 seconds. However, recently, it has been demonstrated that the diffusion model also provides a good fit and valid results for more complex tasks with mean response times that are notably above 1.5 seconds (Lerche et al., 2018 ; Lerche & Voss, 2017). In the present study, we will draw upon these findings and analyze the cognitive processes underlying age-related slowing based on a much larger variation of task complexity. Furthermore, previous diffusion model studies (e.g., Ratcliff et al., 2004, 2010, 2001; McKoon & Ratcliff, 2012; R. Ratcliff, 2008) typically used a group design, comparing young adults (i.e., college age) to old adults with a mean age above 60. It is an open question whether these results are generalizable to other age classes, that is whether there are linear age trends for the model parameters across the whole span of adulthood. In our study, we focus on differences across young- and mid-adulthood, employing a continuous measure of age.

After establishing that there are systematic effects of age on the diffusion model parameters, the next step is to assess the role of these effects in age related differences in outcome measures like intelligence. Do changes in diffusion model parameters mediate the link between age and intelligence in the same way as mean response times do? Recently, Schubert et al. (2020) reported first answers to this question. Using two different response time tasks, they found that non-decision time and latencies in event-related potentials (ERP) in the P3 component of the electroencephalogram (EEG) mediated the relationship between age and IQ as measured in a standard intelligence test. In contrast, age did not mediate the correlation between non-decision time and IQ, implying that the link between non-decision time and intelligence is not due to a confounding between age and non-decision time. The model parameter non-decision time is thought to reflect the time needed for encoding processes and motor response execution. As the authors did not find non-decision time to be related to early ERP latencies that might reflect encoding (i.e., N1 and P1), they proposed two possible (contrasting) explanations for the observed mediation effect of non-decision time: First, differences in non-decision time might reflect age-related differences in anterior brain regions that are associated with motor planning and response execution. Importantly, the same anterior brain regions might also affect latencies of ERP components occurring later in the stream of information-processing such as the P3 that are closely related to higher-order processing and intelligence (Schubert et al., 2017 ; Schubert & Frischkorn, 2020). Second, the mediation via non-decision time might reflect the influence of non-decisional processes on the intelligence test scores, because the test used (Berlin Intelligence Structure Test; Jäger et al., 1997) has strict time limits for each task and scores are thus affected by the speed of motor response execution (i.e., hand-writing). One limitation of the results reported by Schubert et al. (2020) is the low number of response time tasks that were used in their study. The authors applied the Sternberg memory task and the Posner letter matching task, two well established paradigms. However, based on solely these two tasks, Schubert et al. (2020) could not examine influences of different intelligence components, content domains, or task complexities. It thus remains an open question whether the results of the mediation analyses hold (a) across different response time tasks from different content domains and from different complexity and (b) across different aspects and domains of intelligence. Both varieties should be studied using a one and the same sample, to ensure full comparability and offer a clear picture of the relations between age, the diffusion model parameters, and intelligence. This paper aims precisely at closing this gap.

The present study

The present study reanalyzes data from a large study on the structure of cognitive speed (Lerche et al., 2020). In the original publication, no age effects are reported. The study uses 18 response time tasks from three different content domains (figural, numeric, verbal). Half of the tasks are fast tasks (with mean RTs below 1 second) and half are more complex (mean RTs > 2 seconds). We tested a sample of adults with an age range of 18 to 62 years, thus spanning all of young- and mid-adulthood, as well as the beginning of (young) old adulthood. To investigate whether response times and diffusion model parameters mediate age differences in a range of cognitive abilities, we used the same intelligence test as Schubert et al. (2020) to obtain a score of *g*, but we also computed scores for three intelligence process domains (processing capacity, psychometric speed, and memory), and three intelligence content domains (figural, numeric, and verbal). For each of these intelligence scores, we analyzed whether the diffusion model parameters, aggregated across tasks, account for age effects. We expected to find positive age correlations for boundary separation and non-decision time, indicating that older adults use more conservative decision criteria and take longer for encoding and motor response execution processes. We did not expect to find any age correlations for drift rates, except for the verbal domain, where — according to the results from the meta-analysis of Theisen et al. (2020) — older adults might have an advantage because verbal abilities involve a strong element of knowledge that might increase across a large part of adulthood. We had no specific hypotheses about the impact of task complexity. Following the results of Schubert et al. (2020), we expected age effects in the intelligence scores to be mediated by non-decision time. We also tested the other main diffusion model parameters (threshold separation and drift rate) as possible mediators, as well as mean logarithmized response times. In addition, differentiating between the process domains allowed us to compare the different explanations of the non-decision time mediation offered by Schubert et al. (2020). If age-related differences in non-decision time reflect age-related changes in anterior brain regions linked to both response preparation and higher-order processes such as intelligence, the mediation should occur equally across process domains. However, if the mediation is based on the fact that the intelligence test tasks have strict time limits, the mediation via t_0 should be especially strong for the psychometric speed intelligence tasks, and be less pronounced for the processing capacity intelligence tasks, which have more lenient time limits and are therefore less based on quick response execution and more similar to a power test.

Materials and Methods

Analyses based on the data of this study have also been reported by Lerche et al. (2020). The authors examined relationships between diffusion model parameters and intelligence and found both domain-general and domain-specific relationships between drift rate and intelligence. Age effects were not analyzed in their paper. Next, we will report the main aspects regarding sample characteristics, procedure and material of the study. More details can be found in Lerche et al. (2020).

Participants

We determined our sample size based on a power analysis for structural equation model analyses reported in Lerche et al. (2020). We had a sample of 125 participants, leading to a power of .81 to detect correlations of $r = .25$ ($\alpha = .05$). We recruited participants by means of a newspaper article, via fliers distributed at public places and by means of an online participant pool. All participants provided informed consent and received 35€ as well as feedback on their performance after completing the study. Our final sample (see below for a description of the proportion of missing data) consisted of 123 participants. The proportion of women amounted to 62.60 % and 50.41 % were students. The mean age was 35.85 years ($SD = 14.13$), with a range of

18 to 62 years. 59 participants were 18-29 years old, 15 were 30-39 years old, 19 were 40-49 years old, and 30 were 50-62 years old, with five of them being 60 or older.

Procedure

Participants completed three data collection sessions. In Session 1, participants filled in the BIS intelligence test (see below), while in Sessions 2 and 3 they worked on response time tasks (nine in each session). The order of the tasks was identical across participants. Table 1 gives on overview of the RT tasks and their order in the study. In each response time task session, participants took a three-minute break after the third and sixth task.

In all RT tasks, people started with four practice trials with feedback on the correctness of their responses (green checkmark vs. red cross shown for 1.5 seconds), followed by one warmup trial and 100 test trials.

Material

Intelligence Assessment

As a measure of intelligence, we used the Berlin Intelligence Structure Test (BIS; Jäger et al., 1997) that is based on the bimodal intelligence structure model (Jäger, 1982). The test provides tasks for three different content domains (figural, numeric, verbal) and four different process domains (processing capacity, psychometric speed, memory, and idea fluency). We used a short version of the test and disregarded the three idea fluency tasks, leading to a final set of 12 intelligence test tasks. We excluded the idea fluency tasks in the current analyses as we were not interested in creativity. Four tasks stemmed from each of the content domains. The processing capacity scale consisted of six tasks (two from each content domain), while psychometric speed and memory were both measured with one task from each content domain. We computed scale means for general intelligence *g* (including all tasks used), the four process domains, and the three content domains. Please note that the BIS manual only gives scoring rules for processing capacity and *g* when the short version of the test is used - we computed the scale means for the other scales correspondingly. For three participants, we could not use the scores from two tasks due to disturbances during data collection.

RT tasks

We used three fast tasks (mean RT ca. 800 ms) and three slow and more complex tasks (mean RT ca. 3000 ms) for each of the three content domains (figural, numeric, verbal), leading to a total of 18 RT tasks (see Table 1).

In the fast figural tasks, people had to determine whether a dot was within or outside of a rectangle (FF1, dot-rectangle task), which of two rectangles shown on the left and right side of the screen covered the greater area (FF2, simple area task), and whether a polygon shown was a triangle or a rectangle (FF3, polygon task). Among the slow figural tasks was a maze task (SF1), where participants had to judge whether a way out of the maze could be found from a marked spot. Another slow figural task was an extended version of the simple area task: Participants now had to judge whether three rectangles marked in blue or three rectangles marked in red covered the greater total area (SF2, complex area task). Finally, in the pie task (SF3), people judged whether three "slices" of a pie plot added up to less or more than a total pie.

In the fast numeric tasks, people had to determine whether a number was greater or smaller than 500 (FN1, number discrimination task), whether it was odd or even (FN2, odd-even task), or whether a number shown on the left side of the screen was larger than a number on the right side (FN3, simple inequation task). Among the slow numeric tasks was the mean value computation task (SN1) where people had to determine whether the mean of 16 numbers was greater or smaller than 500. In the equation task (SN2), participants judged whether equations were correct or wrong (e.g., 5*7 = 25). Finally, in the complex inequation task (SN3), people had to decide whether the solution of an equation shown on the left side of the screen was larger than the solution of an equation shown on the right side (e.g., "9 - 6" vs. "19 - 17"). In the fast verbal tasks, people judged whether a word was an adjective or noun (FV1, word category task), whether a letter combination was a word or not (FV2, lexical decision task), and whether a noun denoted a living versus a non-living entity (FV3, animacy task). Among the slow verbal tasks was a grammar task (SV1). People had to decide if the grammatical error in a five-word sentence was in the possessive pronoun or in the noun. In the statement task (SV2), in each trial we presented four to six words scattered across the screen. People's task was to determine whether a true statement could be formed from these words. Finally, in the semantic category task (SV3), people saw a list a five nouns (e.g., chair, sun, armchair, sofa, bench). People had to decide whether one or two of the items belonged to a different semantic category than the others. In the example, one of the nouns, i.e., "sun", differs from the dominant category (i.e., furniture). A more detailed description of all tasks is provided by Lerche et al. (2020).

Data preparation

For all RT tasks, we excluded data from trials faster than 300ms. In a second step, we also excluded intra-individual outliers, separately for each participant and each task. We defined outliers as RTs more than three interquartile ranges (IQRs) above the third quartile or three IQRs below the first quartile of the intra-individual RT distribution (Tukey, 1977). One participant accidentally skipped two tasks, introducing some missing response time data. We removed diffusion model parameters from model estimations that resulted in an inadequate fit according to a simulation study, separately for each participant and task [for a description of the simulation study, see Lerche et al. (2020); 0.93% of data excluded]. In the next step, we excluded the data separately for each task and participant, if the mean RT or accuracy rate were more than 3 IQRs away from the first or third quartile for this task. Finally, we excluded two participants as multivariate outliers, because their Mahalanobis distance, based on all diffusion model parameter estimates, mean RTs, and the intelligence content domains scores, exceeded the critical value of $\chi^2 = 140.89$ ($df = 93,\ p = .001$). The resulting sample thus contained 123 people.

Parameter estimation

We used the maximum likelihood estimation procedure provided of fast-dm (Voss et al., 2015; Voss & Voss, 2008, 2007) for obtaining estimates of diffusion model parameters. Simulation studies show that this procedure provides reliable parameter estimates for 100 trials (Lerche et al., 2017). We estimated parameters separately for each participant and each task. We used a simple model, estimating drift rate (v), boundary separation (a), non-decision time (t_0), and the inter-trial variability of non-decision time (st_0). The starting point (z) was fixed at the center between the two boundaries, as we associated the boundaries with correct and erroneous responses and thus did not except an a priori bias. We fixed all other parameters to zero, following recommendations by Lerche and Voss (2016). Across all tasks, model fit was good according to a graphical analysis and a simulation study (see Lerche et al., 2020, for a detailed description of model fit).

Data analysis

We used R (Version 4.0.3; R Core Team, 2020) and the R-packages *lavaan* (Version 0.6.7; Rosseel, 2012), *papaja* (Version 0.1.0.9997; Aust & Barth, 2020), *psych* (Version 1.8.12; Revelle, 2018), *scales* (Version 1.1.1; Wickham & Seidel, 2020), and *tidyverse* (Version 1.3.0; Wickham et al., 2019) for all analyses. All data and the analysis script are available on the Open Science Framework (https://osf.io/xpbwe/).

In a first step, we examined the bivariate correlations with age and general intelligence (g) for the diffusion model parameters and mean logarithmized response times, separately for each task. In the next step, we computed the means of the z-standardized values across all tasks and separately for each content domain for the same variables (v, a, t_0, mean log RT). Our sample size was too small to fit a structural equation model including the three diffusion model parameters for all 18 tasks. Hence, we used scale means for the mediation analyses. Additionally, we also examined the task-specific age correlations. Cronbachs's alpha values across all 18 tasks were good for threshold separation ($\alpha = 0.86$) and acceptable for drift rate ($\alpha = 0.76$) and non-decision time ($\alpha = 0.71$).

Mediation models

We formulated and tested several different mediation models to examine the interplay between age, intelligence, and the diffusion model parameters in depth. For the mediation analyses, we used the *R* package *psych*, that provides bootstrapped confidence intervals for the indirect effects. Specifically, we analyzed whether the diffusion model parameters can account for the age effects on different intelligence measures, that is *g* and the scores of the three different process domains (processing capacity, psychometric speed, and memory), and the three content domains (figural, numeric, verbal). Accordingly, in all models, age was the primary predictor variable. In the first two models the outcome variable was *g*. As we wanted to first confirm the established finding that mean response times mediate the age/intelligence relation, we used mean logarithmized RTs as a mediator in Model 1. Then, the three diffusion model parameters (v, a , t_0) served as mediators in Model 2, testing the assumption that these parameters can jointly account for the age-intelligence associations. In the next step, we tested mediation models for each of the process domains (processing capacity: Model 3; psychometric speed: Model 4; memory: Model 5), also using the diffusion model parameters as mediators. Finally, we examined whether the mediation was robust across content domains. We used content-domain specific diffusion model parameters (figural, numeric, verbal) as mediators of the relation of age to the respective intelligence domain scores (figural: Model 6; numeric: Model 7; verbal: Model 8). Model figures are given in the Results section (Figures 4 to 6).

For all significance tests, we used a strict alpha level of $\alpha = .005$ to account for multiple testing.

Results

Descriptive statistics and simple correlations

Figures 2 and 3 show boxplots of the mean response times for the final data set. Mean RT ranged between 527 and 792 ms (M = 647 ms) for the fast tasks and between 2380 and 4189 ms (M = 3225 ms) for the slow tasks. Table 2 shows descriptive statistics for mean RT, accuracy rate, drift rate, boundary separation and non-decision time for all tasks. Figures A1 to A4 show boxplots of mean response times and accuracy rate, split by age group.

Table 3 contains the bivariate age correlations of all diffusion model parameters for all tasks. Age correlations ranged from -.34 to .25 for drift rate, from .11 to .49 for boundary separation, and from .13 to .62 for non-decision time.

In general, there were medium positive age correlations for boundary separation, medium to strong positive age correlations for non-decision time, and no significant correlations between age and drift rate. In addition, it is important to note that there are substantial task-specificities. Some drift rates showed negative age correlations (i.e., the simple area task, the maze task, and the statement task), but in the word category task, older participants had higher drift rates. Also, for non-decision time and boundary separation, two of the correlations were very low ($|r| < .15$) and several values did not reach statistical significance, although the overall trend was clear.

As we did not find linear age correlations for most of the drift rates, we explored the age-drift rate relation by fitting cubic models. Figure A5 in the Appendix shows the scatter plots across all tasks. Across many tasks, drift rates seemed to rise until about the age of 30 years, declining thereafter. The rise in drift rate above the age of 60 found in some tasks has to be interpreted with caution, given that we had fewer than 5 participants of that age group.

Table 4 shows the correlations of mean log RT and the diffusion model parameters with general intelligence. All variables were substantially correlated with intelligence, with great variability across tasks. Generally, drift rates in slow tasks showed stronger correlations with intelligence than drift rates in fast tasks.

Table 5 shows the correlations of the aggregated diffusion model parameters, age, and the content-general outcome variables (g, processing capacity, psychometric speed, and memory). Table 6 shows the correlations of the aggregated diffusion model parameters, age, and the content-specific outcome variables (figural, numeric, and verbal intelligence).

Mediation analyses

Mediation models with g *as outcome (Models 1 and 2)*

Figure 4 shows the results for the mediation Models 1 and 2, that used g as outcome variable and either mean log RT (Model 1) or the diffusion model parameters (Model 2) as mediators. In both models, after introducing the mediating variables the relation of age and g was no longer statistically significant. In Model 1, mean log RT was linked to both age and g. The bootstrapped 99.5% confidence interval for the indirect effect of age via mean log RT did not include zero. Mean log RT accounted for 80% of the total effect. In Model 2, age was linked to t_0 and a, but not v, while g was linked to t_0 and v, but not a. The only indirect effect with a bootstrapped 99.5% confidence interval that did not include zero was for t_0 (non-decision time). The diffusion model parameters accounted for 59% of the total effect.

Mediation models with process domains as outcomes (Models 3, 4, and 5)

First, we checked whether mean logarithmized RTs mediated the relation of age and the respective outcome scores. This was the case for all three outcome measures. Accordingly, in the next step, the diffusion model parameters were examined as mediators of the link between age and the intelligence process domains. Figure 5 shows the results for the Models 3, 4, and 5. In these mediation models, the intelligence process domains processing capacity (Model 3), psychometric speed (Model 4), and memory (Model 5) were used as outcomes, respectively. In all three models, the correlations of age and the intelligence process domains were no longer statistically significant after introducing the mediating variables. In Model 3, processing capacity was linked only to drift rate, but not to boundary separation and non-decision time. Here, all bootstrapped 99.5% confidence intervals of the mediation effects included zero. Still, the diffusion model parameters accounted for 55% of the total effect on processing capacity. In Model 4, psychometric speed was linked to t_0 and v, but not to a. The only indirect effect with a bootstrapped 99.5% confidence interval that did not include zero was observed for t_0. The diffusion model parameters accounted for 66% of the total effect on psychometric speed. In Model 5, memory was linked to t_0 and v, but not to a. The only indirect effect with a bootstrapped 99.5% confidence interval that did not include zero was again t_0. The diffusion model parameters accounted for 56% of the total effect on memory.

Mediation models with content domain scores as outcomes (Models 6, 7, and 8)

First, we checked whether domain-specific mean logarithmized RTs mediated the relation of age and the respective outcome scores. This was the case for all three outcome measures. Accordingly, in the next step, the content domain specific diffusion model parameters were examined as mediators of the link between age and the intelligence content domain scores. Figure 6 shows the results for the Models 6, 7, and 8. In these mediation models, the figural (Model 6), numerical (Model 7), and verbal (Model 8) intelligence scores were used as outcomes, respectively. For figural intelligence (Model 6), the age correlation remained significant even after introducing the mediators. Figural intelligence was linked only to drift rate, but not to boundary separation and non-decision time. Age was correlated only to figural non-decision time and figural boundary separation, but not to figural drift rate. All bootstrapped 99.5% confidence intervals of the mediation effects included zero. The diffusion model parameters accounted for 30% of the total effect on figural intelligence. In the verbal and numerical models, the correlation of age and the intelligence scores was no longer statistically significant after introducing the mediating variables. In Model 7, numerical intelligence was linked to numerical t_0, a, and v. Age was correlated only to numerical non-decision time and numerical boundary separation, but not to numerical drift rate. The only indirect effect with a bootstrapped 99.5% confidence interval that did not include zero was for t_0. The diffusion model parameters accounted for 96% of the total effect on numerical intelligence. In Model 8, verbal intelligence was linked to verbal t_0 and v, but not to a. Age was correlated only to verbal non-decision time and verbal boundary separation, but not to verbal drift rate. The only indirect effect with a bootstrapped 99.5% confidence interval that did not include zero was for t_0. The diffusion model parameters accounted for 59% of the total effect on verbal intelligence.

Discussion

Results from several studies show that response times from elementary cognitive tasks substantially mediate the relation of age and cognitive abilities (Finkel et al., 2007; Salthouse, 1996), suggesting that age differences in intelligence might be (partly) based on age differences in processing speed. However, response times are not process-pure measures, as they reflect not only the speed of information processing, but also - for example - speed-accuracy trade-offs or the time needed for sensory encoding and motor response execution. The diffusion model (Ratcliff, 1978) provides separate estimates for these different components of the decision process. A previous study demonstrated that not processing speed but non-decision time mediates the relation of age and general intelligence (Schubert et al., 2020). The present study builds upon this finding and aims at testing which components of information processing mediate the link of age and decline in a range of intelligence content domains and intelligence process domains.

For the present study, we used a wide range of response time tasks across different content domains. In previous studies on the age effects in diffusion model parameters, only a limited number of tasks have been examined simultaneously so that it was not possible to examine effects of content domain systematically (e.g., Ratcliff et al., 2004, 2010, 2001; McKoon & Ratcliff, 2012; R. Ratcliff, 2008). Of the 18 response time tasks employed in our study, six belonged to the figural, numeric, and verbal domain, respectively. Furthermore, half of the tasks were based on fast decisions, while the other half were more complex tasks and required much longer processing times. As outcomes, we did not only examine *g*, but also different intelligence scores (processing capacity, psychometric speed, and memory). Thus, we could examine the generalizability of the non-decision time mediation reported by Schubert et al. (2020) across content domains, task complexities, and intelligence process domains. An additional important difference between our study and most previous studies on age differences in diffusion model parameters is that we studied a broad age range from 18 to 62 years, whereas most previous studies had compared only two age groups, college age people and old adults (65+ years old). In contrast, our sample included 66 persons from mid-adulthood, aged 30-60 years, an age group that is understudied in diffusion model analyses so far. Previous studies found compelling evidence for an age-related increase of boundary separations and non-decision times (for a meta-analysis of age-effects on diffusion-model parameters, see Theisen et al., 2020). • That is, elder adults are more cautious decision-makers and they are slower in encoding and/or motoric response execution. In our study, we could assess whether age differences found for the group comparisons map onto linear age correlations across a wider range of adulthood. For most of the 18 employed RT tasks, we found strong age correlations of mean logarithmized response times reflecting slower responses for elder participants. Correlations between age and RT tended to be higher for fast than for slow tasks, and among the slow tasks correlations were more heterogeneous. This last finding might reflect greater task complexity of the slow tasks, which might lead to greater between-task variability in the cognitive processes and in the abilities required for solving the tasks, thus resulting in different age correlations. Non-decision times – as estimated by the diffusion model – showed medium to strong correlations with age for most tasks. This implies that older participants take longer for encoding and/or motor processes. As expected, age was also related to boundary separation, though to a smaller degree. This implies that for most tasks, older participants tend to apply more conservative decision criteria, indicating that they gather more information before making a decision. These results are perfectly in line with results from the recent meta-analysis by Theisen et al. (2020), although it should be noted that the meta-analysis compared young adults and old adults, while our study focused on young- and mid-adulthood.

Our pattern of results suggests a continuous developmental increase in cautiousness - elder people get more conservative and take more time for encoding and motor execution. Of course, our cross-sectional design does not allow for a direct test of this hypothesis. Regarding speed of information processing (drift rate), we found no age correlations for most of the tasks. We also did not find a clear pattern of differences in the age-drift correlation between the three different content domains or for fast vs. complex tasks. Younger people had higher drift rates in some, but not in all figural tasks. Regarding drift rates in verbal tasks, we had expected older people to have an advantage, as Theisen et al. (2020) report that task content moderates the age effects on drift, with an age-related increase for lexical decision tasks. In our sample, older people had higher drift rates in one verbal task (the noun-adjective task, but not in the lexical decision task). For the other verbal tasks, we found no correlations between drift rate and age, except for a negative age correlation in the statement task. In this regard, our findings regarding the drift rate are not in line with the effects reported by Theisen et al. (2020). In exploratory analyses we fitted cubic models to examine a possible nonlinear relationship of age and drift. Interestingly, for drift rates from many tasks as well as for the composite drift rate across tasks, we found evidence for a positive age trend from age 18 until about the age of 30. After that, drift rates showed a linear negative age trend until about the age of 60. These findings suggest that many previous studies might not have found significant age effects in drift rate because they compared very young people (i.e., in their early twenties) to old adults (65+ years). A similar interpretation has also been proposed for findings on different cognitive abilities like, for example, working memory (Hartshorne & Germine, 2015). Indeed, when excluding our youngest participants (i.e., people aged 18-29), we found small to medium negative age correlations for drift rates across several tasks - most of them were fast tasks. As excluding these young adults made our sample considerably smaller, the findings should be interpreted with caution. Still, future studies might be well advised to include people in the mid-adult age range to get a clearer picture on where the turning point in the development of processing speed lies. Ideally, one could study the trends longitudinally, measuring participants repeatedly from college age into middle or even old adulthood.

Our main research question was whether diffusion model parameters could explain age differences in intelligence. First of all, we replicated the finding that logarithmized mean response times fully mediated the correlation between age and general intelligence (Finkel et al., 2007; Salthouse, 1996). The models using the diffusion model parameters as mediators of the age effect on intelligence showed a robust indirect effect for non-decision time, indicating that the age related decline in intelligence test scores is mediated by the duration of encoding and/or motor processes. Drift rates were clearly linked to *g*, but not to age, and thus did not show a significant indirect effect. Boundary separation was linked to age, but not to *g*, also leading to an insignificant indirect effect. The three diffusion model parameters jointly fully mediated the relation between age and *g*.

These findings replicated across most of the analyses using the process domain scores (processing capacity, psychometric speed, memory) and the content domain scores (figural, numerical, verbal) as outcomes. Drift rates where linked to the intelligence outcomes, but not to age. The only exception for the latter was in the figural content domain, where figural drift rates showed a small negative correlation to age and the indirect effect via drift rate accordingly approached statistical significance. Boundary separation was not linked to the intelligence outcomes, except for numerical intelligence, where numerical boundary separation showed a small negative correlation to numerical intelligence and the indirect effect via boundary separation accordingly approached statistical significance. Finally, non-decision time was linked to both age and the intelligence outcomes in all cases except processing capacity and figural intelligence, which showed no significant correlations to the respective non-decision times. These findings suggest that age differences in processing capacity and figural intelligence are not based on age differences in any of the diffusion model parameters.

The correlation of intelligence with non-decision time was particularly strong for the psychometric speed scores, indicating that this intelligence scale is strongly influenced by speed in sensory encoding and/or motor response execution, but not necessarily by speed of information processing, as drift rates showed no correlation. Schubert et al. (2020) offered two different possible explanations for the mediation of the age to *g* relation through non-decision time. On the one hand, age-related variation in non-decision time might reflect age-related variation in anterior brain areas associated both with response preparation and other higher-order processes, implying that the non-decision time mediation generalizes across process domains. On the other hand, the indirect effect might be overestimated, as performance in intelligence test tasks involves a component of motoric speed. The degree to which motor speed is involved differs between intelligence tests - the psychometric speed tasks of the BIS, that rely extensively on quick hand-writing, should in this case be strongly related to non-decision time. Our finding that non-decision time was particularly closely related to scores in the psychometric speed tasks of the BIS test could thus be viewed as support to this latter notion, implying that speed of motor response execution plays an important role in determining the relationship of non-decision time, age, and intelligence test scores. On the contrary, reasoning tasks that are closer to a power test and rely less on time pressure, should show strong relations to processing speed, and be less correlated to non-decision time. This is exactly the pattern we find in our data, with the processing capacity tasks being the closest to a power test among the BIS tasks. These results bring up the question whether the mediation of age differences in *g* scores via non-decision time truly informs us about intelligence, or is partly an artifact of the speeded intelligence test tasks. Using a power test without any time limit as an outcome might be the next step to further investigate this issue.

One important issue when studying age differences in cognition is whether these differences and developmental patterns are general or domain-specific. Given our finding that non-decision times mediated age differences in intelligence for the verbal and numerical content domains, we conducted additional analyses to investigate whether it were the domain-general or domain-specific parts of variance in non-decision time that accounted for the mediation. To this end, we estimated a simple structural equation model, using the three non-decision time values from figural, numerical, and verbal non-decision times as indicators of a general non-decision time factor. We then used this general non-decision time and domain-specific non-decision time as mediators of the relationship between age and domain-specific intelligence. It turned out that it was the general non-decision time factor, but not the domain-specific non-decision time residuals that accounted for the indirect effect, both in the numerical and in the verbal domain. The domain-specific non-decision time residuals were not related to age. This suggests that the processes eliciting age differences in non-decision time generalize across domains.

Taken together, our findings suggest that the often reported age-related slowing in response time tasks, which mediates the relationship between age and a wide range of cognitive abilities, can mostly be attributed to the fact that older people take longer for non-decisional processes. This finding proved to be robust across a range of cognitive ability outcomes including general intelligence and memory, with the exception of processing capacity and figural intelligence.

It is important to note that all our outcome tasks were speeded and our findings might therefore be partly overstating the relationship between non-decision time and general cognitive ability. In the least speeded intelligence tasks - namely, those assessing processing capacity and thus probably most closely reflective of reasoning ability - non-decision time was not a mediator, but neither was processing speed (i.e., drift rate). In this sense, all our findings contradict the idea that a decline in processing speed is the basis of cognitive decline in general. Our results are more easily reconcilable with the assumption of a „common cause“ (Christensen et al., 2001) that is related to decline in a wide range of cognitive abilities, including response times – the age relationship of the latter being, according to our analyses, in large parts defined by the time taken for motor processes. At the same time, the variability in correlations of non-decision time with age and IQ across tasks implies the importance of domain-specific factors. The literature on the relation between age differences in cognition and in brain structure suggests correlated change, but findings greatly differ regarding the strength of this relationship (for a review, see Oschwald et al., 2019). Findings on processing speed are also inconclusive in this regard. According to the Scaffolding Theory of Aging and Cognition (STAC-r; Reuter-Lorenz & Park, 2014), people employ different compensatory scaffolding techniques (e.g., strategy use, activation of additional brain networks) to counter the detrimental effects of age-related alterations in brain structure.

Differences in coping abilities might thus influence the relations between brain structure and cognitive abilities. Regarding the diffusion model parameters, drift rates might reflect a type of processing speed that is open to compensatory scaffolding techniques and thus relatively stable across a large part of the life-span, while the more basic processes contributing to non-decision times might be less malleable and thus show clearer age correlations.

One important additional finding is the great variability of age correlations for the diffusion model parameters across the 18 tasks employed in the present study. For drift rates, no age correlations were found in most of the tasks. Yet in two figural tasks, elder people showed lower drift rates. At the same time, in one verbal task (the noun-adjective task), elder persons had higher drift rates. These findings underline the importance of using a wider variety of tasks when studying age differences in diffusion model parameters. Had we only used one or two tasks, the general picture might have looked quite different, maybe implying age-related decline in drift rates. The same holds true for the age correlations in boundary separation and non-decision time. Even though the general picture is quite clear in both cases—medium to large age correlations—there are several tasks where either boundary separation, non-decision time, or both parameters are not related to age. Thus the wide range of response time tasks employed proves to be an important strength of this study.

Limitations

For diffusion modeling, the number of trials per task and participant was rather low. We decided to employ a wide range of tasks instead of just a few tasks with high numbers of trials. Simulation studies suggest that the diffusion model yields adequate estimates for 100 trials (Lerche et al., 2017). We also examined model fit, which was good for all tasks in our study. A second important limitation of our study is that also the sample size is limited. This has implications for the modeling approach employed. One could argue that aggregating parameters across tasks simply by computing the mean of the standardized values is an oversimplification of the structure of drift rates, boundary separation values, non-decision times, and response times.

The procedure implies the assumption of parallel measurement, that is, the presumption that all items contribute equally and fully to a common latent factor. This is a strong assumption that cannot be tested in the modeling approach we used. Unfortunately, investigating the mediations through latent variable structural equation modeling including all task-specific diffusion model parameters, such as in the approach used by Schubert et al. (2020), was impossible due to our restricted sample size. To address this issue, we estimated principal component analyses, separately for each of the diffusion model parameters and mean log RTs (across all 18 tasks, and separately for each content domain). In each principal component analysis, we assumed one general factor, to mirror the factor structure from our main analyses. We then extracted factor scores and used these as mediators in the mediation models (Models 1-8). This did not alter the interpretation of any of the results. In fact, factor scores were highly correlated (often $r = .99$) to the means of standardized task scores. This implies that our simple aggregation procedure (means of standardized values across tasks) is justified. At the same time, the range of age and IQ correlations across tasks hints at task-specific aspects and/or sub-factors. We also estimated separate structural equation mediation models for each diffusion model parameter and each content domain, for example, a mediation model with age as predictor, numerical intelligence as outcome, and numerical non-decision time as the only mediator – the latter being a latent factor linked to non-decision times in all numerical tasks. Though several of these models suffered from inadequate model fit and results from these models must thus be interpreted with caution, these additional analyses did not indicate a different pattern of results than our main analyses. All these analyses can be replicated using the scripts on the paper's OSF page (https://osf.io/xpbwe). It is critical to note that mediation models cannot provide a test of causal relations. In fact, one could think of a number of different models that would show identical model fit, but assume a completely different causal relationship of the variables. While the models tested in our study are based in theory, there is no way to tell if they reflect the "true" causal relationships between age, the diffusion model parameters, and cognitive abilities. Another important limitation is the fact that age and cohort effects are confounded in our study - a problem that could only be fully remedied by following several different cohorts longitudinally.

Future studies might shed light on the question what accounts for the age differences in processing capacity and figural intelligence, as they were unrelated to non-decision time in our sample. These studies should include measures of working memory capacity and executive functions, as well as neuro-cognitive data, to disentangle the non-speed related processes that might account for age differences in cognition.

A final limitation of our study is the fact that we did not include people older than 62 years. Thus, we cannot examine the developmental patterns that occur in old age. R. Ratcliff, Thapar, et al. (2006b) found significant differences in diffusion model parameters between people aged 60-74 and those older than that. In comparison with participants aged 60-74, the eldest participants (aged 75-85) had more conservative decision criteria, longer non-decision times, and lower drift rates, though all these findings differed between tasks. It would be highly interesting to expand the mediation analyses to this age group to assess whether the correlational patterns are qualitatively different here.

Conclusion

Cognitive slow-down is thought to contribute to the age-related decline found for a wide range of cognitive abilities, including general intelligence. We investigated the relationships between age, three main diffusion model parameters calculated from 18 different response time tasks, and different measures of intelligence. Older people in our sample (ranging from young adulthood to the beginning of old age) used more conservative decision criteria and needed more time for extra-decisional processes, but no linear age effect was found for processing speed. Individual differences in non-decision times fully mediated the relation between age and intelligence for most measures of intelligence. Only scores of processing capacity and figural intelligence did not show a significant relationship to non-decision time. Our findings support the account that, already in mid-adulthood, age differences in intelligence test scores are based on age differences in non-decisional processes, in particular motor execution time.

References

Arnold, N. R., Bröder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research*, *79*(5), 882–898. https://doi.org/10.1007/s00426-014-0608-y

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. https://github.com/crsh/papaja

Ball, B. H., & Aschenbrenner, A. J. (2018). The importance of age-related differences in prospective memory: Evidence from diffusion model analyses. *Psychonomic Bulletin & Review*, *25*(3), 1114–1122. https://doi.org/10.3758/s13423-017-1318-4

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Christensen, H., Mackinnon, A. J., Korten, A., & Jorm, A. F. (2001). The" common cause hypothesis" of cognitive aging: Evidence for not only a common factor but also specific associations of age with vision and grip strength in a cross-sectional analysis. *Psychology and Aging*, *16*(4), 588.

Dully, J., McGovern, D. P., & O'Connell, R. G. (2018). The impact of natural aging on computational and neural indices of perceptual decision making: A review. *Behavioural Brain Research*, *355*, 48–55. https://doi.org/10.1016/j.bbr.2018.02.001

Finkel, D., Reynolds, C. A., McArdle, J. J., & Pedersen, N. L. (2007). Age changes in processing speed as a leading indicator of cognitive aging. *Psychology and Aging*, *22*(3), 558–568. https://doi.org/10.1037/0882-7974.22.3.558

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the lifespan. *Psychological Science*, *26*(4), 433–443. https://doi.org/10.1177/0956797614567339

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta*

Psychologica, *26*, 107–129. https://doi.org/10.1016/0001-6918(67)90011-X

Janczyk, M., Mittelstädt, P., & Wienrich's, C. (2018). Parallel dual-task processing and task-shielding in older and younger adults: Behavioral and diffusion model results. *Experimental Aging Research*, *44*(2), 95–116. https://doi.org/10.1080/0361073X.2017.1422459

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen [Multimodal classification of intelligent performance]. *Diagnostica*, *28*, 195–225.

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: BIS-Test*. Hogrefe. https://madoc.bib.uni-mannheim.de/14578/

Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences*. Elsevier.

Lerche, V., Christmann, U., & Voss, A. (2018). Impact of Context Information on Metaphor Elaboration: A Diffusion Model Study. *Experimental Psychology*, *65*(6), 370–384. https://doi.org/10.1027/1618-3169/a000422

Lerche, V., Krause, M. von, 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*. https://doi.org/10.1037/xge0000774

Lerche, V., & Voss, A. (2017). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research*. https://doi.org/10.1007/s00426-017-0945-8

Lerche, V., & Voss, A. (2016). Model Complexity in Diffusion Modeling: Benefits of Making the Model More Parsimonious. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01324

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods*, *49*(2), 513–537. https://doi.org/10.3758/s13428-016-0740-2

Lindenberger, U., von Oertzen, T., Ghisletta, P., & Hertzog, C. (2011). Cross-sectional age variance extraction: What's change got to do with it? *Psychology and Aging*, *26*(1), 34–47. https://doi.org/10.1037/a0020525

Madden, D. J., Costello, M. C., Dennis, N. A., Davis, S. W., Shepler, A. M., Spaniol, J., Bucur, B., & Cabeza, R. (2010). Adult Age Differences in Functional Connectivity during Executive Control. *NeuroImage*, *52*(2), 643–657. https://doi.org/10.1016/j.neuroimage.2010.04.249

McGovern, D. P., Hayes, A., Kelly, S. P., & O'Connell, R. G. (2018). Reconciling age-related changes in behavioural and neural indices of human perceptual decision-making. *Nature Human Behaviour*, *2*(12), 955–966. https://doi.org/10.1038/s41562-018-0465-6

McKoon, G., & Ratcliff, R. (2012). Aging and IQ effects on associative recognition and priming in item recognition. *Journal of Memory and Language*, *66*(3), 416–437. https://doi.org/10.1016/j.jml.2011.12.001

McKoon, G., & Ratcliff, R. (2013). Aging and predicting inferences: A diffusion model analysis. *Journal of Memory and Language*, *68*(3), 240–254. https://doi.org/10.1016/j.jml.2012.11.002

Oschwald, J., Guye, S., Liem, F., Rast, P., Willis, S., Röcke, C., Jäncke, L., Martin, M., & Mérillat, S. (2019). Brain structure and cognitive ability in healthy aging: A review on longitudinal correlated change. *Reviews in the Neurosciences*, *31*(1), 1–57.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging*, *23*(4), 900–916. https://doi.org/10.1037/a0013930

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence*, *36*(1), 10–17. https://doi.org/10.1016/j.intell.2006.12.002

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A Diffusion Model Analysis of the Effects of Aging in the Lexical-Decision Task. *Psychology and Aging*, *19*(2), 278–289. https://doi.org/10.1037/0882-7974.19.2.278

Ratcliff, R., Thapar, A., & McKoon, G. (2006a). Aging, practice, and perceptual tasks: A diffusion model analysis. *Psychology and Aging*, *21*(2), 353–371. https://doi.org/10.1037/0882-7974.21.2.353

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*(3), 127–157. https://doi.org/10.1016/j.cogpsych.2009.09.001

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*(2), 323–341.

Ratcliff, R., Thapar, A., & McKoon, G. (2006b). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review*, *13*(4), 626–635. https://doi.org/10.3758/BF03193973

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it stac up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology Review*, *24*(3), 355–370.

Revelle, W. (2018). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. https://CRAN.R-project.org/package=psych

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. http://www.jstatsoft.org/v48/i02/

Salthouse, T. (2004). Localizing age-related individual differences in a hierarchical structure.

Intelligence, *32*(6), 541–561. https://doi.org/10.1016/j.intell.2004.07.003

Salthouse, T. A. (1996). *The Processing-Speed Theory of Adult Age Differences in Cognition*. 26. https://doi.org/0033-295X/96/$3.00

Salthouse, T. A. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science*, *13*(4), 140–144. https://doi.org/10.1111/j.0963-7214.2004.00293.x

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society : JINS*, *16*(5), 754–760. https://doi.org/10.1017/S1355617710000706

Schaie, K. W. (2005). What can we learn from longitudinal studies of adult development? *Research in Human Development*, *2*(3), 133–158.

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*, 096372141989636. https://doi.org/10.1177/0963721419896365

Schubert, A.-L., Hagemann, D., & Frischkorn, G. T. (2017). Is general intelligence little more than the speed of higher-order processing? *Journal of Experimental Psychology: General*, *146*(10), 1498–1512. https://doi.org/10.1037/xge0000325

Schubert, A.-L., Hagemann, D., Löffler, C., & Frischkorn, G. T. (2020). Disentangling the Effects of Processing Speed on the Association between Age Differences and Fluid Intelligence. *Journal of Intelligence*, *8*(1), 1. https://doi.org/10.3390/jintelligence8010001

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, *44*(3), 535–551. https://doi.org/10.1016/j.paid.2007.09.015

Sliwinski, M. J., & Hall, C. B. (1998). Constraints on general slowing: A meta-analysis using hierarchical linear models with random coefficients. *Psychology and Aging*, *13*(1), 164–175. https://doi.org/10.1037/0882-7974.13.1.164

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 101–117. https://doi.org/10.1037/0278-7393.32.1.101

Spaniol, J., Voss, A., Bowen, H. J., & Grady, C. L. (2011). Motivational incentives modulate age differences in visual perception. *Psychology and Aging*, *26*(4), 932–939. https://doi.org/10.1037/a0023297

Spieler, D. H. (2001). Modelling age-related changes in information processing. *The European Journal of Cognitive Psychology*, *13*(1), 217–234. https://doi.org/10.1080/09541440042000287

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A Diffusion Model Analysis of the Effects of Aging on Letter Discrimination. *Psychology and Aging*, *18*(3), 415–429. https://doi.org/10.1037/0882-7974.18.3.415

Theisen, M., Lerche, V., Krause, M. von, & Voss, A. (2020). Age differences in diffusion model parameters: A meta-analysis. *Psychological Research*. https://doi.org/10.1007/s00426-020-01371-8

Tucker-Drob, E. M. (2011). Global and domain-specific changes in cognition throughout adulthood. *Developmental Psychology*, *47*(2), 331–343. https://doi.org/10.1037/a0021361

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co.

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, *122*(3), 231–249. https://doi.org/10.1037/0033-2909.122.3.231

Voskuilen, C., Ratcliff, R., & McKoon, G. (2018). Aging and confidence judgments in item recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 1–23. https://doi.org/10.1037/xlm0000425

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology*, *60*(6), 385–402. https://doi.org/10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*(7), 1206–1220. https://doi.org/10.3758/BF03196893

Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology*, *52*(1), 1–9. https://doi.org/10.1016/j.jmp.2007.09.005

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775. https://doi.org/10.3758/BF03192967

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00336

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Seidel, D. (2020). *Scales: Scale functions for visualization*. https://CRAN.R-project.org/package=scales

Zimprich, D., & Martin, M. (2002). Can longitudinal changes in processing speed explain longitudinal age changes in fluid intelligence? *Psychology and Aging*, *17*(4), 690–695.

https://doi.org/10.1037/0882-7974.17.4.690

Table 1

Overview of the 3 (domain: figural vs. numeric vs. verbal) × 2 (speed: fast vs. slow) × 3 (number of tasks) = 18 RT tasks

	Fast	Slow
Figural	FF1: dot-rectangle task (1.9)	SF1: maze task (2.1)
	FF2: simple area task (2.4)	SF2: complex area task (1.6)
	FF3: polygon task (1.3)	SF3: pie task (2.7)
Numeric	FN1: number discrimination task (2.2)	SN1: mean value computation task (1.8)
	FN2: odd-even task (1.5)	SN2: equation task (2.5)
	FN3: simple inequation task (2.8)	SN3: complex inequation task (1.2)
Verbal	FV1: word category task (2.6)	SV1: grammar task (1.4)
	FV2: lexical decision task (1.1)	SV2: statement task (2.3)
	FV3: animacy task (1.7)	SV3: semantic category task (2.9)

Note. The first letter indicates the task complexity (F = fast, S = slow); the second letter denotes the domain (N = numeric, V = verbal, F = figural). The numbers in parentheses indicate the time point of assessment (session and number in sequence).

Table 2

Means (M) and Standard Deviations (SD) of Mean RTs in ms, Accuracy Rates in % and the Diffusion Model Parameters for all 18 Tasks

Task	M_{RT}	SD_{RT}	$M_{\text{Acc.}}$	$SD_{\text{Acc.}}$	M_v	SD_v	M_a	SD_a	M_{t0}	SD_{t0}
FF1	560	96	93.65	2.88	3.16	0.73	0.91	0.21	0.42	0.07
FF2	620	176	98.68	1.60	3.26	1.02	1.53	0.53	0.36	0.07
FF3	551	96	97.71	1.90	4.27	0.96	1.16	0.61	0.41	0.06
FN1	527	78	98.03	2.26	4.97	1.82	1.47	1.31	0.39	0.07
FN2	590	107	97.68	2.03	3.95	0.97	1.20	0.51	0.43	0.06
FN3	670	135	97.17	2.74	3.97	1.39	1.36	1.03	0.50	0.10
FV1	792	164	96.22	3.76	2.81	0.88	1.52	0.73	0.51	0.08
FV2	781	162	95.11	3.97	2.68	0.78	1.33	0.44	0.53	0.07
FV3	737	124	97.18	2.41	3.21	0.89	1.35	0.55	0.52	0.07
SF1	3234	1091	95.53	2.91	0.94	0.20	3.75	1.44	1.29	0.49
SF2	4189	2009	86.69	6.50	0.58	0.17	3.71	1.37	1.48	0.92
SF3	2856	906	80.47	9.10	0.50	0.18	3.06	0.81	0.91	0.40
SN1	4168	1904	90.76	8.11	0.70	0.22	4.00	1.53	1.63	1.21
SN2	2761	1098	91.16	5.48	0.80	0.25	3.25	0.92	0.84	0.31
SN3	2805	885	93.51	3.71	1.08	0.33	2.85	0.92	1.50	0.42
SV1	2380	709	96.36	2.39	1.17	0.20	3.08	0.84	1.09	0.35
SV2	3030	1002	95.11	2.61	1.03	0.29	3.19	0.87	1.45	0.42
SV3	3600	895	94.24	4.77	0.90	0.23	3.69	1.23	1.64	0.41

Note. See Table 1 for an explanation of the task names. Diffusion Model parameters: a: boundary separation; v: drift rate; t_0: non-decision time.

Table 3

Age correlations of RTs, accuracy rates and diffusion model parameters for all 18 RT tasks

Task	Mean RT	Mean log. RT	Accuracy Rate	Drift Rate	Boundary Sep.	Non-Decision Time
FF1	.64**	.66**	.41**	-.16	.43**	.62**
FF2	.54**	.57**	.27*	-.29*	.37**	.50**
FF3	.56**	.60**	.37**	.01	.38**	.49**
FN1	.61**	.62**	.43**	.02	.16	.37**
FN2	.32**	.37**	.39**	.01	.25	.35**
FN3	.59**	.60**	.50**	.09	.34**	.40**
FV1	.28*	.32**	.46**	.25	.36**	.25
FV2	.37**	.40**	.48**	.02	.49**	.17
FV3	.46**	.48**	.34**	-.07	.21	.44**
SF1	.50**	.51**	.28*	-.31**	.33**	.25*
SF2	.25	.32**	.23	-.08	.22	.28*
SF3	.24	.31**	.18	.05	.22	.19
SN1	.26	.27*	.17	-.05	.22	.13
SN2	.25*	.28*	.29*	.01	.25	.29*
SN3	.25*	.30**	.20	.02	.11	.42**
SV1	.31**	.32**	.35**	.00	.25	.31**
SV2	.48**	.51**	.19	-.34**	.45**	.32**
SV3	.45**	.47**	.24	-.09	.32**	.30**

Note. See Table 1 for an explanation of the task names.

* $p < .005$, ** $p < .001$

Table 4

IQ correlations of RTs, accuracy rates, and diffusion model parameters for all 18 RT tasks

Task	Mean RT	Mean log. RT	Accuracy Rate	Drift Rate	Boundary Sep.	Non-Decision Time
FF1	-.46**	-.47**	-.33**	.13	-.34**	-.44**
FF2	-.46**	-.44**	-.19	.32**	-.35**	-.25
FF3	-.62**	-.63**	-.21	.25	-.29*	-.45**
FN1	-.57**	-.57**	-.13	.18	-.07	-.36**
FN2	-.60**	-.64**	-.28*	.33**	-.33**	-.48**
FN3	-.67**	-.69**	-.27*	.15	-.27*	-.48**
FV1	-.48**	-.50**	-.12	.21	-.28*	-.29*
FV2	-.49**	-.50**	-.12	.22	-.38**	-.34**
FV3	-.51**	-.53**	-.08	.32**	-.18	-.41**
SF1	-.54**	-.54**	-.04	.46**	-.38**	-.21
SF2	-.35**	-.40**	.03	.37**	-.28*	-.21
SF3	-.22	-.24	.25	.34**	-.07	-.23
SN1	-.26*	-.23	.24	.41**	.00	-.25
SN2	-.66**	-.71**	.10	.60**	-.55**	-.44**
SN3	-.67**	-.72**	-.06	.44**	-.52**	-.49**
SV1	-.54**	-.55**	-.20	.29*	-.34**	-.51**
SV2	-.56**	-.57**	-.02	.42**	-.45**	-.42**
SV3	-.62**	-.64**	.01	.42**	-.41**	-.25

Note. See Table 1 for an explanation of the task names.

* $p < .005$, ** $p < .001$

Table 5

Correlations of all the variables used for the general mediation analyses.

	1	2	3	4	5	6	7	8
1 - Age								
2 - *g*	-.47**							
3 - Processing Cap.	-.37**	.91**						
4 - Psy. Speed	-.44**	.78**	.55**					
5 - Memory	-.39**	.75**	.55**	.48**				
6 - mean log RT	.58**	-.70**	-.59**	-.63**	-.55**			
7 - t_0	.57**	-.60**	-.46**	-.57**	-.51**	.78**		
8 - *a*	.50**	-.51**	-.44**	-.47**	-.38**	.89**	.50**	
9 - *v*	-.08	.60**	.57**	.40**	.47**	-.52**	-.23	-.34**

Note. * $p < .005$, ** $p < .001$

Table 6

Correlation matrix of the variables used for the content-domain specific mediation analyses.

	1	2	3	4	5	6	7	8	9	10	11	12
1 - Age												
2 - Verbal IQ	-.41**											
3 - Figural IQ	-.52**	.53**										
4 - Numerical IQ	-.26*	.56**	.53**									
5 - t_0 Verbal	.44**	-.59**	-.31**	-.43**								
6 - t_0 Figural	.60**	-.40**	-.40**	-.33**	.70**							
7 - t_0 Numerical	.50**	-.54**	-.41**	-.59**	.66**	.72**						
8 - *v* Verbal	-.04	.53**	.24	.37**	-.28*	-.08	-.25*					
9 - *v* Figural	-.21	.38**	.52**	.38**	-.07	-.16	-.27*	.49**				
10 - *v* Numerical	.03	.39**	.27*	.60**	-.10	.01	-.29*	.50**	.53**			
11 - *a* Verbal	.49**	-.52**	-.38**	-.35**	.43**	.51**	.48**	-.41**	-.33**	-.15		
12 - *a* Figural	.50**	-.47**	-.38**	-.25	.35**	.36**	.39**	-.24	-.39**	-.13	.78**	
13 - *a* Numerical	.36**	-.47**	-.36**	-.35**	.42**	.43**	.29*	-.25	-.34**	-.09	.73**	.73**

Note. * $p < .005$, ** $p < .001$

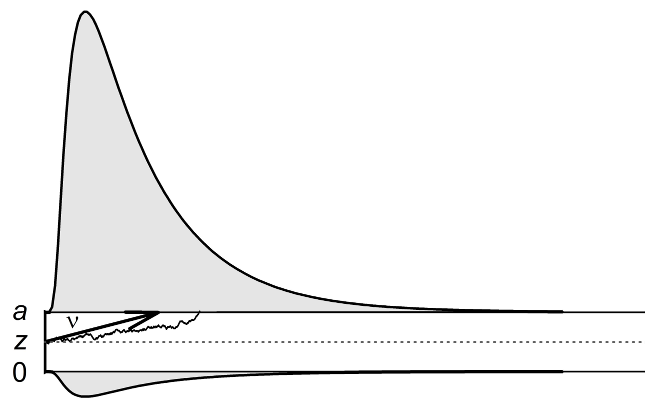

Figure 1

The diffusion model. The accumulation process starts at starting point z, moves with average slope v, and terminates when one of the two thresholds (0 or a) has been reached.

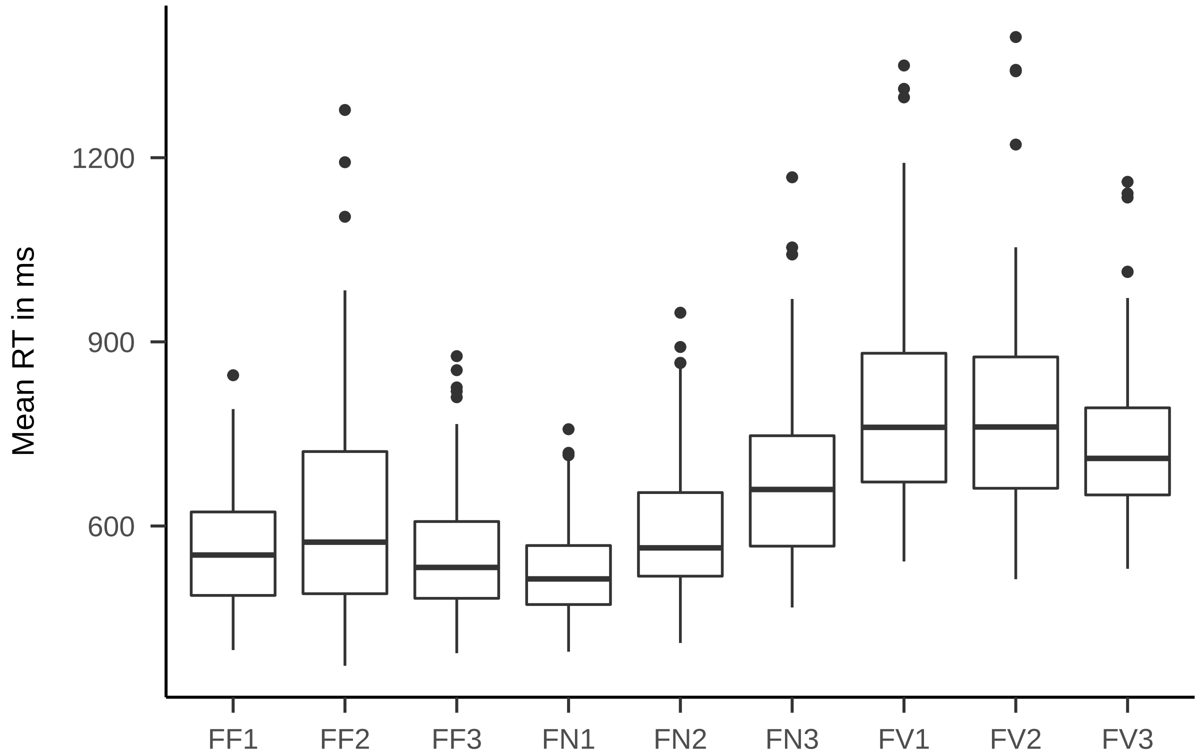

Figure 2

Boxplots of mean response times for all fast tasks. See Table 1 for an explanation of the task names.

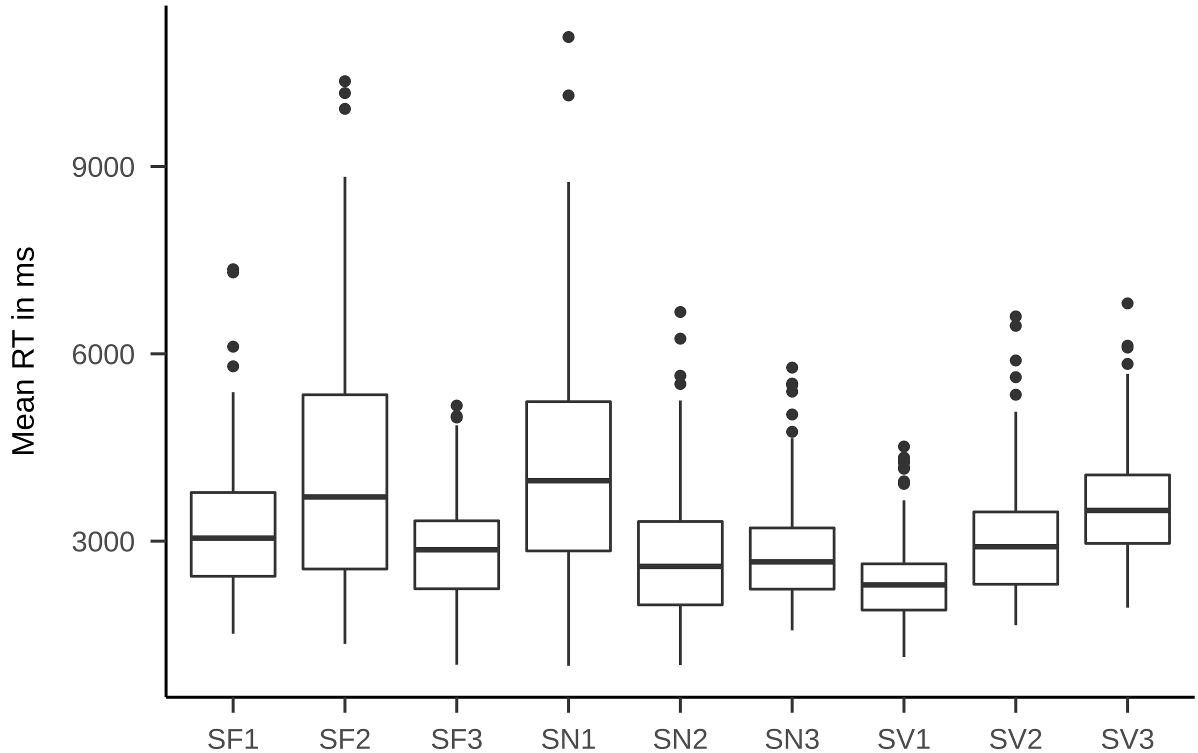

Figure 3

Boxplots of mean response times for all slow tasks. See Table 1 for an explanation of the task names.

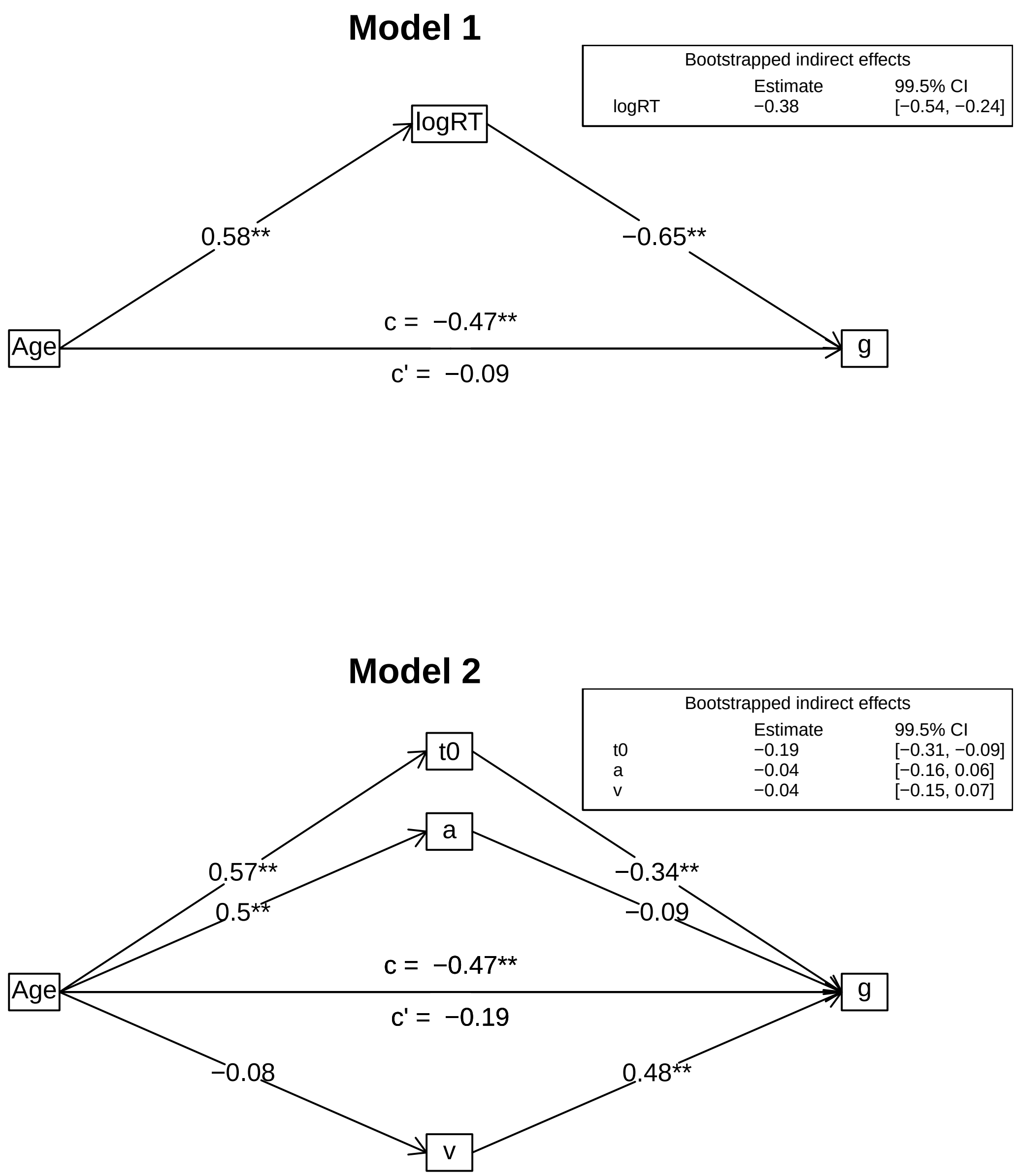

Figure 4

Mediation models for general intelligence. Standardized estimates are reported.

** p < .005, ** p < .001.*

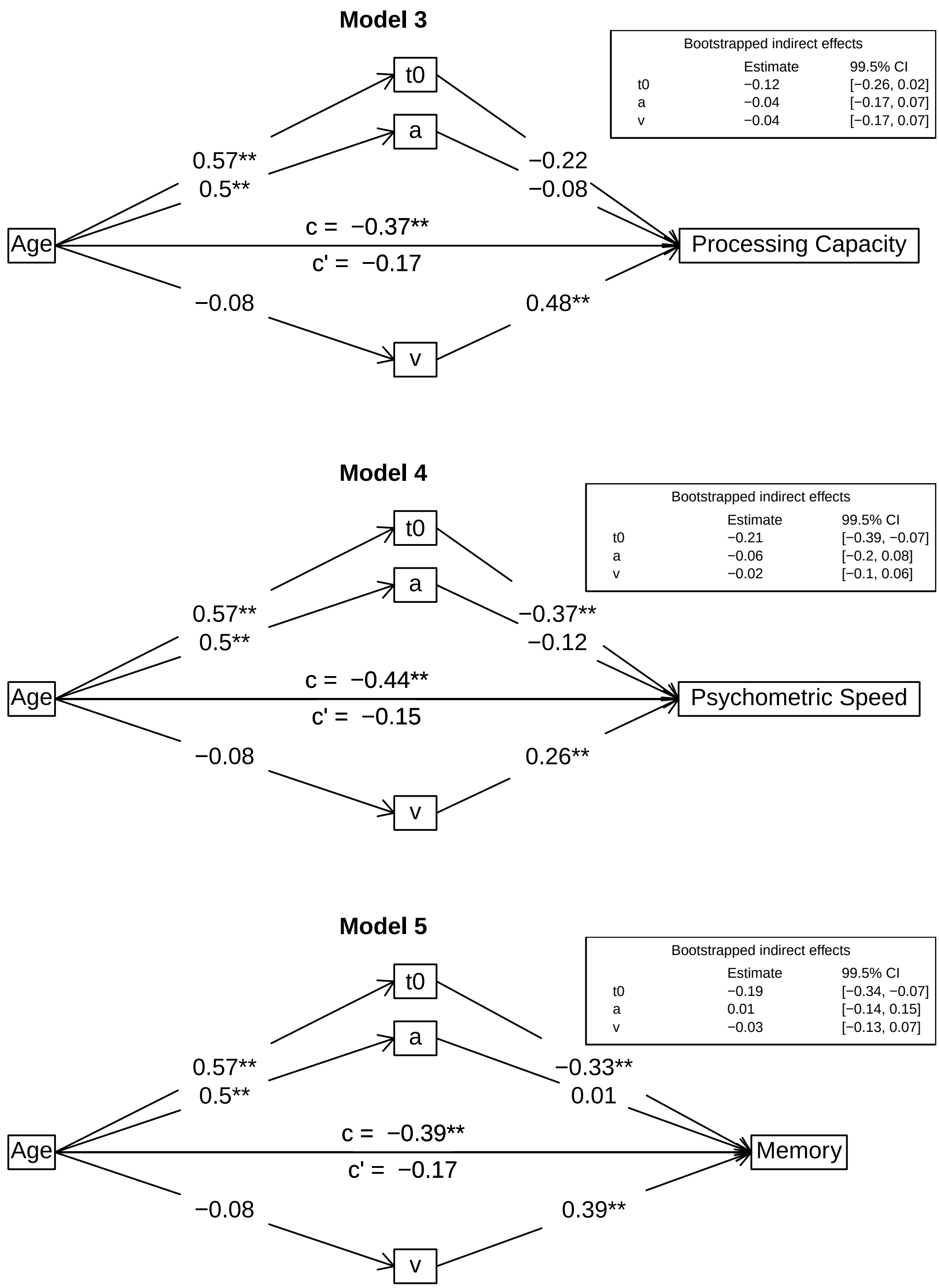

Figure 5

Mediation models for intelligence process domains. Standardized estimates are reported.

** p < .005, ** p < .001.*

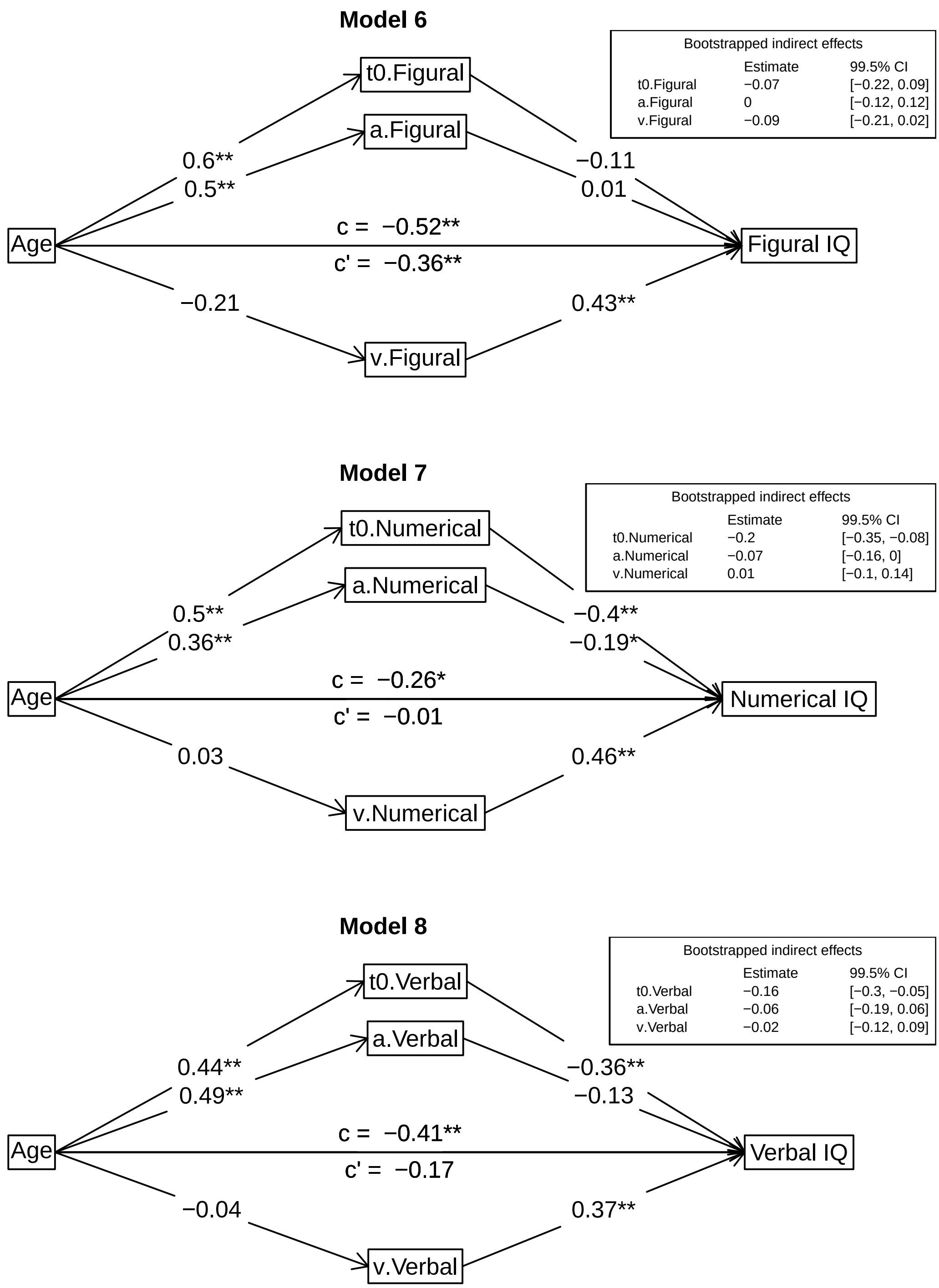

Figure 6

Mediation models for content domains. Standardized estimates are reported.

** p < .005, ** p < .001.*

Appendix

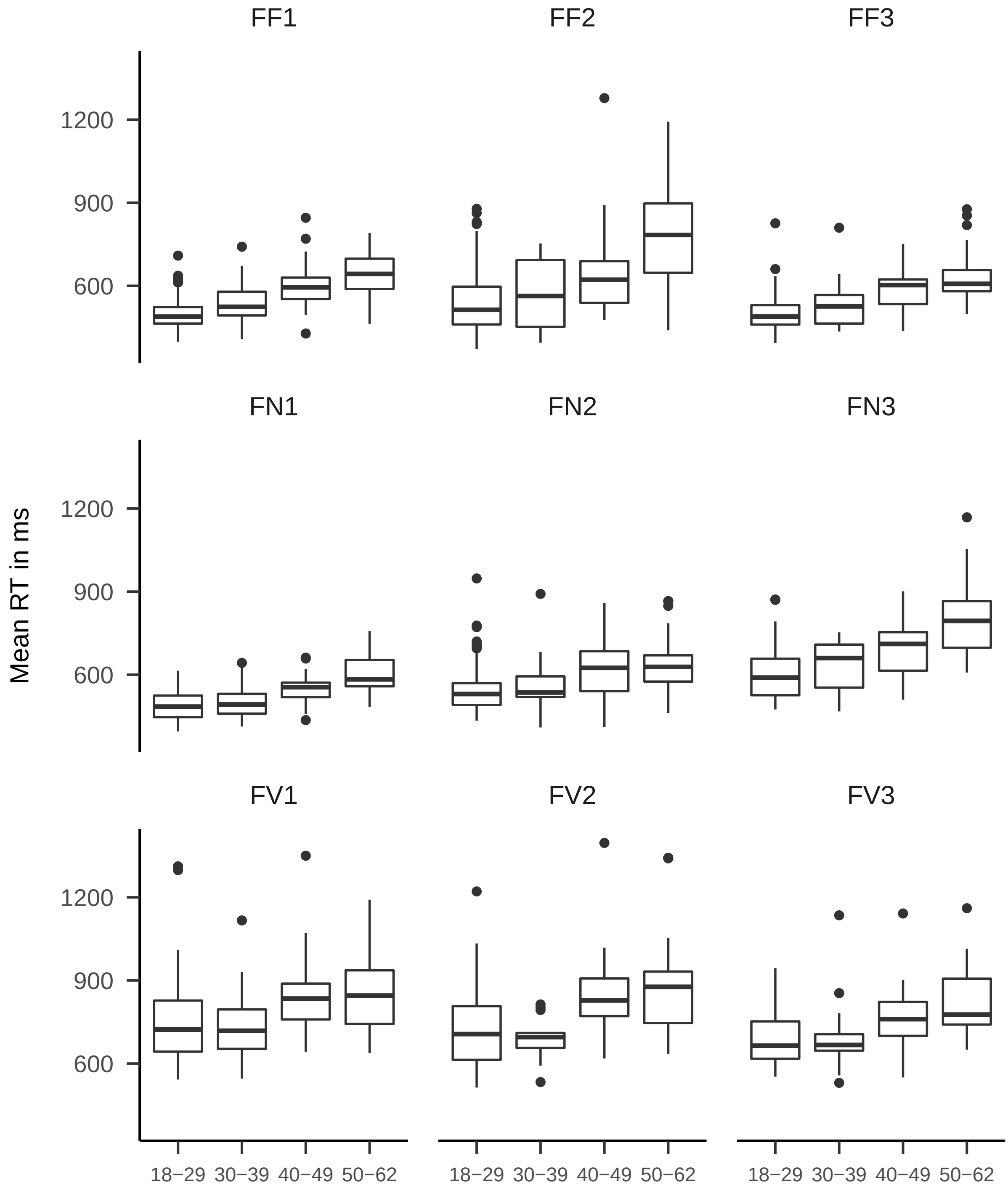

Figure A1

Boxplots of mean response times for all fast tasks, split by age groups. See Table 1 for an explanation of the task names.

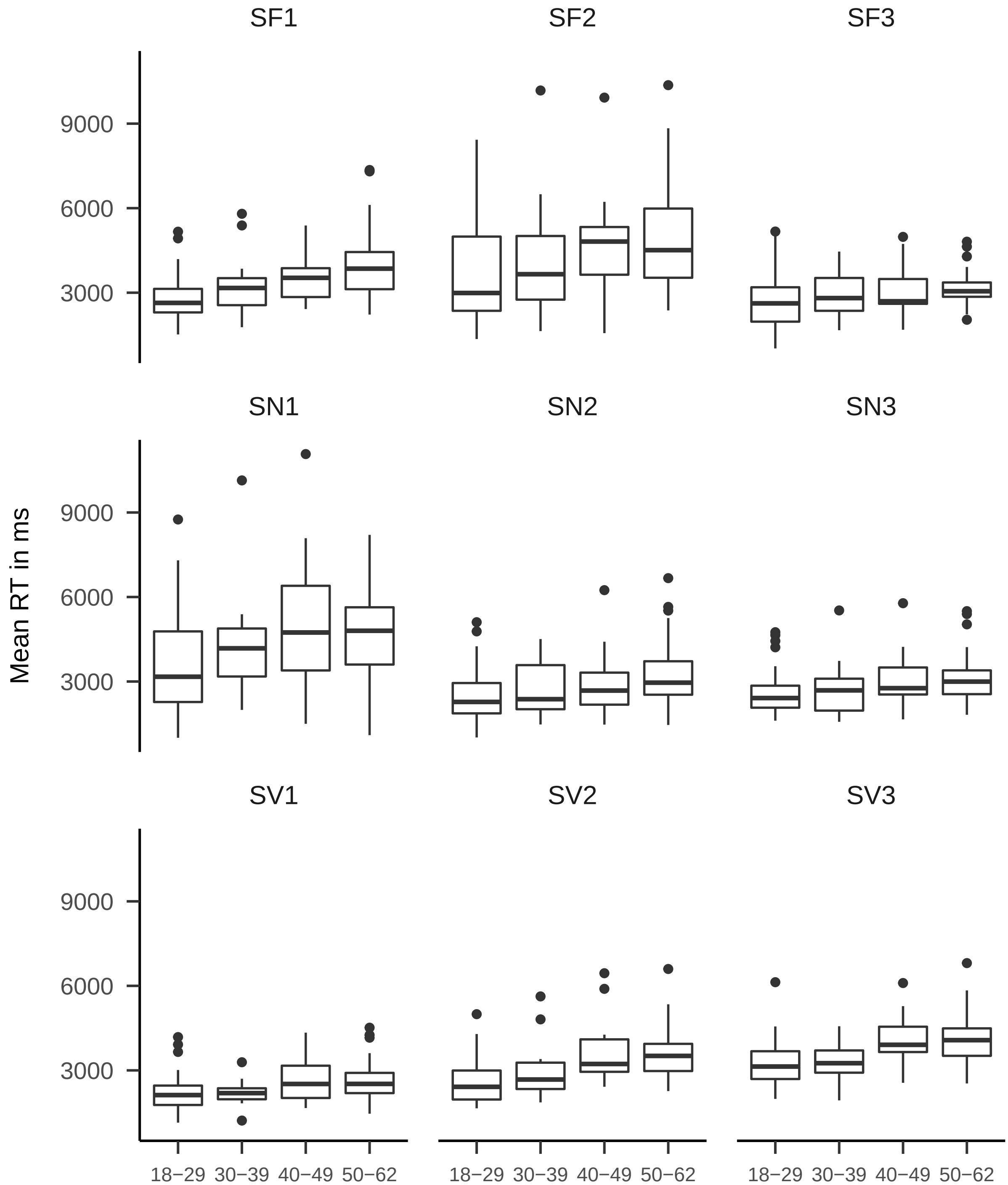

Figure A2

Boxplots of mean response times for all slow tasks, split by age groups. See Table 1 for an explanation of the task names.

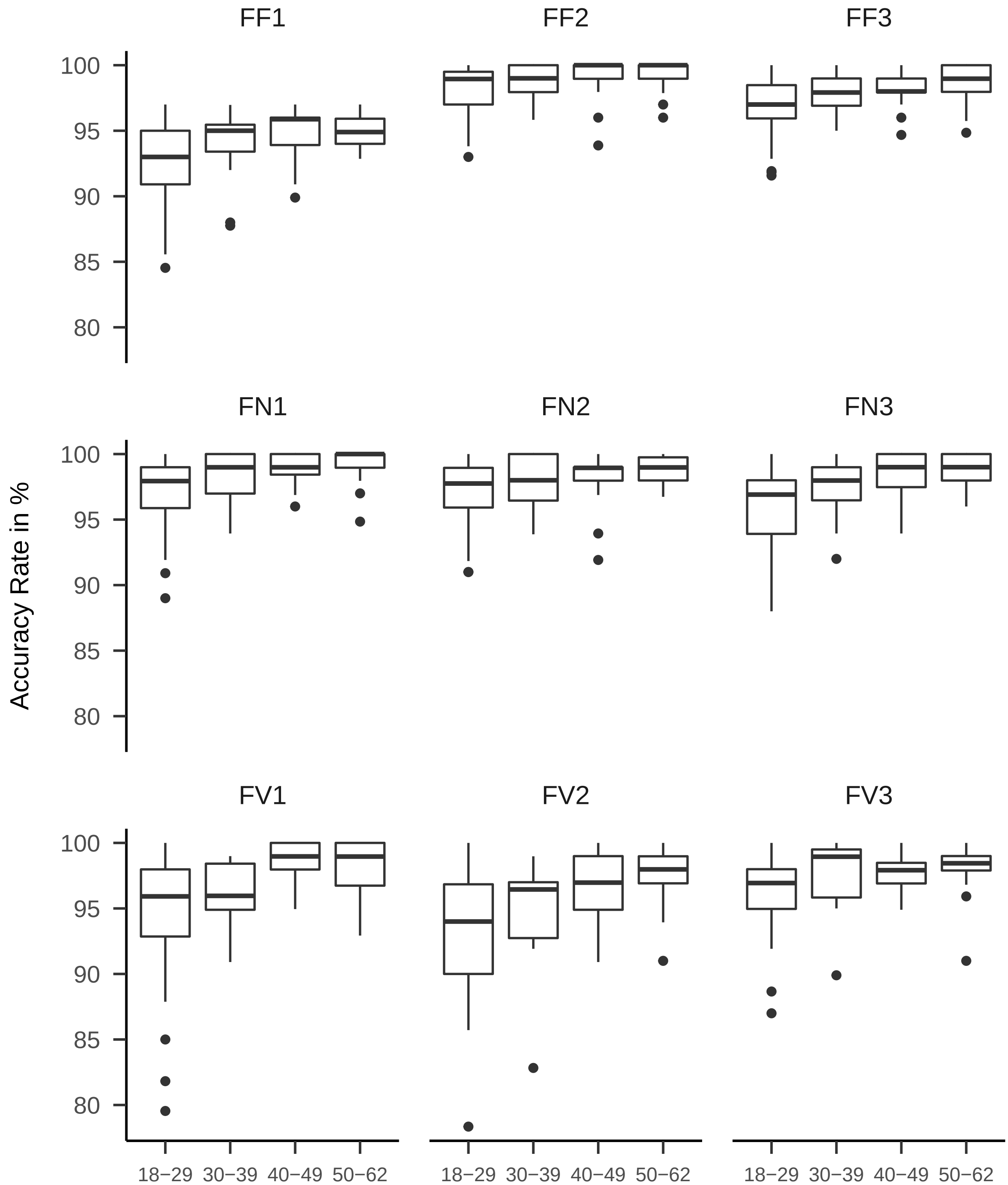

Figure A3

Boxplots of accuracy rates for all fast tasks, split by age groups. See Table 1 for an explanation of the task names.

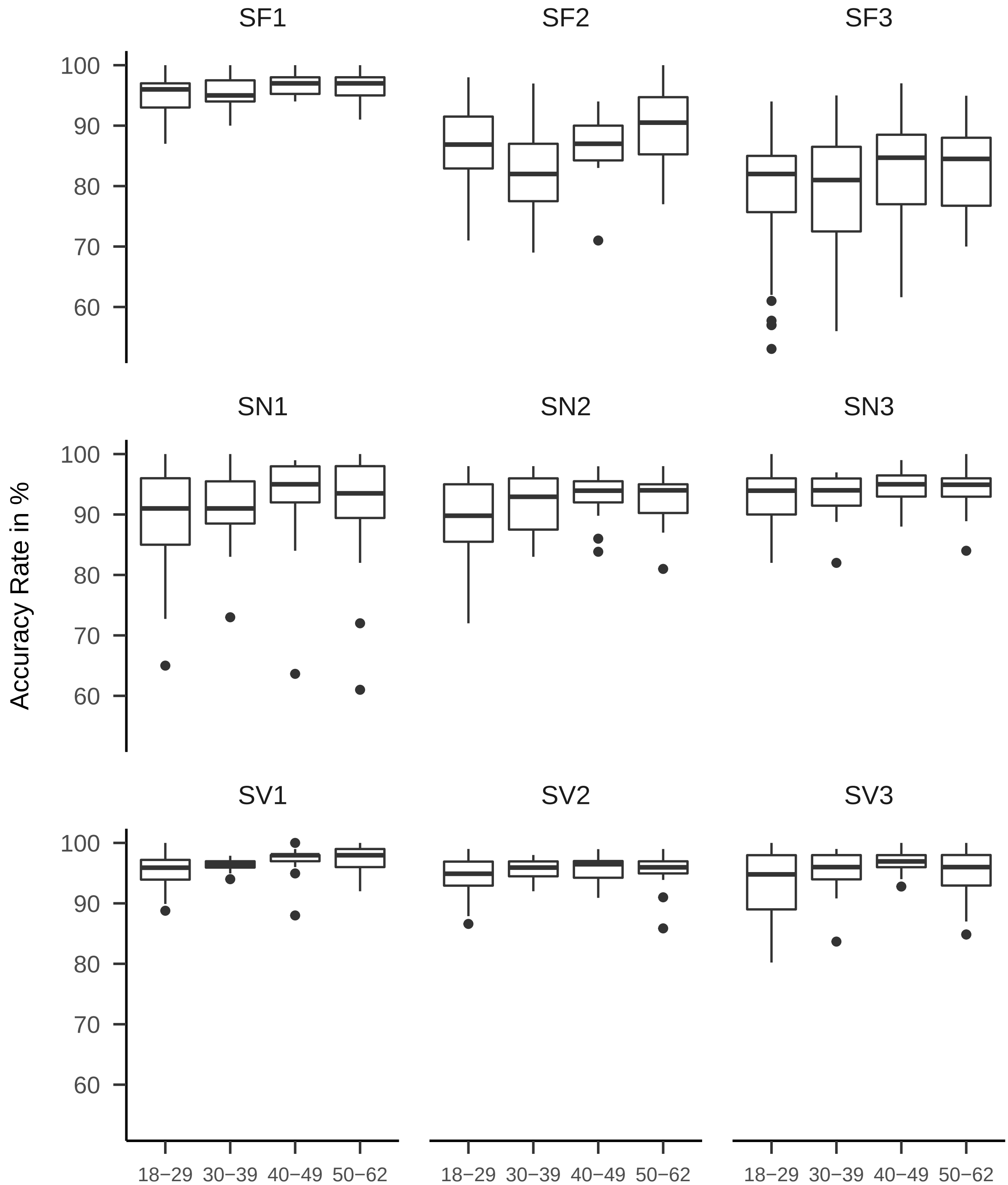

Figure A4

Boxplots of accuracy rates for all slow tasks, split by age groups. See Table 1 for an explanation of the task names.

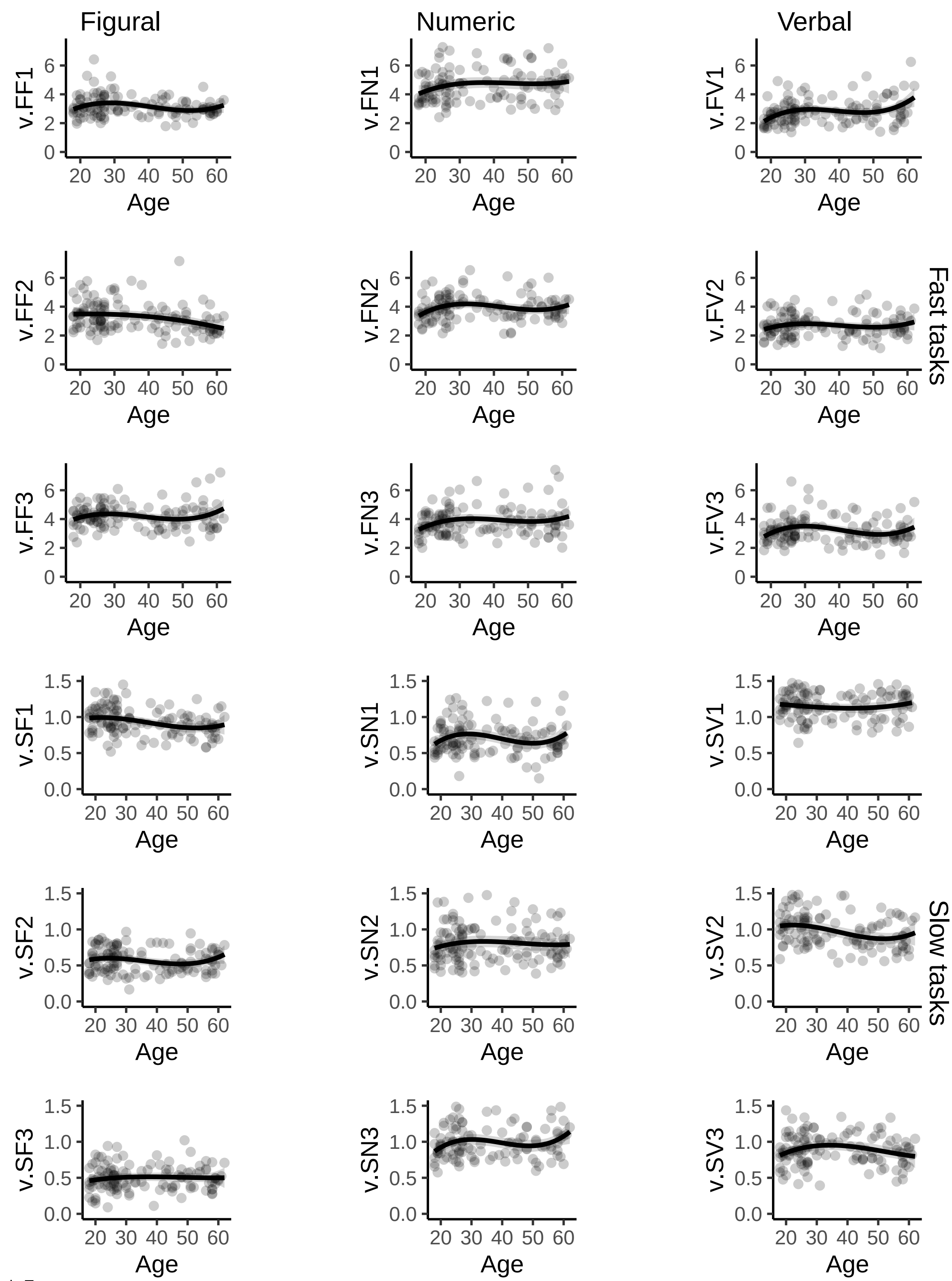

Figure A5

Scatterplots of drift rate vs. age in the 18 different tasks. See Table 1 for task descriptions. The trends show results from a polynomial regression of third degree.

Appendix A 5

Manuscript 5:

von Krause, M., Radev, S.T., & Voss, A. (submitted). Processing speed is high until age 60 - Insights from Bayesian modeling in a one million sample (with a little help of deep learning). *Proceedings of the National Academy of Sciences of the United States of America.*

Processing speed is high until age 60: Insights from Bayesian modeling in a one million sample (with a little help of deep learning)

Mischa von Krause[a1], Stefan T. Radev[a], and Andreas Voss[a]

[a]Heidelberg University

This manuscript was compiled on February 24, 2021

Processing speed is a fundamental aspect of human cognition and intelligence. Many studies from the last decades report that processing speed, typically measured as mean reaction time in simple cognitive tasks, significantly slows down in old age and already declines in young and middle adulthood. Our study employs a Bayesian diffusion model approach to disentangle different cognitive components involved in simple decision-making. We apply our model to a massive data set of more than one million participants, which allows us to provide fine-grained and robust analyses of age differences. Since standard Bayesian methods are not suitable to data sets of this size, we use a novel deep learning method for parameter estimation. Our results indicate that processing speed is stable from young adulthood until an age of about 60. The typical age-related slowdown in mean response times in this age range seems attributable to increases in decision caution and slower non-decisional processes – like encoding and motor response – but not to differences in cognitive processing speed. Our research has important implications for the study of cognitive aging.

Cognitive Aging | Cognitive Modeling | Processing Speed | Big Data | Deep Learning

Speed of information processing is a fundamental property of cognitive agents and an important prerequisite for timely and adequate responses in complex environments. Over the past decades, a large body of research has consistently found a negative relation between processing speed and age, that is, older people tend to be slower than younger people across a wide variety of cognitive tasks and contexts (1, 2). This approximately linear trend starts already in young adulthood, at ages 20 to 30 (1, 3–5), and has been reported in both cross-sectional and longitudinal studies (1, 5–7). The notion that processing speed already declines over young and middle adulthood has important implications for the study of human cognition. Since developmental patterns of cognitive abilities are linked to changes in the brain (8), studying the former can also provide insights into the neurophysiological basis of cognition.

The vast majority of findings on age and processing speed rely on mean response times (RTs) in elementary cognitive tasks (e.g., comparison of two letters) as a measure of basic speed of information processing (2, 3, 9). However, this approach has two major shortcomings. First, the solitary use of mean RTs does not utilize the full information contained in empirical response time distributions and ignores accuracy data also obtainable from experimental paradigms. Second, mean RTs are not pure measures of processing speed, but instead represent the sum total of disparate cognitive processes (10). For instance, speed-accuracy trade-offs (i.e., different settings of response caution that affect both speed and accuracy of responses) or the time taken up for encoding and motor processes contribute to the observed response time, although they are unrelated to cognitive processing speed. Thus, the extent to which mean RTs reflect processing speed is, at the very least, debatable (11–13).

Mathematical models of cognition strive to decompose behavior in interpretable and neurophysiologically plausible constructs. One of the most popular process models for explaining RT data is the diffusion model (14–18, DM, see **Materials and Methods** section for a more detailed description of the model). By employing the DM, it is possible to obtain a process-pure estimate of processing speed through the model's *drift rate* parameter. This measure of processing speed is independent of decision caution (*boundary separation*) and the time required for encoding and motor processes (*non-decision time*). Moreover, the parameters of the DM have been extensively validated both experimentally (19–21) and neurophysiologically (22–24).

In the past two decades, a growing number of diffusion modeling studies on age differences in a great variety of experimental environments has been published (12, 21, 25–36). Most of these studies compared groups of young adults, around age 20, with old adults, aged 65 and older, with respect to the model's parameters. Interestingly, it has often been reported that processing speed exhibits no differences between young and old adults. Conversely, decision caution and non-decision

Significance Statement

We present the first study to apply Bayesian diffusion modeling with generative neural networks on a massive data set of human response times. Our analysis implies that cognitive processing speed declines much later in life than previously assumed. Since processing speed is a central aspect of human cognition and intelligence, this finding has far-reaching implications for all fields concerned with human information processing and its developmental patterns. The age-related increase in response times observed in young and middle adulthood can be attributed to greater decision caution and slower encoding and motor times.

Author contributions: Conceptualization, M.v.K; methodology, M.v.K. and S.T.R; formal analysis, M.v.K. and S.T.R.; investigation, M.v.K..; data curation, M.v.K.; writing—original draft preparation, M.v.K. and S.T.R.; writing—review and editing, M.v.K., S.T.R, and A.V.; visualization, S.T.R. and M.v.K; supervision, A.V. All authors have read and agreed to the final version of the manuscript.

The authors declare no conflict of interest.

[1]Correspondence should be addressed to Mischa von Krause. E-mail: mischa.vonkrausepsychologie.uni-heidelberg.de

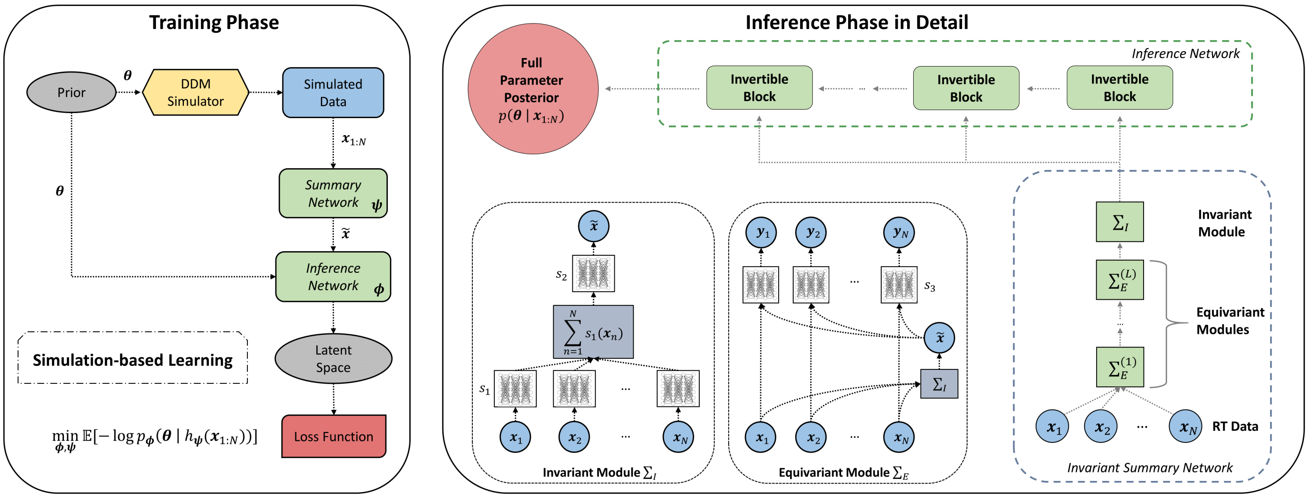

Fig. 1. The BayesFlow framework used for individual parameter estimation on a million data sets. Left panel: During training, the computational model serves as an "instructor", which, by means of simulations, guides the summary (ψ) and the inference network (ϕ) to become "experts" in inverting the model and recovering plausible estimates of cognitive parameters (θ). Right panel: During inference, the trained networks efficiently process each observed data and estimate the full posterior over parameters of interest. The training effort thus "amortizes" over multiple estimation passes, as no further training of the networks is required. Specialized *invariant* and *equivariant* networks (37) are required for processing the *i.i.d.* response times and accuracy data obtained in each IAT experiment.

times were often markedly increased in old age.

Although model-based analyses of cognitive aging have many advantages over the direct analysis of raw data, many model-based studies have two serious shortcomings, both related to the samples used. First, sample sizes were comparably small in most studies, which is especially problematic for research on individual differences seeking to increase reliability through larger samples. For instance, a recent meta-analysis summarizing 25 studies had a total sample size of only 1,503 observations, indicating an average sample size of 60 participants per study (38). Second, most studies only compared two age groups, typically college-age students and older adults aged 65 to 75. Taken together, these two aspects severely limit the generalizability of previous results, especially with regard to the age span between 25 and 65 years, that is, large parts of young and middle adulthood.

There are two main reasons for the small sample sizes common for diffusion modeling studies. First, data collection for such studies is tedious, given the large number of trials per person that were long thought to be required for diffusion modeling (15). However, such requirements are now considered as largely overstated (39, 40). Second, and more importantly, fitting the diffusion model to observed data is computationally expensive, especially when employing sampling-based Bayesian estimation methods. Thus, obtaining individual parameters even from moderately large samples is often infeasible for practical reasons. Yet, in order to provide a robust analysis of individual differences in processing speed in relation to age, a rather large data set including participants across the entire lifespan seems imperative.

In recent years, Bayesian methods have become the gold-standard for model-based inference in cognitive modeling (41). Bayesian methods allow for principled uncertainty quantification in the form of full posterior distributions over quantities of interest (e.g., the parameters of a cognitive model). Once estimated, the posterior distribution can be used to extract credibility intervals or to perform posterior predictions to assess the quality of model fit. Moreover, posterior correlations between model parameters can be extracted and used as a measure of (linear) disentanglement between parameters at an individual-level. However, a major disadvantage of standard Bayesian methods for cognitive models (e.g., Markov chain Monte Carlo methods) is their computational slowness, which makes them impractical or even impossible to apply in data-rich contexts. In this work, we therefore demonstrate the utility of a novel deep-learning framework developed to scale up model-based Bayesian inference to millions of data sets (42).

We present an analysis of cross-sectional age differences in diffusion model parameters estimated from a massive data set ($N > 1,000,000$), utilizing response times and accuracy rates collected in an online implicit association test (43). Notably, this sample is multiple orders of magnitude larger than the samples used in all previous diffusion model studies combined. Our deep-learning architecture for parameter estimation is based on a a two-stage inference framework which is illustrated in Figure 1 and further described in the **Materials and Methods** section (42). Regarding chronological age, our sample covers childhood till late adulthood (ages 10 to 80), with a sufficient depth for fine-grained and robust year-by-year analysis.

Our study is the first to derive substantial insights into individual differences in cognitive parameters by applying Bayesian diffusion modeling to a large sample with the help of modern deep learning methods. Accordingly, our approach yields unique and robust findings on age-related patterns of different aspects of cognition, separating processing speed, decision caution, and non-decision parts of response times.

We observe a clear non-linear association between drift rate (as an index of processing speed) and age, which is strikingly different than the one implied by mean RTs and far more informative than the age differences found in previous diffusion model studies. Thus, our model-based analysis reveals a picture of age differences in cognitive parameters yielding a

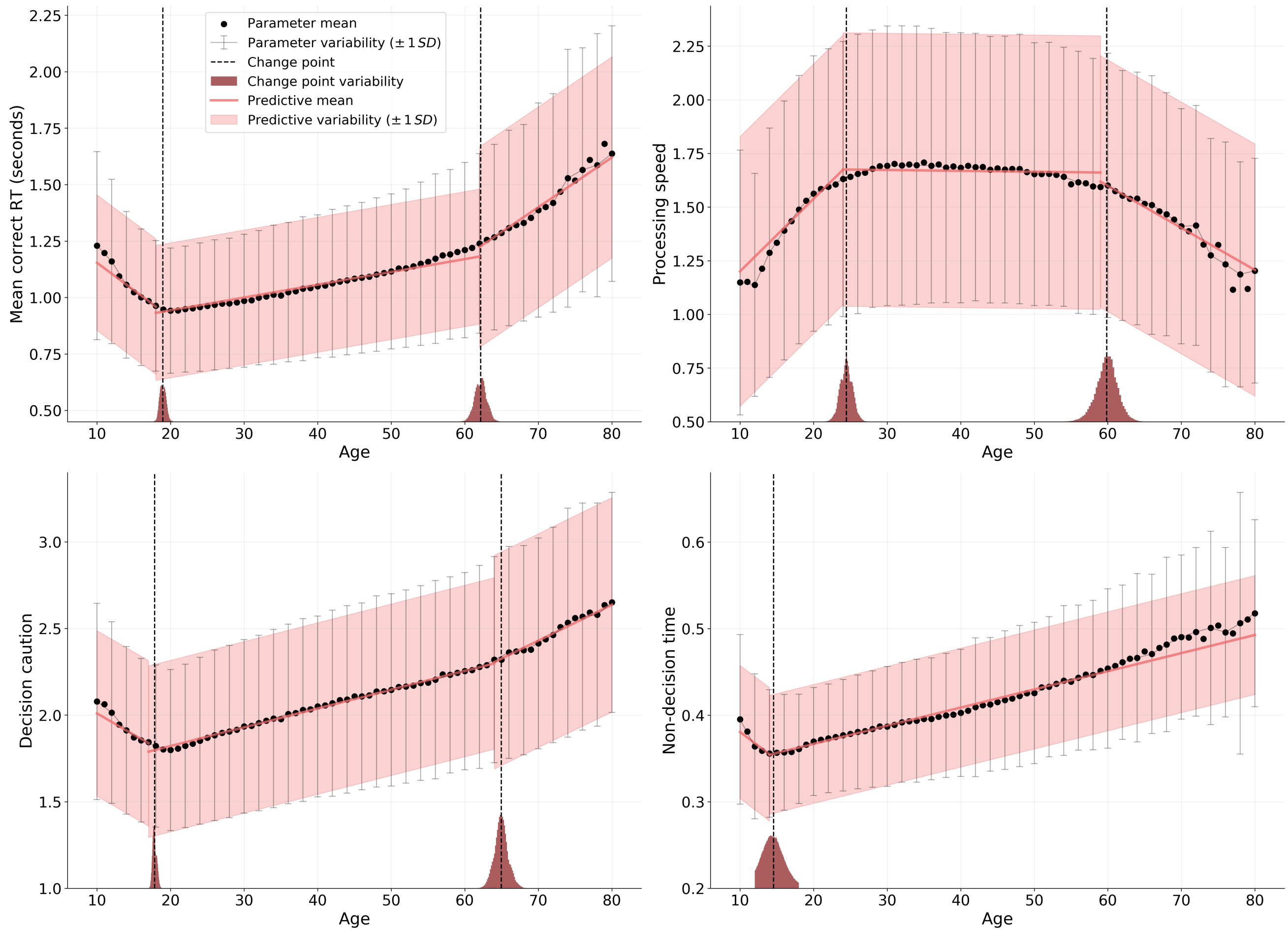

Fig. 2. Mean correct response times (RTs) and diffusion model parameters as a function of age. Black points indicate means computed separately for each year of age. Bars indicate standard deviations (only shown for every second year for better clarity). Red lines denote the Bayesian piece-wise ridge regression model's mean predictions, which describe the observed means fairly well. The shaded red region denotes the uncertainty (standard deviation) of the piece-wise model's predictions. The dashed lines indicate the mean change points estimated from the per-age-group averaged data, with the full posterior distributions (scaled for readability) of the change points shown at the bottom of each plot. Both the data- and model-implied standard deviations highlight the great variability within each year of age. Nevertheless, the year-specific means suggest a clear and consistent pattern for mean correct RT and each parameter. The figure depicts drift rates and boundary separations for the incongruent condition and non-decision times obtained from correct responses. Very similar trends for the congruent condition and non-decision times from incorrect responses can be found in the **SI Appendix**

radically different implication than the one based on analyses of raw RT data.

Results

Figure 2 depicts our main findings. Mean correct response times, processing speed, decision caution, and correct non-decision time are plotted against age in years. The figure shows results for one of the two experimental conditions (incongruent trials). The other condition (congruent trials) yields very similar patterns, which are presented in the **SI Appendix**. Each dot represents the mean of the individual posterior parameter means for one year of age. The vertical bars represent one standard deviation within each year of age. Descriptive statistics for all parameters can be found in the **SI Appendix**. To better describe the age-related patterns we found, we estimated linear Bayesian change-point models combined with piece-wise Bayesian ridge regressions (see **Materials and Methods**). The estimated change points and piece-wise regression lines together with their respective uncertainties are also depicted in Figure 2.

As evident from Figure 2, cross-sectional mean correct RTs decrease sharply from the age of 10 to about 20, with the change point regarding the age trend estimated at age 19. After that, mean correct RTs show a quasi-linear increasing trend until the estimated change point of age 62. Thereafter, the average increase in response times per year accelerates, although data become more sparse when approaching age 80 (e.g., $n = 169$ for age 80).

Drift rates, that is, our proxy for processing speed, increase notably from age 10 to 30 in our cross-sectional data. After this, mean levels in drift rates remain fairly stable until an age of 60, showing little age-related differences during middle adulthood. At around the age of 60, an accelerated negative trend in cognitive processing speed commences, which holds until an age of 80. Importantly, this inverted U-shaped pattern does not mirror the age trends found for the other diffusion model parameters or mean RTs. Our change points are estimated at ages 25 and 60. The change point model misses the minor increase in drift rates that continues until age 30, as well as the slight decrease in drift rates starting at age 50.

Boundary separation, that is, estimates of decision caution, decrease from age 10 to about age 20, after which they show a quasi-linear increase until an age of 65. Thereafter, the average increase in response times per year accelerates. Change points are estimated at ages 18 and 65. It should be noted that in the congruent experimental condition (see **SI Appendix**), the change point for boundary separation was already estimated at age 50, and the subsequent increasing trend was less pronounced there.

Finally, non-decision time estimates, that is, the time taken for encoding and motor response, decrease from age 10 to the estimated change point of age 15, after which show a quasi-linear increase until an age of 80. The age differences for decision caution and non-decision times closely mirror the pattern found for RTs, suggesting that these components could have a large impact on the mean levels of response latencies over the life course.

As can further be observed, variability in mean correct RTs increases across the lifespan. The trend is paralleled by the increase in variance found for non-decision times. Conversely, the between-person variability for boundary separation and drift rates shows no age-related increase.

In order to ensure that our findings hold across a wide range of conditions, we conducted several robustness checks. Figure 3 shows that the mean level trend for drift rates is robust across genders, levels of education, and experimental conditions (congruent vs. incongruent). However, the increased decline in drift rates after age 60 is more pronounced for the incongruent condition, and women show higher mean levels of drift rates also in the incongruent condition. The vertical bars in Figure 3 indicate standard errors of the means. Due to the very large sample size, standard errors are very small for all age groups except for the very old participants. This guarantees that the differences in processing speed across the lifespan were assessed very accurately. We performed additional robustness checks by comparing the trends in age effects across different sub-samples. For this purpose, we first divided the sample into four almost evenly sized sub-samples. Across these sub-samples, mean-level patters were virtually identical. The same was true when comparing participants born in the United States with those originating from other countries, as well as for the comparison between participants working on tasks with different classes of stimuli (i.e., "Black/White" or "African American/European American"). All these additional analyses can be found in the **SI Appendix** where we also report correlations between the different diffusion model parameters, both across participants and within each person - the latter by utilizing the individual posterior distributions.

Discussion

In this work, we presented a cross-sectional study of age differences in mean response times and cognitive processes as measured by the diffusion model. We applied the diffusion model to a massive data set, containing response time and accuracy data from the implicit association test (IAT). Our sample covers large parts of the human lifespan (ages 10 to 80) in sufficient depth for a fine-grained analysis of age differences at a year-specific level. To our knowledge, this is the first study to apply diffusion modeling on data of this magnitude. Given the sample size, our analyses would have been infeasible using standard parameter estimation procedures. Thus, the novel deep learning approach utilized in this work for parameter estimation was both necessary and extremely efficient for the task at hand. Our findings stand in pronounced contrast to previous findings on age differences in processing speed. The implications of our work are thus relevant for all research domains studying cognitive development across the lifespan. We will now discuss the implications of our findings in turn.

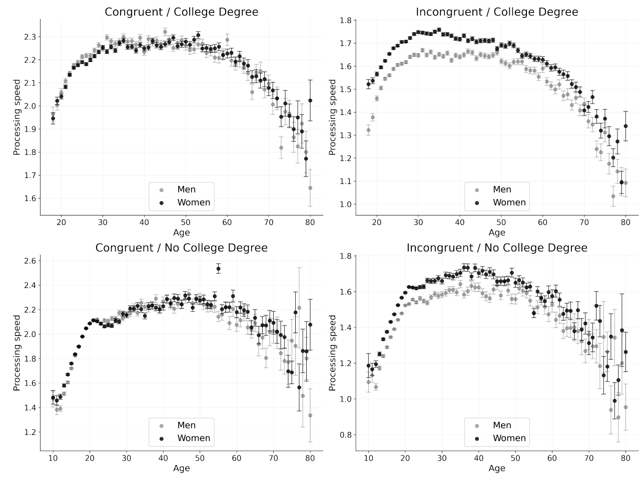

Fig. 3. Processing speed as a function of age, experimental condition, and demographic variables. We observe the same inverted *U*-shape as in our main analysis. Vertical bars denote standard errors of the mean (SEMs).

Mean response times. Our results replicate the age-related decline in mean response times previously reported (1–5). In our sample, mean response times showed a negative trend during the teenage years, were fastest around age 20, and showed a nearly linear increase thereafter, which further accelerated after the age of 60. It is important to note that these findings are in line with earlier response time studies. This indicates that the diverging patterns found for the diffusion model parameters are not based on qualitatively different raw data than previously collected in the field.

Decision caution. In the diffusion model, decision caution, that is, the amount of information sampled before making a decision, is represented by the boundary separation parameter (16). Our results suggest that, on average, boundary separation declines from age 10 to approximately age 18, indicating that people at college age were the least cautious in our sample - they were most willing to trade off accuracy for speed. After age 18, decision caution increases linearly until about age 65, with a greater increase per year thereafter until age 80. In the congruent condition, the increasing trend in old age was less pronounced, maybe due to lower task difficulty. For both conditions the implication is that, on average, the older a person in the sample was, the less likely she was to make rapid decisions based on sparse information. Moreover, the trend towards higher decision cautiousness becomes noticeable already very early in adult life. Thus, the increase in the amount of information sampled before making a decision provides a first explanation for the age-related increase in RTs starting in young adulthood.

Encoding and motor processes. Non-decision time is the diffusion model parameter that represents all processes beyond information sampling and evidence accumulation in a decision task. These processes are typically thought to encompass the time taken for encoding of stimuli and motor response execution (20, 44). Interestingly, in our sample, non-decision times were, on average, fastest around age 14 to 16, with people outside this range needing more time for non-decision processes. It seems that these processes, should they represent a trait-like ability, reach their peak earliest among all cognitive abilities typically studied in the literature (5). After age 16, non-decision times exhibit a linear increase that continues up until age 80. Thus, the increase in the time needed for non-decision processes provides a second explanation for the slower RTs found with increasing age, already among young adults and in middle adulthood.

Processing speed. Our most significant finding concerns the drift rate, that is, the parameter representing processing speed in the diffusion model framework. The drift rate denotes the average rate of information sampling per time unit and, theoretically, represents a process-pure measure, because speed-accuracy trade-offs and non-decision aspects have been controlled for by the other two main diffusion model parameters. During early adulthood, drift rates showed, on average, a continuous positive age trend, that is, processing speed became faster from age 10 to age 30. Processing speed thus peaks notably later than the lowest points of decision caution (around age 20) and non-decision times (around age 15). This result partly mirrors previous findings reporting that processing speed is still high around age 30 (4). Yet, in our sample, processing speed showed a slight increase from the age of 20 to 30, which is in contrast to previous findings based on the analysis of mean RTs. It should be noted that our change point analysis indicated that the positive age trend in drift rates is weaker from age 25 to age 30 than in the years before that, with the corresponding change point estimated at age 25.

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

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

Our key findings also explain the age-related findings reported in previous diffusion modeling studies. Typically, these studies compared two groups of participants: college-aged students and people aged 65 and older (12, 21, 25–33, 36). A consistent result of these studies was that older participants show higher boundary separations and non-decision times, but comparable drift rates.

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

Our results are also in line with recently reported results on age differences in diffusion model parameters using a continuous assessment of age (34). In this study, for a wide variety of different tasks a peak in processing speed around age 30 was also observed. However, sample size across later young adulthood and middle adulthood was too small to reveal clear age trends.

Another interesting finding emerging from our study is the fact that diffusion model parameters showed different cross-sectional patterns of across-person variability over the lifespan. While the variances of boundary separation and drift remained roughly the same even into old age, non-decision times showed an increase in variability after the age of 60. The latter pattern is also present for mean RTs. Thus, it seems plausible that the greater spread in mean RTs observed for older people is attributable to greater inter-individual differences in encoding and motor processes, not in processing speed or decision caution.

Finally, this is one of the first studies to report age differences in diffusion model parameters in late childhood and adolescence (but see 45), thus allowing the study of differentiable temporal patterns in these age periods. Most notably, the fastest non-decision times were observed already at ages 14 to 16, with mean RTs, processing speed, and decision caution all showing much later turning points.

The differing age-related patterns of the diffusion model parameters become more plausible when viewed in the context of the literature linking changes in cognitive abilities with changes in their neurophysiological basis (8). According to the Scaffolding Theory of Aging and Cognition (STAC-r; 46), people differ in their use of different compensatory techniques (e.g., activation of additional neural networks), all of which aim to counter the detrimental effects of age-related changes in brain structure. While such compensatory strategies might be well-suited to keep the level of processing speed in simple decision-making tasks high across large parts of the lifespan, more basal processes such as the ones captured in non-decision time might be less adaptable (34).

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

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

A second limitation concerns the cross-sectional nature of our findings. Thus, it remains an open question whether the age differences and trends found in our data represent within-person developmental processes. We did not study longitudinal change, and neither did we account for cohort effects. However, given the clear age trends (with the majority of means almost perfectly aligned across age groups) found for the cognitive parameters of interest, we argue that our data provide as clear a picture of developmental patterns as is reasonably achievable using cross-sectional data. We also want to note that the IAT data made publicly available by Project Implicit (43) includes data-sets from the years 2002 to 2020, thus making it possible to study cohort effects, and also participant IDs, making it possible to study longitudinal change in participants taking the task several times. Such analyses were beyond the scope of this paper, but might be well worthwhile in future endeavours.

For anyone interested in replicating or expanding our analyses using similar data and estimation methods, we provide open source code and pre-trained deep learning networks for pre-processing and obtaining Bayesian diffusion model parameter estimates on our GitHub page (https://github.com/stefanradev93/DataSizeMatters).

To conclude, according to our analysis of a massive data set of human response times, processing speed increases until the age of 30, remains at a plateau until an age of around 60, and declines thereafter. Furthermore, the slowdown in mean response times found already in young and middle adulthood seems largely attributable to age-related changes in decision caution and non-decision times. In old age, cumulative effects of all three cognitive parameters - processing speed, decision caution, non-decision times - contribute to an accelerated slowdown that is also evident from the raw RT data.

Materials and Methods

Our analyses are based on publicly available race IAT data provided by Project Implicit (43). We extracted raw response time and accuracy data, as well as demographics, all of which were collected from September 2016 to December 2018. All data are openly available on the Project Implicit OSF page: https://osf.io/y9hiq/. In addition, all analysis scripts for reproducing the results are available at: https://github.com/stefanradev93/DataSizeMatters

Participants. Our original sample contained $1,804,325$ people. We excluded cases that did not complete the task, did not provide their year of birth, were older than 80 or younger than 10 years at time of data collection, or had more than 10% response times under 300 ms. In order to be able to obtain full data-informed parameter sets for each person, including the error non-decision time, we excluded all participants with 100% accuracy. Further, we excluded trials faster than 300 ms or slower than 10 seconds. After fitting the diffusion models, we excluded cases with estimates for drift rates, boundary separations, or non-decision times beyond the borders of our respective (very broad) prior ranges (see below: **The diffusion model**). This left us with a final sample of $1,185,898$ people. Of these, 38.27% were female and 61.29% were male (the question asked was about the sex assigned at birth). Mean age was 27.41 years ($SD = 12.33$), with a robust sample size across the entire age span of 10 to 80 years. About halve of the participants (46.89%) had completed at least college level education. The majority (84.06%) of the participants indicated that they were born in the USA, with the rest reporting different countries of origin.

Task. The race IAT is a quasi-standard cognitive task originally designed to measure implicit racial bias (47). In a series of binary decisions, people have to classify words and images as belonging to one of two categories, for example "good/bad" or "Black person/White person". Across the two different main blocks of the experiment, the mappings of the categories to the same response button change. "Good" might share a common response key (e.g., left) with "Black person" in the first condition, and then be paired with "White person" in the second condition. 60 trials are completed in each of the two conditions. The difference in mean response times is then used to obtain a measure of implicit bias (48). The exact procedure and materials can be found on the Project implicit OSF page (https://osf.io/y9hiq/, 43). We did not use the IAT as an instrument to study implicit cognition, but instead as an example of a simple binary decision task.

The diffusion model. In the present work, we employ the diffusion model (DM), a prominent mechanistic model of neurocognitive dynamics designed to explain human performance in simple decision-making tasks (16). The DM is embedded in the larger model class of evidence accumulator models (EAMs), which conceptualize information processing as a gradual, temporally-ordered, and noisy process (49). A core assumption of the DM is that task-relevant information is integrated at multiple neurocognitive levels in which sensory evidence for one of the alternatives is dynamically accumulated at a constant rate. A categorical decision for one of the alternatives is determined as soon as a pre-defined threshold is reached. Moreover, the key parameters of the DM are well-validated in experimental settings (19–21) and well-grounded in biological neural-network theory (49).

In order to decompose performance in the race IAT into meaningful cognitive constructs, we formulate and fit a DDM with six parameters: $\boldsymbol{\theta} = (v_1, v_2, a_1, a_2, \tau_c, \tau_n)$. Here, v_1 and v_2 denote the speed of information processing (drift rates) in the two experimental conditions; a_1 and a_2 denote the decision thresholds (boundary separation); τ_c and τ_n denote the additive non-decision time constants for correct and incorrect responses, respectively. We estimate separate drift rates and boundary separations for the congruent and incongruent conditions, because these parameters have been shown to differ across the IAT conditions in previous studies (50). We estimate separate non-decision times for correct and incorrect trials due to the nature of the race IAT task (i.e., trials do not terminate immediately following a wrong response but require an additional response from the participants).

Our choice of Bayesian priors for the DM parameters reflects the goal to cover meaningful parameter ranges, as known from previous studies (51). However, we also place uniform priors over the plausible numerical ranges in order to render the data maximally informative for posterior inference. We place broad uniform priors over both drift rates, that is, $v \sim \mathcal{U}(0.1, 7)$, which we deem sufficient to cover the entire range of realistically observable processing speeds. On the basis of similar considerations, we place a broad uniform prior over the boundary separation parameters, $a \sim \mathcal{U}(0.1, 4)$. For the non-decision constants, we use $\tau_c \sim \mathcal{U}(0.1, 3)$ and $\tau_n \sim \mathcal{U}(0.1, 7)$, incorporating our expectation of longer non-decision times for incorrect responses in the particular task.

Parameter estimation. Performing Bayesian estimation on hundreds of thousands participants is not feasible with current gold-standard Markov chain Monte Carlo (MCMC) methods. We therefore resort to *amortized Bayesian inference* via specialized neural networks, which nevertheless guarantee correct posterior inference under perfect convergence (42). The term amortized inference refers to an approach which reduces the computational cost of Bayesian estimation by splitting the analysis into a costly upfront training phase, followed by an extremely efficient inference phase (42).

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

$$p(\boldsymbol{\theta}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = p(\boldsymbol{\theta}) \prod_{n=1}^{N} p(\boldsymbol{x}_n \,|\, \boldsymbol{\theta}) \quad [1]$$

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

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

$$\min_{\boldsymbol{\phi}, \boldsymbol{\psi}} \mathbb{E}_{p(\boldsymbol{\theta}, \boldsymbol{x})} \left[-\log p_{\boldsymbol{\phi}}(\boldsymbol{\theta} \,|\, h_{\boldsymbol{\psi}}(\boldsymbol{x}_{1:N})) \right] \quad [2]$$

which corresponds to minimizing the discrepancy between the true and the approximate amortized posterior induced by the networks. To train the networks, we performed approximately 50 000 simulations from the DM model with the priors for the parameters as described in the previous paragraph. Training the networks took approximately 8 hours on a GPU-accelerated laptop. Inference on the entire data set took approximately 24 hours on a machine without GPU-acceleration.

Bayesian Workflow. To further enhance the transparency and trustworthiness of our Bayesian pipeline, we follow the steps pertaining to a *principled Bayesian workflow*, as advocated by (52). Accordingly, we partition our pipeline into the following steps: i) prior predictive checks; ii) checks of computational faithfulness; iii) checks of model adequacy/sensitivity; iv) posterior predictive checks. These validation results, along other robustness analyses, are described and visualized in the **SI Appendix**.

Curve fitting. Given the massive data set available for data mining, inferential statistics were of minor importance to our analyses. Due to the non-linear age-related patterns of cognitive parameters, we computed separate piece-wise Bayesian ridge regressions of each quantity of interest (mean correct RT and DM parameters) on age as the simplest and yet reasonable approximation of the observed age trends.

Our statistical analyses followed a two-step approach. First, we performed a linear Bayesian change-point regression on the age-group averaged data using the *R*-package for multiple change points *mcp* (53). Note, that this step ignores all variability within an age group and thus focuses on fast change point detection, which otherwise would have been infeasible if executed on the full data. In a second step, we extracted the posterior distributions of each change points and used the corresponding posterior means for a piece-wise Bayesian ridge regression on the full data set. In this way, the piece-wise model's predictive means and uncertainty account for the full variability in the estimated parameters.

We placed the following priors over change points to broadly reflect the trends visible in the data: mean correct response times - $t_1 \sim \mathcal{U}(15, 25)$, $t_2 \sim \mathcal{U}(50, 70)$; drift rates - $t_1 \sim \mathcal{U}(20, 40)$, $t_2 \sim \mathcal{U}(50, 70)$; boundary separations - $t_1 \sim \mathcal{U}(15, 25)$, $t_2 \sim \mathcal{U}(50, 70)$; non-decision times - $t_1 \sim \mathcal{U}(12, 18)$, where the scales of measurement correspond to chronological age. For the Bayesian ridge regression, we used the default priors available through the *scikit-learn* implementation in the Python programming language.

ACKNOWLEDGMENTS. This research was supported by a grant from the German Research Foundation to the Graduate School SMiP (GRK 2277; Statistical Modeling in Psychology).

1. TA Salthouse, Selective review of cognitive aging. *J. Int. Neuropsychol. Soc. JINS* **16**, 754 (2010).
2. AR Jensen, *Clocking the mind: Mental chronometry and individual differences.* (Elsevier), (2006).
3. TA Salthouse, The processing-speed theory of adult age differences in cognition. *Psychol. review* **103**, 403 (1996).
4. TA Salthouse, What and when of cognitive aging. *Curr. directions psychological science* **13**, 140–144 (2004).
5. JK Hartshorne, LT Germine, When does cognitive functioning peak? the asynchronous rise and fall of different cognitive abilities across the life span. *Psychol. science* **26**, 433–443 (2015).
6. KW Schaie, What can we learn from longitudinal studies of adult development? *Res. human development* **2**, 133–158 (2005).
7. D Zimprich, M Martin, Can longitudinal changes in processing speed explain longitudinal age changes in fluid intelligence? *Psychol. aging* **17**, 690 (2002).
8. J Oschwald, et al., Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. *Rev. Neurosci.* **31**, 1–57 (2019).
9. GT Frischkorn, AL Schubert, Cognitive models in intelligence research: Advantages and recommendations for their application. *J. Intell.* **6**, 34 (2018).
10. RG Pachella, The interpretation of reaction time in information processing research, (Michigan University Ann Arbor Human Performance Center), Technical report (1973).
11. AL Schubert, GT Frischkorn, Neurocognitive psychometrics of intelligence: How measurement advancements unveiled the role of mental speed in intelligence differences. *Curr. Dir. Psychol. Sci.* **29**, 140–146 (2020).
12. R Ratcliff, A Thapar, G McKoon, Individual differences, aging, and iq in two-choice tasks. *Cogn. psychology* **60**, 127–157 (2010).
13. V Lerche, et al., Diffusion modeling and intelligence: Drift rates show both domain-general and domain-specific relations with intelligence. *J. Exp. Psychol. Gen.* **1** (2020).
14. R Ratcliff, A theory of memory retrieval. *Psychol. review* **85**, 59 (1978).
15. R Ratcliff, G McKoon, The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation* **20**, 873–922 (2008).
16. R Ratcliff, JN Rouder, Modeling response times for two-choice decisions. *Psychol. science* **9**, 347–356 (1998).
17. A Voss, M Nagler, V Lerche, Diffusion models in experimental psychology: A practical introduction. *Exp. psychology* **60**, 385 (2013).
18. D Fudenberg, W Newey, P Strack, T Strzalecki, Testing the drift-diffusion model. *Proc. Natl. Acad. Sci.* **117**, 33141–33148 (2020).
19. V Lerche, A Voss, Experimental validation of the diffusion model based on a slow response time paradigm. *Psychol. research* **83**, 1194–1209 (2019).
20. A Voss, K Rothermund, J Voss, Interpreting the parameters of the diffusion model: An empirical validation. *Mem. & cognition* **32**, 1206–1220 (2004).
21. NR Arnold, A Bröder, UJ Bayen, Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychol. research* **79**, 882–898 (2015).
22. DP McGovern, A Hayes, SP Kelly, RG O'Connell, Reconciling age-related changes in behavioural and neural indices of human perceptual decision-making. *Nat. human behaviour* **2**, 955–966 (2018).
23. R Ratcliff, YT Hasegawa, RP Hasegawa, PL Smith, MA Segraves, Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J. neurophysiology* **97**, 1756–1774 (2007).
24. S Kühn, et al., Brain areas consistently linked to individual differences in perceptual decision-making in younger as well as older adults before and after training. *J. Cogn. Neurosci.* **23**, 2147–2158 (2011).
25. BH Ball, AJ Aschenbrenner, The importance of age-related differences in prospective memory: Evidence from diffusion model analyses. *Psychon. Bull. & Rev.* **25**, 1114–1122 (2018).
26. J Dully, DP McGovern, RG O'Connell, The impact of natural aging on computational and neural indices of perceptual decision making: A review. *Behav. brain research* **355**, 48–55 (2018).
27. M Janczyk, P Mittelstädt, C Wienrich's, Parallel dual-task processing and task-shielding in older and younger adults: Behavioral and diffusion model results. *Exp. aging research* **44**, 95–116 (2018).
28. G McKoon, R Ratcliff, Aging and iq effects on associative recognition and priming in item recognition. *J. Mem. Lang.* **66**, 416–437 (2012).
29. R Ratcliff, A Thapar, G McKoon, The effects of aging on reaction time in a signal detection task. *Psychol. aging* **16**, 323 (2001).
30. R Ratcliff, P Gomez, G McKoon, A diffusion model account of the lexical decision task. *Psychol. review* **111**, 159 (2004).
31. A Thapar, R Ratcliff, G McKoon, A diffusion model analysis of the effects of aging on letter discrimination. *Psychol. Aging* **18**, 415 (2003).
32. J Spaniol, DJ Madden, A Voss, A diffusion model analysis of adult age differences in episodic

and semantic long-term memory retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* **32**, 101 (2006).

33. J Spaniol, A Voss, HJ Bowen, CL Grady, Motivational incentives modulate age differences in visual perception. *Psychol. aging* **26**, 932 (2011).
34. M von Krause, V Lerche, AL Schubert, A Voss, Do non-decision times mediate the association between age and intelligence across different content and process domains? *J. Intell.* **8**, 33 (2020).
35. AL Schubert, D Hagemann, C Löffler, GT Frischkorn, Disentangling the effects of processing speed on the association between age differences and fluid intelligence. *J. Intell.* **8**, 1 (2020).
36. G McKoon, R Ratcliff, Aging and predicting inferences: A diffusion model analysis. *J. Mem. Lang.* **68**, 240–254 (2013).
37. B Bloem-Reddy, YW Teh, Probabilistic symmetries and invariant neural networks. *J. Mach. Learn. Res.* **21**, 1–61 (2020).
38. M Theisen, V Lerche, M von Krause, A Voss, Age differences in diffusion model parameters: a meta-analysis. *Psychol. Res.*, 1–10 (2020).
39. R Ratcliff, R Childers, Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision* **2**, 237 (2015).
40. V Lerche, A Voss, M Nagler, How many trials are required for parameter estimation in diffusion modeling? a comparison of different optimization criteria. *Behav. Res. Methods* **49**, 513–537 (2017).
41. MD Lee, EJ Wagenmakers, *Bayesian cognitive modeling: A practical course.* (Cambridge university press), (2014).
42. ST Radev, UK Mertens, A Voss, L Ardizzone, U Köthe, Bayesflow: Learning complex stochastic models with invertible neural networks. *IEEE Transactions on Neural Networks Learn. Syst.*, 1–15 (2020).
43. K Xu, B Nosek, A Greenwald, Psychology data from the race implicit association test on the project implicit demo website. *J. Open Psychol. Data* **2** (2014).
44. R Ratcliff, Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychol. Aging* **23**, 900 (2008).
45. R Ratcliff, J Love, CA Thompson, JE Opfer, Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child development* **83**, 367–381 (2012).
46. PA Reuter-Lorenz, DC Park, How does it stac up? revisiting the scaffolding theory of aging and cognition. *Neuropsychol. review* **24**, 355–370 (2014).
47. AG Greenwald, DE McGhee, JL Schwartz, Measuring individual differences in implicit cognition: the implicit association test. *J. personality social psychology* **74**, 1464 (1998).
48. AG Greenwald, BA Nosek, MR Banaji, Understanding and using the implicit association test: I. an improved scoring algorithm. *J. personality social psychology* **85**, 197 (2003).
49. M Usher, JL McClelland, The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. review* **108**, 550 (2001).
50. KC Klauer, A Voss, F Schmitz, S Teige-Mocigemba, Process components of the implicit association test: A diffusion-model analysis. *J. Pers. Soc. Psychol.* **93**, 353 (2007).
51. D Matzke, EJ Wagenmakers, Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis. *Psychon. bulletin & review* **16**, 798–817 (2009).
52. DJ Schad, M Betancourt, S Vasishth, Toward a principled bayesian workflow in cognitive science. *Psychol. Methods* (2020).
53. JK Lindeløv, mcp: An r package for regression with multiple change points. (2020).

PNAS
www.pnas.org

Supplementary Information for

Processing speed is high until age 60: Insights from Bayesian modeling in a one million sample (with a little help of deep learning)

Mischa von Krause, Stefan T. Radev, Andreas Voss

Mischa von Krause.
E-mail: mischa.vonkrause@psychologie.uni-heidelberg.de

This PDF file includes:

Figs. S1 to S15 (not allowed for Brief Reports)
Tables S1 to S3 (not allowed for Brief Reports)
SI References

Bayesian Workflow

In the following, we describe and visualize the details of our Bayesian pipeline, which we separate into four main steps, as advocated by the notion of a principled Bayesian workflow (1).

A. Domain Expertise Consistency. Prior predictive checks are designed to test whether a model is consistent with the relevant domain expertise. Typically, prior predictive checks are carried out either in i) prior space, or ii) in data space. In the latter case, more aptly termed a *prior pushforward check*, one obtains a random draw from the prior and simulates a synthetic data set using the sampled parameter configuration. We resort to the following plausibility considerations and visualizations to ensure consistency with domain expertise:

1. We place uniform priors over all diffusion model parameters. Concerning the choice of prior space, we select a broad prior domain which is sufficient to cover the range of plausible parameter values obtainable from human participants (2–5). The priors are visualized in Figure S1. Our priors encode no initial beliefs in parameter differences between the experimental conditions (congruent/incongruent), so any differences emerging after estimation will be due to the information contained in the data. Moreover, the broader prior for non-decision times in incorrect trials compared to correct trials reflects the nature of the IAT task, as described in the main body of our paper.

2. In order to explore the data space implied by our choice of priors, we perform a prior pushforward check with respect to mean response times. The resulting pushforward densities are depicted in Figure S2 and comply with our expectation that incorrect trials are more likely to result in higher response times.

3. Further, we perform a prior pushforward check with respect to accuracy. The resulting pushforward densities are depicted in Figure S3. These results imply a high expected accuracy prior to observing data (since the IAT is a relatively easy task) and no prior belief about differences in accuracies between conditions.

4. Finally, we perform prior pushforward checks with respect to higher moments of the implied RT distribution, namely, skewness and variance (standard deviation). The resulting pushforward densities are depicted in Figures S5 and S4, respectively. These densities imply a positive expected skewness, as observed in numerous empirical RT studies. Moreover, our prior choice implies a moderate expected variability in individual RTs, which nevertheless is broad enough to include participants exhibiting very high variability in their responses. A larger variability is expected in incorrect responses.

B. Computational Faithfulness. Computational faithfulness refers to the ability of a Bayesian method to recover the correct target posterior in a particular modeling scenario. Since BayesFlow enables fully amortized inference, we can efficiently compute simulation-based calibration (SBC, (6)), which allows us to visually detect potential biases in posterior estimation. After the training phase of BayesFlow, we performed 5000 simulations from the diffusion model and obtained 250 posterior samples for each simulated data set. SBC histograms of the resulting rank statistics are depicted in Figure S6 and indicate no apparent global biases in posterior location and dispersion. The marginal histograms for non-decision times in incorrect trials imply a slight underestimation of the *true* parameter values by the posterior means. This small distortion is probably caused by simulated data sets with zero incorrect trials in both conditions, which render estimation of τ_w impossible.

C. Model Adequacy/Sensitivity. Model sensitivity asks whether the parameters of a model can be recovered given the model's prior specification, generative scope, and particular algorithmic from. To evaluate model sensitivity, we first perform a simulation-based recovery study and plot the known true parameters vs. the corresponding posterior means (as summaries of the full posteriors). The recovery result obtained from 300 simulations are depicted in Figure S7. The plots indicate very good point-estimate recovery, with R^2 metrics ranging from 0.999 (non-decision time in correct trials) to 0.678 (non-decision time in incorrect trials). The low R^2 for the non-decision time parameter in incorrect trials is due to simulated data sets having zero or very few errors. These data sets thus provide no information for Bayesian updating and BayesFlow returns the prior. Second, we compute posterior contraction and posterior z-score on 300 simulated data sets, and visualize these on a 2D Cartesian plane (1). Accordingly, an adequate model exhibits high posterior contraction and a posterior z-score symmetrically distributed around 0. Indeed, such behavior is depicted in Figure S8, which also represents the estimation problems for the non-decision time parameter in incorrect trials posed by data sets with zero or very few errors.

D. Posterior Predictive Checks. Finally, posterior predictive checks assess whether the model captures the relevant structure of the assumed true data generating process. We randomly selected $100,000$ cases from the IAT sample to simulate response times and accuracy data based on the diffusion model parameter posterior means. Due to the large sample size studied, it was sufficient to use posterior means as representatives of the parameter values. 60 trials each were generated for both experimental conditions per person, as is also the case in the empirical data. Figures S9 and S10 show scatterplots of the empirical response time quartiles plotted against the respective empirical data. Model fit is good on average, although simulated slow error response times show some larger deviations from a perfect correlation, explainable by unstable summary statistics for both empirical and simulated data due to the low number of error trials available for analysis.

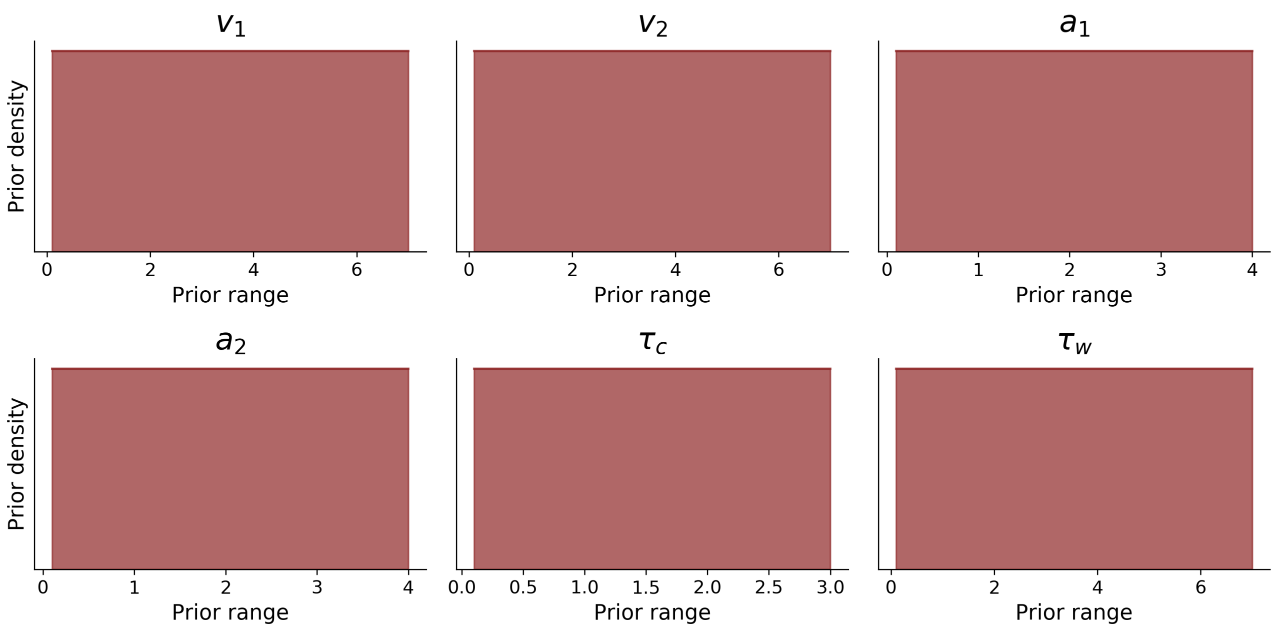

Fig. S1. Uniform prior distributions over diffusion model parameters.

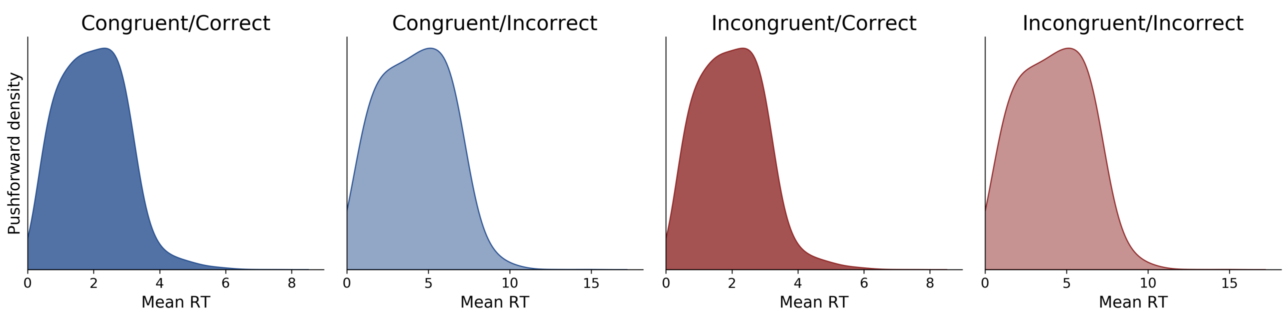

Fig. S2. Simulated distributions of mean response times (RTs), as implied by our prior specification.

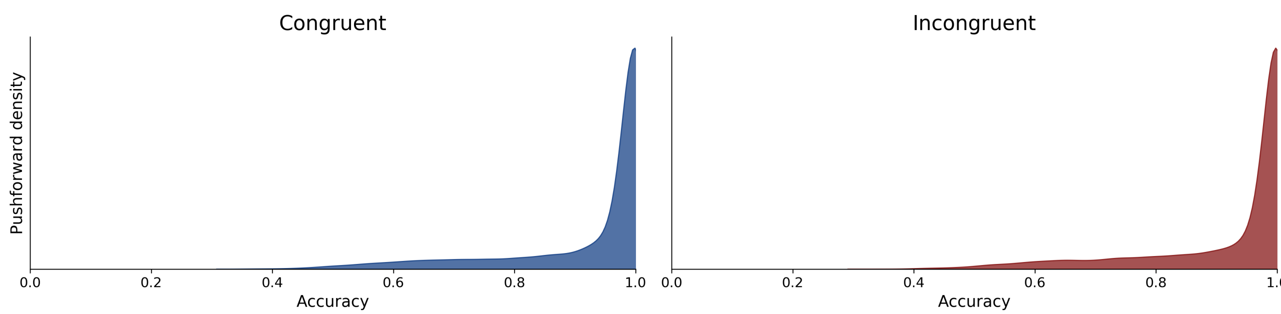

Fig. S3. Simulated distributions of accuracies, as implied by our prior specification.

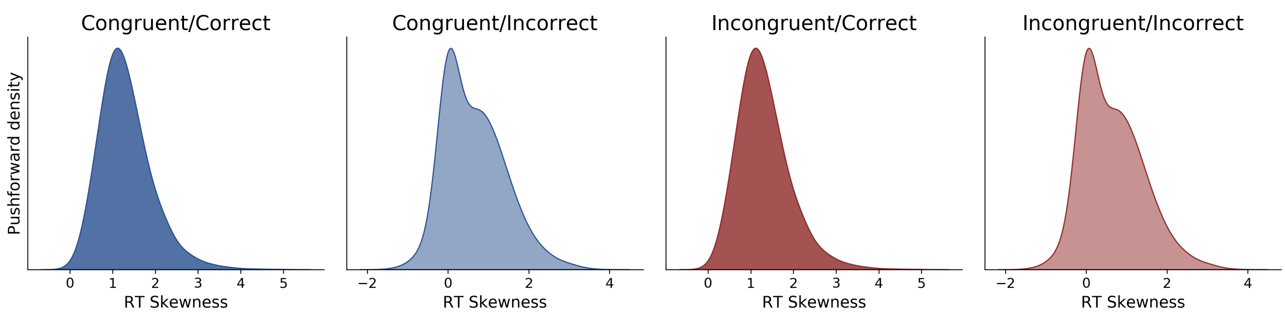

Fig. S4. Simulated distributions of skewness, as implied by our prior specification.

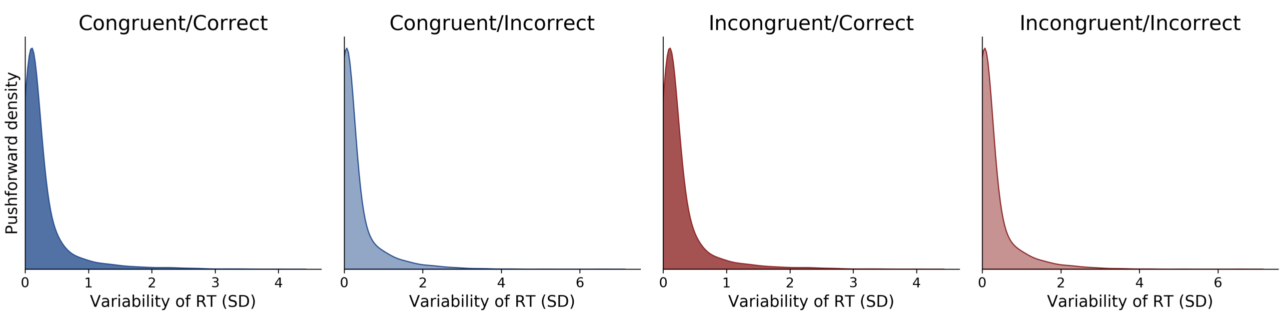

Fig. S5. Simulated distributions of variability (standard deviations from the mean RT), as implied by our prior specification.

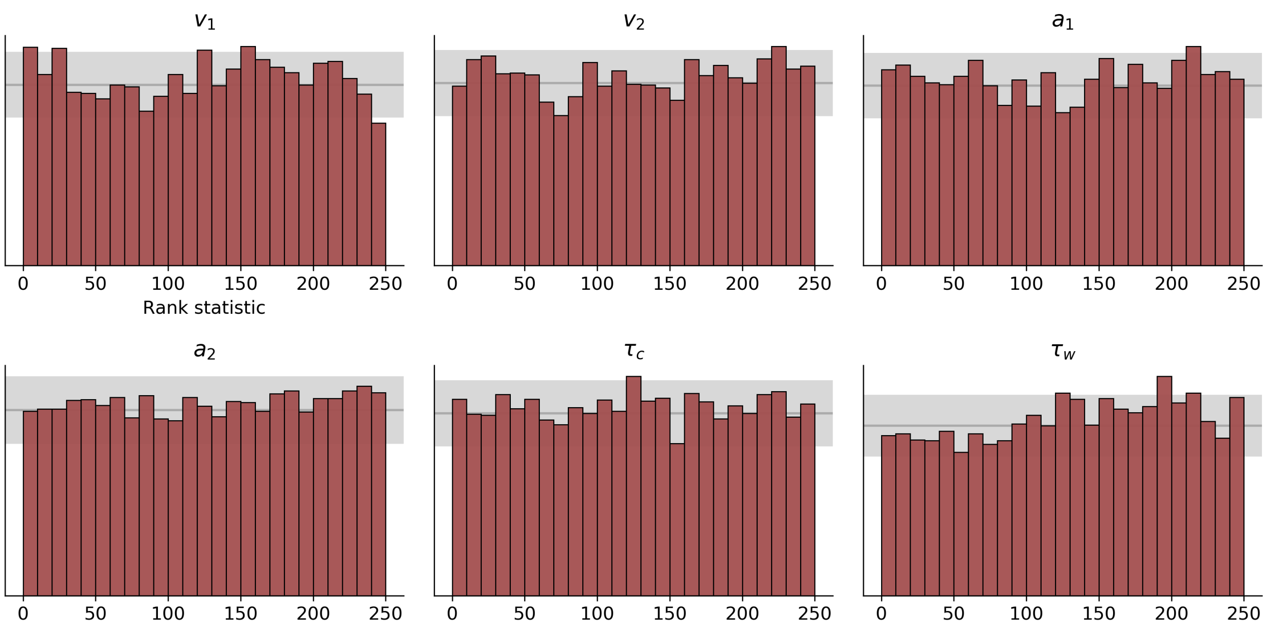

Fig. S6. SBC histograms for each marginal parameter posterior. Horizontal gray bars depict confidence intervals for a binomial distribution, as recommended by (6).

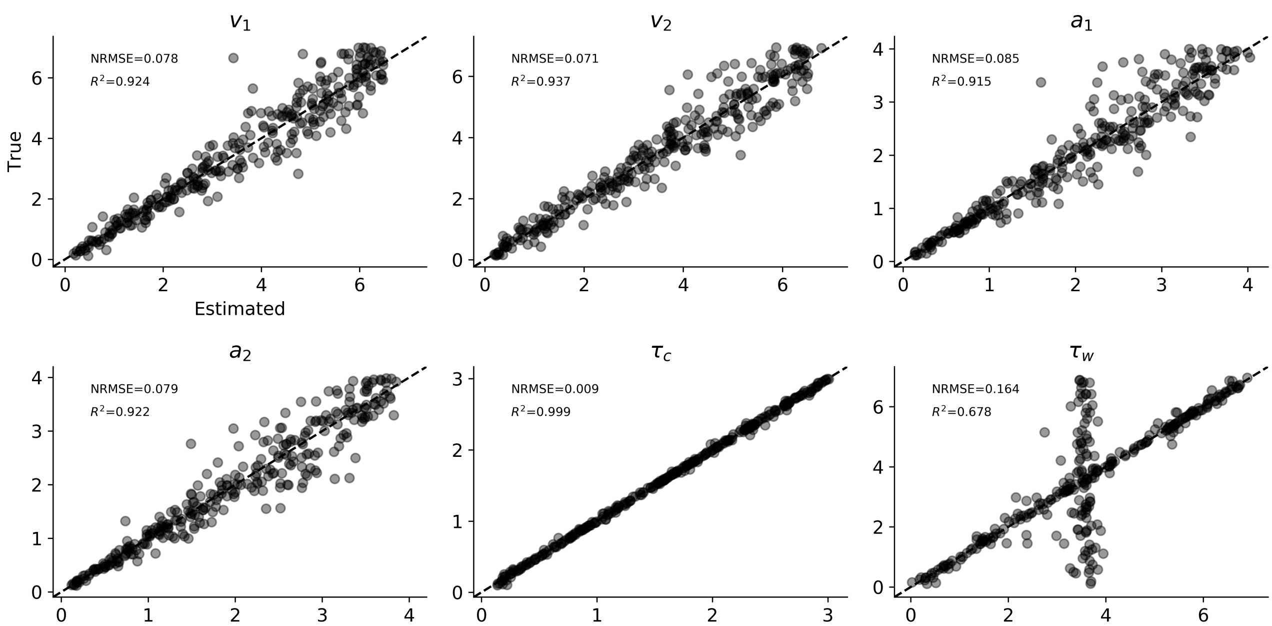

Fig. S7. True parameter values vs. estimated posterior means. The plots indicate excellent recovery of the parameters, with the vertical line in the last plot indicating unrecoverable non-decision time parameters for data sets with zero or very few errors.

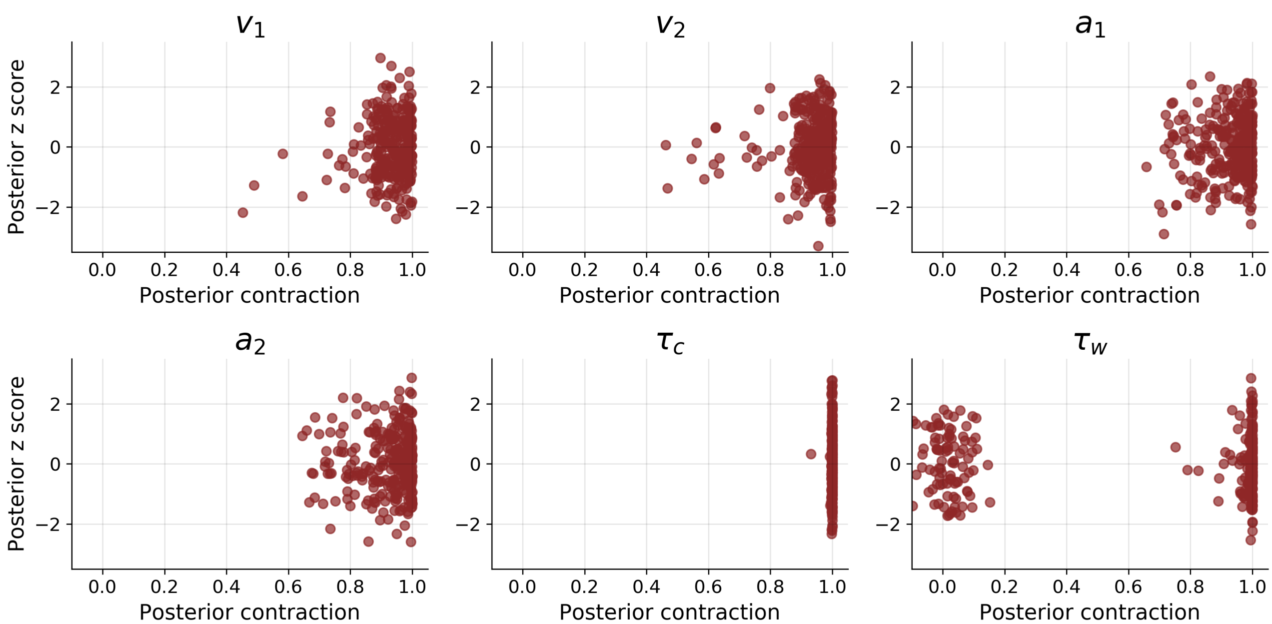

Fig. S8. Posterior z-scores and posterior contraction for each marginal posterior obtained from simulations from the Bayesian model. The plots indicate high posterior contraction and a z-score centered around 0, but also no contraction for τ_w in data sets with zero or very few errors.

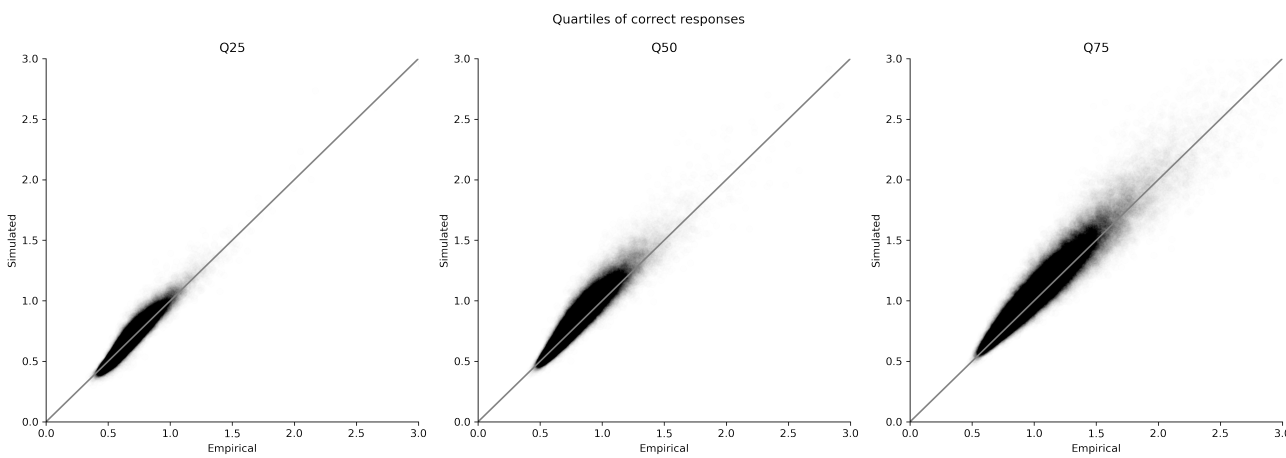

Fig. S9. Posterior predictive checks for correct response times. The plot shows the quantiles of the correct response RTs for empirical and simulated data plotted against each other in a scatterplot. The alpha level in the graph is very low because of the great number of plotted points.

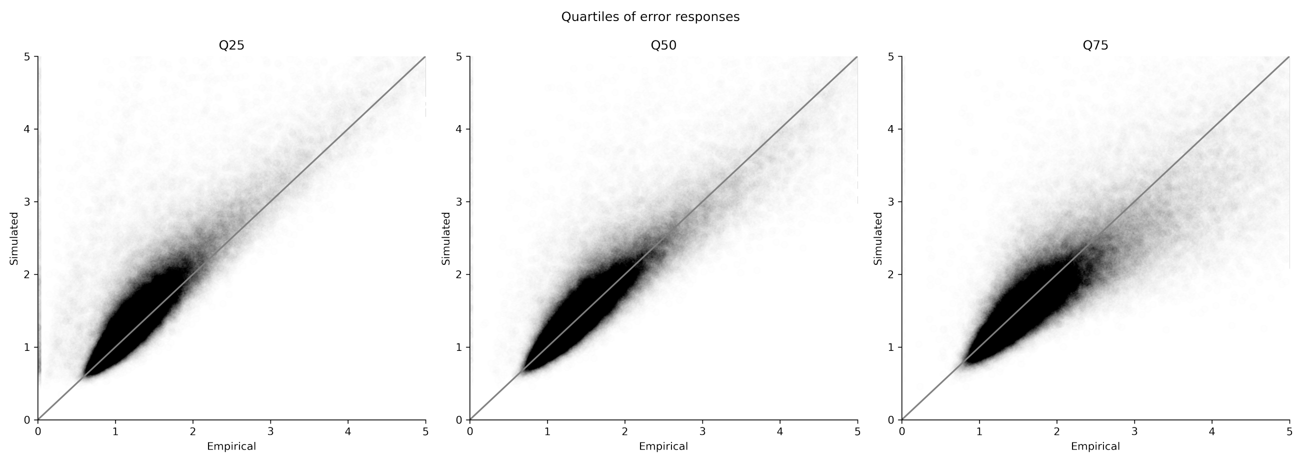

Fig. S10. Posterior predictive checks for error response times. The plot shows the quantiles of the error response RTs for empirical and simulated data plotted against each other in a scatterplot. As accuracy generally was high (median accuracy 95%), error quantiles are based on very low numbers of trials and thus unstable. The alpha level in the graph is very low because of the great number of plotted points.

Additional Analyses

Table S1 shows descriptive statistics of age, mean correct RTs in both experimental conditions (incongruent/congruent), and the posterior means of all estimated diffusion model parameters. In Table S2 we report across-person correlations of the diffusion model parameters, while in Table S3 we present the summarized within-person correlations based on the individual joined posterior distributions.

In Figure S11 we show our main analyses for mean correct RTs, drift rates, and boundary separations for the congruent experimental condition, as well as for error non-decision times. Results are very similar to those obtained for the incongruent condition. For boundary separation, the second change point was estimated to lie already at age 50, with the age-related increase in old age being less pronounced compared to the incongruent condition.

Figures S12, S13, S14, and S15 show different robustness checks. Our main results on age trends in drift rates were consistent across sub-samples (S12, S13), participant countries of origin (S14), and different experimental stimuli (S15), all in regard to both experimental conditions. Across all checks, the age-related decline in drifts rates after age 60 was less pronounced in the congruent compared to the incongruent experimental condition.

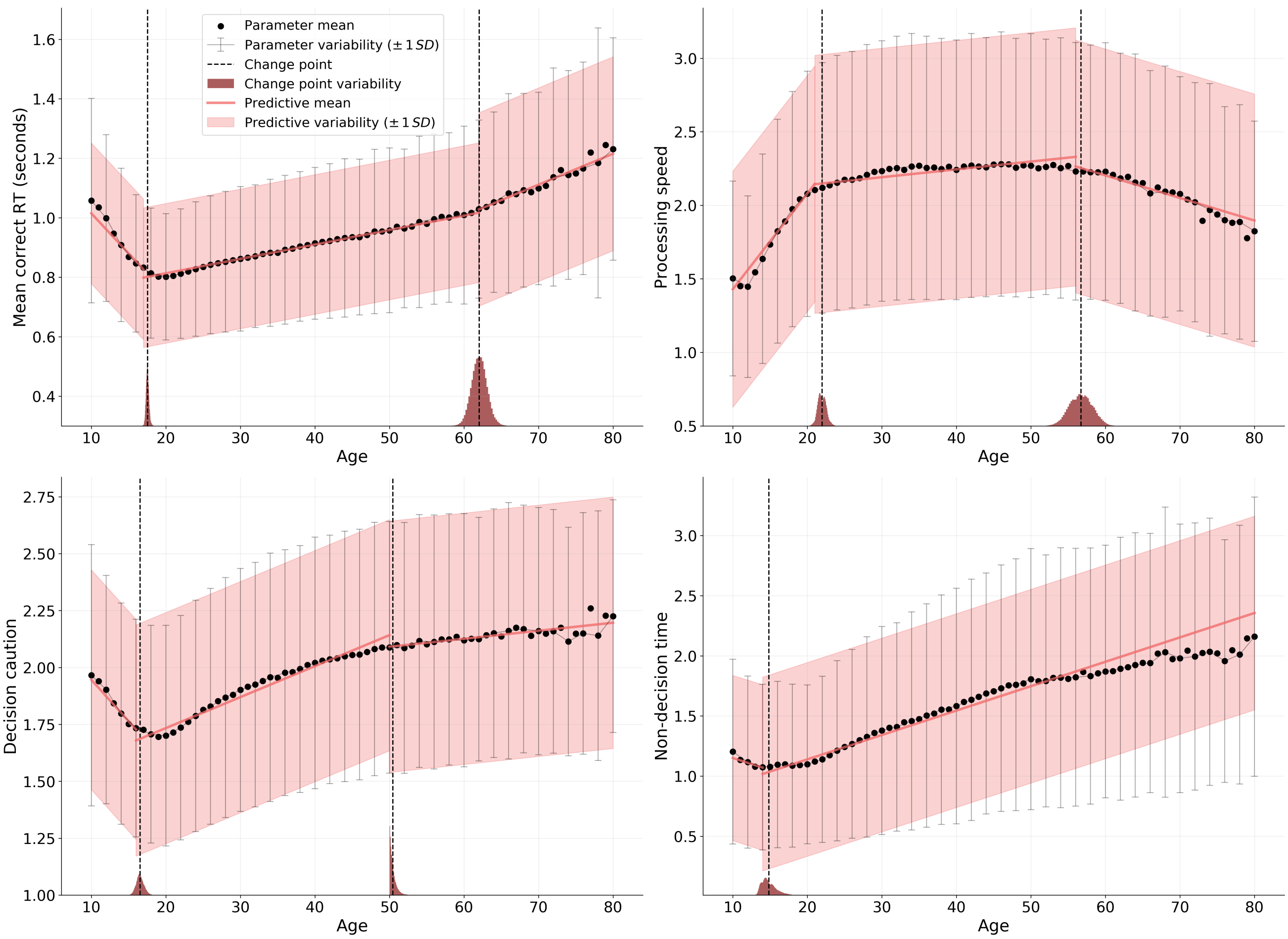

Fig. S11. Results for the congruent experimental condition and error non-decision times. Mean response times (RTs) and diffusion model parameters as a function of age. Black points indicate means computed separately for each year of age. Bars indicate standard deviations (only shown for every second year for better clarity). Red lines denote the Bayesian piece-wise ridge regression model's mean predictions, which describe the observed means fairly well. The shaded red region denotes the uncertainty (standard deviation) of the piece-wise model's predictions. The dashed lines indicate the mean change points estimated from the per-age-group averaged data, with the full posterior distributions (scaled for readability) of the change points shown at the bottom of each plot. Trends are very similar to the ones found for the incongruent condition for mean correct RTs and drift rates; the same holds true for error non-decision times in relation to correct non-decision times. Boundary separation values show a slightly earlier second change point and less pronounced increasing trend in old age than is found for the incongruent condition.

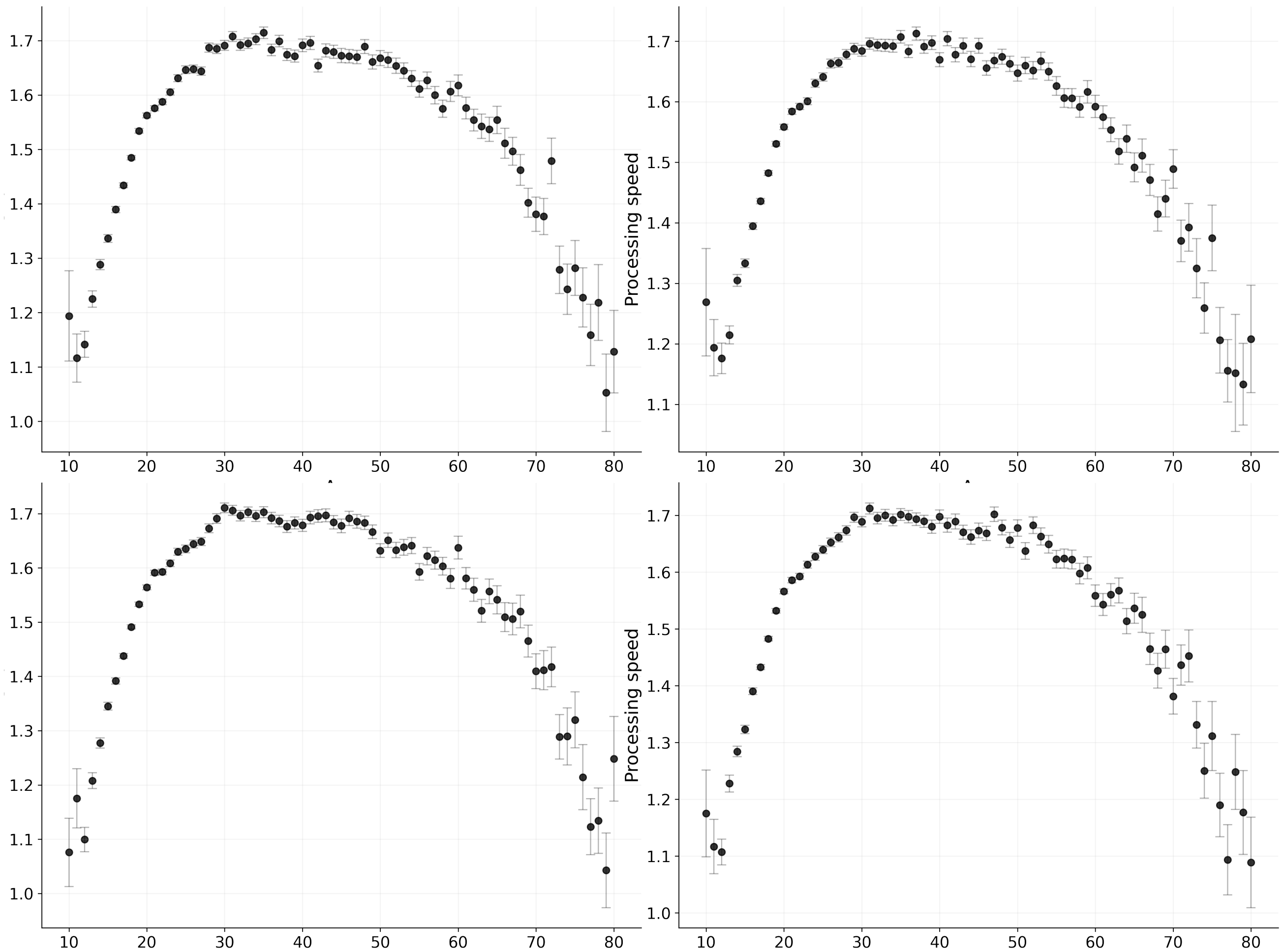

Fig. S12. Robustness check I. We randomly divided our sample in four sub-groups, with three groups containing 300,000 participants each, and one containing the rest. The figure shows the trends in drift rates for the incongruent condition. Trends are very similar across the sub-groups. Dots indicate means per year of age, while the bars indicate the standard errors of the means.

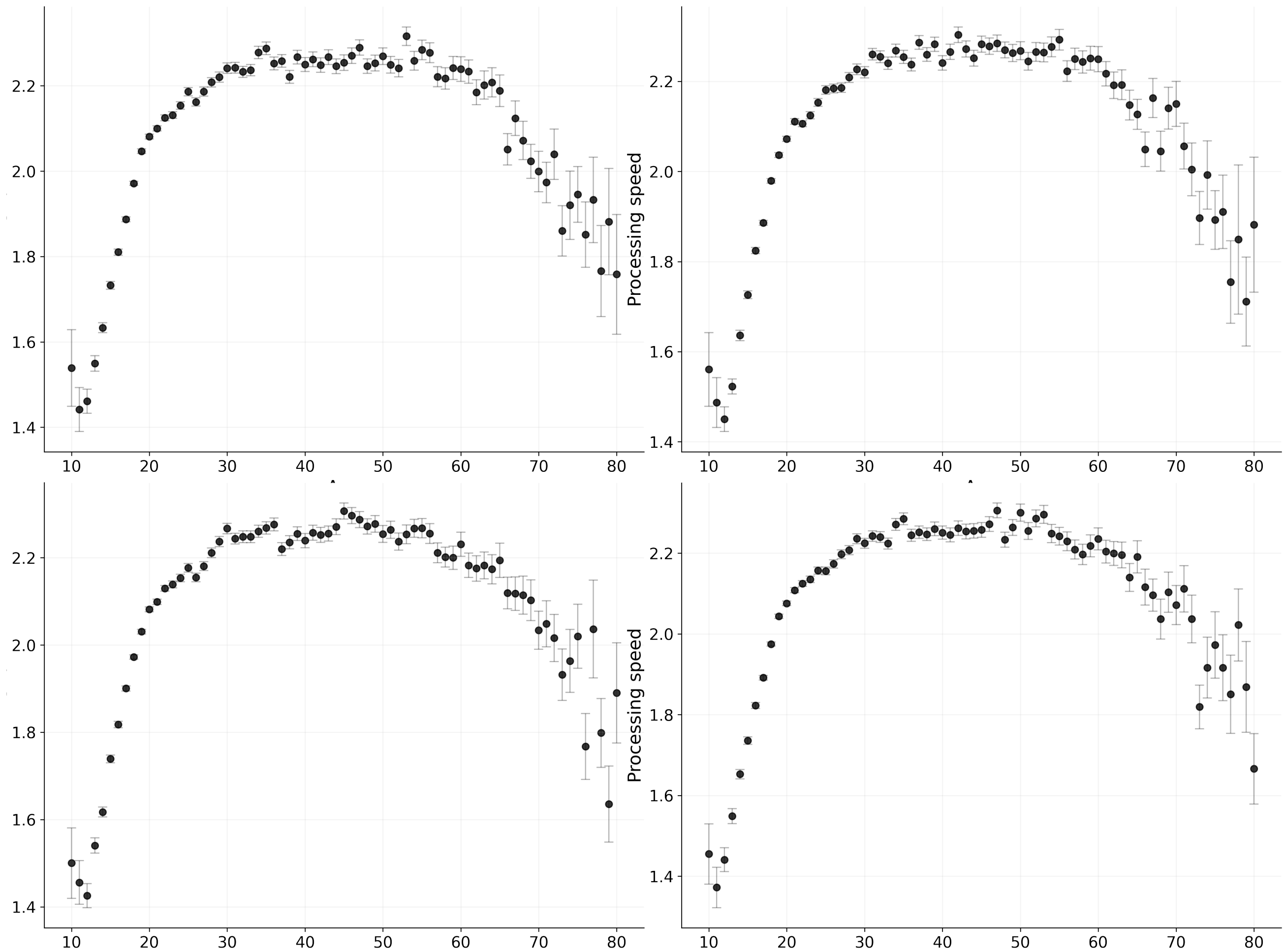

Fig. S13. Robustness check II. We randomly divided our sample in four sub-groups, with three groups containing 300,000 participants each, and one containing the rest. The figure shows the trends in drift rates for the congruent condition. Trends are very similar across the sub-groups. Notably, the decrease in drift rates after age 60 is less pronounced than for the incongruent condition. Dots indicate means per year of age, while the bars indicate the standard errors of the means.

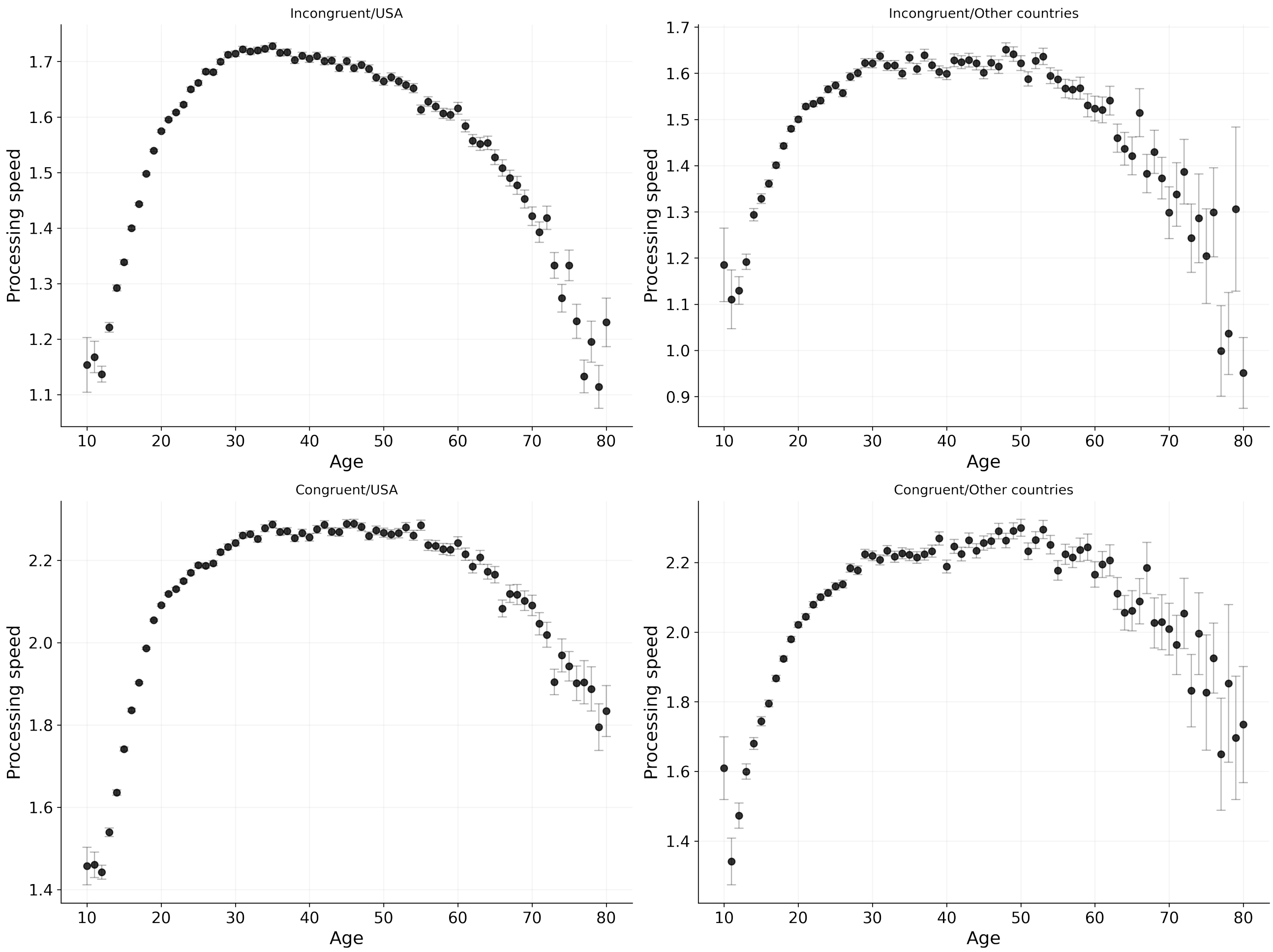

Fig. S14. Robustness check III. We compared the age trends in processing speed between participants indicating that they were born in the United States vs. all other countries. Trends were similar both in the congruent and the incongruent experimental conditions. Dots indicate means per year of age, while the bars indicate the standard errors of the means.

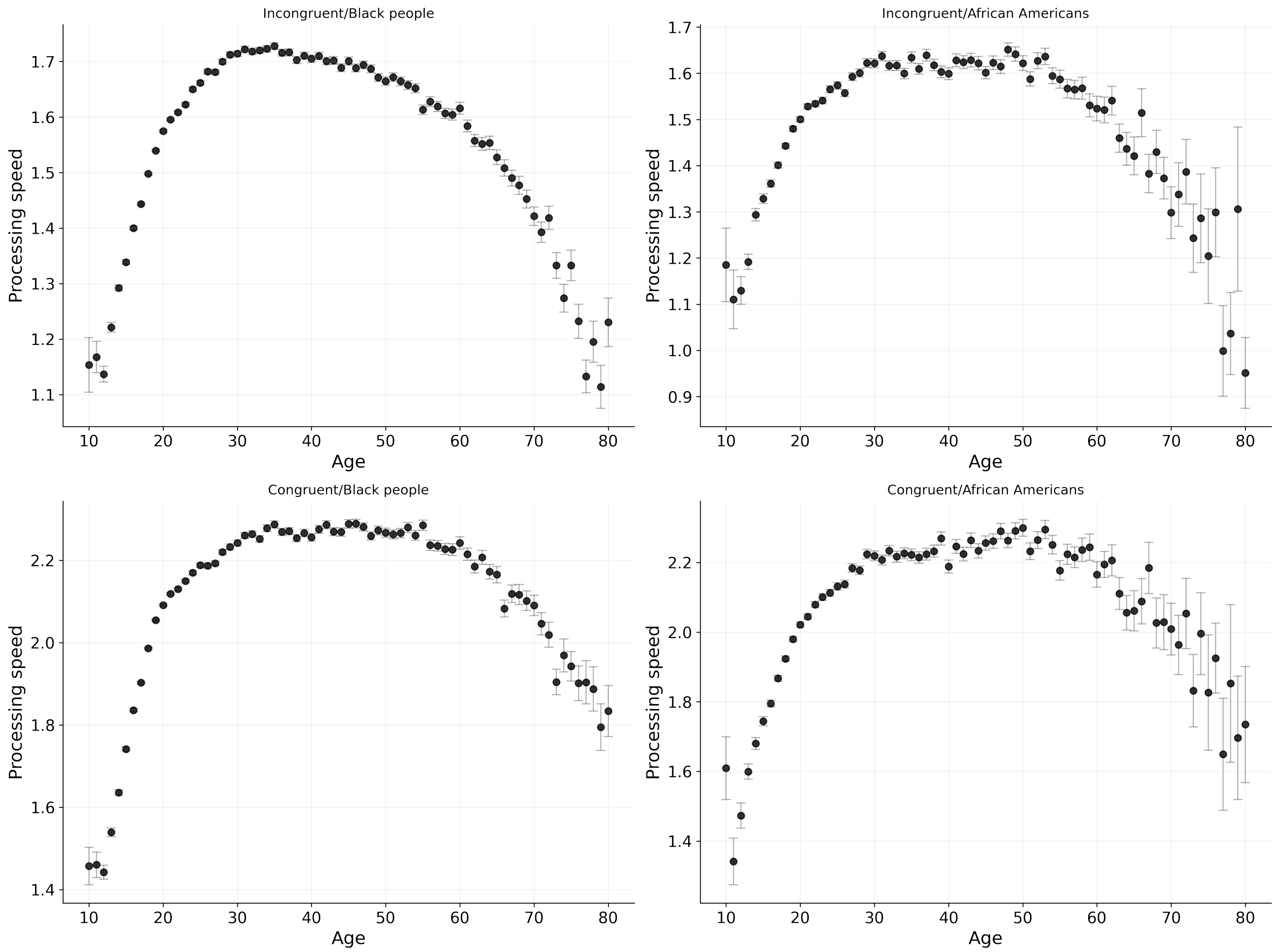

Fig. S15. Robustness check IV. We compared the age trends in processing speed between experimental sessions where "Black/White" were the stimuli and those that used "African American/European American". Trends were similar both in the congruent and the incongruent experimental conditions. Dots indicate means per year of age, while the bars indicate the standard errors of the means.

Table S1. Descriptive Statistics

	Mean	*SD*	Minimum	Maximum
Age	27.42	12.33	10.00	80.00
Mean correct RT (incongruent)	1.00	0.31	0.40	5.75
Mean correct RT (congruent)	0.86	0.24	0.38	5.16
Processing Speed (incongruent)	1.58	0.64	0.10	6.99
Processing Speed (congruent)	2.10	0.86	0.10	6.99
Decision Caution (incongruent)	1.91	0.51	0.33	4.00
Decision Caution (congruent)	1.83	0.53	0.46	4.00
Non-decision Time (correct)	0.38	0.07	0.10	2.89
Non-Decision Time (error)	1.29	0.84	0.10	7.00

Note. Processing speed = drift rate; decision caution = boundary separation; *SD* = standard deviation. Age was computed as year of data collection minus year of birth.

Table S2. Across-person correlations of diffusion model parameters

	$\nu_{incongruent}$	$\nu_{incongruent}$	$a_{incongruent}$	$a_{congruent}$	$\tau_{correct}$
$\nu_{congruent}$	0.32				
$a_{incongruent}$	-0.19	-0.23			
$a_{congruent}$	-0.11	0.04	0.56		
$\tau_{correct}$	0.15	0.17	-0.03	-0.15	
τ_{error}	0.10	0.04	0.39	0.37	0.19

Note. ν = drift rate; a = boundary separation; τ = non-decision time. Correlation estimates based on the entire sample. While drift rates and boundary separations show the strongest correlations with the respective parameter from the other experimental condition, error non-decision times are related to boundary separations, highlighting the distinct interpretation of τ_{error} in relation to $\tau_{correct}$

Table S3. Within-person correlations of diffusion model parameters

	$\nu_{incongruent}$	$\nu_{congruent}$	$a_{incongruent}$	$a_{congruent}$	$\tau_{correct}$
$\nu_{congruent}$	-0.02				
$a_{incongruent}$	0.37	0.07			
$a_{congruent}$	0.01	0.53	0.38		
$\tau_{correct}$	-0.05	-0.19	-0.62	-0.66	
τ_{error}	0.08	0.07	-0.20	-0.05	0.13

Note. ν = drift rate; a = boundary separation; τ = non-decision time. Person-specific correlations were computed based on the individual joined posterior distributions of diffusion model parameters. Individual correlations were then Fisher-z transformed, averaged, and transformed back to a correlation estimate.

References

1. DJ Schad, M Betancourt, S Vasishth, Toward a principled bayesian workflow in cognitive science. *Psychol. Methods* (2020).
2. D Matzke, EJ Wagenmakers, Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis. *Psychon. bulletin & review* **16**, 798–817 (2009).
3. A Voss, M Nagler, V Lerche, Diffusion models in experimental psychology: A practical introduction. *Exp. psychology* **60**, 385 (2013).
4. V Lerche, A Voss, M Nagler, How many trials are required for parameter estimation in diffusion modeling? a comparison of different optimization criteria. *Behav. Res. Methods* **49**, 513–537 (2017).
5. V Lerche, A Voss, Experimental validation of the diffusion model based on a slow response time paradigm. *Psychol. research* **83**, 1194–1209 (2019).
6. S Talts, M Betancourt, D Simpson, A Vehtari, A Gelman, Validating Bayesian inference algorithms with simulation-based calibration. *arXiv preprint* (2018).

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

**FAKULTÄT FÜR VERHALTENS-
UND EMPIRISCHE KULTURWISSENSCHAFTEN**

Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg
Doctoral Committee of the Faculty of Behavioural and Cultural Studies of Heidelberg University

Erklärung gemäß § 8 (1) c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften
Declaration in accordance to § 8 (1) c) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe.
I declare that I have made the submitted dissertation independently, using only the specified tools and have correctly marked all quotations.

Erklärung gemäß § 8 (1) d) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften
Declaration in accordance to § 8 (1) d) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.
I declare that I did not use the submitted dissertation in this or any other form as an examination paper until now and that I did not submit it in another faculty.

Vorname Nachname
First name Family name ______________Michael von Krause_____________________________

Datum, Unterschrift
Date, Signature ____